Characterizing and revealing biomarkers on patients with Cerebral Amyloid Angiopathy using Artificial Intelligence

Fátima Solange Silva

UMinho | 2020

Universidade do Minho
Escola de Engenharia

Fátima Solange Lima Rezende da Silva

Characterizing and revealing biomarkers on patients with Cerebral Amyloid Angiopathy using Artificial Intelligence

September 2020

Universidade do Minho
Escola de Engenharia

Fátima Solange Silva

Characterizing and revealing biomarkers on patients with Cerebral Amyloid Angiopathy using Artificial Intelligence

Master Dissertation
Master's degree in bioinformatics

Dissertation supervised by
Victor Alves
Tiago Gil Oliveira

September 2020

DECLARATION

Name: Fátima Solange Silva

Dissertation Title: Characterizing and revealing biomarkers on patients with Cerebral Amyloid Angiopathy using Artificial Intelligence.

Supervisors: Victor Manuel Rodrigues Alves, Tiago Gil Oliveira

Conclusion Year: 2020

Master Designation: Mestrado em Bioinformática

Master Branch: Tecnologias de Informação

Universidade do Minho, 11 /September /2020

Assinado por : **Fátima Solange Lima Rezende da Silva**
Num. de Identificação: BI15616589
Data: 2020.09.11 18:43:32 +0100

Signature: _____ CHAVE MÓVEL _____

Acknowledgements

I would like to thank Rui Lebre, who was always there to support and push me to go further and share his wisdom. To António Neves and Alina Trifan, for lighting the way and never give up on me. The cheers from the three of you made the path easier and enjoyable.

I would also like to thank my parents and siblings for the support and patience.

To all my friends, thank you for your presence and friendship, which were essential during this period.

Last, but not least, I would like to thank my supervisors at University of Minho, Victor Alves and Tiago Gil Oliveira, I am grateful for the support and availability throughout the development of the project.

To my beloved ones, I dedicate this dissertation.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Universidade do Minho, 11/ September/2020

Signature: Assinado por : **Fátima Solange Lima Rezende da Silva**
Num. de Identificação: BI15616589
Data: 2020.09.11 18:44:52 +0100

CHAVE MÓVEL

[This page intentionally left blank]

ABSTRACT

Cerebral Amyloid Angiopathy is a cerebrovascular disorder resulting from the deposition of an amyloidogenic protein in small and medium sized cortical and leptomeningeal vessels. A primary cause of spontaneous intracerebral haemorrhages, it manifests predominantly in the elder population. Although CAA is a common neuropathological finding on itself, it is also known to frequently occur in conjunction with Alzheimer's disease, being sometimes misdiagnosed.

Currently, CAA diagnosis is generally conducted by *post-mortem* examination or, in live patients by the examination of an evacuated hematoma or brain biopsy samples, which are typically unavailable. Therefore, a reliable and non-invasive method for diagnosing CAA would facilitate the clinical decision making and accelerate the clinical intervention.

The main goal of this dissertation is to study the application of Machine Learning (ML) to reveal possible biomarkers to aid the diagnosis and early medical intervention, and better understand the disease. Therefore, three scenarios were tested: Classification of four neurodegenerative diseases with annotation data obtained from visual rating scores, age and gender; Classification of the diseases with radiomic data derived from the patient's MRI; and a combination of the previous experiments. The results show that the application of Artificial intelligence in the medical field brings advantages to support the physicians in the decision-making process and, at some point, make a correct prediction of the disease label.

Although the results are satisfactory, there are still improvements to be done. For instance, image segmentation of cerebral lesions or brain regions and additional clinical information of the patients would be of value.

Keywords: Machine learning, CAA, Medical imaging, MRI, biomarkers, Artificial intelligence

RESUMO

Angiopatia Amiloide Cerebral (AAC) é uma doença vascular cerebral resultante da deposição de matéria amiloide. Principal causa de hemorragias cerebral espontâneas, a AAC manifesta-se predominantemente na população idosa. Embora a AAC seja uma doença que por si só tem um grande impacto no grupo etário referido, ocorre em simultâneo com inúmeras outras doenças neurodegenerativas, como a doença de Alzheimer. Atualmente, o diagnóstico de AAC realiza-se quer em *post-mortem*, quer em pacientes vivos. No entanto, o diagnóstico em vida é conseguido por meio de biópsias de tecidos cerebrais, sendo um método invasivo, o que dificulta a intervenção clínica. Deste modo, torna-se imperativa a procura de alternativas fiáveis e não invasivas em vida para auxiliar o diagnóstico da doença e permitir a melhoria da qualidade de vida do paciente. Perante os progressos na área da tecnologia e medicina, esta dissertação propõe o estudo da aplicação de algoritmos de Machine Learning (ML) para revelar possíveis biomarcadores para auxiliar o diagnóstico e permitir uma intervenção precoce. Deste modo, foram testados três cenários distintos: a classificação de quatro doenças neurodegenerativas com dados anotados obtidos a partir de métricas visuais de avaliação da atrofia, idade e sexo; a classificação das doenças com dados gerados a partir de métodos radiómicos; e uma combinação das duas abordagens anteriores.

Neste documento apresenta-se e discute-se os resultados obtidos com a aplicação de quatro diferentes algoritmos de ML que visam a deteção automática da doença associada à imagem testada. Adicionalmente, é feita uma análise crítica de quais as características mais relevantes que levaram à tomada de decisão por parte do algoritmo. Os resultados demonstram que através de aplicação de metodologias automáticas é possível o auxílio ao diagnostico médico por especialistas e, no limite, a realização de diagnostico automático com elevada precisão. Finalmente, são apresentadas possíveis alternativas de trabalho futuro para que os resultados possam ser aperfeiçoados, como por exemplo, a segmentação das regiões de interesse, i.e., identificação das lesões, aquando da anotação por especialistas. Mediante a inclusão dessa segmentação, uma vez que será mais especifica, os resultados serão, por sua vez, aprimorados.

Palavras-Chave: *Machine learning*, AAC, Imagiologia médica, Imagem por Ressonância Magnética, biomarcadores, Inteligência Artificial

[This page intentionally left blank]

Table of Contents

LIST OF FIGURES

List of Tables

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
CE	Cross-Entropy
CNN	Convolution Neural Networks
CSF	Cerebrospinal Fluid
CSO	centrum semiovale
CT	Computer Tomography
DL	Deep Learning
FPR	False Positive Rate
FTD	Frontotemporal Dementia
GM	Gray Matter
ISLES	Ischemic Stroke Lesion Segmentation
ML	Machine Learning
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
PVS	Perivascular spaces
RF	Radiofrequency
TPR	True Positive Rate
WM	White Matter
WMH	White Matter Hyperintensities

GLOSSARY

Artificial Intelligence (AI)	Computer science programs developed to simulate the human and animal intelligence, particularly the ability to learn and make decisions, to perform tasks and solve problems.
Artificial Neural Networks (ANN)	A model inspired by the activity of the brain and nervous system, composed by layers, consisting of processing units are interconnected by nodes.
Class	In machine learning, class refers to the category of the data, or the labels in a dataset.
Dataset	A collection of data. The data contains features and, in the case of supervised training, labels.
Deep Learning (DL)	A subfield of the Machine Learning methods, capable of unsupervised learning to classify the data provided.
DICOM	Digital Imaging and Communications in Medicine (DICOM) is the international standard to transmit, store, retrieve, print, process, and display information in medical imaging.
Feature	In Machine Learning, a feature expresses the information of an attribute.
Machine Learning (ML)	An approach to achieving Artificial Intelligence, where a computer is trained using specific algorithms and a large amount of data to gain the ability to adapt to new situations and predict patterns.
Magnetic Resonance Imaging (MRI)	A non-invasive imaging technique based on the absorption and emission of energy at a specific frequency. MRI is widely used to assist the clinical and research fields.
NIFTI	Neuroimaging Informatics Technology Initiative (NIFTI) is a format commonly adopted in neuroimaging to store brain imaging data.
Pixel	Picture element that defines a unit of information in two-dimensional space.
Region of Interest (ROI)	Contours or surfaces outlining the boundaries of an object on an image or volume.
Skull Stripping	Method of removing non-brain tissues (e.g. bone, scalp, veins) from MRI.

Soft Tissue Tissues that support other structures and organs of the body, such as tendons, fibrous tissues, fat, muscles, nerves and blood vessels.

Test set The subset of the dataset used to test the model after the training phase.

Train set The subset of the dataset used to train a model.

Voxel Volume element that represents a value in the three-dimensional space.

[This page intentionally left blank]

1 INTRODUCTION

1.1 CONTEXT AND MOTIVATION

Cerebral amyloid angiopathy (CAA) is a disease with inflammatory characteristics resulting from amyloid-beta (Aß) deposition in the small and medium calibre vasculature of the cerebral cortex and leptomeninges. It is the primary cause of spontaneous intracranial haemorrhagic events. Accumulation of these amyloid deposits occurs mostly in the posterior cortical arterial microvasculature and larger vessels such as cerebral veins and venules [1], [2]. CAA is a common cause of lobar haemorrhage, and in most cases is sporadic. Its clinical course is slow and often silent, manifesting predominantly in the elderly population with haemorrhagic events [3]. CAA predominates in an elderly population, and co-exists in about 80% of Alzheimer's patients, and may play a key role in the pathogenicity of this dementia, in addition to intracerebral and infarct haemorrhages [2]. The symptoms of CAA include changes in the nervous system, which may suddenly manifest, leading to delirium, confusion, weakness, paralysis, or joint problems [4], [5].

With the advances on the medical and technological fields, a higher impact of imaging and biochemical approaches in medical decision making would be expected, providing a spectrum of accessible and valuable information on the amyloid deposits and consequences in the cerebral vasculature. However, currently tests for cerebral amyloid angiopathy are still at an early stage, and the diagnosis is therefore based only post-mortem tissue analysis. Given the inaccessibility of brain tissue in life, diagnostic approaches will have to undergo indirect methods such as magnetic resonance imaging (MRI) analysis [4], [5].

In order to standardize the diagnosis of CAA, the Boston criteria were proposed in 1995, based on imaging and clinical data. These comprise a combination of clinical, imaging, and pathological parameters that allow a probable diagnosis to be made in living patients in the absence of brain tissue. The criterion for identifying CAA was therefore divided into three categories: possible CAA, probable CAA, and definitive CAA. Definitive autopsy based CCA requires an extraordinary level of neuropathological severity, including features of advanced vasculopathy, revealing lobar, cortical or subcortical haemorrhagic evidence. Probable CAA supported by MRI is the most certain level of diagnosis without the use of neuronal tissues. It allows the evaluation and counting of microhaemorrhages in the cortical and lobar grooves and intracerebral haemorrhages. Finally, probable CAA may also be supported by pathological evidence from biopsies [5].

Pittsburgh Compound Positron Emission Tomography (PET-PIB) and magnetic resonance assisted with mapping such as GRE (Gradient - Recall Echo) and SWI (Susceptibility - Weighted Imaging) stand out in the line of complementary diagnostic directed examinations although they do not yet offer objective and effective results in its diagnosis. The cerebrospinal fluid analysis, namely Aß-40 and tau protein levels also stand out as auxiliaries for the diagnosis of this disease, but its specificity is low [6].

Although the medical and technological fields have been witnessing extraordinary advances, currently the approaches used for the detection of Cerebral Amyloid Angiopathy are still at an early stage, as the diagnosis is still focussed on the analysis of post-mortem tissues. An evolution on the complementary diagnostic tools, either in imageology or biochemistry, would be expected. Furthermore, the existing studies focus either on the MRI scans or on the clinical information and sum scores of visual metrics.

1.2 THE PROBLEM

Over the last years, Artificial Intelligence algorithms have been extensively used in the medical field. AI approaches are used for image segmentation, diagnostic prediction, and to provide tools to aid in the medical decision making.

Making accurate and life-time diagnoses becomes increasingly mandatory. Early diagnosis is crucial for the patient to be followed in life. To this end, the present study is proposed to identify clinical markers based on brain atrophy assessment metrics (proposed by Harper et al.) [7], quantification of hypertensive lesions in the white matter [8], evaluation enlargement of perivascular spaces in the basal ganglia, quantification of microhaemorrhages, evaluation of relative density of perivascular spaces, and finally, identification of new imaging biomarkers using artificial intelligence (AI) tools. These metrics will be compared using as control groups other diagnosed neurodegenerative diseases (Alzheimer's disease, Mild Cognitive Impairment, or Parkinson's disease).

This study brings the novelty of the association with clinical data obtained from the evaluation of MRI scans by trained research physicians, and Radiomics data obtained after image segmentation.

1.3 OBJECTIVES

The main goal of this thesis is to study the application of AI tools to aid in CAA diagnostic, providing data to aid in the early prediction and diagnostic of the disease, thus enabling the patient to have access to appropriate and early medical intervention.

For instance, the work will go through the development of several machine learning models to predict four diagnosed diseases: Cerebral Amyloid Angiopathy, Alzheimer's disease, Parkinson's disease dementia, and Mild cognitive impairment. The outcome and accuracy of the models will be assessed, to determine the biomarkers that better describe the data.

The work will address the following topics:

- Evaluate and classify clinically and analytically MRI of patients diagnosed with Cerebral Amyloid Angiopathy (CAA), Alzheimer's disease (AD), Parkinson's Disease Dementia (PDD), and Mild cognitive impairment (MCI), followed mostly at the Braga Hospital (Internal Medicine Service, Neuroradiology);
- Identify pathological factors predicting the occurrence of cerebral haemorrhagic events, in addition to vascular changes and cerebral atrophy;
- Identify associations between disease manifestation profile and possible markers such as gender, age and brain injury;
- Compare findings with other neurodegenerative diseases' studies.

1.4 STRUCTURE OF THE DISSERTATION

The manuscript is structured in 5 chapters, starting with the current one introducing the reader to the disease, the motivation and objectives of the study. Next, follows a theoretical chapter regarding the disease's general features, morphological aspects, pathology, and diagnosis.

The third chapter starts by making a brief overview of the contents and technologies used in the work, including the data acquisition, the visual ratings for evaluation of neurodegenerative diseases, and magnetic resonance imaging. Also, some insights about machine learning approaches can be read.

On the fourth chapter, the materials and methods are presented in three experiments, describing the data used for the studies, and the techniques carried on for data acquisition and creation of Datasets, and all of the steps from data preparation to the creation of ML

models for classification purposes, ending with the results for each experiment. Finally, the fifth chapter ends with the conclusions and future work.

2 Cerebral Amyloid Angiopathy

Cerebral amyloid angiopathy (CAA), is the term used to define a cerebral small vessel disease (SVD) characterised by the presence of amyloid-beta (Aβ) [9] protein within blood vessel walls of small and medium-sized cortical and leptomeningeal arteries, arterioles, capillaries, and veins [10]. Described pathologically in the early 20th century, CAA is a condition of increasing clinical importance.

CAA is a pathological hallmark of Alzheimer's disease, being present in almost 90% of AD cases [9]. This type 2 small vessel disease also occurs in rare hereditary diseases and in other disorders such as Down's syndrome [10]. This angiopathy is frequently found in the general elder population, with its incidence increasing with age [10], [11]. Large lobar haemorrhages, microbleeds, and ischemic changes such as white matter lesion and microinfarcts, are typically associated with the development of the disease [4], [12].

It occurs as sporadic or familiar forms with several amyloid proteins involved, being firmly associated with intracerebral haemorrhage (ICH). As mentioned before, amyloid proteins are cleaved by a precursor before they are deposited [12], [13]. In hereditary disease forms, mutations lead to amino acid substitutions or elongation of the precursor proteins, resulting in the mutation of amyloid proteins with different aggregation properties or the increase in proteolytic cleavage of the amyloid protein from its precursor, increasing Aβ production [14]. The most common type of CAA is caused by Aβ deposition, thus termed as Aβ-CAA. This form, is associated with sporadic and familial AD, occurring also as a hereditary cerebral haemorrhage disorder, as amyloidosis-Dutch type (HCHWA-D) and similar familial disorders, or in normal aging in the elderly [11], [12].

2.1 Morphological Aspects

CAA results from a chronic degenerative process by which the media of parenchymal arterioles undergoes progressive loss of its muscle cells with simultaneous accumulation of an eosinophilic hyaline material [15], [16], mostly composed of the more soluble, amyloid-β40 Aβ species, contrasting with the amyloid plaques found in AD, predominantly composed of amyloid-β42 species [17].

The vessels affected by amyloid-β accumulation undergo secondary vasculopathy changes, from fibrinoid necrosis, loss of smooth muscle cells, wall thickening, and microaneurysm formation, to the deposition of blood breakdown products in perivascular vessels [10], [16].

In the most severe form of cerebral amyloid angiopathy, the vessels become dilated and disrupted [17], with focal wall fragmentation and blood extravasation, with or without micro-aneurysmal dilatation, and sometimes show luminal occlusion [18].

Light microscopically, a green appearance under polarized light when stained with Congo red dye and the fluorescent appearance under ultraviolet light when stained with thioflavin S are specific histological features of cerebral amyloid angiopathy. Another typical feature on light microscopy is the “double barrel” appearance given by the splitting of the internal elastic lamina caused by the deposition of hyaline material in the vessel wall [19], [20].

The cortex, particularly the occipital lobe, is frequently the most affected cerebral region by Aβ-CAA [21]. Hippocampus, cerebellum and basal ganglia are less affected while deep central grey matter, subcortical white matter and brain stem usually show no vascular amyloid alterations [22]. Veins and capillaries tend to be less frequently affected by vascular Aβ deposits, while leptomeningeal and cortical small and medium sized arteries and arterioles are the most affected [21], [23]. Figure 2.1 represents cerebral blood vessels with amyloid deposition (A- C) and amyloid progressive accumulation, resulting in smooth muscle cell loss (D–F) in a sporadic CAA case with multiple cerebral haemorrhages. The arrow at F points at a preserved smooth muscle cell [24].

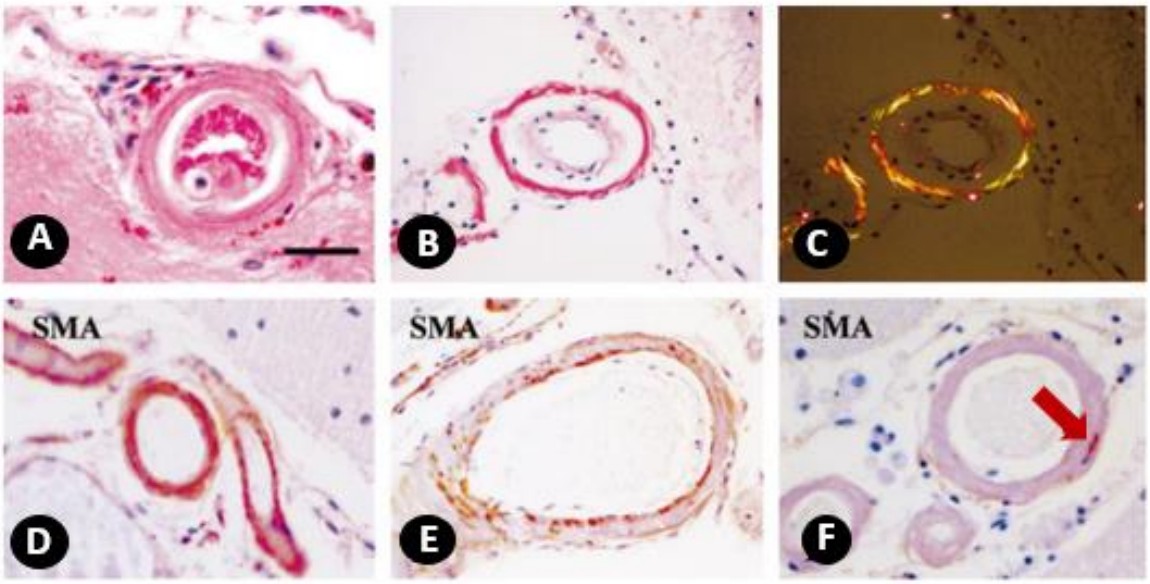


Figure 2.1- Deposition of amyloid in Cerebral blood vessels followed by progressive loss of smooth muscle cells (Adapted from Revesz et al.)

2.2 PATHOLOGICAL SUBTYPES OF CAA

Cerebral amyloid angiopathy can be divided into sporadic (spontaneous) and familial or hereditary forms. Mutations in the gene encoding the APP account for some rare (usually autosomal dominant) forms of CAA, including CAA-Dutch type [25]. Familial non-amyloid-β forms of CAA include familial British dementia, familial Danish dementia and Icelandic cystatin C mutation. In general, hereditary forms of CAA have an earlier onset and more severe clinical manifestations than sporadic CAA [25], [26].

Although exceptionally rare, familial CAAs have provided significant insights on how mutations in the coding region of the APP contribute to CAA pathogenesis: for example, the Iowa, Dutch, Italian, and Arctic mutations render amyloid-β highly toxic to vessel wall components, and more resistant to proteolytic degradation or clearance from the brain [10], [27], [28].

At least two distinct pathological subtypes of sporadic CAA have been described: CAA-type 1, characterized by amyloid-β in cortical capillaries (with or without involvement of other vessels) [29], [30]; and CAA-type 2, where amyloid-β deposits are restricted to leptomeningeal and cortical arteries, arterioles and, rarely, veins [30]. Amyloid-β deposition in the wall of capillaries may cause luminal obstruction in the most severe stages [18]. The APOE ε4 allele is most strongly associated with CAA-type 1, while APOE ε2 is more associated with CAA-type 2. CAA-type 1 appears to be more closely associated with parenchymal amyloid deposition in Alzheimer's disease [30].

2.2.1 HEREDITARY AB- CAA

Hereditary CAA can be classified in Aβ and non- Aβ forms, based on the peptide accumulated. The first β-amyloid precursor protein (APP) mutation was discovered in 1990, in two Dutch families. A familial occurrence of CAA had already been described in 1964, considering severe CAA the pathological cause of the disease [31]. Familial cerebral amyloid angiopathy describes a group of rare disorders usually autosomal dominant disorders. Many of these disorders are specific to a few families, striking at an earlier age of onset, differing from the spontaneous CAA, typically affecting at middle to late middle age [2], [32].

Another type of hereditary CAA, Amyloidosis-Dutch type (HCHWA-D) occurs due to the heterozygous mutation at the codon 693 of the APP gene, corresponding to the amino acid 22 of Aβ [33]. The first nucleotide of the triplet is mutated, resulting in the substitution of the

original codon GAA to CAA, thus resulting in the substitution of glutamine amino acid for glutamic acid [34].

Furthermore, experimental studies on the Dutch variant revealed an increase in mutated Aβ toxicity towards vascular cells, reflecting on large lobar intracerebral haemorrhages, cognitive deterioration associated with white matter abnormalities, and small ischemic infarctions and haemorrhages [34]. Besides the Dutch mutation, four other Aβ related mutations located between the 21st and 23rd amino acids have been described [33].

The Flemish (A692G Flemish mutation), Italian (APP E693K, caused by the substitution of glutamic acid by lysine at residue 22), Arctic, Iowa, and Spanish types were found to have an association with severe CAA, and also dementia compatible with AD, clinically and neuropathologically [35]–[38]. In the Italian type, amyloid deposits are found in cerebral parenchyma and meningocortical vessels. Patients suffer from stroke, cognitive decline, and in some cases seizures [39].

The Artic type, from Northern Sweden, is affected by the glutamic acid substitution for glycine, reflecting on subjects with clinical features of early AD and no signs of strokes or vascular lesions on brain imaging [35]. Iowa mutation type occurs at position 23 of Aβ and causes severe amyloid angiopathy, dementia, occipital calcifications, and small ischemic infarctions [37], [39]. A similar mutation has been recently found on a Spanish family, showing similar pathological features, besides the development of symptomatic intracerebral haemorrhagic stroke [38]. CAA has also been described as a key feature in a German familial AD mutation in the PS2 gene, reporting cases of cerebral haemorrhages [40].

Although severe cases of CAA may not result of APP mutations, they can be associated with AD in result of mutations in the presenilin-1 (PS1) and presenilin-2 (PS2) genes [41]. Extensive and widespread CAA are caused by The PS1 mutations and PS1 deletions, with the abundance of amyloid plaques having an important role as well [42], [43].

CAA pathology has been shown to be more severe when the mutation occurs beyond codon 200 of the PS1 gene [44].

2.2.2 Hereditary forms of CAAs caused by non-Aβ proteins

Besides Aβ, the deposition of other amyloid proteins has originated several forms of familial CAAs. Hereditary forms of CAA are generally more severe, and present earlier age of onset

and neurodegeneration/death. These rare hereditary forms tend to present in the form of autosomal dominant disorders.

The most common disease forms caused by these amyloid proteins are the HCHWA-Icelandic type, Familial British and Danish dementia, CAA related to Prion Protein Amyloidosis, and Finnish familial amyloidosis [45], [46].

The non-Aβ Icelandic type has an early onset and results in fatal cerebral haemorrhages on half of the mutated patients. The ones who survive the haemorrhage may be left with cognitive decline and dementia. Brain imaging shows severe amyloid deposits within small arteries and arterioles of leptomeninges, cerebral cortex, basal ganglia, brainstem, and cerebellum. Furthermore, amyloid deposits can be found in salivary glands, and peripheral and lymphoid tissues [45].

The truncated N-terminal amyloid protein bears a substitution mutation (Alanine for Tyrosine) at codon 68 of the Cystatin C, resulting in the substitution of glutamine for leucine amino acid [45], [46]. A cysteine protease inhibitor present in cortical neurons and the cerebrospinal fluid (CSF), Cystatin C level in the CSF of Icelandic disease type patients is about half of those measured in control patients [47].

Since Cystatin C is also present in parenchymal and vascular Aβ deposits, it may present a role in the pathogenesis of other amyloidosis, AD and cerebral amyloid lesions familial dementias [18].

Familial British dementia (FBD) and familial Danish dementia (FDD) are autosomal dominant neurodegenerative diseases characterized by the elongation of the protein BriPP to genetic abnormalities [48], [49]. Hippocampal degeneration and Parenchymal and vascular amyloid deposits lesions lead to progressive cognitive decline, spasticity and cerebellar ataxia [10].

FBD is caused by a Tyrosine to alanine point mutation in the stop codon of BRI2, resulting in an extended 277 amino acid long protein, the ABriPP. Characterized by progressive dementia, ataxia, spastic tetraparesis, and in some rare cases, cerebral haemorrhage, FBD patients suffer from progressive memory loss, with disease onset around the sixth decade [50].

Vascular amyloid accumulates in small arteries and arterioles in the leptomeninges, grey and white matter throughout the Central nervous system, and the striatum. In the retina, blood vessels are heavily affected by CAA, along with several other organs.

It has been associated to AD, with severe and widespread amyloid plaques, neurofibrillary degeneration, and phosphorylated tau proteins. These findings suggest that FBD might be a phenocopy of AD, with Abri being the equivalent to the Aβ protein [51].

BRI2 is a type II transmembrane protein with a 100 residue long conserved domain. This domain is found in 309 proteins. Although its exact physiological role is still unknown,

it has been suggested to have several functions related to neuronal differentiation, stress response, and receptor on the cell surface [52].

FDD presents in an early stage of life with cataract and ocular haemorrhages, followed, some years later, with severe hearing loss. After the mid-age, the patient starts suffering from cerebellar ataxia followed by psychiatric disturbance and progressive dementia [53].

FDD is associated with a 10-nucleotide duplication between codons 265 and 266 of the BRI2 gene. The resulting frameshift lacks the normal stop codon, leading, as well, to the production of a 277 amino acid long precursor protein (ADanPP) [50], [54].

Wild type BRI2 and FBD-BRI2 are cleaved close to their C terminal by furin, protein convertases, releasing 23- and 34-aminoacid long mutated peptides, termed Bri and ABri, respectively [55]. Although the Bri peptide is not known to be amyloidogenic, the ABri peptide is found in perivascular and parenchymal deposits in the brain of affected individuals. Deposition of Aβ, either isolated or combined with ADan, in vessels and brain parenchyma are a feature of FDD [53], [56].

CAA related to Prion protein (PrP) amyloidosis is known as the GerstmannSträussler-Scheinker syndrome (GSS) variant, caused by a Tyrosin to Guanine mutation in the PRNP gene, resulting in a newly formed stop codon and truncated N- and C- terminal of PrP.

In this case, PrP-immunoreactive CAA, prominent perivascular PrP deposits and neurofibrillary tangle pathology are the main findings of GSS neuropathology [24].

At last, familial amyloidosis-Finnish type (FAF) is a rare condition carrying mutated G654A or G654T gelsolin gene. Amyloid deposition in spinal, cerebral and meninges are traits of the disease, along with extravascular deposits in the spinal nerve roots and sensory ganglia [57], [58].

Amyloidosis caused by mutations on the transthyretin gene (TTR) result in the deposition of the transthyretin protein in extracellular spaces of several organs, progressing into late- onset autosomal dominant systemic diseases [59]. The Hungarian and Ohio pedigrees are examples

of the involvement of the meninges and brain parenchyma [60], [61]. Table 2.1 sums up both the Aβ and non- Aβ Hereditary forms of CAA.

Table 2.1 - Hereditary CAA forms caused by Aβ and non- Aβ proteins.

HEREDITARY CAAS

DISEASE	HCHWA -I	HCHWA-D	FBD	FDD	FAF	PrP-CAA
GENE	CYST C	APP	BRI2	BRI2	GEL	PRNP
PRECURSOR PROTEIN	Cystatin C	Amyloid Precursor Protein (APP)	Abri Precursor Protein (AbriPP)	Abri Precursor Protein (AbriPP)	Gelsolin (GEL)	Prion Protein (PRP)
AMYLOID PROTEIN	ACys	Aβ	Abri	ADan	AGel	APrP
NOTES	Recurrent lobar Haemorrhages; Age at onset: 20-30 years.	Lobar haemorrhages, dementia; Age at onset: 50 years;	Progressive dementia, cerebral ataxia, spastic tetraparesis; Age of onset: 45 -50 years	Cataracts, deafness, progressive ataxia, dementia; Age of onset: 30 years.	Progressive Lattice corneal dystrophy, chronic and peripheral neuropathy, cutaneous amyloidosis.	Progressive cognitive decline.

2.2.3 Sporadic forms of CAA

There are two main forms of sporadic small vessel diseases. One is the sporadic cerebral Amyloid angiopathy (SCAA or Sporadic CAA), a chronic neurodegenerative disease that occurs with progressive deposition of amyloid-β in the media and adventitia of small arteries, arterioles and capillaries in the cerebral cortex [62]. The term used to name the remaining sporadic small vessel disease is “hypertensive arteriopathy”, an age-related disease affected by non-amyloid processes, related to hypertension, diabetes mellitus and several other risk vascular factors [63].

Sporadic CAA is a term used to describe a group of age-related neuropathological processes affecting the small vessels in the brain. Small vessels encompass all small vascular structures (from 5 µm up to 2 mm), including arteries, arterioles, capillaries, venules and small veins located in the brain parenchyma [63], [64]. These vessels supply the brain cortex superficially,

reaching the grey matter and subcortical white matter, and the stem at the base of the brain, supplying the basal ganglia, thalami, and brainstem structures. Pathologies can affect differentially these two systems, or a range of vessels within each system [63].

Most CAA cases are sporadic and a common neuropathological finding in elderly individuals. The disease incidence and severity may vary, depending on the study and age of the patient [15]. Several reports have shown moderate to severe CAA to be present in up to 80% of AD cases [16], [17].

Age is the strongest known clinical risk factor for sporadic CAA [13]. Evidence has shown, from autopsies, that cortical vascular Aβ deposition progresses from the 7th to the 9th decades [14], and ICH affected patients older than 60 years. Other than age, although hypertension is not a risk factor, it can contribute to CAA related cerebral bleeding [65], [66].

ApoE is a protein encoded by the APOE gene with crucial roles in lipoprotein complexes, responsible for the lipid metabolism regulation by binding to cell-surface receptors and proteins associated with lipid transfer and lipolysis [67], [68].

Apolipoprotein E (ApoE) is the only known genetic risk factor for both sporadic and familial CAA [67]. The ApoE is a gene located on chromosome 19 that exists as three alleles. There are three major polymorphisms in the APOE gene, ε4, ε2 and ε3, resulting in an amino acid change which alters the functional properties of APOE isoforms [68], [69].

Data from the analysis of both post-mortem and clinical series has shown that APOE ε4, a known risk factor for AD, increases the risk of sporadic CAA-related lobar ICH by promoting amyloid-β deposition. Moreover, the number of ε4 alleles have a significant relation to clinical severity [67], [69].

ApoE ε2 promotes CAA- related haemorrhage due to the rupture of amyloid-laden vessels, thus increasing the risk of CAA-related lobar ICH, independent of AD [68]. Both alleles are also associated with a younger age of ICH onset, greater likelihood of hematoma expansion, poor clinical outcome, and a higher risk of recurrence. Furthermore, the two allelic variants may interact, resulting in patients with both APOE ε2 and ε4 alleles having the earliest disease onset and highest risk of early ICH recurrence [67], [70]. Other yet to identify genetic polymorphisms related to amyloid metabolic pathways may also play a role in sporadic CAA, (for example presenilin-1, neprilysin and transforming growth factor beta-1) [67], [71].

Depending on the type of vessel involved, at least two distinct pathological subtypes of CAA have been described: CAA-type 1 and CAA-type 2. CAA-type 1 is distinguished by the presence

of amyloid -β in meningeal and cortical arterioles, capillaries, veins and venules, while in CAA-type 2, also known as large-vessel CAA, amyloid-β deposits are restricted to leptomeningeal and cortical arteries, arterioles, and in some cases, veins, sparing capillaries [29], [72]. Amyloid-β deposition in the wall of capillaries (capillary CAA) may cause luminal obstruction in the most severe stages. The APOE ε4 allele is most strongly associated with CAA-type 1, while APOE ε2 is more associated with CAA-type 2 [72]–[74]. Table 2.2 sums up the sporadic forms of CAA.

Table 2.2 - Sporadic forms of CAA.

SPORADIC CAAS

DISEASE	SCAA	SAD
GENE	APP	APP
PRECURSOR PROTEIN	Amyloid Precursor Protein (APP)	Amyloid Precursor Protein (APP)
AMYLOID PROTEIN	Aβ	Aβ
NOTES	Related to the increased risk for lobar ICH.	CAA associated with AD; Associated to presenilin-1 and presenilin-2 mutations.

2.3 CLINICAL CONSEQUENCES OF CAA

CAA is the most common cause of spontaneous, lobar ICH, particularly in patients over the age of 75. Approximately up to 10% of the lobar haemorrhages in non-hypertensive elderly individuals are related to CAA [13]. CAA affects preferentially the cortical or cortico-subcortical regions (especially the occipital and temporal lobes), the cerebellum and, less commonly, deep or brainstem structures [15], [16].

The clinical manifestation of CAA, regarding ICH incidents, depends on the haemorrhage size and location. Brains with intracerebral bleeding caused by CAA show extensive amyloid deposition in blood vessel walls and evidence of disruption of vascular architecture by amyloid accumulation, such as cracking, microaneurysms and fibrinoid necrosis [17]. Vessel dysfunction, reduction in the cerebral blood flow and ischemia are evidence of CAA. Ischemic lesions are characteristic of Sporadic CAA and several hereditary CAA disease types, including

HCHWA-D and the disorder caused by the APP D694N Iowa mutation [18]. Although less common, CAA has been linked to perivascular inflammation, though to cause vascular dysfunction, and giant cell arteritis. CAA also has serious effects on the blood vessels supply of the brain and interstitial fluid, leading to abnormalities in the white matter.

Patients experience headache, nausea, vomiting, seizures and consciousness [19]. A history of head trauma might be a predisposition to ICH in individuals with CAA. The first episode of ICH might have a mild clinical effect, but it is counterbalanced by the risk of recurrent haemorrhagic episodes, which are usually more severe [20]. Survivors of lobar ICH present a higher risk of recurrence compared to deep ICH, often in the same lobe as the first CAA related bleed episode, and in some cases, multiple simultaneous lobar bleeding can occur [21]. Advanced age and larger haematoma size are factors that contribute negatively to the disease prognostic [22]. CAA is also associated with ICH related to oral anticoagulant use. The poor and fragile condition of the cerebral vasculature due to the presence of CAA, is an aggravating factor for the haemorrhagic episode [23].

Evidence suggest that cognitive impairment is a chronic effect of CAA due to CAA-induced haemorrhages, such as ischaemia or perivascular inflammation and the co-existence with Alzheimer's disease or other age-related pathologies (e.g. hypertensive arteriopathy) [24]. However, studies showed that CAA presence in demented patients was much higher than non-demented [25], even when controlling for age and neuropathologies related to dementia [26]. Moreover, CAA together with Alzheimer's disease may worsen the severity of cognitive performance, compared to Alzheimer's disease on its own, even when, again, controlling for age of death, education, neurofibrillary tangles and neuritic plaques number, infarcts, haemorrhage and APOE genotype [27]. Moderate to severe CAA is associated with lower performance in the perceptual speed and episodic memory cognitive domain, but not with semantic and working memory, visuospatial skills or global cognition [28].

Finally, progressive dementia has also been observed in association with perivascular tau pathology but in absence of neuritic-cored plaques in patients with the APOE-ε4 allele. CAA is thus emerging as an important link between neurodegenerative diseases and cerebrovascular pathologies [29]. Figure 2.2 represents Cerebral haemorrhagic lesions and White matter degeneration in patients with CAA and AD related cases [24].

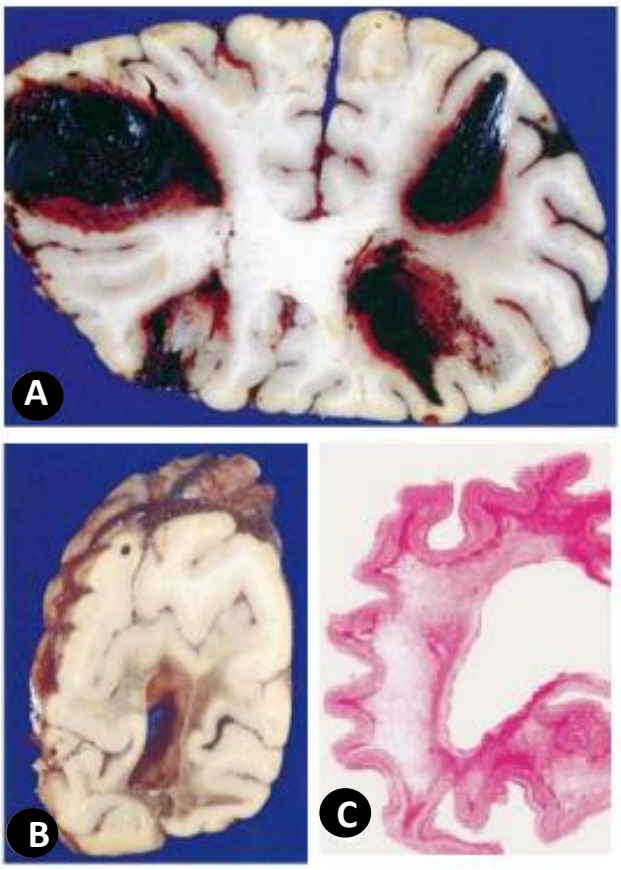


Figure 2.2 - Cerebral haemorrhages in a patient with severe sporadic CAA (A, B), and White matter atrophy and degeneration (C) in a patient suffering with familial AD. (Adapted from Revesz et al., 2003).

2.4 PATHOGENESIS OF CAA

The pathophysiological mechanisms by which the amyloid proteins deposit in the cerebral vasculature is still poorly understood. At least three hypotheses for mechanisms leading to CAA have been suggested: The systemic, the vascular and the drainage hypothesis [30]. These mechanisms are not necessarily exclusive and might even occur at the same time. The systemic hypothesis proposes that Aβ is transferred from the blood to the cerebral vascular structures [31]. This hypothesis is based on the fact that almost all cell types of the body are able to express the β-amyloid precursor protein and potentially secrete Aβ into the circulation. Moreover, in vivo studies demonstrated the bidirectional mediated transportation of Aβ across the blood brain barrier (BBB) using several receptors (e.g. RAGE, LRP-1, SR) [32],[33]. The presence of Aβ in the central nervous system and cerebrospinal fluid is determining of its concentration in the brain, which may happen when the BBB is weakened, which happens to be the case of AD [34], [35]. Several arguments refute the systemic hypothesis based on

evidence of Aβ deposits in the abluminal basement membrane of vessels, suggesting that Aβ has its origin within the central nervous system instead. The presence of Aβ in both AD and non-demented patients supports claims that the source of human Aβ is neuronal [36], [37].

The vascular hypothesis suggests local production of Aβ from cerebrovascular cells. The presence of APP in several cell types found in vessels of AD and HCHWA-D patients' brains and in different layers of the cerebral vessel support this hypothesis [38]. Among such cells, there are the smooth muscle cells (SMCs), human brain pericytes (HBPs) and endothelial cells. In vitro data also suggest that SMCs and HBPs overexpress APP and SMCs to overproduce Aβ [39]. In addition to the aforementioned cell types, adventitial, myocytes, pericytes and perivascular cells have also shown APP overexpression. Arguments against the vascular hypothesis are the fact that large arteries with several layers of SMCs are less severely affected by CAA than small ones, and that Aβ deposits are also present in capillaries that lack SMC. These arguments indicate that neural factors may be important to trigger vascular Aβ deposition [40]–[43].

Finally, the drainage hypothesis suggests that interstitial fluid and solutes, among which is the neuronally produced, Aβ drains out of the brain along perivascular spaces in the capillary walls and between smooth muscle cells in the tunica media of small arteries [44], [45]. When this and other clearance pathways fail in the brain of elderly patients or under other neuronal pathological conditions, such as sporadic CAA, Aβ deposits in the wall of small vascular structures [46], [47].

With the artery walls narrowing caused by the build-up of plaque, the passage of Aβ along the vessel walls is slowed, causing the once soluble Aβ to precipitate forming amyloid plaques, which results in CAA. The amyloid plaques then block the elimination of Aβ, increasing the concentration of Aβ in the brain and, thus, its precipitation [48].

To support this hypothesis, mouse models of hereditary CAA or AD have been crucial to understanding the mechanisms under the cerebral amyloidosis. This has highlighted the presence of CAA in mice which express human APP in the brain under the control of neuron-specific promoters [48], [49]. Based on the drainage hypothesis and the follow up work, several mechanisms for the formation of CAA have been proposed [50], [51].

Evidence show that cerebrovascular disease may contribute to CAA pathogenesis. Furthermore, the impairment of the perivascular drainage may lead to the dilatation of the perivascular spaces within the lobar region and white matter. These enlarged perivascular

spaces can be visible on appropriate brain imaging, enabling its use as a potential neuroimaging marker of CAA [45].

2.5 VISUAL MRI FINDINGS

Since the small vessels and the alterations in consequence of degenerative diseases cannot be seen in vivo by the usual clinical practice, neuroimaging techniques have been adopted to visualize small vessel disease damages [4]. Advances in neuroimaging allow to investigate the hallmarks of small vessel disease in vivo, using relevant disease manifestations applying magnetic resonance imaging (MRI) to visualize cerebral microbleeds (CMBs), microinfarcts and changes in the white matter [4], [52]. Figure 2.3 illustrates the presence of cerebral microbleeds, a visual feature of small vessel disease, in particular CAA [4].

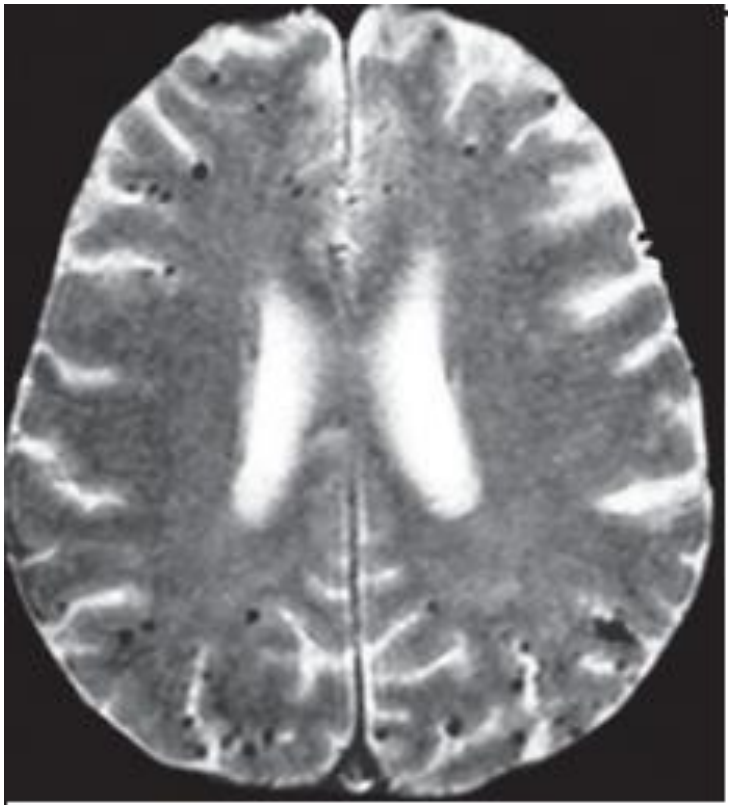


Figure 2.3 - Multiple microbleeds in the cortex of a patient with possible CAA.

2.5.1 CEREBRAL MICROBLEEDS

An important concept to bear in mind is that, when using the neuroimaging approach, the pathogenic interpretation is not uniform for all the disease markers. For instance, microbleed

have a preferential location, either in deep brain regions (associated to hypertensive vasculopathy) or in the lobar region (in the case of patients affected by CAA) [53], [54].

Regarding CAA, studies have proven the association between lobar CMBs and ApoE ε4. Moreover, patients of hypertensive arteriopathy with deep ICH, are more likely to suffer CMBs, whereas patients affected by CAA are more likely to have strictly lobar CMBs [55]. CMBs are also correlated with the risk of lobar ICH recurrence, an important key in prognosis ad diagnosis [56], [57]. The validation of strictly lobar CMBs as a diagnostic marker of CAA makes it a potential predictor of future intracerebral haemorrhage risk, a possible contribution to the study of cognitive impairment and dementia, a potential key link between vascular and degenerative pathologies, and an imaging tool to understand vascular pathologies [58]. Such findings allow the early assistance of asymptomatic individuals with new therapeutic agents to target the disease progression.

Histopathological studies show cerebral microbleeds to be tiny and rounded perivascular haemorrhage lesions, composed of hemosiderin (a magnetic susceptible sub-product of blood-breakdown), between 5 and 10 mm, with high concentration of amyloid, seen on MRI sequences that are sensitive to magnetic susceptibility [59]–[61]. When identifying microbleeds, some artefacts may be mislabelled. To minimize this problem, contiguous slices, different imaging modalities (T2*-GRE, CT, T2*MRI, FLAIR or DWI), lesion shape and location must be assessed [58]. An example of a brain affected with CMB lesions (pointed out with the arrows) is seen in Figure 2.4, adapted from [54].

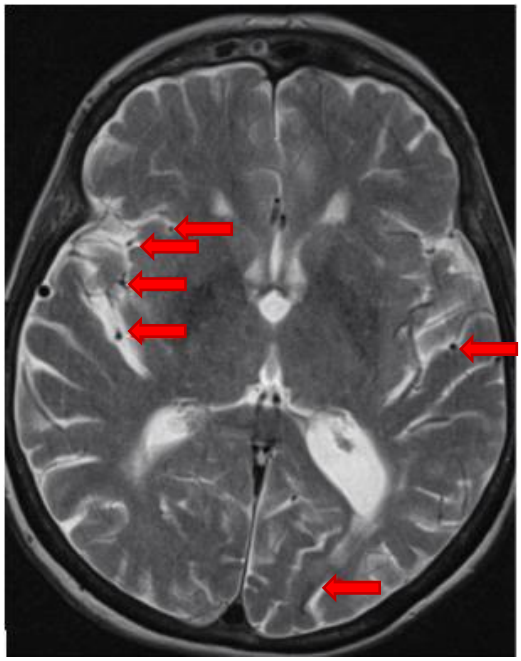


Figure 2.4 – Cerebral microbleeds lesions in a patient with possible small vessel disease.

CMBs count, rating scale, anatomical distribution in the brain (lobar or deep), and the Microbleed Anatomic Rating Scale (MARS) are examples of data used to investigate neurodegenerative diseases and dementias [61], [62].

As MRI techniques have become more sophisticated, CMBs analysis is being increasingly studied either in vascular impaired or healthy older people. Although the most commonly used imaging sequences in the neurological area are the T2 and T1 -weighted, in the last decade the T2*-weighted and T2*-GRE MRI have been shown to produce remarkable data and allow the correct visualization of the hypointense CMBs lesions, for the magnetic susceptibility effects are not corrected with the application of spin-echo techniques [58], [63].

2.5.2 Cerebral microinfarcts

Brains of patients with advanced CAA have evidence of cerebral microinfarcts. Recent investigation found a high prevalence of positive diffusion-weighted imaging (DWI) lesions, indicative of small infarcts, in patients with advanced CAA [19], [64]. These lesions were associated with microbleed burden, suggesting that haemorrhagic processes and ischemia share the same pathophysiological pathways [65]. Another study has stablished that acute, subclinical ischemic brain lesions are recurrent, but previously underestimated, after recent acute intracerebral bleeding, and are three times more frequent in CAA-related ICH than other spontaneous ICH [66]. The microinfarcts lesions were associated with leukoaraiosis and lobar

microbleeds, suggesting that they result from a CAA-related occlusive arteriopathy [66]. However, recent data does not support this fact, suggesting that both CAA and hypertensive arteriopathy trigger high rates of infarction, and proposing the interaction between the haemorrhagic and ischaemic components in CAA and small vessel disease [65],[67]. Figure 2.5, adapted from [75], represents microinfarcts lesions on an elder patient with CAA.

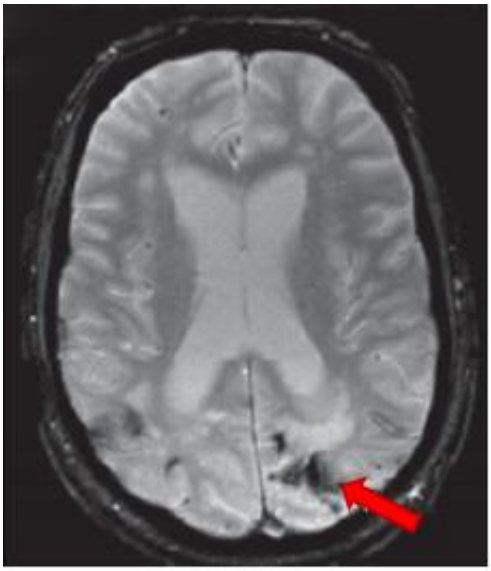


Figure 2.5 - Example of MRI imaging showing small acute infarction.

2.5.3 WHITE MATTER HYPERINTENSITIES

White matter hyperintensities, also termed leukoaraiosis, describes areas of bright signal imaging changes in deep cerebral white matter observed in CT scans as low attenuation, or hyperintense on T2-weighted or FLAIR sequences [13], [68]. White matter lesions are a potential predictor of ICH, a possible contributor to cognitive impairment and dementia and its imaging may provide an understanding of the links between vascular and degenerative pathologies [68]. The presumed pathogenicity of white matter lesions on CAA involves disruption of the BBB due to amyloid accumulation on the small vessels, as described before [4], [54], [69]. Leukoaraiosis is present in CAA patients, affecting the arteries, arterioles and capillaries of the cerebral cortex and junction of the grey-white matter, and patients affected by hypertensive vasculopathy, that affects small arterial penetrators to the white matter and the deep grey nuclei [70], [71]. Recent studies suggest that patients with CAA related lobar ICH have a higher incidence of white matter hyperintensity compared to normal elder

controls, that tends to build up over time [72]. Also, leukoaraiosis volume seems to be greater in patients with CAA and hypertension, suggesting that lowering the hypertension might reduce disabilities related to leukoaraiosis in CAA [71]. Figure 2.6, adapted from [71], represents a MRI of a patient affected with white matter lesions.

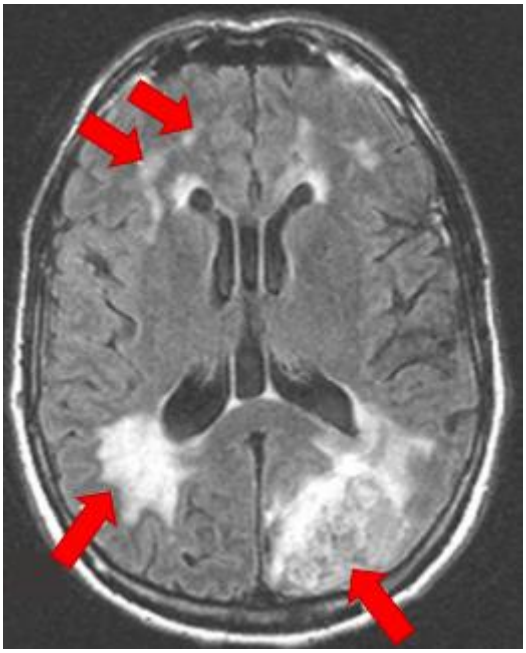


Figure 2.6 - Example of a MRI image showing White matter damage.

2.6 DIAGNOSIS OF CAA

CAA has generally been diagnosed by post-mortem examination [73]. Although the disease may be evaluated by biopsy or evacuated hematoma, in vivo samples are typically unavailable [13], [74]. The Boston criteria was created with the goal of presenting the community with reliable and non-invasive methods for diagnosing CAA, to facilitate the decision making and fasten the clinical interventions. The criteria are based on the probability of multiple haemorrhagic occurrences and the incidence region: lobar or deep [74]. Following these criteria, the diagnosis of probable CAA is made on elderly patients after the middle age, with at multiple acute or chronic lobar haemorrhagic lesions and with no other definite cause of ICH [13]. Also, Probable CA may be diagnosed with supporting clinical data and pathologic tissue (from biopsy or evacuated hematoma) when there is some CAA in the specimen [75].

The possible CAA diagnostic, similarly to Probable CAA, is attributed to elderly patients (older than 55 years old), with no other cause of haemorrhage. However, these patients show evidence of single lobar, cortical, or cortico-subcortical haemorrhage [75].

Finally, the Definite CAA is made post-mortem with evidence of lobar, cortical, or cortico-subcortical haemorrhage, absence of other diagnostic lesions, and severe CAA vasculopathy. The Boston criteria is supported by studies comparing the prevalence of the ϵ2 and ϵ4 ApoE alleles in groups of patients with clinically and pathologically diagnosed CAA [73], [74]. It has previously been validated against the well stablished and standard neuropathological diagnosis [76]. However, its sensitivity is strongly affected by the quality of the patients' scan [77]. The recent introduction of susceptibility weighted imaging (SWI), a three-dimensional T2*-GRE technique, Pittsburgh Compound Positron Emission Tomography (PET-PIB), Gradient - Recall Echo (GRE), used in the assistance of diagnostic, increases the sensibility of microbleed visualization, resulting in the improvement of lesion count accuracy [78]–[80]. The Boston Criteria can be consulted at Table 2.3 [76].

Table 2.3 - Classic and modified Boston criteria, stablished by the Boston Cerebral Amyloid Angiopathy Group, for diagnosis of CAA.

1. Definite CAA Full post-mortem examination demonstrating: • Lobar, cortical, or cortical-subcortical haemorrhage • Severe CAA with vasculopathy • Absence of another diagnostic lesion
2. Probable CAA with supporting pathology Clinical data and pathologic tissue demonstrating: • Lobar, cortical, or cortico-subcortical haemorrhage • Some degree of CAA in specimen • Absence of other diagnostic lesion
3. Probable CAA Clinical data and MRI or CT demonstrating: • Multiple haemorrhages restricted to lobar, cortical, or cortico-subcortical regions • Age ≥ 55 years • Absence of other cause of haemorrhage

4. Possible CAA Clinical data and MRI or CT demonstrating: • Single lobar, cortical, or cortico-subcortical haemorrhage • Age ≥ 55 years • Absence of other cause of haemorrhage
Other causes of intracerebral haemorrhage: Excessive warfarin with International normalisation ratio (INR) above 3, antecedent head trauma or ischemic stroke, haemorrhagic tumour, vascular malformation, or vasculitis and blood dyscrasia.

3 TECHNOLOGIES

The study of the application of Artificial Intelligence tools to aid in CAA diagnostic presupposes the usage of an appropriate dataset. However, there is a lack of open and annotated datasets regarding CAA. Part of this work focused on the acquisition of Magnetic Resonance images of patients previously diagnosed with this disease, and annotation by a neuroradiologist specialist. This chapter explores how the data is acquired and which technologies and procedures were involved in the process.

3.1 MEDICAL IMAGING

3.1.1 DIGITAL IMAGING AND COMMUNICATIONS IN MEDICINE

Digital imaging and communications in medicine (DICOM) is a globally accepted standard that had an early beginning in 1990 by the hands of the American College of Radiology (ACR) and the National Manufacturers Association (NEMA) [77]. Created to represent, store and exchange efficiently medical image data as well as associated metadata, DICOM is also a network communication protocol. The added value of DICOM data usability, access, and the continuous adoption of DICOM compliant equipment enables the reliable communication and exchange of data with other DICOM medical equipment, thus triggering the implementation of the Picture Archiving and Communication Systems (PACS) [78]. The DICOM standard specifies a set of protocols for network communications, for media communications, a file format and medical directory structure to facilitate access to the image data and metadata, and information for its implementation [79].

3.1.1.1 DICOM INFORMATION MODEL

DICOM was designed to mimic real-world scenarios in the healthcare world. The DICOM Information Model (DIM) describes the structure and organization of information related to the communication of medical imaging data. Every data and network operations are represented as a DICOM object [80]. This approach is made by setting the Information Object Definitions (IODs), that contains the necessary attributes that any real-life objects should have for its description, and the Information Entities (IEs), that specifies a collection of attributes typically present in an entity. For instance, IODs are specified for each modality (e.g., CT or MR) [79], [81]. The IOD is organized by data elements as attributes. Each attribute is identified

by a tag in the format (group.element), a value representation, value length and a data field, as represented in the Figure 3.1.

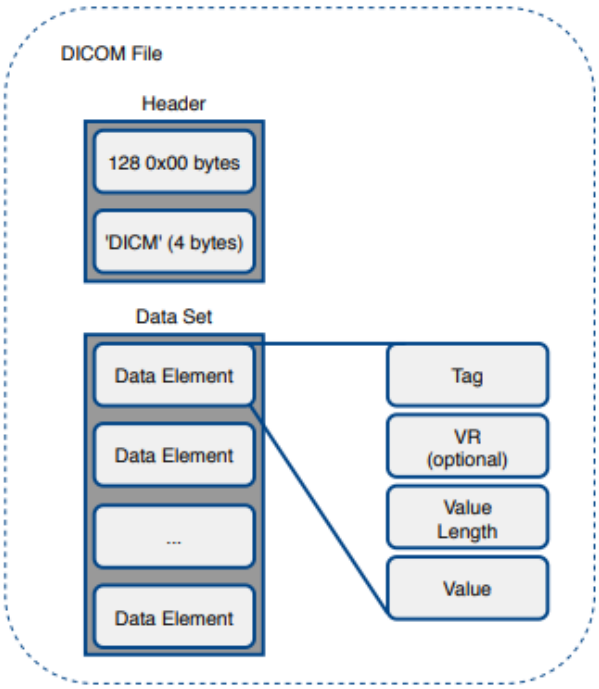


Figure 3.1 - Representation of a DICOM object.

Figure 3.2 represents the Patient-Study-Series-Image DICOM hierarchy. A patient can have multiple studies. Each study can have multiple series from different modalities and each modality can produce a different number of images [81].

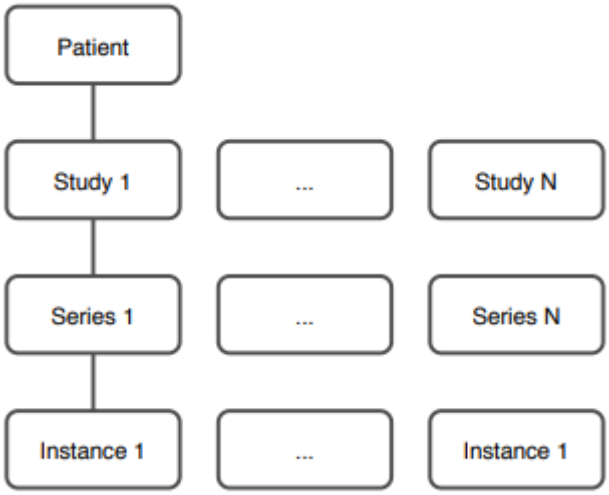


Figure 3.2 - DICOM information hierarchy: Patient-Study-Series-Image.

3.1.1.2 DICOM SERVICES

DICOM defines also the way the communication between nodes in the network is achieved. The most common services are the storage service and the query/retrieve service which can be used between the association of a Service Class Provider (SCP) and a SCU (Service Class User), similar to a server-client architecture. The implementation of DICOM services in a DICOM-compliant infrastructure is particularly important to store acquired images or query an archive for a specific patient, study, series or instance. Additionally, services are also crucial so that one may retrieve DICOM objects (for instance, request a DICOM object from the archive to be shown in a workstation viewer).

3.1.1.3 PICTURE ARCHIVING AND COMMUNICATION SYSTEM

The volume of medical imaging data generated in the healthcare institutions has increased substantially over the last years. The data and metadata from the different modalities acquired within the healthcare institution like, for instance, angiography (XA), ultrasound (US), multi-slice computed tomography (CT) or magnetic resonance (MR), need to be stored and distributed for diagnostic or treatment purposes. It is fundamental to create a deploy an efficient and robust system to deal with the storage and distribution of the high number of medical data.

The adoption of DICOM standard across the different equipment manufacturers and end-users lead to the implementation of the Picture Archiving and Communication Systems (PACS) [78]. PACS is the name given to describe a system that orchestrate the hardware and software that compose the imaging laboratory network. It is responsible for processing, storing and distributing medical images inside or outside the healthcare institution [79], [82]. The infrastructure is normally supported by a Local Area Network (LAN) or a Wide Area Network (WAN). The coming of PACS allowed the fast find and retrieve of patient data, improving the patient medical treatment as it removed the need to use film jackets and the high probability of losing studies. The workflow that the PACS handle may be divided in three major steps: acquisition, the procedures that lead to the capturing of the image; distribution, according to [83] the process of move or copy the images and their metadata to another one of the network (for instance, from a modality to a storage server); and visualization, the process where a workstation shows the image acquired.

3.1.2 Magnetic Resonance Imaging

MRI stands for Magnetic Resonance Imaging. It is the de facto technique for brain tumour diagnosis [84], [85]. The literature points several advantages in the use of the MRI techniques in the clinical environment, among them, the non-invasive characteristics and the adequate sampling quality, providing good tissue contrast without the use of unhealthy radiation [86]–[89]. Due to its benefits, MRI is being used for disease tracking down and treatment monitoring.

MRI scans are especially suited for soft tissues. Particularly in the brain, MRI scans output shows notable contrast between White Matter and Grey Matter, allowing to identify tumours, for instance [90].

Figure 3.3 depicts the main high-level components of a general MRI scanner. It is composed by a magnet that generates a strong electromagnetic field which goal is to polarize the sample; by shim coils, that are responsible for correcting shifts of magnetic field homogeneity; the gradient system, responsible for detecting the MR signal; and the Radio Frequency (RF) system, which goal is to excite the sample [91].

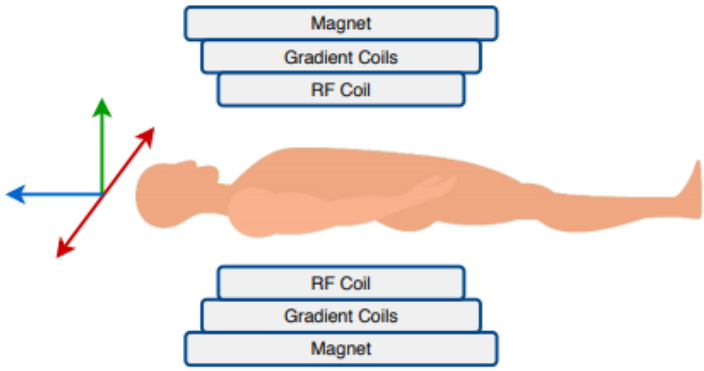
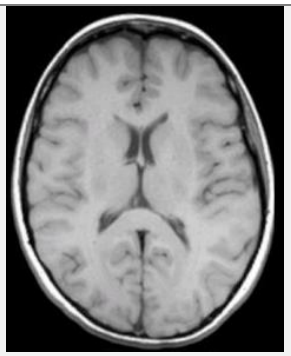
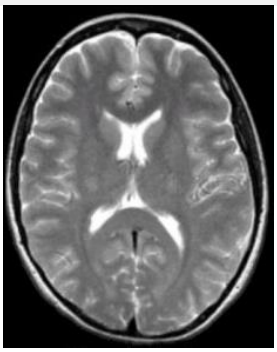
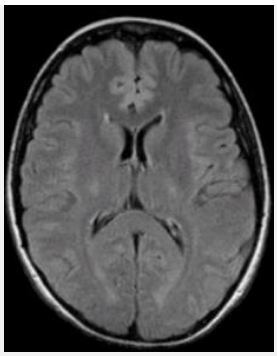


Figure 3.3- MRI scanner component.

MRI relies on the hydrogen atoms magnetic properties to produce diagnostic readable images. A hydrogen atom is essentially a proton that produces an electromagnetic field [92]. When an MRI scanner emanates energy to the human body, the proton spinning is excited, thus generating a magnetic field, and consequently shifting the protons from a random state to an aligned one. A radio frequency (RF) pulse is then applied. The application of the RF produces an alignment discomposure and the absorption of energy by the protons, allowing the shift to higher energy states. The reduction to a lower level stage is followed by the release of energy.

These disturbances are scanned, and the time read until the equilibrium state (relaxation), resulting in the image contrast [93]. The variation of the relaxation time is used to generate different types of MRI sequences. A longer TE, will reveal a large difference in signal from the tissues, resulting in T2 sequences, whereas short TE and TR time reveals T1 sequences [94]. Table 3.1 summarizes the most common MRI sequences.

Table 3.1 - Description of the T1, T2 and FLAIR MRI sequences.

SEQUENCE	MRI	PROPERTY		APPLICATION
T1		**CSF**	Dark	• Study of vascular structures • Study the breakdown in the BBB • Anatomical detail
		WM	Light	
		Cortex	Gray	
		Fat	Bright	
		Inflammation	Dark	
T2		**CSF**	Bright	• Visualization of lesions (appear bright) • Anatomical detail
		WM	Dark Gray	
		Cortex	Light Gray	
		Fat	Light	
		Inflammation	Bright	
FLAIR		**CSF**	Dark	• Similar to T2, except the CSF appear Dark • sensitive to pathologies
		WM	Dark Gray	
		Cortex	Light Gray	
		Fat	Light	
		Inflammation	Bright	

3.1.3 NEUROIMAGING INFORMATICS TECHNOLOGY INITIATIVE

The Neuroimaging Informatics Technology Initiative (NIFTI) [95] is a file format that surfaced in the early 2000s by the hands of a committee from the National Institutes of Health in the USA [95]. The primary purpose for its creation was to support a coordinated and targeted service, training, and research to speed the development and maximize the utility of informatics on neuroimaging, maintaining the advantages of the ANALYZE file format (another format for medical imaging), with several improvements [96].

Moreover, the primary focus of NIFTI are the tools used in imaging informatics, usually for neuroscience and neuroradiology research [95]. Although DICOM files are standard in clinical care environments, NIFTI format was adopted as the default format for many software packages, as SPM, and supported by several image viewers and data analysis software, like 3D slicer, ImageJ, and OsiriX [96].

Technically there are NIFTI-1 and NIfTI-2 file formats. The NIFTI-1 format is a direct upgrade of the ANALYZE 7.5, with the innovation of Image orientation, spatial location, spatio-temporal slice ordering for FMRI, frequency, phase and slice encoding axis, standardized way to add vector values to the header, and many more [96], [97].

The NIFTI-2 format is an update of the NIfTI-1 to allow the management of larger data sets. This new version maintains almost all characteristics of the previous one, enabling the storage of large images and matrices, encoded with 64-bit integer, eliminating the size limit restriction [95].

Finally, although the format allows the storage of the header (.hdr) and pixel data (.img) in separate files, like in the ANALYZE format, data can also be stored in a single file (.nii), containing both the header and image data and, if needed, additional metadata [98]. The mentioned MRI formats are illustrated and summarized in Figure 3.4.

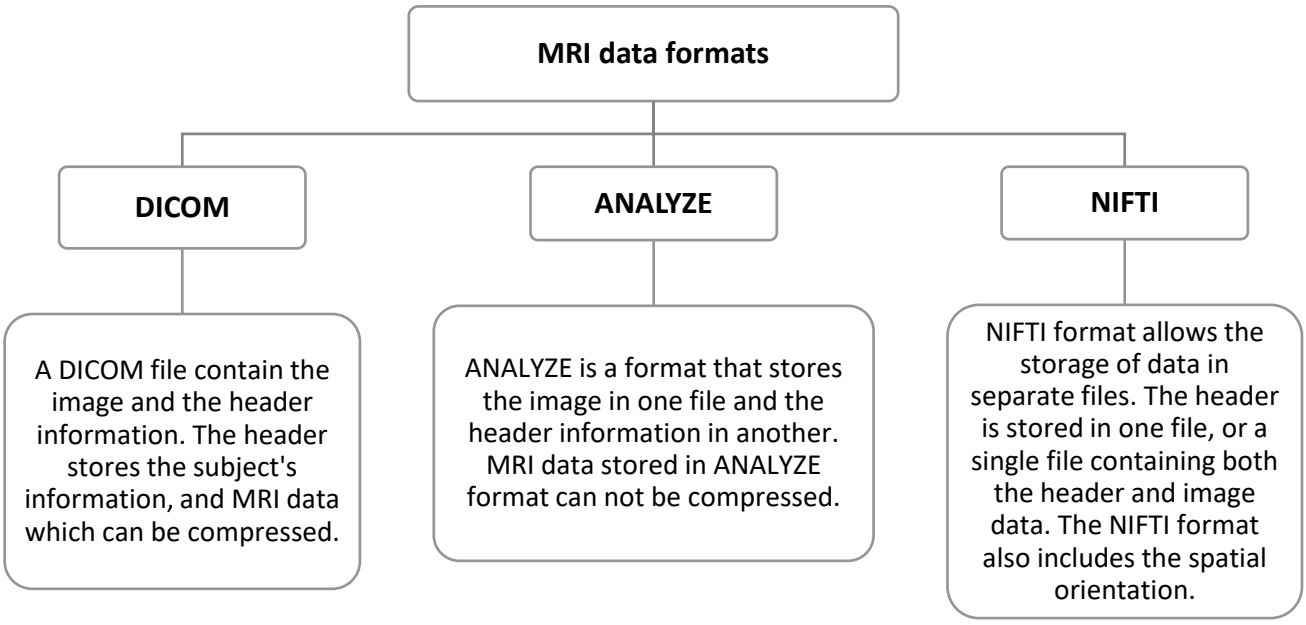


Figure 3.4 - Summary of MRI file formats characteristics.

3.1.4 MRI VISUAL RATING SCALES IN THE DIAGNOSIS OF DEMENTIA

The distinction and diagnosis of neurodegenerative diseases is crucial for the early access to support and targeted treatment for the affected individuals. As explained before, except for rare autosomal dominant forms of dementia, accurate and definitive conclusions during life can be challenging. Besides the difficulties in evaluating tissues for complementary diagnosis, the overlapping clinical symptoms for distinct pathologies is an obstacle to accurate disease identification. By contrast, structural neuroimaging is widely available and integrated in clinical and research environments. By focusing on brain regions of patients susceptible to dementia, visual ratings improve sensitivity and reliability of diagnosis-based image interpretation, and findings of value for differential diagnosis of dementia. Furthermore, since visual rating scales are quick to apply, they can be easily performed by trained physicians and introduced into clinical practice to extract valuable information.

Reports have shown positive impacts of brain imaging markers visible on MRI, that reflect small vessel injuries associated with CAA. However, its implementation in routine clinical assessments is still taking small steps, due to the requirement of special hardware, long processing time and specific acquisition techniques. Furthermore, the use of patients with several pathologies and risk factors, different MRI protocols, data from different centres,

different technological resources and Physicians expertise may introduce variations in the study [99].

Among these markers, medial temporal lobe atrophy, lobar CMB, cortical superficial siderosis (cSS), centrum semiovale PVS (CSO-PVS) and WMH of vascular are highlighted.

3.1.4.1 WHITE MATTER LESIONS

The occurrence of white matter abnormalities on MRI has been noted to increase with age. However, extensive damage is distinctive on patients suffering from presumed vascular dementia. Gouw et al. created a simple grading criterion for the development of WMH, with the same significance as a complex one that evaluates the WMH localization and WMH volume quantification [100]. The use of simple grading measurements reduces subjectivity and obviates the need of expert instruction and standard reference image [100], [101]. On the other side, the proposed scale is limiting as it does not incorporate periventricular WMH assessment. King et al. applied this visual metric to determine a global disease score based on the largest lesion identified, either deep or periventricular, for use as a baseline evaluation in clinical practice [101].

Later, DeCarli et al. evaluated a visual method for WMH grading using axial images and found classification of lesions as periventricular or deep to be inaccurate, as they are highly correlated with each other [102]. This claim supports the rating of deep and periventricular WMNH together [101], [102]. Brant-Zawadzki et al. used a rating scale to grade the severity of changes in the WM in patients with non-Alzheimer dementia and non-demented elderly patient groups. Results suggested a higher incidence of lesions in the demented group and cognitively normal elderly patients. However, given the small number of participants, these results may not entirely reflect the reality [103]. Zimmerman et al. created a 5-level rating scale for assessment of PVH, starting from no PVH (labelled 0) to diffuse WM abnormality-hyperintensity involving almost all WM (labelled 4). This study also evaluated the presence and number of focal areas of WMH independent of PVH [104].

To simplify the evaluation of WMH, Fazekas et al. suggested a modified rating scale, based on the ones proposed by Zimmerman et al. and Brant-Zawadzki et al., to describe the severity of hyperintense signal abnormalities in the periventricular and deep white matter regions. Each region is given a grade depending on the size and confluence of lesions. Periventricular

hyperintensity (PVH) was graded as 0 = indicating absent, 1 = “caps” or pencil-thin lining, 2 = indicating smooth halo, 3 = irregular PVH extending into the DWM. Deep white matter hyperintense signals (DWMH) were rated as 0 = absent, 1 = punctate, 2 = early confluence of foci, 3 = indicating large confluent areas [105]. The visual rating scales for the evaluation of Periventricular and White matter lesions proposed by Fazekas et al. is summarized at Figure 3.5.

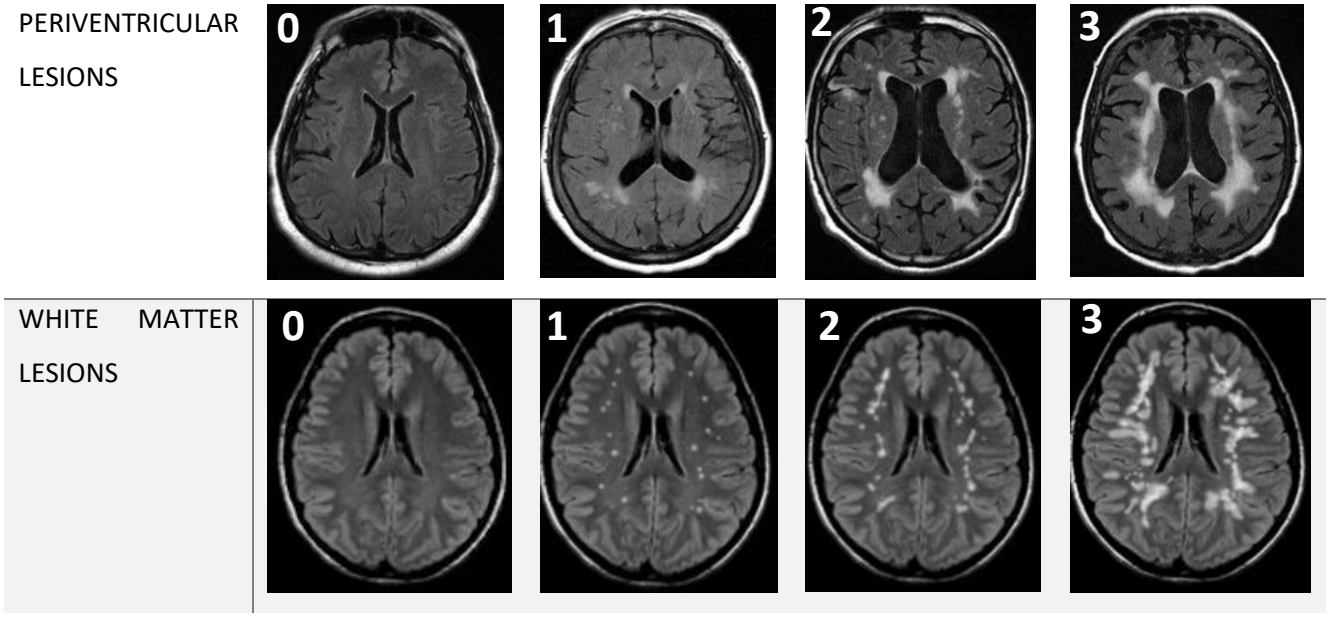


Figure 3.5 - Visual representation of the Fazekas et al. rating scale for Periventricular and White matter lesions.

Although the work first proposed the evaluation of the two brain regions, only the DWM score is useful in the assessment of the condition of patients with possible dementia. This rating scale is still one of the most widely used and well validated visual rating scales, regarding WM lesions, although some studies have reported limitations regarding definite conclusions about the association of WMH and cognition [99], [106].

3.1.4.2 Cerebral Atrophy

The occurrence of cerebral atrophy has been noted to increase with age in both non-demented and demented individuals.

Pasquier et al [107], proposed a scale to assess the global cortical atrophy (GCA) in patients with stroke, to test the hypothesis of a relationship between stroke and atrophy. To do so, 13 brain regions were evaluated separately in each hemisphere by several neurologists, including

the frontal, parieto-occipital, medial temporal lobe, and the dilatation of the ventricles. The final score corresponded to the sum of all 13 regions. Although the hypothesis was not validated, the scale was used as a valuable tool as a diagnostic marker toll for dementia and strongly associated with vascular damage [107], [108]. Figure 3.6 represents the visual rating scale proposed by Pasquier et al.

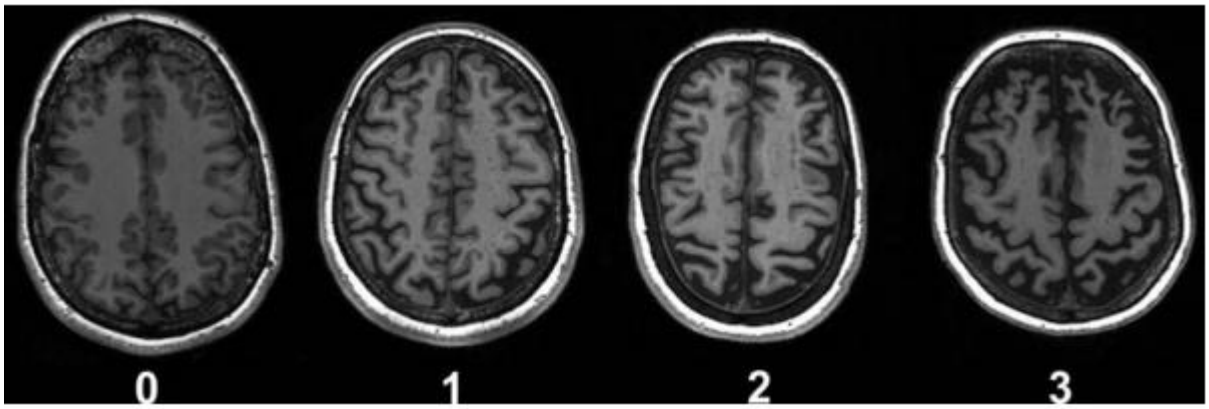


Figure 3.6 - Example of the scale proposed by Pasquier et al for the global cortical trophy.

O'Donovan et al [109] developed a rating scale to evaluate ventricular enlargement, in order to distinguish between AD and dementia with Lewy bodies. Once again, to determine lateral changes, each hemisphere was rated separately, and the scores summed up to get a global atrophy score. Using the ventricles to assess atrophy proved to be reliable, however, the scale may not be useful for differential diagnosis.

To understand the brain atrophy progress in patients with Frontotemporal Dementia (FTD), Davies et al [110] proposed a scale devised to be applied to the anterior temporal lobe and the lateral geniculate nucleus. Although sex, age, symptom duration and other clinical records were used in the study, atrophy proved to be significant in the prognosis prediction. Kipps et al [76] adapted the rating scale to include the rating of the posterior temporal lobe, and a larger group of patients diagnosed with FTD and control participants. Later, **Davies** et al included 15 frontotemporal brain regions. The scale is intended for the diagnosis and localisation of function in neurodegenerative diseases and other brain abnormalities. Insula, Anterior hippocampus, Orbitofrontal gyri and Temporal pole were reported to be relevant in discriminating AD from controls, and the rating of the orbitofrontal cortex later confirmed by Hornberger et al [111] as good discriminators between AD and bvFTD. While the results were

informative and provide resources for subsequent investigation, these ratings scales were not easily introduced in the clinical routine practice.

Medial temporal lobe visual rating scales were developed first by De Leon et al to rate hippocampal fissure dilatation, and later changed by Scheltens et al [112], which lead a positive impact on the field and inspired several other researchers, like Galton et al [113] and Urs [114]/Duara et al [115], who later developed scales to study the sulci surrounding the hippocampus, and who operationalized the rating scale to be suited for clinical practice. Figure 3.7 displays the visual rating scale proposed by Scheltens et al for the medial temporal atrophy.

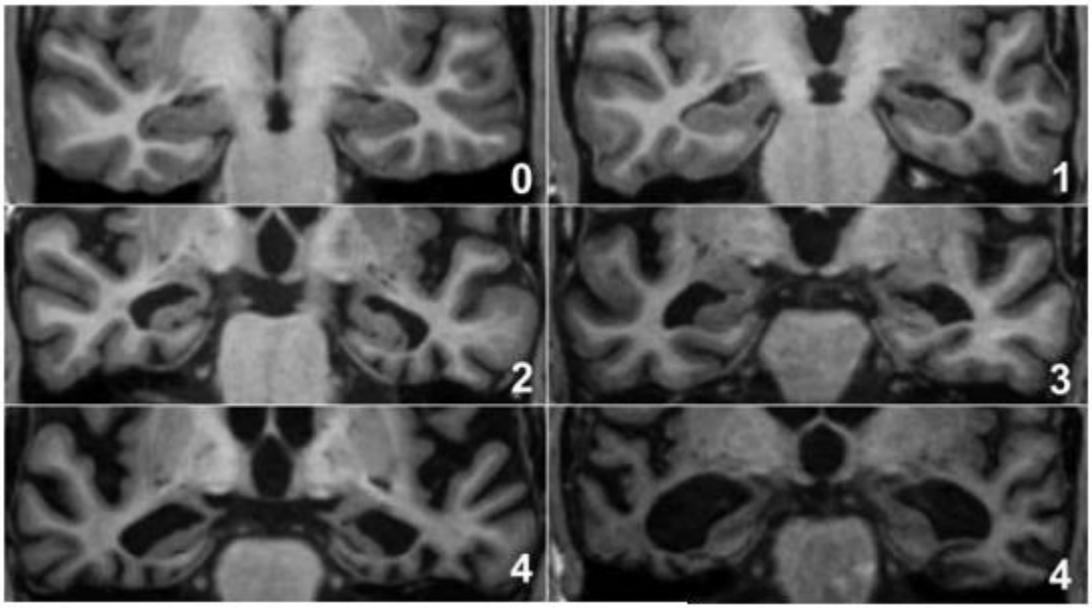


Figure 3.7 - Representation of the visual rating scale developed by Scheltens et al.

Regarding the Posterior lobe, Koedam et al [116] propose a scale to rate the atrophy on the posterior cingulate sulcus, precuneus, parieto-occipital, and the parietal lobe. Contrary to the above-mentioned scales, on this one a separate score for the hemispheres is given for each region and imaging plane.

At last, Harper et al [7] proposed a scale, based on the previous ones, and developed on post-mortem confirmed cases, for the visual assessment of atrophy in six different areas: medial temporal, posterior, anterior temporal, orbito-frontal, anterior cingulate and fronto-insula lobes. The visual rating scales applied to structural MRIs proved to be reliable and highly elucidative of the cerebral atrophy in regions vulnerable to dementia. The present scale has

also demonstrated great results and promising application in the automated classification of dementias to aid the clinical decision making. Figure 3.8 represents the visual rating scale proposed by Harper at al.

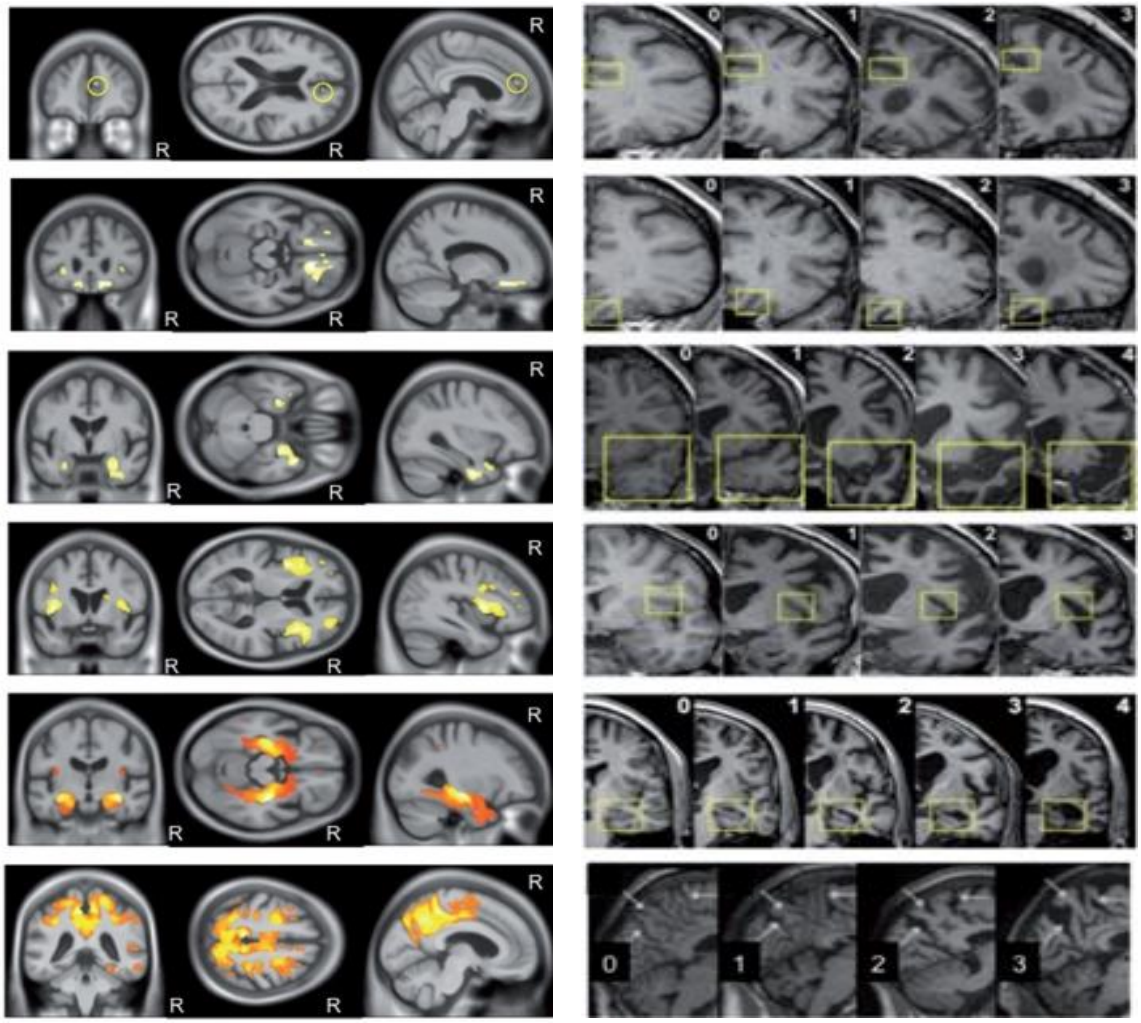


Figure 3.8 - Representation of the scale proposed by Harper et al.

3.1.4.3 CEREBRAL MICROBLEED

Cerebral microbleeds have been proven to be a useful biomarker for small vessel diseases, with potential relevance for diagnosis, prognosis, and the study of the disease mechanism. For this purpose, two major validated scales were proposed to describe their presence, number and distribution in the brain. The first one, Brain Observer Microbleed Scale (BOMBS) [117], considered the size of the lesions (<5 mm, 5–10 mm), the side of brain (left, right), and the location: lobar (cortex, subcortical white matter); deep (basal ganglia grey matter, internal and external capsules, thalamus); and the posterior fossa (brain stem, cerebellum). The BMBs

were measured manually by physicians blinded to the other one's ratings. Figure 3.9 represents the BOMBS rating form.

	Certain	Uncertain	Certain	Uncertain
► **Cortex / grey-white junction**[1]				
Number of BMBs <5mm	☐	☐	☐	☐
Number of BMBs 5-10mm	☐	☐	☐	☐
► **Subcortical white matter**[2]				
Number of BMBs <5mm	☐	☐	☐	☐
Number of BMBs 5-10mm	☐	☐	☐	☐
► **Basal ganglia grey matter**[3]				
Number of BMBs <5mm	☐	☐	☐	☐
Number of BMBs 5-10mm	☐	☐	☐	☐
► **Internal and external capsule**				
Number of BMBs <5mm	☐	☐	☐	☐
Number of BMBs 5-10mm	☐	☐	☐	☐
► **Thalamus**				
Number of BMBs <5mm	☐	☐	☐	☐
Number of BMBs 5-10mm	☐	☐	☐	☐
► **Brainstem**				
Number of BMBs <5mm	☐	☐	☐	☐
Number of BMBs 5-10mm	☐	☐	☐	☐
► **Cerebellum**				
Number of BMBs <5mm	☐	☐	☐	☐
Number of BMBs 5-10mm	☐	☐	☐	☐

Figure 3.9- The BOMBS rating form.

The second rating scale, the Microbleed Anatomical Rating Scale (MARS) [118] has the potential to help distinguish CAA from hypertensive small vessel disease. The lesions counting, by size, is done by the deep region (including the basal ganglia, thalamus, internal capsule, external capsule, corpus callosum, and deep and periventricular white matter), lobar region (including cortical and superficial subcortical white matter regions) and infratentorial regions (which includes the brainstem and cerebellum). The main difference between the two scales is that MARS allows the categorization of CMB distribution by different brain lobes. Furthermore, the fact that MARS considers the lobar anatomical description, is an important key to investigate the impact of microbleeds on cognitive functions and degenerative diseases and for the diagnosis of CAA. Figure 3.10 displays the MARS rating form, for the lobar, deep and infratentorial regions and the cerebral hemispheres.

		DEFINITE		POSSIBLE	
		R	L	R	L
Infratentorial TOTAL	Brainstem (B)				
	Cerebellum (C)				
	Basal ganglia (Bg)*				
	Thalamus (Th)				
Deep TOTAL	Internal capsule (Ic)				
	External capsule (Ec)				
	Corpus callosum (Cc)				
	Deep and periventricular WM (DPWM)				
	Frontal (F)				
	Parietal (P)				
Lobar** TOTAL	Temporal (T)				
	Occipital (O)				
	Insula (I)				
	TOTALS				

Figure 3.10 - The MARS rating form.

3.1.4.4 PERIVASCULAR SPACES DENSITY

Enlarged perivascular spaces (PVS) in the brain may reflect underlying cerebral small vessel disease, being, thus considered markers of the disease and the health of the brain. Several scales for the assessment of the enlargement of perivascular spaces were proposed (MacLullich et al. [119], Potter et al. [120], Zhu et al. [121]). However, these scales share the same limitations in the varying number of PVS on different slices, the poor scan quality and acquisition, asymmetry in PVS, presence of dilated PVS that could be mistaken for lacunes, and differences between the scale for severe cases. Doubal et al. [122], proposed a rating scale with the purpose of determining whether the enlargement of PVS were associated with lacunar strokes and WMH, associated with small vessel disease. To do so, enlarged perivascular spaces in the basal ganglia and centrum semiovale were rated according to Table

3.2. The scores were rated for each hemisphere to obtain a total EPVS score (0 to 8). Due to the easy understanding and application, this scale is still widely used and adapted.

Table 3.2 - Rating scale for the evaluation of the EPVS, by Doubal et al.

ENLARGED PERIVASCULAR SPACES

NO EPVS	0
< 10	1
11 - 20	2
21 - 40	3
> 40	4

Table 3.3 - Description of the study variables.

VARIABLES	DESCRIPTION
CEREBRAL ATROPHY	Evaluation of cerebral atrophy according to the scale proposed by Harper et al, for five brain regions: Anterior Cingulate cortex, Orbitofrontal Cortex, Anterior temporal lobe, Medial temporal lobe, and Posterior temporal lobe [108].
WHITE MATTER LESIONS	Fazekas et al., in 1987, proposed a scale to describe the widening of perivascular spaces, resulting from injuries to the white matter and periventricular regions, evidence of cerebral vascular disease. The used scale is divided into 4 levels, according to the MRI or CT hyperintense signal abnormality level in white matter [123]: • Fazekas 0: Absent; • Fazekas 1: Punctate lesions; • Fazekas 2: Early confluence of foci; • Fazekas 3: Large confluent areas.
MICROHAEMORRHAGES	Quantification of microhaemorrhages is made from the evaluation of haemorrhagic patterns detected in t2W MRI or FLAIR. Microhaemorrhage distribution may be divided by brain region occurrence as lobar (frontal, parietal, occipital cortex and insulate), mixed (lobar and deep), or deep (cerebellum, corpus callosum, basal ganglia, white matter perivascular, etc).
PERIVASCULAR SPACES DENSITY	Metrics for the measurement of perivascular space's density allows the estimation of the enlargement of perivascular spaces in patients with confirmed vascular pathologies. The density (PVS) may vary between four levels: 0, ≤ 10, 11-20, and> 40.

3.2 RADIOMICS

The development of medical imaging in medical sciences and its introduction on the clinical practice paved the way to its generalized use in decision making, thus aiding the diagnosis, prognosis, and the research for therapeutic interventions.

Radiomics introduces a way to mine and transform qualitative and quantitative information encrypted in digital images. It relies on mathematical algorithms to uncover image characteristic, details, and lesions on medical images that may be neglected by humans.

The process of feature extraction starts with image segmentation to reduce the image to the desired portion for the study. After this step, a great amount of features are extracted, such as size and shape related features, image descriptors obtained from intensity histograms and the relationship between voxels, to texture related features, obtained by analysing the grey level variation within the defined region [124].

Given the wide variety of features extracted, studies based on Radiomic features present a promising approach to aid the recognition of visual patterns or lesions on medical data.

Using a radiomic approach, Johansen et al, Baek et al. and Peng et al. showed that histogram-based features are useful to predict treatment responses and outcomes [125]–[127]. Also, Tixier et al. and Yang et al. demonstrated the potential of radiomic textural features on the distinction of tumour phenotypes, and even outperform the traditional approaches on the prediction of the treatment outcome [128], [129]. Furthermore, radiomic texture features have drawn attention due to its value when predicting the pathologic response, thus improving the patients' survival rate, identifying cases with higher risk of developing metastases, and distinguish different disease stages [124].

Although the first radiomic studies were mainly focused on oncological researches and prediction of treatment response, recently it has been extended to other medical fields, such as the study of neurological disorders. Shinde et al. applied radiomics to data obtained from patients suffering from PD, achieving a good performance for the prognosis prediction [130]. Moreover, Salvatore et al. reported outstanding results when combining non-imaging and radiomics variables to predict the outcome of neurodegenerative diseases, suggesting that radiomics analysis can be used as an alternative approach to the visual AI classification or prediction methodologies [131].

Despite its potential application, the extracted features may be sensible to variations on the MRI intensities, as this is not standardized and highly dependent on the sequence type, acquisition parameters, and even on the manufacturer. Consequently, a high variation on image intensities from patient to patient is expected, thus requiring a standardization step before further analysis [132]. Additionally, to achieve accurate results, a dataset of ten to fifteen patients per disease is recommended [124].

3.3 CLASSIFICATION USING ML APPROACHES

In 1959 Arthur Samuel created the first self-learning program, built to play checkers by calculating the scoring at any position, measuring the chance of winning, and eventually become better at it. This was a pioneer AI program, introducing a new way to handle technology and simulate the human capacity of analysing the information, predicting possible outcomes, learn from experience, and change the learning process throughout practice [133]. Besides being a subset of AI, ML can be defined as a supervised or unsupervised science. The Supervised learning relies on human labelled data. Given a training set (which usually corresponds to 70-80% of the original data), the machine successively utilizes each entry to train the model. The results obtained from this process are used by the machine to make accurate predictions and improve its performance, and, if needed, repeat the process to return the best output. In the unsupervised learning occurs when the machine is fed a dataset with no predefined classes. The data fed into the model is grouped together by a similarity metric, forming clusters. From this approach, it is possible to get some insight on the data and reduce dimensionality.

In Classification models, a training and test set is defined. The algorithms applied take the classes and the features that describe each of them to learn and distinguish the attributes and behaviour of the data. In the testing phase, the knowledge obtained from the data on the previous step is used to classify the new feature vectors [134].

The goal of ML methods is to automate learning and therefore classify, and eventually be used to simulate the behaviour of the data. It has been used for the detection of facial features, speech recognition, personalize advertisements, and even to facilitate the medical decision making. When it comes to the medical field, ML is helpful on the prognosis, diagnosis, search for treatments, and the evaluation of the evolution of a certain disease or recovery process,

presenting high accuracy and reliability. Although the use of ML in the health environment is still in its infancy, results suggest that ML models may eventually surpass the human decision-making abilities, and facilitate precise and early detection of diseases, enabling the improvement of the patients' life quality [135].

Although Neurodegenerative diseases have been intensely studied in the late years, numerous questions are yet to be answered. In this context, ML provides tools to efficiently study and understand the disease mechanisms, and eventually change the way the prognosis and diagnosis is conducted, by identifying abnormalities on images or clustering diseases and biomarkers. In search for effective therapeutic interventions, large collections of data have been gathered and stored in curated databases to supply studies and help expand the intervention of computer-aided approaches [134].

4 MODELLING OF NEURODEGENERATIVE DISEASES

4.1 MATERIALS

As the goal of the project is to study the biomarkers for the complementary diagnosis of CAA, MRI images and non-image clinical information were used to assess the optimal conditions and data to maximize the score of the classification models and, thus, determine the most important features to provide information for an accurate diagnosis. The image and non-image data acquisition steps and materials to pursue the objectives are described below.
Figure 4.1 represents the steps followed on the development of the project.

4.1.1 DATA ACQUISITION

The data acquisition relied on the collaboration with the neuroradiology service of Hospital de Braga, EPE, after the study approval by the Health Ethics Committee, and all methods were adjusted in accordance with the guidelines of the ethics board. Hospital de Braga adopted a Vendor Neutral Archive and DICOM compliant PACS, Sectra IDS7. The acquisition workflow was based on the usage of the software to query and retrieve the stored DICOM objects.
Patients admitted between 2014 and 2017 with non-traumatic lesions, who underwent brain MRI, were sorted and diagnosed with CAA. The goal was to retrieve the last study MRI images of each diagnosed patient. The outcome of the filtering resulted in the selection of 138 patients.
Additional data, with patients from the same hospital, was used in the study to enrich the dataset. These data contained patients with 3 diagnosed neurodegenerative diseases different from CAA: Alzheimer's disease, with 90 patients, Mild cognitive impairment, with 56 patients, and Parkinson's disease dementia, with 19 patients. The acquired information was stored by disease in four Comma Separated Values (CSV) files.
Using the software capabilities, the studies were anonymised so the patients could not be identified by analysing the metadata contained in the DICOM objects. Finally, a Patient Name alias, Patient Age, and Patient gender were kept in the DICOM metadata so the retrieved DICOM objects could be identified in the annotation process.

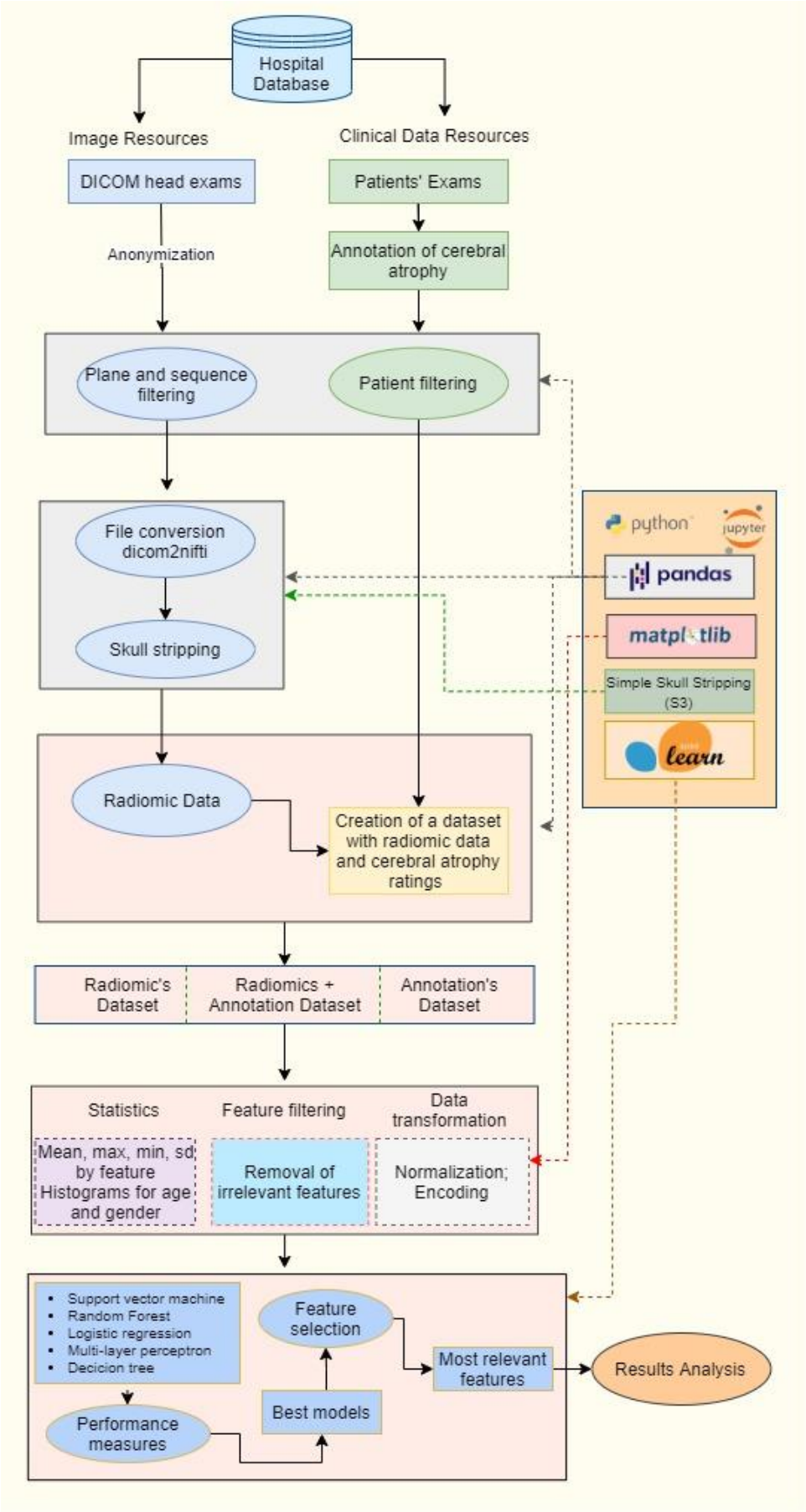
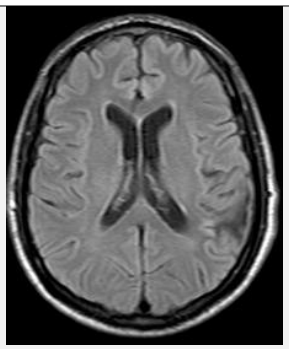
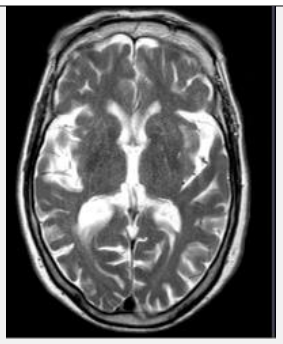
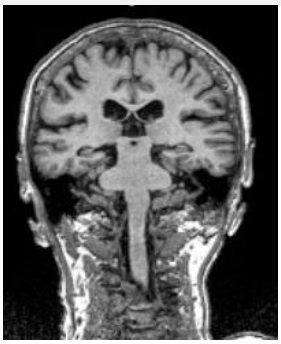
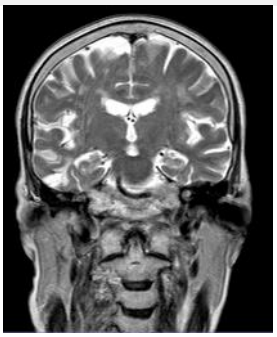
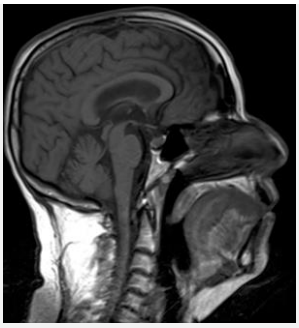
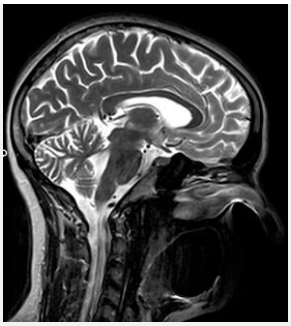


Figure 4.1- Overview of the project's architecture.

4.1.2 NEURODEGENERATIVE IMAGE DATA

Some examples of images contained in the dataset are presented in Table 4.1. The images had different resolutions and were also not spatially aligned. Several magnetic pulse sequences for enhancing brain lesions were applied, to create contrast between normal and pathologic structures. In particular, T1, T2-, T2*-weighted and FLAIR, along with different planes (axial, coronal, sagittal) are available. However, not all patients have the same acquisition sequences and number of MRI data.

Table 4.1 - Example of the data-set images and Sequences/ Orientations.

SEQUENCE/ ORIENTATION	***MR SAMPLE***	**SEQUENCE/ ORIENTATION**	***MR SAMPLE***
AXIAL FLAIR		Axial T2	
CORONAL T1 MPRAGE		Coronal T2	
SAGITTAL T1		Sagittal T2	

Contrast images are obtained by exploiting the differences between tissues, due to the different behaviour within the magnetic field. Parameters can be manipulated to generate the signal to maximize the tissue contrast. T1 weighted sequences enhance fat matter (hyperintense), and fluids have a low signal intensity, appearing dark. Also, in the brain, grey matter appears with an intermediate signal intensity (grey), and white matter is hyperintense (white). With T2 weighted sequences, fluids (e.g. water, CSF) and fat matter have high signal intensity, appearing bright, and muscle is seen grey (intermediate signal intensity). In the brain, grey matter appears grey (intermediate signal intensity), and white matter appears darker (hypointense signal intensity).

4.2 METHODS

The following subsections describe the steps to prepare a new dataset with annotations based on the clinical data, and the classification of neurodegenerative diseases. Also, a similar classification approach was conducted with the acquired MRI images. Jupyter Notebooks with Pandas library were used to carry this project.

A description of dataset cleaning, format conversion, image segmentation, and feature generation are also outlined.

4.2.1 IMAGE-BASED BIOMARKER EXPLORATION

The purpose of this experiment is the evaluation of the application of several imaging-based scores to complement the diagnosis of CAA. It focuses on the retrieval of clinical data from the available patient's information. This step depends on the analysis of data by trained physicians, followed by data processing, and data modelling by AI tools.

4.2.1.1 DATA ANNOTATION

The resulted datasets described in Section 4.1.1 were annotated. Under the collaboration described in Section 1, the MRI data was independently analysed by two neuroradiologist physicians, blinded to the patient's identity. The specialists reviewed a set of high-level characteristics to be annotated, and rated or counted, depending on the location, the possible

biomarkers. Whenever disagreements about values were significant, the ratings were reviewed, and a consensus was reached. Furthermore, additional clinical variables were collected (e.g. occurrence of death and existence of cerebral tumours) to better understand the data. Among the original retrieved dataset, series with low image quality were discarded due to the impossibility of getting a rigorous evaluation. Not all datasets were annotated with the same metrics. In other words, some datasets, namely AD, PDD, and MCI are missing annotations compared to CAA.

Due to some constraints, the annotation of data for the AD, MCI and PDD diseases according to the rating scales used to study the White matter lesions, Cerebral microbleed, and Perivascular spaces density was not possible. Consequently, the dataset for the CAA disease was the only one annotated with the four ratings scales. Therefore, only the Cerebral atrophy annotations were kept on the study.

4.2.1.2 Data Preprocessing

After the data acquisition and annotation, the datasets were loaded into a Jupyter notebook environment to ease the visualization, data processing, and future feature engineering. Using the pandas library, the datasets were concatenated keeping all the axis and columns. For this experiment a Dataset of non-image was created, containing only the clinical information and metric annotations. The resulting DataFrame contained a set of unfiltered and untreated patient information. Among the original list of patients diagnosed with one of the dementias, subjects with corrupted files or unavailable data were removed from the experiment. Moreover, patients with extreme cerebral lesions, malformations and tumours were also excluded from the dataset, as the MRI images were in most cases impossible to classify. Finally, a disease label, (CAA, AD, MCI, and PDD) was attached to each patient entry. The final dataset contains the ratings for 252 total diagnosed patients (98 with CAA, 84 with AD, 53 with MCI, and 17 with PDD), age, sex, label, and an anonymized ID.

To proceed to the analysis of the data, a filtering by MRI sequence and imaging plane was conducted. Since the visual biomarkers, as microhaemorrhages are seen in T2 sequences, these were the sequences selected for each patient. Furthermore, as the visual rating was performed on the coronal MRI planes, only patients with available studies containing the desired planes were kept on the study.

The last step of the filtering was the removal of the remaining patients' ID, keeping unique entries. Moreover, only the annotations available for all the patients were kept. The characteristics of patients by Neurodegenerative disease is summarized in Table 4.2.

After the filtering, the dataset must be prepared to build the classification models. To do so, encoding of the categorical information into numeric values, such as gender. This way, Male gender corresponds to the value "0" and Female gender corresponds to "1".

Table 4.2 - Patients' attributes according to Neurodegenerative disease.

	CAA	**AD**	**MCI**	**PDD**
AGE	66.09 (± 13.9) [31 – 89]	72.48 (± 10.13) [33 – 89]	73.08 (± 9.30) [51– 88]	68.29 (± 8.50) [58 – 82]
SEX	Male: 63 (61.17%) Female: 40 (38.83%)	Male: 37 (44.05%) Female: 47 (55.95%)	Male: 31 (58.49%) Female: 22 (41.51%)	Male: 7 (41.18%) Female: 10 (58.82%)

4.2.1.3 HOLDOUT

The prepared Dataset is then split into train, a subset which will be used to train a model, and test, a subset used to test the model accuracy. The data was randomly split by a function available in the Scikit-learn python package at the 80/20 (%) ratio. Figure 4.2 represents the cases per disease in the train and test subsets.

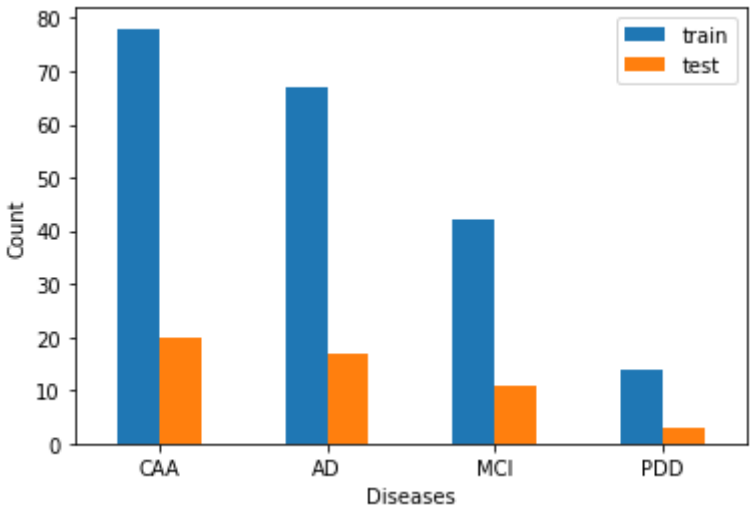


Figure 4.2 - Representation of the distribution of the data by the train and test subsets.

4.2.1.4 CREATING AND TRAINING MODELS

Classification models using Machine learning algorithms were created using the annotation DataFrame generated above. Choosing the most suitable machine learning approach is a defining step for the success of the work, hence the need to weight the advantages and disadvantages of each method. Multiple approaches were tested, the results analysed, and adjustments made to optimize the algorithms' parameters and data features. The following ML methods were considered suitable for the current task:

- Decision Trees: The main advantage of Decision Trees is the easy to interpret outcomes, fast response, and robustness when it comes to outliers and missing values. In contrast to that, Decision trees are prone to overfitting, especially on deep trees. The chosen algorithms were the Random Forest and Decision Trees, due to its ability to limit overfitting without substantially increasing error.
- Support Vector Machines: The advantage of this classification-oriented ML is the capacity to map data to a higher dimensional space, thus enabling an easier label separation. However, it requires an understanding of the kernels, and the learning and testing process, which can substantially affect the classification. Also, Support Vector Machines does not perform well on large datasets due to the longer training time.
- Artificial Neural Networks (ANN): The ability to perceive nonlinear relationships between attributes and interactions between variables is a great advantage of ANN, resulting in higher prediction accuracies. On the other hand, the computation times and tendency to overfitting are the main disadvantages. The implemented ANN algorithm was the Multilayer perceptron, a supervised learning algorithm capable of learning non-linear models, although sensitive to feature scaling hyperparameter tuning.
- Logistic Regression: Logistic Regression usually is quick and easy to implement and does well with unscaled data. Also, it works better on data after feature engineering. The disadvantage is its weakness when compared to other classification algorithms and the vulnerability to overfitting.

Following the algorithm selection, hyper-parameter tuning is an essential step to attain an accurate classification of the diseases. To optimize the hyper-parameters, Grid-search was employed, an optimization process that exhaustively searches over the optimal parameter

values from a set of specified hyperparameter for the chosen model algorithm, fitted on the train subset.

This provides the best set of hyper-parameters for a given model and dataset, ensuring the optimization of the project, with a high computational cost, depending on the data and number of parameters.

4.2.2 Classification of diseases based on Radiomic features

The second experiment consists of the classification of Neurodegenerative diseases, based on MRI images. With this approach, the aim is to test the accuracy of an AI classification in comparison to the non-image clinical data approach. To meet this goal, two major steps were taken: data preparation and patient outcome prediction.

Since the number of images is not sufficient to make reliable conclusions using Deep Learning (DL) approaches, an alternative had to be found. Thus, low level radiomic features were generated. As both morphological and functional clinical images contain valuable mineable data, advances in medicine and technology suggest the association of such features with underlying pathology of a tissue.

The extracted quantitative features may be characterized by one of the following subgroups [136]:

- Shape features: describes the 3D geometric characteristics of a defined region of interest, like volume or maximum diameter along different orthogonal directions, for instance.
- First-order features: describes the distribution of voxel intensities, within the region of interest, through statistic metrics based on histograms. Examples of these metrics are the mean, minimum value of the voxel intensity on the image, uniformity, and entropy.
- Second-order features: include descriptors of the shape and texture, obtained by statistical analysis of neighbour voxels. It provides a measure of the spatial arrangement of the voxel intensities, giving insight about the tissue heterogeneity.
- Higher-order features: describe repetitive or non-repetitive patterns and highlight details on the region of interest, by applying mathematic transforms of the images.

The quality of the extracted features depends on the image quality, processing, and segmentation, which will also reflect on the accuracy of the modelling outcome and its depiction of the original data.

4.2.2.1 Pre-processing

The obtained MRI data was unfiltered. It had several MRI sequences acquisitions, and the available data per patient was not consistent. Also, some images were distorted, had different contrast, intensity, and image noise. To minimize the effect of these artifacts, the following measures were taken: first, a manual selection of images and sequences was performed. Patients who did not undergo the sequences mentioned in Section 4.1.2, whose images were not available, whose images had low quality for analysis, or were previously rejected by the physicians, were excluded; Second, similarly to the procedure applied in the previous study, for subjects that had several sequences, only the T2 sequences were keep. Patients, whose images of the mentioned sequences were not available or were removed from the study. Examples of the removed data or patients are shown on Figure 4.3: (A) and (E) represent images with extreme noise and impossible to visualize any microbleeds or other visual MRI Biomarkers. These images would also difficult the skull stripping process. Images (B) and (C) belong to patients with deep haemorrhages on the left side, making it impossible to evaluate the lateral brain atrophy. Finally, (D) and (F) represent images with destructive lesions in the Frontoparietal and Temporal lobes, impossible to evaluate and classify according to the visual rating scales.

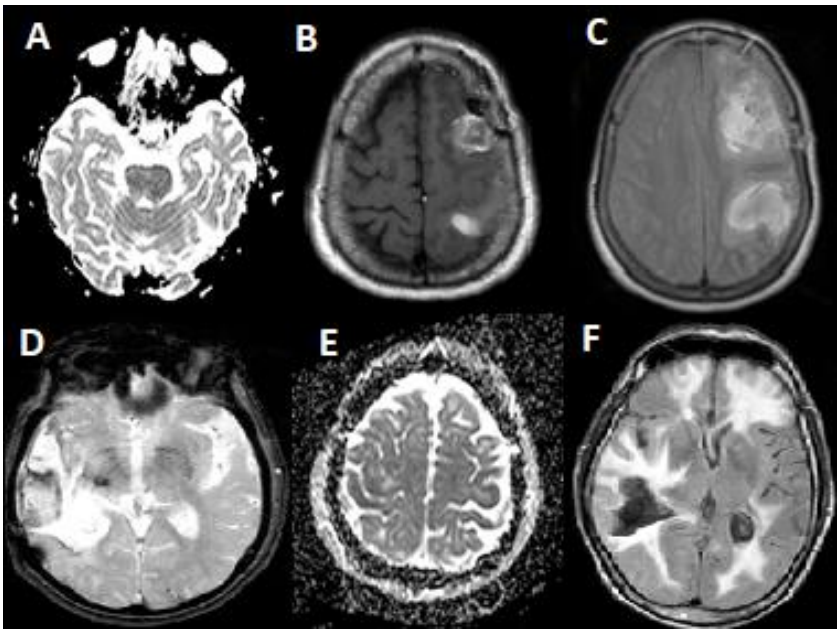


Figure 4.3 - Example of brain MRI images excluded from the study.

4.2.2.2 IMAGE PROCESSING

The first step for the neuroimaging analysis was the conversion of data on the DICOM format to NIFTI. The NIFTI format provides a simple way to handle the data and simplifies the next processing steps, since a great portion of the tools used for image processing require inputs in the NIFTI format. The chosen tool to transform the data was the dicom2nifti [1] converter. It sorts out the slices within each series into 3D structures.

Thereafter, an image segmentation process was conducted, to extract the non-brain and soft tissues (e.g. skull, fat). To remove the skull, S3 [137], a free software toll was used, based on Atlas registration, an image processing technique used to align multiple images into one, and detect morphological differences between images.

The basic procedure comprises the application of a series of geometric transformations for feature alignment and detection, quantification of similarity between the images, for estimation of a registration function, and the optimization algorithm. Often used to overcome issues such as image rotation, scale, and skew that are common when overlaying images. S3 [137] takes the patient head scan as input (Figure 4.4- A), and returns a skull stripped scan, a binary mask (Figure 4.4 -B), and a probabilistic segmentation of GM, WM and CSF (Figure 4.4

[1] https://pypi.org/project/dicom2nifti/

-C). Next, the processed images were checked, to assess the skull stripping quality. Examples of processed images, and the outputs are presented in Figure 4.4.

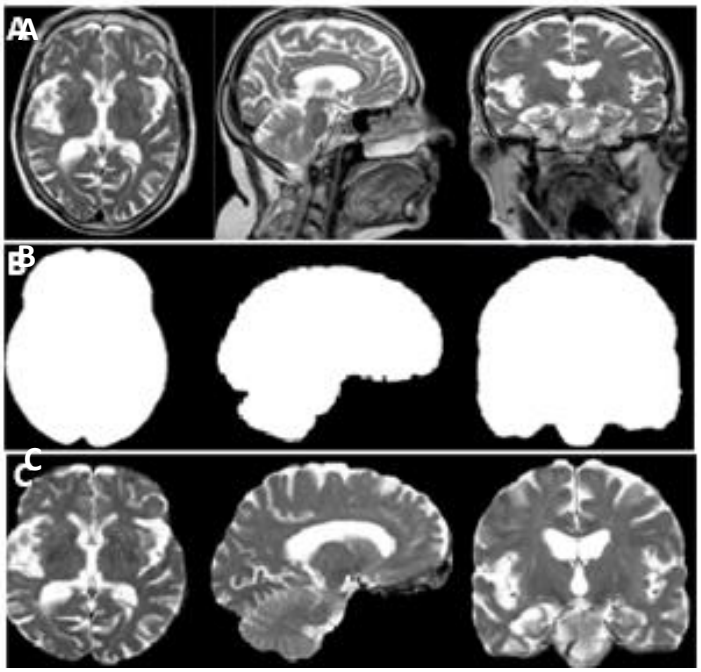


Figure 4.4 - Example of skull stripping on MRI image

4.2.2.3 LOW-LEVEL FEATURE EXTRACTION

To quantify the radiomic characteristics on regions of interest on medical images, an opensource tool was used, the PyRadiomics [2]. The tool carries on the four following steps to extract features from images:

1. Loading of medical images and correspondent mask of the region of interest. The handling and processing (filtering operations, image segmentation and registration) of images is done using SimpleITK[3];
2. Image filtering applying selected filters. The available ones include wavelet and Laplacian of Gaussian filters, implemented using PyWavelets [4]and SimpleITK, as well as several simple ones, including square, square root, logarithm, and exponential

[2] https://pyradiomics.readthedocs.io/
[3] https://simpleitk.org/
[4] https://pywavelets.readthedocs.io/

filters, implemented with NumPy[5]. The desired filters and parameters were specified on a file. Besides the filters, the normalization option was chosen;

3. Feature generation using five different feature classes: first-order statistics, shape descriptors, texture classes with grey level co-occurrence matrix, with grey level run length matrix, and grey level size zone matrix. Feature extraction is supported for both 2D and 3D segmentations;

The resulting features are stored in a csv file, organized by the anonymized patient ID associated with the original image. Besides the calculated features, the file also contains additional information, including the version, applied filters, settings, and image spacing. These entries were later removed before the data classification process, as long as other features with zero variance. The process of Radiomic Data acquisition is portrayed in Figure 4.5, starting with the acquisition of MRI images, followed by the definition of the ROI, and ending with the feature extraction.

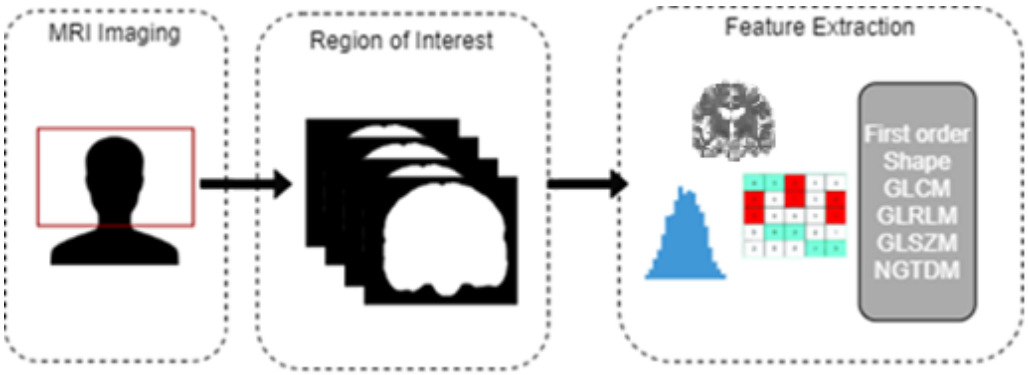


Figure 4.5- Overview of the pipeline of Radiomic feature .

4.2.2.4 HOLDOUT

To start the AI study, the generated csv file is converted into a DataFrame and split into train and test as explained before in 4.2.1.3.

[5] https://numpy.org/

4.2.2.5 Creating and Training Models

The chosen models, hyper-parameter adjustment and training procedures follow the guidelines described above at 4.2.1.4. The difference is in the input dataset. Figure 4.6 is an example of the application of Grid search for the generation of the best set of hyper-parameters, model fitting, and application of the Logistic Regression algorithm.

```
param_LR_radiomics={'penalty' : ['l1', 'l2', 'elasticnet', 'None'],
    'C' : np.logspace(-4, 4,10, 20),
    'solver' : ['newton-cg', 'lbfgs', 'sag'],
    'multi_class':['ovr','multinomial']}

grid_LR = GridSearchCV(LogisticRegression(), param_grid=param_LR_radiomics,scoring='accuracy',cv=5,n_jobs=-1)
best_grid_LR = grid_LR.fit(X_train, y_train)
```

```
print('Best score:', best_grid_LR.best_score_)
print('Best parameters: ',best_grid_LR.best_params_)
```

```
Best score: 0.49743589743589745
Best parameters:  {'C': 0.046415888336127774, 'multi_class': 'ovr', 'penalty': 'l2', 'solver': 'newton-cg'}
```

```
LR = LogisticRegression(C=0.046415888336127774, solver='newton-cg', multi_class='ovr',penalty= 'l2').fit(X_train, y_train)
y_pred_LR=LR.predict(X_test)
```

Figure 4.6 - Result of fitting parameters and Model training

4.2.3 Classification of diseases based on Radiomic features and Annotations

The third and final experiment approaches the combination of medical images data (in the form of radiomics) and non-image data. The purpose of this experiment is to determine whether the combination of both datasets brings any advantage to aid the diagnosis of CAA, and what biomarkers would be used to differentiate the disease in hands from the rest.

4.2.3.1 Dataset preparation

In this step, a combination dataset containing the clinical and annotation data and the radiomics data is created. For instance, the “clinical_annotations” and “radiomics_features” datasets are loaded to the Jupyther Notebook. To proceed the merging of the datasets, first the correspondence between the ID’s from each dataset must be done. The pandas package allows the creation of a new Dataframe with both data, keeping all the different columns from the original datasets. A summary of assembling of the combined data is represented at Figure 4.7.

Figure 4.7- Summary of the development of the combined Dataset

4.2.3.2 HOLDOUT

Once again, the train and test splitting of the data for further analysis is done as described in 4.2.1.3. The train subset was used to fit the grid-search and fed to a training model. The test sub-set was used to assess the model classification's accuracy.

4.2.3.3 CREATING AND TRAINING MODELS

As the purpose of this experiment is to compare the outcome of the classification model of the combined Dataset and the original ones, the chosen algorithms, hyper-parameters passed through and training procedures follow the ones described at 4.2.1.4. Figure 4.8 demonstrates an example of the returned dictionary for the best hyper-parameters for the Logistic Regression model and the model training.

```
param_complete_LG={'penalty' : ['l2'],
    'C' : np.logspace(-4, 4,10, 20),
    'solver' : ['newton-cg', 'lbfgs', 'sag'],
    'multi_class':['ovr','multinomial']}

grid_LG = GridSearchCV(LogisticRegression(), param_grid=param_complete_LG,scoring='accuracy',cv=5,n_jobs=-1)

best_grid_LG = grid_LG.fit(X_train, y_train)
```

```
print('Best score:', best_grid_LG.best_score_)
print('Best parameters: ',best_grid_LG.best_params_)
```

```
Best score: 0.5008403361344538
Best parameters:  {'C': 21.54434690031882, 'multi_class': 'ovr', 'penalty': 'l2', 'solver': 'newton-cg'}
```

```
LR = LogisticRegression(C=21.54434690031882,multi_class= 'ovr', penalty= 'l2',
                        solver= 'newton-cg').fit(X_train, y_train)
y_pred_LG=LR.predict(X_test)
LR_score=LR.score(X_test,y_test)
```

Figure 4.8 - Example of Logistic Regression model training

4.2.4 Feature Engineering

Feature engineering is the process of extraction relevant features from the raw Dataset, an important step for preparing the input dataset and improve the performance of ML models. Without relevant features, accurate predictions or classifications are hard to achieve. Initial efforts should be focused on the identification of appropriate pathways to pick features with the most potential for the purpose in hands. The most important step to take is the dimensionality reduction and feature selection into a much smaller set that can be modelled, with less redundant entries and noise information. For tabular data, Principal Component Analysis and unsupervised clustering methods are usually used. For image data, this might include pattern detection or segmentation.

A high correlation between variables means that they may have a similar behaviour and affect the data in the same way, which worsens the model performance. Therefore, for this study, redundant Radiomic features were removed to improve the model performance. Correlation matrices were created to visualize the relationship between the features, and a threshold was determined. When the correlation coefficients were higher than the threshold, only one of the features was kept, resulting in a much smaller set of variables. For instance, 0.99 and 0.90. were the values chosen to start the experiment. No further values where tested as a threshold from 0.90 to 0.6 doesn't seem to have a big influence on the accuracy. After this step, histograms were also plotted to compare the data distribution by variable for the four

diseases. The variables which presented a similar distribution for the four classes did not contribute to the patient's disease labelling, being thus removed from the study. Figure 4.9 presents the value distribution of a feature (diagnostics_Image-original_Minimum) for the four diseases. As this feature is not discriminatory of the classes, it was removed from the Dataset.

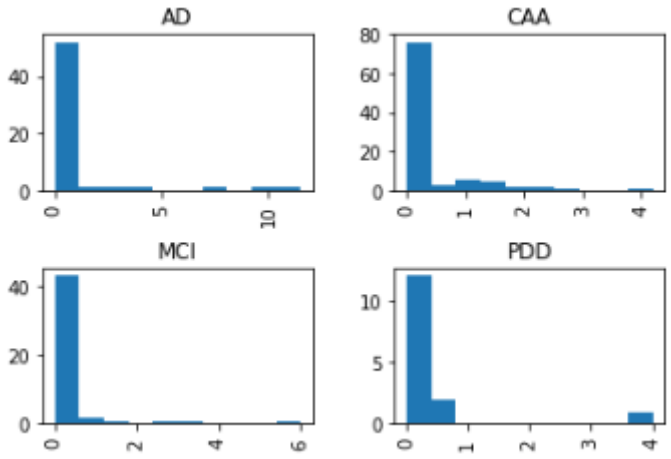


Figure 4.9 - Histogram of the "diagnostics_Image-original_Minimum" data distribution.

4.2.5 EVALUATION METRICS

To evaluate the performance of the different supervised ML algorithms from the three experiments, several measures were used to evaluate the performance of the classifiers and learning algorithms. The diagnostic capability of classifiers is usually determined by the confusion matrix and the receiver operating characteristic (ROC) curve. The confusion Matrix measures the quality of the classification, returning a matrix with the correctly and incorrectly identified diseases, giving insight about the errors affecting the classifier. The matrix can be useful on multi-class studies, assigning the rows to true event and the columns to the predictions, thus having a N x N matrix dimension. From this framework results a set of parameters summarized at Table 4.3:

- True positive (TP): The correctly classified events, that is, the disease was correctly classified;
- False Positive (FP): The cases where the classifier incorrectly identified the event as positive. This is also called type I error ;
- True Negative (TN): The correctly predicted negative cases;

- False Negative (FN): Incorrectly classified positive events. It is also called type II error.

Table 4.3- The framework of the confusion matrix

		PREDICTED CLASS	
		P	N
ACTUAL CLASS	P	True Positive (TP)	False Negative (FN)
	N	False Positive (FP)	True Negative (TN)

From this matrix, several metrics to evaluate the performance can be calculated. The classifier accuracy can be measured as follows:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{TP + FP + FN + TN} \quad (1)$$

From the accuracy, two other measures of the model's performance may be estimated: the precision and recall.

$$Precision = \frac{\mathrm{TP}}{TP + FP} \quad (2)\ ;\ Recall = \frac{\mathrm{TP}}{TP + FN} \quad (3)$$

Precision (Equation 2) is obtained by the division of the correctly labeled entries by the total positive cases. It may be seen as a measure of exactness or quality of the model and capability to predict of classify correctly the positive cases.

Recall (Equation 3), also called sensitivity, is obtained by dividing the true positive cases by the Sum of true positives and false negatives. As it estimates the proportion of positive cases identified correctly, the two ways to get a large recall is by increasing the number of TP or lowering the number of FN.

The combination of both measures results in the F1 score, allowing an accurate evaluation of the classifiers test accuracy. As represented in Equation 4, it consists on the harmonic mean of precision and recall.

$$F_1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (4)$$

When considering such metrics in multiclass problems, two distinct measures may be applied depending on the scenario: the “macro-average” and “micro-average”. The first one calculates for each metric the mean of the binary metrics, giving equal weight to the classes. On the other hand, the “micro-average” considers the weight of each class on the overall metric. Using this scenario in imbalanced classifications tends to overlook the impact of the predominant class.

4.3 RESULTS AND DISCUSSION

When evaluating the disease classification models, accuracy metrics were used. Table 6 9 contains the parameters used in this experiment.

4.3.1 EXPERIMENT 1: IMAGE-BASED BIOMARKER EXPLORATION

The results of the classification models using the disease labels with the annotation clinical data is revealed below at Table 4.4:

Table 4.4 - Shows the results obtained from the classifiers for the first experiment.

ALGORITHM	ACCURACY	RECALL		PRECISION		F1 SCORE	
		Macro-recall	Micro-recall	Macro-precision	Micro-precision	Macro-F1	Micro-F1
Logistic Regression	0.4118	0.29	0.41	0.23	0.33	0.24	0.35
Support vector machine	0.4902	0.34	0.49	0.31	0.44	0.16	0.29
Random Forest	0.5490	0.39	0.55	0.37	0.49	0.37	0.51
Multi-layer Perceptron	0.3922	0.25	0.39	0.10	0.15	0.14	0.22
Decision Tree	0.5450	0.28	0.55	0.33	0.53	0.29	0.51

After testing the models, it is obvious that the one that better classifies the four neurodegenerative diseases is the Random Forest, although the DT and SVM were not far behind. RF also presents the best values of precision and recall of all the trained models, meaning that the proportion of True positives is this case is slightly higher than the false negatives.

To determine the features from which these values resulted, feature importance from sklearn is used. The List of the most important features for the correct classification of 55% of the

diseases with the Random Forest classifier is seen at Table 4.5 and the visual representation at Figure 4.10.

Table 4.5 - Feature importance for the Random forest classifier using the annotation data.

FEATURE	WEIGHT
AGE	0.4198 ± 0.0696
SUMPOST	0.2412 ± 0.0516
GENDER	0.2382 ± 0.0296
SUMMTA	0.2015 ± 0.0700
SUMANTTEMP	0.1389 ± 0.0203
SUMFRONTINSUL	0.1206 ± 0.0614
SUMANTCING	0.1053 ± 0.0541
SUMORBFRONT	0.0977 ± 0.0114

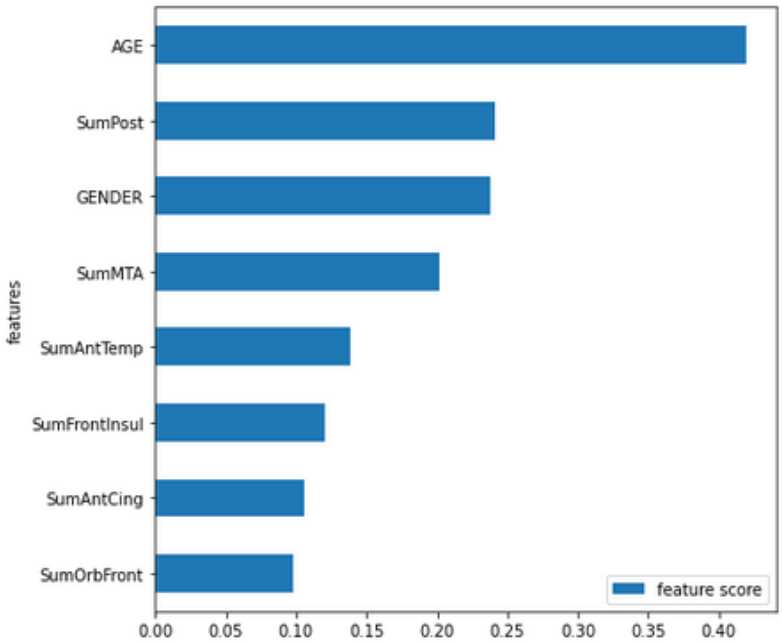


Figure 4.10 - Feature importance for the annotation data.

From the Table 4.5, it is clear thar Age is a major feature for the differentiation of the classes. An histogram of the age distribution by disease, Figure 4.11, was plotted to observe the differences in the Age range and incidence by each disease. CAA and AD seem to have a wider age distribution ([31-89] and [33-89] respectively). In the AD case, this patient seems to be an outlier, which could be removed from the study. As for CAA, the early onset may be due to a

hereditary CAA form. MCI and PDD have a different data distribution around Age, when compared to the other two diseases.

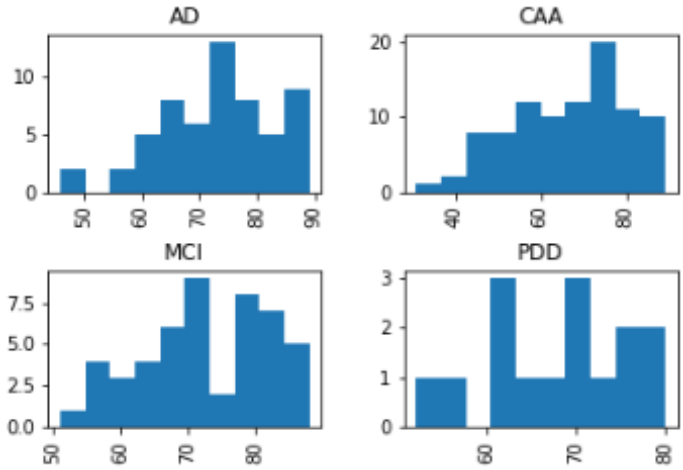


Figure 4.11 - Histogram of Age distribution for the annotated data

Regarding the Posterior lobe, the behaviour of the data for the patients with AD and CAA seems quite similar. Observing the barplot of the CAA disease, it is obvious that the rating of two is predominant (accounts for about 60% of the cases), while 30% is distributed between the ratings of one, three, and four. When it comes to AD, the rating of two takes around 50% of the cases, while 20% are rated with zero, and 10% with three ad four each. For MCI, almost 60% of the cases are rated with the score two, followed by the four rating (30%). Regarding PDD, 55% of patients were rated two and almost 30% with zero.

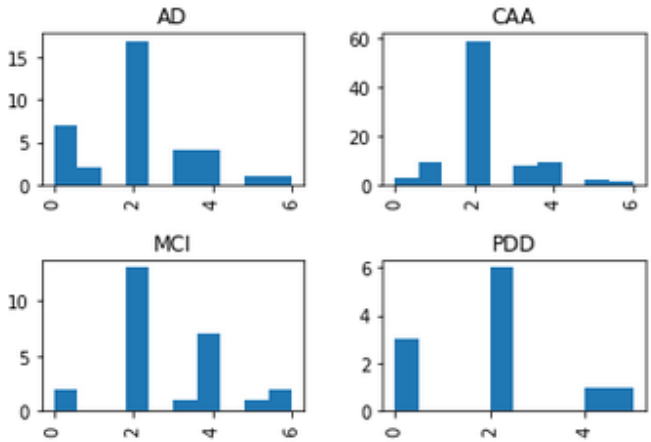


Figure 4.12 - Histogram of the Posterior lobe distribution for the annotated data.

4.3.2 EXPERIMENT 2: CLASSIFICATION OF DISEASES BASED ON RADIOMIC DATA

For the second experiment, feature engineering was applied, starting with a total of 1722 features. For this step, the accuracy values for the classification models are presented at Table 4.6.

Table 4.6- Results obtained from the classification of Radiomic Data.

ALGORITHM	ACCURACY	RECALL		PRECISION		F1 SCORE	
		Macro-recall	Micro-recall	Macro-precision	Micro-precision	Macro-F1	Micro-F1
Logistic Regression	0.4848	0.21	0.48	0.14	0.32	0.17	0.38
Support vector machine	0.4583	0.25	0.46	0.11	0.21	0.16	0.29
Random Forest	0.6670	0.43	0.67	0.56	0.66	0.46	0.63
Multi-layer Perceptron	0.5757	0.25	0.58	0.14	0.33	0.18	0.42
Decision Tree	0.5440	0.45	0.55	0.45	0.55	0.45	0.55

Once again, RF is the classifier with the best Precision and recall, although the values of True positives and False negatives may not be much different. From these results, the feature importance was determined, resulting in a list with the most important features for the RF classifier. The ordered weights for each feature is presented in the Table 4.7 and a data histogram in Figure 4.13 .

Table 4.7 - Feature importance for the Random forest classifier using the Radiomic data.

FEATURE	WEIGHT
LBP-3D-M2_GLRLM_SHORTRUNLOWGRAYLEVELEMPHASIS	0.1634 ± 0.0449
WAVELET-HHL_GLDM_HIGHGRAYLEVELEMPHASIS	0.1450 ± 0.0273
LBP-3D-M1_FIRSTORDER_MEAN	0.1160 ± 0.0379
ORIGINAL_FIRSTORDER_10PERCENTILE	0.0824 ± 0.0366
WAVELET-HHH_GLCM_CLUSTERTENDENCY	0.0565 ± 0.0248
LOG-SIGMA-5-0-MM-3D_GLSZM_SMALLAREAHIGHGRAYLEVELEMPHASIS	0.0565 ± 0.0207
WAVELET-HLH_FIRSTORDER_SKEWNESS	0.0504 ± 0.0299

LBP-3D-M1_FIRSTORDER_MEANABSOLUTEDEVIATION	0.0504 ± 0.0183
WAVELET-HLL_GLCM_CORRELATION	0.0504 ± 0.0075
SQUAREROOT_GLCM_MCC	0.0443 ± 0.0224
WAVELET-HLH_GLSZM_ZONEPERCENTAGE	0.0412 ± 0.0122
LOGARITHM_FIRSTORDER_ENERGY	0.0412 ± 0.0122
LOG-SIGMA-1-0-MM-3D_GLRLM_LONGRUNEMPHASIS	0.0366 ± 0.0150
LBP-3D-K_GLRLM_GRAYLEVELNONUNIFORMITY	0.0321 ± 0.0178
LBP-3D-M2_FIRSTORDER_90PERCENTILE	0.0260 ± 0.0248
WAVELET-HHL_GLCM_DIFFERENCEENTROPY	0.0229 ± 0.0097
LOGARITHM_FIRSTORDER_TOTALENERGY	0.0214 ± 0.0114
WAVELET-LLL_GLRLM_SHORTRUNLOWGRAYLEVELEMPHASIS	0.0214 ± 0.0203
WAVELET-HHH_FIRSTORDER_ROOTMEANSQUARED	0.0198 ± 0.0075
LOG-SIGMA-4-0-MM-3D_GLCM_MAXIMUMPROBABILITY	0.0153 ± 0.0097

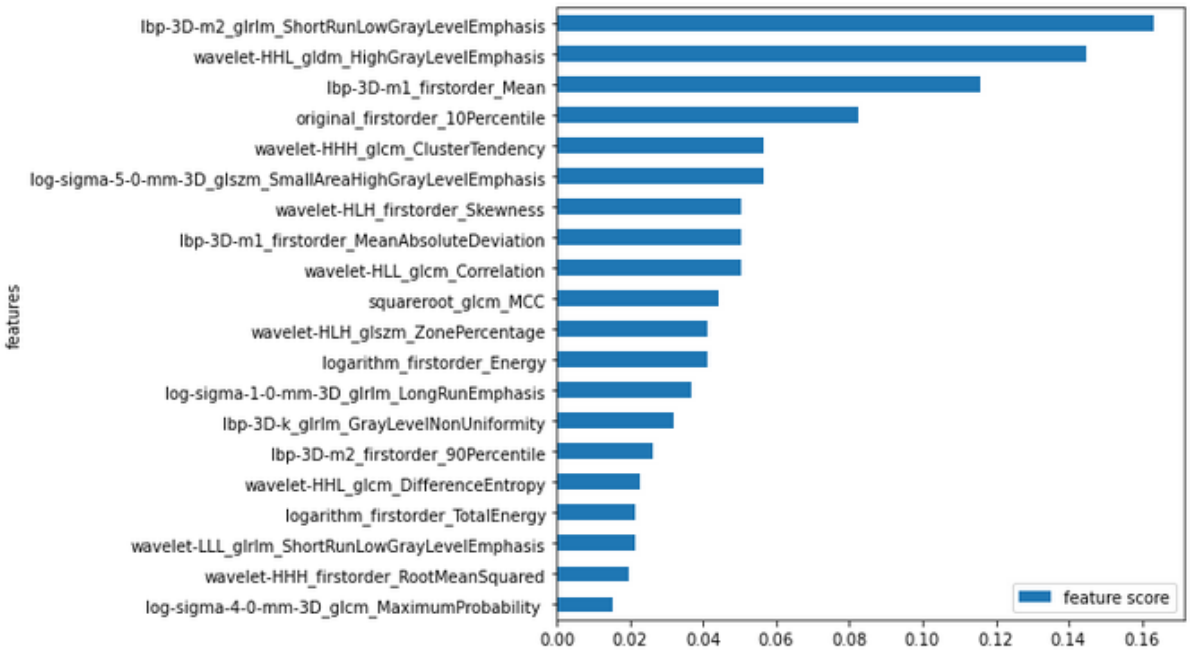


Figure 4.13 - Feature importance for the Radiomic data.

In order to determine whether the results improve by selecting the most important features, a threshold of 0.05 was chosen. The features with a score above the threshold were kept on the study. Table 4.8 presents the results of a new classification with the selected features.

Table 4.8 - Results obtained from the classification of the Radiomic data after feature selection.

ALGORITHM	ACCURACY	RECALL		PRECISION		F1 SCORE	
		Macro-recall	Micro-recall	Macro-precision	Micro-precision	Macro-F1	Micro-F1
Logistic Regression	0.4375	0.21	0.48	0.14	0.32	0.17	0.38
Support vector machine	0.5758	0.25	0.58	0.14	0.33	0.18	0.42
Random Forest	0.6970	0.47	0.70	0.54	0.66	0.48	0.66
Multi-layer Perceptron	0.6360	0.35	0.64	0.40	0.50	0.33	0.52
Decision Tree	0.6060	0.56	0.61	0.70	0.66	0.59	0.61

Taking a look at Table 4.9, the features who have a higher impact on the classification score are the Root-Mean-Squared, from the first order features with a wavelet filter, the 90-Percentile feature with a lbp filter, Long-Run-Emphasis feature from the glrlm group, with the log filter, and the Short-Run-Low-Gray-Level-Emphasis from the glrlm group, with a lbp filter. The Root-Mean-Squared (Equation 4), or quadratic mean is a measure of the magnitude of the image values. Corresponds to the square of all the numbers in a set, followed by Its arithmetic mean, and then the square root of the result.

$$RMS = \sqrt{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)+C)^2} \qquad (4)$$

The 90Percentile corresponds to the 90th percentile of the values on the Local Binary Pattern filter.

The Long Run Emphasis (Equation 5) feature is a measure of the longer pixels in an Image that have the same grey level value, with a greater value indicating a coarser image texture. N In the equation corresponds to the number of homogeneous pixel values in the Volume of Interest.

$$LRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j|\theta)j^2}{N_r(\theta)} \qquad (5)$$

Finally, Short Run Low Gray Level Emphasis (equation 6) corresponds to the shorter distance between pixels with the same lower grey-level values.

$$SRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j|\theta)}{i^2 j^2}}{N_r(\theta)} \qquad (6)$$

Table 4.9 - Feature importance for the Random forest classifier using the Radiomic data, after feature selection.

FEATURE	WEIGHT
WAVELET-HHH_FIRSTORDER_ROOTMEANSQUARED	0.3099 ± 0.0156
LBP-3D-M2_FIRSTORDER_90PERCENTILE	0.2611 ± 0.0636
LOG-SIGMA-1-0-MM-3D_GLRLM_LONGRUNEMPHASIS	0.2565 ± 0.0479
LBP-3D-M2_GLRLM_SHORTRUNLOWGRAYLEVELEMPHASIS	0.2122 ± 0.0244
WAVELET-HHL_GLCM_DIFFERENCEENTROPY	0.1679 ± 0.0348
LBP-3D-M1_FIRSTORDER_MEANABSOLUTEDEVIATION	0.1435 ± 0.0311
WAVELET-HLL_GLCM_CORRELATION	0.1160 ± 0.0203
LOGARITHM_FIRSTORDER_ENERGY	0.0916 ± 0.0237
WAVELET-HHL_GLDM_HIGHGRAYLEVELEMPHASIS	0.0809 ± 0.0156
ORIGINAL_FIRSTORDER_10PERCENTILE	0.0748 ± 0.0296
LOG-SIGMA-5-0-MM-3D_GLSZM_SMALLAREAHIGHGRAYLEVELEMPHASIS	0.0336 ± 0.0122
SQUAREROOT_GLCM_MCC	0.0260 ± 0.0156
WAVELET-HLH_GLSZM_ZONEPERCENTAGE	0.0229 ± 0.0137
LOGARITHM_FIRSTORDER_TOTALENERGY	0.0183 ± 0.0122
WAVELET-HLH_FIRSTORDER_SKEWNESS	0.0168 ± 0.0061
LBP-3D-M1_FIRSTORDER_MEAN	0.0137 ± 0.0114
LBP-3D-K_GLRLM_GRAYLEVELNONUNIFORMITY	0.0122 ± 0.0183
WAVELET-HHH_GLCM_CLUSTERTENDENCY	0 ± 0.0000
WAVELET-LLL_GLRLM_SHORTRUNLOWGRAYLEVELEMPHASIS	0 ± 0.0000
LOG-SIGMA-4-0-MM-3D_GLCM_MAXIMUMPROBABILITY	0 ± 0.0000

4.3.3 EXPERIMENT 3: CLASSIFICATION OF DISEASES BASED ON RADIOMIC AND ANNOTATION DATA

On the last experiment, the remaining radiomic features were combined with the annotated data. After this step, the aforementioned models were applied, generating the results presented at Table 4.10:

Table 4.10 - Results obtained from the classification of the combined datasets

ALGORITHM	ACCURACY	RECALL		PRECISION		F1 SCORE	
		Macro-recall	Micro-recall	Macro-precision	Micro-precision	Macro-F1	Micro-F1
Logistic Regression	0.4318	0.28	0.43	0.25	0.34	0.24	0.36
Support vector machine	0.4320	0.25	0.43	0.11	0.19	0.15	0.26
Random Forest	0.6970	0.44	0.70	0.58	0.70	0.47	0.65
Multi-layer Perceptron	0.4318	0.25	0.43	0.11	0.19	0.15	0.26
Decision Tree	0.5450	0.36	0.55	0.38	0.51	0.36	0.52

Once again, the RF presents the best classification results, correctly classifying the diseases 70% of the cases. The higher Micro-F1 score stands out from the other models and is quite different from the Macro average of F-score, suggesting that the classes representation on the data is unbalanced. Also, it Is clear that the score of the RF for the combined Dataset is higher than the one obtained for the radiomic and annotations data individually. Moreover, the feature importance was determined and listed as follows on Table 4.11, and Figure 4.14.

Table 4.11 - Feature importance for the Random forest classifier using the combined data.

FEATURE	WEIGHT
LOG-SIGMA-1-0-MM-3D_FIRSTORDER_UNIFORMITY	0.2427 ± 0.0769
LBP-3D-M2_GLRLM_SHORTRUNHIGHGRAYLEVELEMPHASIS	0.1695 ± 0.0203
LOG-SIGMA-3-0-MM-3D_GLCM_IMC1	0.1206 ± 0.0353
WAVELET-HHH_GLCM_JOINTAVERAGE	0.1069 ± 0.0255
LBP-3D-M2_GLDM_LARGEDEPENDENCELOWGRAYLEVELEMPHASIS	0.0718 ± 0.0356
SUMANTCING	0.0672 ± 0.0263

WAVELET-HLH_GLRLM_SHORTRUNLOWGRAYLEVELEMPHASIS	0.0595 ± 0.0178
LOG-SIGMA-4-0-MM-3D_GLDM_LARGEDEPENDENCELOWGRAYLEVELEMPHASIS	0.0565 ± 0.0183
SQUARE_FIRSTORDER_MEAN	0.0550 ± 0.0403
LOG-SIGMA-3-0-MM-3D_FIRSTORDER_MAXIMUM	0.0489 ± 0.0248
LOG-SIGMA-3-0-MM-3D_FIRSTORDER_KURTOSIS	0.0351 ± 0.0122
WAVELET-LLL_GLSZM_GRAYLEVELVARIANCE	0.0336 ± 0.0122
LOG-SIGMA-5-0-MM-3D_GLCM_SUMSQUARES	0.0305 ± 0.0097
SQUARE_FIRSTORDER_MEDIAN	0.0305 ± 0.0137
WAVELET-HHH_GLSZM_GRAYLEVELNONUNIFORMITYNORMALIZED	0.0260 ± 0.0156
ORIGINAL_SHAPE_MAXIMUM2DDIAMETERROW	0.0214 ± 0.0224
WAVELET-HHL_FIRSTORDER_TOTALENERGY	0.0198 ± 0.0156
WAVELET-HHH_GLRLM_SHORTRUNHIGHGRAYLEVELEMPHASIS	0.0198 ± 0.0075
WAVELET-HHH_GLSZM_GRAYLEVELVARIANCE	0.0183 ± 0.0156

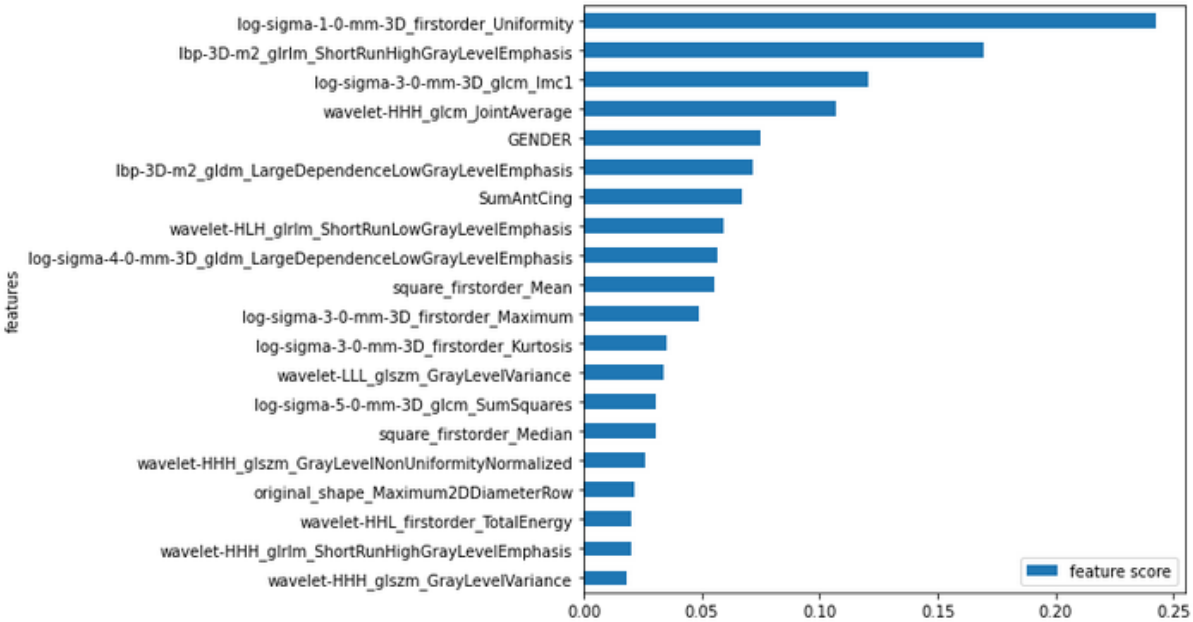


Figure 4.14 - Feature importance for the Radiomic data.

Following the steps applied before, a threshold of 0.05 was determined, and the features with a score above the threshold were kept on the study. Table 4.12 presents the results of a new classification with the selected features.

Table 4.12 - Results obtained from the classification of the combined datasets, after feature selection.

ALGORITHM	ACCURACY	RECALL		PRECISION		F1 SCORE	
		Macro-recall	Micro-recall	Macro-precision	Micro-precision	Macro-F1	Micro-F1
Logistic Regression	0.5760	0.25	0.58	0.14	0.33	0.18	0.42
Support vector machine	0.5758	0.25	0.58	0.14	0.33	0.18	0.42
Random Forest	0.6970	0.47	0.70	0.54	0.66	0.48	0.66
Multi-layer Perceptron	0.4849	0.21	0.48	0.14	0.32	0.17	0.38
Decision tree	0.6060	0.52	0.61	0.51	0.61	0.51	0.60

From Table 4.12, it is clear that, although the results of the classification model with the Random forest algorithm did not change, the overall results improved. Furthermore, observing the data present on Table 4.13, Short Run High Gray Level Emphasis and Joint Average may have an Important weight on the classification model used with the Random forest algorithm.

The Short Run High Gray Level Emphasis (Equation 7) corresponds to the shorter distance between pixels with the same hight grey-level values.

$$SRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j|\theta)i^2}{j^2}}{N_r(\theta)} \qquad (7)$$

Joint Average (equation 8) is the mean grey level intensity of a distribution.

$$joint\ average = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)i \qquad (8)$$

Table 4.13 - Feature importance after feature selection.

FEATURE	WEIGHT
LBP-3D-M2_GLRLM_SHORTRUNHIGHGRAYLEVELEMPHASIS	0.1786 ± 0.0568
WAVELET-HHH_GLCM_JOINTAVERAGE	0.0824 ± 0.0203
LOG-SIGMA-1-0-MM-3D_FIRSTORDER_UNIFORMITY	0.0733 ± 0.0283
LOG-SIGMA-3-0-MM-3D_GLCM_IMC1	0.0718 ± 0.0248
LOG-SIGMA-4-0-MM-3D_GLDM_LARGEDEPENDENCELOWGRAYLEVELEMPHASIS	0.0718 ± 0.0183
SQUARE_FIRSTORDER_MEAN	0.0580 ± 0.0207
SUMANTCING	0.0519 ± 0.0150

LBP-3D-M2_GLDM_LARGEDEPENDENCELOWGRAYLEVELEMPHASIS	0.0473 ± 0.0178
WAVELET-HLH_GLRLM_SHORTRUNLOWGRAYLEVELEMPHASIS	0.0351 ± 0.0156
GENDER	0.0305 ± 0.0193

5 CONCLUSIONS

The purpose of this work was to determine biomarkers to complement the diagnosis of CAA. To achieve this goal, AI algorithms were applied after the creation of a dataset and MRI image retrieval, data pre-processing, image handling, and feature engineering.

AI has the potential to revolutionize the medical field, by assisting the diagnosis of patients by performing tasks that currently rely on human observation, detection, or quantification. The technological advancement allied to research on the health domain comes to assist the decision making and to identify patterns or findings that better describe or distinguish diseases.

Several machine learning models were developed for three experiments: Classification of CAA and three other Neurodegenerative diseases using Data annotated with visual metrics, Classification of the aforementioned diseases using Radiomic Data, and finally the same classification using a combined Dataset, created by aggregation of the two previous ones.

The first step was the creation of the Annotation Dataset and MRI image acquisition, using adult patients admitted to a University Hospital with non-traumatic ICH, later diagnosed with CAA. The annotation was obtained in collaboration with trained physicians who rated the brain atrophy by region, according to proposed metrics. To enrich the Dataset, control data should be added. As healthy elder patients' data were not available, the control data was replaced with AD, MCI and PDD data.

Another step necessary to complete the work was the generation of Radiomics Features. It required image segmentation of the skull and brain and the construction of 3D images from the 2D slices. This step also required an understating of the DICOM and NIFTI format structure to handle the data. Also, all the procedures for the project were carried on in a Jupyter Notebook container, optimizing the processes, organization, and time management of the processes. This also assisted the development of the ML models.

The ML approach was carried on by applying four algorithms: Logistic Regression, Support Vector Machine, Random Forest, Multi-layer Perceptron, and Decision Trees. The five algorithms were first run with the annotation data after pre-processing, achieving a score of 0.5490 with RF. As for the classification based on radiomic data, RF performed better than the four other models, with a score of 0.697, a better result than the one achieved with the clinical data. Finally, by combining the Dataset (radiomics) and clinical data (annotation), the aim is to determine whether the classification is favoured by the medical annotations, the image derived data, or both. In this scenario, RF achieved a score of 0.697, which is better than the

first experiment and equal to the second one. From these results It is safe to affirm that the hypothesis of the use of the combined Dataset and the Radiomic approach for complementary diagnosis of CAA should not be discharged and studied even further.

After classifying the patient's data, feature importance was determined to give some insight into the variables used to distinguish the neurodegenerative diseases, thus being a possible biomarker for the diagnosis of CAA. Age was pointed out as being a crucial feature when classifying the annotation Dataset. The generated results suggested that the age distribution for AD and CAA were different from MCI and PDD. Although the Age histogram distribution is somewhat similar to the AD's, the early onset on the CAA case may be due to an hereditary CAA form, as it occurs in more than one subject, while the single patient under 40 on the AD dataset may be an outlier with a malformation or hereditary disease. Additionally, the atrophy pattern on the Posterior region is demonstrated to be also important for the classification of the diseases in the Dataset with the visual rating scores. The Posterior region was referred to be affected by Cerebral microinfarct and white matter lesions in CAA, an evidence that supports this work findings.

When it comes to the radiomic data experiment, features related to the voxel intensities within the image ROI, voxel intensity symmetry, and features related to image texture seem to weight on the classification accuracy. Histograms were also plotted and once again, AD and CAA seemed to have a similar value distribution, contrasting with MCI and PDD. These findings may be explained by the fact that patients with CAA or AD may suffer from similar cerebral lesions, for example, cerebral bleedings which in the late years has been seen as a strong biomarker for the diagnosis of CAA.

Finally, the combined data is affected by features that measure the length of consecutive pixels that have the same grey level value, thus related to the texture, and the homogeneity of the pixel intensity.

Although the results were satisfactory, there is still some place for improvement. The use of healthy patients on future works would be a major contribution to better understand the features that weight the most on the diagnosis of CAA, considering the normal human aging process. Unravelling the independent behaviour of CAA could lead to new therapeutic strategies and develop the medical intervention on neurodegenerative diseases.

As for the data acquisition and handling, it would be interesting to implement the annotation of the ROI in the image, as it could be used as a reference for the computer to know were to look at, and what points are used by physicians to examine a patient's MRI scan.
Finally, a bigger clinical dataset would introduce more variables in the study, and a follow up of the living patients and acquisition of MRI scans on the following years would provide insight about the evolution and impact of the disease on the subject.

References

[1] J. J. Gilbert and H. V. Vinters, “Cerebral amyloid angiopathy: Incidence and complications in the aging brain. I. Cerebral hemorrhage,” *Stroke*, 1983, doi: 10.1161/01.STR.14.6.915.

[2] M. Tanskanen *et al.*, “Prevalence and severity of cerebral amyloid angiopathy: A population-based study on very elderly Finns (Vantaa 85+),” *Neuropathol. Appl. Neurobiol.*, 2012, doi: 10.1111/j.1365-2990.2011.01219.x.

[3] H. H. Tsai *et al.*, “Correlation of Cerebral Microbleed Distribution to Amyloid Burden in Patients with Primary Intracerebral Hemorrhage,” *Sci. Rep.*, 2017, doi: 10.1038/srep44715.

[4] L. Pantoni, “Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges,” *The Lancet Neurology*. 2010, doi: 10.1016/S1474-4422(10)70104-6.

[5] S. M. Greenberg and A. Charidimou, “Diagnosis of cerebral amyloid angiopathy evolution of the Boston criteria,” *Stroke*. 2018, doi: 10.1161/STROKEAHA.117.016990.

[6] A. Rostagno, J. L. Holton, T. Lashley, T. Revesz, and J. Ghiso, “Cerebral amyloidosis: Amyloid subunits, mutants and phenotypes,” *Cellular and Molecular Life Sciences*. 2010, doi: 10.1007/s00018-009-0182-4.

[7] L. Harper *et al.*, “MRI visual rating scales in the diagnosis of dementia: Evaluation in 184 post-mortem confirmed cases,” *Brain*, 2016, doi: 10.1093/brain/aww005.

[8] L. O. Wahlund *et al.*, “A new rating scale for age-related white matter changes applicable to MRI and CT,” *Stroke*, 2001, doi: 10.1161/01.STR.32.6.1318.

[9] D. J. Selkoe, “Alzheimer’s disease is a synaptic failure,” *Science*. 2002, doi: 10.1126/science.1074069.

[10] T. I. Mandybur, “Cerebral amyloid angiopathy: The vascular pathology and complications,” *J. Neuropathol. Exp. Neurol.*, 1986, doi: 10.1097/00005072-198601000-00007.

[11] D. M. A. Mann *et al.*, “Predominant deposition of amyloid-β42(43) in plaques in cases of Alzheimer’s disease and hereditary cerebral hemorrhage associated with mutations in the amyloid precursor protein gene,” *Am. J. Pathol.*, 1996.

[12] C. L. Joachim, L. K. Duffy, J. H. Morris, and D. J. Selkoe, “Protein chemical and immunocytochemical studies of meningovascular β-amyloid protein in Alzheimer’s

disease and normal aging," *Brain Res.*, 1988, doi: 10.1016/0006-8993(88)90673-7.

[13] H. V. Vinters, "Cerebral amyloid angiopathy a critical review," *Stroke*, 1987, doi: 10.1161/01.STR.18.2.311.

[14] F. L. Mastaglia, M. L. Byrnes, R. D. Johnsen, and B. A. Kakulas, "Prevalence of cerebral vascular amyloid-β deposition and stroke in an aging Australian population: A postmortem study," *J. Clin. Neurosci.*, 2003, doi: 10.1016/S0967-5868(02)00317-X.

[15] J. P. G. Vonsattel, R. H. Myers, E. Tessa Hedley-Whyte, A. H. Ropper, E. D. Bird, and E. P. Richardson, "Cerebral amyloid angiopathy without and with cerebral hemorrhages: A comparative histological study," *Ann. Neurol.*, 1991, doi: 10.1002/ana.410300503.

[16] S. M. Greenberg, "Cerebral amyloid angiopathy and vessel dysfunction," *Cerebrovasc. Dis.*, 2002, doi: 10.1159/000049149.

[17] J. M. Olichney, L. A. Hansen, C. R. Hofstetter, M. Grundman, R. Katzman, and L. J. Thal, "Cerebral Infarction in Alzheimer's Disease is Associated with Severe Amyloid Angiopathy and Hypertension," *Arch. Neurol.*, 1995, doi: 10.1001/archneur.1995.00540310076019.

[18] D. Cadavid, H. Mena, K. Koeller, and R. A. Frommelt, "Cerebral beta amyloid angiopathy is a risk factor for cerebral ischemic infarction. A case control study in human brain biopsies," *J. Neuropathol. Exp. Neurol.*, 2000, doi: 10.1093/jnen/59.9.768.

[19] M. Yamada *et al.*, "Association of presenilin-1 polymorphism with cerebral amyloid angiopathy in the elderly," *Stroke*, 1997, doi: 10.1161/01.STR.28.11.2219.

[20] M. Haglund, U. Passant, M. Sjöbeck, E. Ghebremedhin, and E. Englund, "Cerebral amyloid angiopathy and cortical microinfarcts as putative substrates of vascular dementia," *Int. J. Geriatr. Psychiatry*, 2006, doi: 10.1002/gps.1550.

[21] J. Rosand, E. M. Hylek, H. C. O'Donnell, and S. M. Greenberg, "Warfarin-associated hemorrhage and cerebral amyloid angiopathy: A genetic and pathologic study," *Neurology*, 2000, doi: 10.1212/WNL.55.7.947.

[22] S. Kumar-Singh, "Cerebral amyloid angiopathy: Pathogenetic mechanisms and link to dense amyloid plaques," in *Genes, Brain and Behavior*, 2008, doi: 10.1111/j.1601-183X.2007.00380.x.

[23] J. Attems, "Sporadic cerebral amyloid angiopathy: Pathology, clinical implications, and possible pathomechanisms," *Acta Neuropathologica*. 2005, doi: 10.1007/s00401-005-1074-9.

[24] T. Revesz *et al.*, “Cerebral amyloid angiopathies: A pathologic, biochemical, and genetic view,” *Journal of Neuropathology and Experimental Neurology*. 2003, doi: 10.1093/jnen/62.9.885.

[25] L. A. Pfeifer, L. R. White, G. W. Ross, H. Petrovitch, and L. J. Launer, “Cerebral amyloid angiopathy and cognitive function: The HAAS autopsy study,” *Neurology*, 2002, doi: 10.1212/WNL.58.11.1629.

[26] S. M. Greenberg, M. E. Gurol, J. Rosand, and E. E. Smith, “Amyloid angiopathy-related vascular cognitive impairment,” in *Stroke*, 2004, doi: 10.1161/01.STR.0000143224.36527.44.

[27] Z. Arvanitakis, S. E. Leurgans, Z. Wang, R. S. Wilson, D. A. Bennett, and J. A. Schneider, “Cerebral amyloid angiopathy pathology and cognitive domains in older persons,” *Ann. Neurol.*, 2011, doi: 10.1002/ana.22112.

[28] C. Iadecola, “The overlap between neurodegenerative and vascular factors in the pathogenesis of dementia,” *Acta Neuropathologica*. 2010, doi: 10.1007/s00401-010-0718-6.

[29] J. Attems, K. Jellinger, D. R. Thal, and W. Van Nostrand, “Review: Sporadic cerebral amyloid angiopathy,” *Neuropathol. Appl. Neurobiol.*, 2011, doi: 10.1111/j.1365-2990.2010.01137.x.

[30] J. B. Mackic *et al.*, “Circulating amyloid-β peptide crosses the blood-brain barrier in aged monkeys and contributes to Alzheimer’s disease lesions,” *Vascul. Pharmacol.*, 2002, doi: 10.1016/S1537-1891(02)00198-2.

[31] L. Park *et al.*, “Innate immunity receptor CD36 promotes cerebral amyloid angiopathy,” *Proc. Natl. Acad. Sci. U. S. A.*, 2013, doi: 10.1073/pnas.1300021110.

[32] J. B. Mackic *et al.*, “Human blood-brain barrier receptors for Alzheimer’s amyloid-β 1-40. Asymmetrical binding, endocytosis, and transcytosis at the apical side of brain microvascular endothelial cell monolayer,” *J. Clin. Invest.*, 1998, doi: 10.1172/JCI2029.

[33] M. Ujiie, D. L. Dickstein, D. A. Carlow, and W. A. Jefferies, “Blood-brain barrier permeability precedes senile plaque formation in an Alzheimer disease model,” *Microcirculation*, 2003, doi: 10.1038/sj.mn.7800212.

[34] P. A. Stewart, K. Hayakawa, M. A. Akers, and H. V. Vinters, “A morphometric study of the blood-brain barrier in Alzheimer’s disease,” *Lab. Investig.*, 1992.

[35] M. Shoji *et al.*, “Production of the Alzheimer amyloid β protein by normal proteolytic

processing," *Science (80-.).*, 1992, doi: 10.1126/science.1439760.

[36] P. Seubert *et al.*, "Isolation and quantification of soluble Alzheimer's β-peptide from biological fluids," *Nature*, 1992, doi: 10.1038/359325a0.

[37] J. Frackowiak, A. Zoltowska, and H. M. Wisniewski, "Non-fibrillar β-amyloid protein is associated with smooth muscle cells of vessel walls in alzheimer disease," *J. Neuropathol. Exp. Neurol.*, 1994, doi: 10.1097/00005072-199411000-00011.

[38] F. Tagliavini, J. Ghiso, W. F. Timmers, G. Giaccone, O. Bugiani, and B. Frangione, "Coexistence of Alzheimer's amyloid precursor protein and amyloid protein in cerebral vessel walls," *Lab. Investig.*, 1990.

[39] M. Kawai *et al.*, "Degeneration of vascular muscle cells in cerebral amyloid angiopathy of Alzheimer disease," *Brain Res.*, 1993, doi: 10.1016/0006-8993(93)90021-E.

[40] J. D. Fryer *et al.*, "Apolipoprotein E markedly facilitates age-dependent cerebral amyloid angiopathy and spontaneous hemorrhage in amyloid precursor protein transgenic mice," *J. Neurosci.*, 2003, doi: 10.1523/jneurosci.23-21-07889.2003.

[41] H. M. Wisniewski, J. Frackowiak, and B. Mazur-Kolecka, "In vitro production of β-amyloid in smooth muscle cells isolated from amyloid angiopathy-affected vessels," *Neurosci. Lett.*, 1995, doi: 10.1016/0304-3940(94)11129-7.

[42] H. M. Wisniewski, J. Wegiel, and L. Kotula, "Review: David Oppenheimer Memorial Lecture 1995: Some neuropathological aspects of Alzheimer's disease and its relevance to other disciplines," *Neuropathol. Appl. Neurobiol.*, 1996, doi: 10.1111/j.1365-2990.1996.tb00839.x.

[43] J. V. Dorpe *et al.*, "Prominent cerebral amyloid angiopathy in transgenic mice overexpressing the London mutant of human APP in neurons," *Am. J. Pathol.*, 2000, doi: 10.1016/s0197-4580(00)82614-7.

[44] R. O. Weller, A. Massey, T. A. Newman, M. Hutchings, Y. M. Kuo, and A. E. Roher, "Cerebral amyloid angiopathy: Amyloid β accumulates in putative interstitial fluid drainage pathways in Alzheimer's disease," *Am. J. Pathol.*, 1998, doi: 10.1016/S0002-9440(10)65616-7.

[45] M. E. Calhoun *et al.*, "Neuronal overexpression of mutant amyloid precursor protein results in prominent deposition of cerebrovascular amyloid," *Proc. Natl. Acad. Sci. U. S. A.*, 1999, doi: 10.1073/pnas.96.24.14088.

[46] J. A. R. Nicoll, M. Yamada, J. Frackowiak, B. Mazur-Kolecka, and R. O. Weller, "Cerebral

amyloid angiopathy plays a direct role in the pathogenesis of Alzheimer's disease: Pro-CAA position statement," in *Neurobiology of Aging*, 2004, doi: 10.1016/j.neurobiolaging.2004.02.003.

[47] K. Hsiao *et al.*, "Correlative memory deficits, Aβ elevation, and amyloid plaques in transgenic mice," *Science (80-.).*, 1996, doi: 10.1126/science.274.5284.99.

[48] J. P. Melchor, L. McVoy, and W. E. Van Nostrand, "Charge alterations of E22 enhance the pathogenic properties of the amyloid β-protein," *J. Neurochem.*, 2000, doi: 10.1046/j.1471-4159.2000.0742209.x.

[49] W. E. NOSTRAND, J. P. MELCHOR, G. ROMANOV, K. ZEIGLER, and J. DAVIS, "Pathogenic Effects of Cerebral Amyloid Angiopathy Mutations in the Amyloid β-Protein Precursor," *Ann. N. Y. Acad. Sci.*, 2002, doi: 10.1111/j.1749-6632.2002.tb04824.x.

[50] S. Mead *et al.*, "Familial British dementia with amyloid angiopathy. Early clinical, neuropsychological and imaging findings," *Brain*, 2000, doi: 10.1093/brain/123.5.975.

[51] J. L. Holton *et al.*, "Regional distribution of amyloid-Bri deposition and its association with neurofibrillary degeneration in familial British dementia," *Am. J. Pathol.*, 2001, doi: 10.1016/S0002-9440(10)63993-4.

[52] R. Vldal *et al.*, "A stop-codon mutation in the BRI gene associated with familial British dementia," *Nature*, 1999, doi: 10.1038/21637.

[53] J. L. Holton *et al.*, "Familial Danish dementia: A novel form of cerebral amyloidosis associated with deposition of both amyloid-Dan and amyloid-Beta," *J. Neuropathol. Exp. Neurol.*, 2002, doi: 10.1093/jnen/61.3.254.

[54] A. Charidimou and D. J. Werring, "Cerebral microbleeds: Detection, mechanisms and clinical challenges," *Future Neurology*. 2011, doi: 10.2217/fnl.11.42.

[55] A. A. Gouw *et al.*, "Heterogeneity of small vessel disease: A systematic review of MRI and histopathology correlations," *Journal of Neurology, Neurosurgery and Psychiatry*. 2011, doi: 10.1136/jnnp.2009.204685.

[56] S. H. Kim *et al.*, "Furin mediates enhanced production of fibrillogenic ABri peptides in familial British dementia," *Nat. Neurosci.*, 1999, doi: 10.1038/14783.

[57] S. M. Greenberg *et al.*, "Microbleeds versus macrobleeds: Evidence for distinct entities," *Stroke*, 2009, doi: 10.1161/STROKEAHA.109.548974.

[58] S. Kiuru, O. Salonen, and M. Haltia, "Gelsolin-related spinal and cerebral amyloid angiopathy," *Ann. Neurol.*, 1999, doi: 10.1002/1531-8249(199903)45:3<305::AID-

ANA5>3.0.CO;2-E.

[59] C. Cordonnier, R. Al-Shahi Salman, and J. Wardlaw, “Spontaneous brain microbleeds: Systematic review, subgroup analyses and standards for study design and reporting,” *Brain*. 2007, doi: 10.1093/brain/awl387.

[60] F. Garzuly, T. Wisniewski, F. Brittig, and H. Budka, “Familial meningocerebrovascular amyloidosis, hungarian type, with mutant transthyretin (TTR Asp18Gly),” *Neurology*, 1996, doi: 10.1212/WNL.47.6.1562.

[61] R. B. Petersen *et al.*, “Transthyretin amyloidosis: A new mutation associated with dementia,” *Ann. Neurol.*, 1997, doi: 10.1002/ana.410410305.

[62] “A united approach to vascular disease and neurodegeneration,” *The Lancet Neurology*. 2012, doi: 10.1016/S1474-4422(12)70050-9.

[63] L. Pantoni, “Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges,” *The Lancet Neurology*. 2010, doi: 10.1016/S1474-4422(10)70104-6.

[64] J. M. Wardlaw, C. Smith, and M. Dichgans, “Mechanisms of sporadic cerebral small vessel disease: Insights from neuroimaging,” *The Lancet Neurology*. 2013, doi: 10.1016/S1474-4422(13)70060-7.

[65] S. M. Gregoire *et al.*, “Acute ischaemic brain lesions in intracerebral haemorrhage: Multicentre cross-sectional magnetic resonance imaging study,” *Brain*, 2011, doi: 10.1093/brain/awr172.

[66] E. Auriel *et al.*, “Characteristic distributions of intracerebral hemorrhage-Associated diffusion-weighted lesions,” *Neurology*, 2012, doi: 10.1212/WNL.0b013e318278b66f.

[67] Y. W. Chen *et al.*, “Progression of white matter lesions and hemorrhages in cerebral amyloid angiopathy,” *Neurology*, 2006, doi: 10.1212/01.wnl.0000223613.57229.24.

[68] H. C. O’Donnell *et al.*, “Apolipoprotein E genotype and the risk of recurrent lobar intracerebral hemorrhage,” *N. Engl. J. Med.*, 2000, doi: 10.1056/NEJM200001273420403.

[69] S. M. Greenberg *et al.*, “Apolipoprotein E ε4 is associated with the presence and earlier onset of hemorrhage in cerebral amyloid angiopathy,” *Stroke*, 1996, doi: 10.1161/01.STR.27.8.1333.

[70] Y. C. Zhu *et al.*, “Distribution of white matter hyperintensity in cerebral hemorrhage and healthy aging,” *J. Neurol.*, 2012, doi: 10.1007/s00415-011-6218-3.

[71] E. E. Smith *et al.*, “White matter lesions, cognition, and recurrent hemorrhage in lobar intracerebral hemorrhage,” *Neurology*, 2004, doi: 10.1212/01.WNL.0000142966.22886.20.

[72] D. R. Thal, E. Ghebremedhin, U. Rüb, H. Yamaguchi, K. Del Tredici, and H. Braak, “Two types of sporadic cerebral amyloid angiopathy,” *J. Neuropathol. Exp. Neurol.*, 2002, doi: 10.1093/jnen/61.3.282.

[73] S. Brandner, “Histopathology of cerebral microbleeds,” in *Cerebral Microbleeds: Pathophysiology to Clinical Practice*, 2011.

[74] D. R. Thal *et al.*, “Capillary cerebral amyloid angiopathy identifies a distinct APOE ε4-associated subtype of sporadic Alzheimer’s disease,” *Acta Neuropathol.*, 2010, doi: 10.1007/s00401-010-0707-9.

[75] E. E. Smith, J. A. Schneider, J. M. Wardlaw, and S. M. Greenberg, “Cerebral microinfarcts: The invisible lesions,” *The Lancet Neurology*. 2012, doi: 10.1016/S1474-4422(11)70307-6.

[76] C. M. Kipps, R. R. Davies, J. Mitchell, J. J. Kril, G. M. Halliday, and J. R. Hodges, “Clinical significance of lobar atrophy in frontotemporal dementia: Application of an MRI visual rating scale,” *Dement. Geriatr. Cogn. Disord.*, 2007, doi: 10.1159/000100973.

[77] S. Rosslyn, “Digital Imaging and Communications in Medicine (DICOM) Part 1 : Introduction and Overview,” *Access*, 2006.

[78] P. Mildenberger, M. Eichelberg, and E. Martin, “Introduction to the DICOM standard,” *European Radiology*. 2002, doi: 10.1007/s003300101100.

[79] O. S. Pianykh, *Digital imaging and communications in medicine (DICOM): A practical introduction and survival guide*. 2008.

[80] W. D. Bidgood, S. C. Horii, F. W. Prior, and D. E. Van Syckle, “Understanding and Using DICOM, the Data Interchange Standard for Biomedical Imaging,” *Journal of the American Medical Informatics Association*. 1997, doi: 10.1136/jamia.1997.0040199.

[81] National Electrical Manufacturers Assosiation, “Digital imaging and communications in medicine (DICOM) Part 3: Information object definitions,” *PS (Wash. DC).*, 2011, doi: 10.2495/ICIE131662.

[82] J. F. Dyro, *Clinical Engineering Handbook*. 2019.

[83] L. Bastião Silva, “Medical imaging services supported on cloud.,” 2014.

[84] C. Schwindack *et al.*, “Real-time functional magnetic resonance imaging (rt-fMRI) in

patients with brain tumours: Preliminary findings using motor and language paradigms," *Br. J. Neurosurg.*, 2005, doi: 10.1080/02688690500089621.

[85] D. Leung, X. Han, T. Mikkelsen, and L. B. Nabors, "Role of MRI in primary brain tumor evaluation," *JNCCN J. Natl. Compr. Cancer Netw.*, 2014, doi: 10.6004/jnccn.2014.0156.

[86] L. Ollivier, C. Balu-Maestro, and J. Leclèr, "Imaging in evaluation of response to neoadjuvant breast cancer treatment," *Cancer Imaging*, 2008, doi: 10.1102/1470-7330.2008.0008.

[87] D. Bulas and A. Egloff, "Benefits and risks of MRI in pregnancy," *Seminars in Perinatology*. 2013, doi: 10.1053/j.semperi.2013.06.005.

[88] A. M. Owrangi, P. B. Greer, and C. K. Glide-Hurst, "MRI-only treatment planning: Benefits and challenges," *Physics in Medicine and Biology*. 2018, doi: 10.1088/1361-6560/aaaca4.

[89] J. F. Schenck and E. Unger, "Mri-guided noninvasive ultrasound surgery," *Med. Phys.*, 1993, doi: 10.1118/1.597093.

[90] C. Langkammer *et al.*, "Susceptibility induced gray-white matter MRI contrast in the human brain," *Neuroimage*, 2012, doi: 10.1016/j.neuroimage.2011.08.045.

[91] K. Coyne, "MRI: A Guided Tour," 2018. https://nationalmaglab.org/education/magnet-academy/learn-the-basics/stories/mri-a-guided-tour. [Accessed: 22-Jul-2019.

[92] R. W. Brown, Y. C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging: Physical Principles and Sequence Design: Second Edition*. 2014.

[93] R. M. Henkelman, G. J. Stanisz, and S. J. Graham, "Magnetization transfer in MRI: A review," *NMR in Biomedicine*. 2001, doi: 10.1002/nbm.683.

[94] M. L. Lipton, *Totally Accessible MRI*. 2008.

[95] D. F. W. Group, "Neuroimaging Informatics Technology Initiative," 2018. https://nifti.nimh.nih.gov/.

[96] M. Larobina and L. Murino, "Medical image file formats," *Journal of Digital Imaging*. 2014, doi: 10.1007/s10278-013-9657-9.

[97] V. Patel, I. D. Dinov, J. D. Van Horn, P. M. Thompson, and A. W. Toga, "LONI MiND: Metadata in NIfTI for DWI," *Neuroimage*, 2010, doi: 10.1016/j.neuroimage.2010.02.069.

[98] P. K. Mandal, R. Mahajan, and I. D. Dinov, "Structural brain atlases: Design, rationale,

and applications in normal and pathological cohorts," *J. Alzheimer's Dis.*, vol. 31, no. SUPPL. 3, 2012, doi: 10.3233/JAD-2012-120412.

[99] M. D. C. Valdés Hernández *et al.*, "Close correlation between quantitative and qualitative assessments of white matter lesions," *Neuroepidemiology*, 2012, doi: 10.1159/000341859.

[100] A. A. Gouw *et al.*, "Simple versus complex assessment of white matter hyperintensities in relation to physical performance and cognition: The LADIS study," *J. Neurol.*, 2006, doi: 10.1007/s00415-006-0193-5.

[101] K. S. King *et al.*, "Evaluation of a practical visual MRI rating scale of brain white matter hyperintensities for clinicians based on largest lesion size regardless of location," *Am. J. Neuroradiol.*, 2013, doi: 10.3174/ajnr.A3283.

[102] C. DeCarli, E. Fletcher, V. Ramey, D. Harvey, and W. J. Jagust, "Anatomical mapping of white matter hyperintensities (WMH): Exploring the relationships between periventricular WMH, deep WMH, and total WMH burden," *Stroke*, 2005, doi: 10.1161/01.STR.0000150668.58689.f2.

[103] M. Brant-Zawadzki, G. Fein, C. van Dyke, R. Kiernan, L. Davenport, and J. de Groot, "MR imaging of the aging brain: Patchy White-matter lesions and dementia," *Am. J. Neuroradiol.*, 1985.

[104] R. D. Zimmerman, C. A. Fleming, and B. C. P. Lee, "Periventricular hyperintensity as seen by magnetic resonance: Prevalence and significance," *Am. J. Neuroradiol.*, 1986, doi: 10.1007/978-3-662-11149-9_41.

[105] F. Fazekas *et al.*, "Pathologic correlates of incidental mri white matter signal hyperintensities," *Neurology*, 1993, doi: 10.1212/wnl.43.9.1683.

[106] J. Kynast *et al.*, "White matter hyperintensities associated with small vessel disease impair social cognition beside attention and memory," *J. Cereb. Blood Flow Metab.*, 2018, doi: 10.1177/0271678X17719380.

[107] F. Pasquier, D. Leys, J. G. E. Weerts, F. Mounier-Vehier, F. Barkhof, and P. Scheltens, "Inter-and intraobserver reproducibility of cerebral atrophy assessment on mri scans with hemispheric infarcts," *Eur. Neurol.*, 1996, doi: 10.1159/000117270.

[108] L. Harper, F. Barkhof, N. C. Fox, and J. M. Schott, "Using visual rating to diagnose dementia: A critical evaluation of MRI atrophy scales," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 86, no. 11. pp. 1225–1233, 2015, doi: 10.1136/jnnp-

2014-310090.

[109] J. ODonovan *et al.*, “Does posterior cortical atrophy on MRI discriminate between Alzheimer’s disease, dementia with Lewy bodies, and normal aging?,” *Int. Psychogeriatrics*, 2013, doi: 10.1017/S1041610212001214.

[110] R. R. Davies, C. M. Kipps, J. Mitchell, J. J. Kril, G. M. Halliday, and J. R. Hodges, “Progression in frontotemporal dementia: Identifying a benign behavioral variant by magnetic resonance imaging,” *Arch. Neurol.*, 2006, doi: 10.1001/archneur.63.11.1627.

[111] M. Hornberger, S. Savage, S. Hsieh, E. Mioshi, O. Piguet, and J. R. Hodges, “Orbitofrontal dysfunction discriminates behavioral variant frontotemporal dementia from Alzheimer’s disease,” *Dement. Geriatr. Cogn. Disord.*, 2011, doi: 10.1159/000321670.

[112] P. Scheltens *et al.*, “Atrophy of medial temporal lobes on MRI in ‘probable’ Alzheimer’s disease and normal ageing: diagnostic value and neuropsychological correlates.,” *J. Neurol. Neurosurg. Psychiatry*, 1992.

[113] C. J. Galton *et al.*, “Temporal lobe rating scale: Application to Alzheimer’s disease and frontotemporal dementia,” *J. Neurol. Neurosurg. Psychiatry*, 2001, doi: 10.1136/jnnp.70.2.165.

[114] R. Urs *et al.*, “Visual rating system for assessing magnetic resonance images: A tool in the diagnosis of mild cognitive impairment and alzheimer disease,” *J. Comput. Assist. Tomogr.*, 2009, doi: 10.1097/RCT.0b013e31816373d8.

[115] R. Duara *et al.*, “Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer disease,” *Neurology*, 2008, doi: 10.1212/01.wnl.0000336925.79704.9f.

[116] E. L. G. E. Koedam *et al.*, “Visual assessment of posterior atrophy development of a MRI rating scale,” *Eur. Radiol.*, 2011, doi: 10.1007/s00330-011-2205-4.

[117] C. Cordonnier *et al.*, “Improving interrater agreement about brain microbleeds: Development of the Brain Observer MicroBleed Scale (BOMBS),” *Stroke*, 2009, doi: 10.1161/STROKEAHA.108.526996.

[118] S. M. Gregoire *et al.*, “The Microbleed Anatomical Rating Scale (MARS): Reliability of a tool to map brain microbleeds,” *Neurology*, 2009, doi: 10.1212/WNL.0b013e3181c34a7d.

[119] A. M. J. MacLullich, J. M. Wardlaw, K. J. Ferguson, J. M. Starr, J. R. Seckl, and I. J. Deary, “Enlarged perivascular spaces are associated with cognitive function in healthy elderly men,” *J. Neurol. Neurosurg. Psychiatry*, 2004, doi: 10.1136/jnnp.2003.030858.

[120] G. M. Potter *et al.*, “Enlarged perivascular spaces and cerebral small vessel disease,” *Int. J. Stroke*, 2015, doi: 10.1111/ijs.12054.

[121] Y. C. Zhu, C. Dufouil, A. Soumaré, B. Mazoyer, H. Chabriat, and C. Tzourio, “High degree of dilated virchow-robin spaces on MRI is associated with increased risk of dementia,” *J. Alzheimer’s Dis.*, 2010, doi: 10.3233/JAD-2010-100378.

[122] F. N. Doubal, A. M. J. MacLullich, K. J. Ferguson, M. S. Dennis, and J. M. Wardlaw, “Enlarged Perivascular Spaces on MRI Are a Feature of Cerebral Small Vessel Disease,” *Stroke*, 2010, doi: 10.1161/STROKEAHA.109.564914.

[123] L. O. Wahlund *et al.*, “A new rating scale for age-related white matter changes applicable to MRI and CT,” *Stroke*, 2001, doi: 10.1161/01.STR.32.6.1318.

[124] S. S. F. Yip and H. J. W. L. Aerts, “Applications and limitations of radiomics,” *Physics in Medicine and Biology*. 2016, doi: 10.1088/0031-9155/61/13/R150.

[125] R. Johansen *et al.*, “Predicting survival and early clinical response to primary chemotherapy for patients with locally advanced breast cancer using DCE-MRI,” *J. Magn. Reson. Imaging*, 2009, doi: 10.1002/jmri.21778.

[126] H. J. Baek, H. S. Kim, N. Kim, Y. J. Choi, and Y. J. Kim, “Percent change of perfusion skewness and kurtosis: A potential imaging biomarker for early treatment response in patients with newly diagnosed glioblastomas,” *Radiology*, 2012, doi: 10.1148/radiol.12112120.

[127] S. L. Peng *et al.*, “Analysis of parametric histogram from dynamic contrast-enhanced MRI: Application in evaluating brain tumor response to radiotherapy,” *NMR Biomed.*, 2013, doi: 10.1002/nbm.2882.

[128] F. Tixier *et al.*, “Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer,” *J. Nucl. Med.*, 2011, doi: 10.2967/jnumed.110.082404.

[129] F. Yang, M. A. Thomas, F. Dehdashti, and P. W. Grigsby, “Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer,” *Eur. J. Nucl. Med. Mol. Imaging*, 2013, doi: 10.1007/s00259-012-2332-4.

[130] S. Shinde *et al.*, “Predictive markers for Parkinson’s disease using deep neural nets on neuromelanin sensitive MRI,” *NeuroImage Clin.*, 2019, doi: 10.1016/j.nicl.2019.101748.

[131] C. Salvatore, I. Castiglioni, and A. Cerasa, “Radiomics approach in the neurodegenerative brain,” *Aging Clin. Exp. Res.*, 2019, doi: 10.1007/s40520-019-01299-

z.

[132] A. Carré *et al.*, “Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics,” *Sci. Rep.*, 2020, doi: 10.1038/s41598-020-69298-z.

[133] P. Bhavsar, I. Safro, N. Bouaynaya, R. Polikar, and D. Dera, “Machine Learning in Transportation Data Analytics,” in *Data Analytics for Intelligent Transportation Systems*, 2017.

[134] M. A. Myszczynska *et al.*, “Applications of machine learning to diagnosis and treatment of neurodegenerative diseases,” *Nature Reviews Neurology*. 2020, doi: 10.1038/s41582-020-0377-8.

[135] A. Tagaris, Di. Kollias, A. Stafylopatis, G. Tagaris, and S. Kollias, “Machine Learning for Neurodegenerative Disorder Diagnosis - Survey of Practices and Launch of Benchmark Dataset,” *Int. J. Artif. Intell. Tools*, 2018, doi: 10.1142/S0218213018500112.

[136] S. Rizzo *et al.*, “Radiomics: the facts and the challenges of image analysis,” *European Radiology Experimental*. 2018, doi: 10.1186/s41747-018-0068-z.

[137] J. Lipkova *et al.*, “Personalized Radiotherapy Design for Glioblastoma: Integrating Mathematical Tumor Models, Multimodal Scans, and Bayesian Inference,” *IEEE Trans. Med. Imaging*, 2019, doi: 10.1109/TMI.2019.2902044.

ATTACHMENTS

A. APPROVAL OF THE STUDY BY THE HOSPITAL'S ETHICS COMMITTEE

HOSPITAL DE BRAGA, EPE

Comissão de Ética

20 de Novembro de 2019

Refª 210_2019

Relator: Juan R Garcia

Parecer emitido em reunião ordinária de 07 de novembro de 2019

Nos termos dos Nº 1 e 6 do Artigo 16º da Lei Nº 21/2014, de 16 de Abril, a Comissão de Ética para a Saúde do Hospital de Braga (CESHB) emite o seguinte parecer em relação ao estudo **"Uso de ferramentas Estatísticas e de Inteligência Artificial para revelar biomarcadores para a caracterização e previsão de Angiopatia Amilóide Cerebral"**, de que é investigadora principal Fátima Solange Lima Rezende da Silva, aluna do 2º ano de Mestrado em Bioinformática - Escola de Engenharia, Universidade do Minho, sob orientação do Prof. Dr. Tiago Gil Oliveira – Hospital de Braga/ Escola de Medicina, Universidade do Minho. Atendendo às dificuldades de diagnóstico da patologia em estudo, o presente trabalho revela-se pertinente na medida em que se propõe a identificar marcadores clínicos, com base em métricas de avaliação da atrofia cerebral, quantificação de lesões hipertensas na substância branca, avaliação do alargamento dos espaços perivasculares nos gânglios da base, quantificação de microhemorragias, avaliação de densidade relativa de espaços perivasculares, e por fim, identificação de novos biomarcadores imagiológicos a partir de ferramentas de inteligência artificial.

Trata-se de um estudo retrospectivo em que serão avaliados parâmetros clínicos e analíticos de todos os doentes diagnosticados com Angiopatia amilóide cerebral (AAC), através da consulta do seu processo clínico informatizado (programa Glintt®), e extração dos dados de MRI. Neste dataset encontram-se registos de pacientes seguidos no Hospital de Braga. Todos os dados usados serão anonimizados.

A população alvo do estudo será constituída por doentes do sexo feminino e masculino seguidos no Hospital de Braga, com diagnóstico de AAC ou eventos hemorrágicos cerebrais e exames imagiológicos possíveis de avaliar, assim como de outras doenças neurodegenerativas. Serão ainda recolhidos casos para usar como controlo de outras doenças neurodegenerativas (doença de Alzheimer, demência fronto-temporal e doença de Parkinson). Os dados recolhidos

serão organizados numa base de dados informática e analisados, aplicando os métodos estatísticos adequados na análise dos dados. Metodologias de Inteligência artificial serão aplicadas com o intuito de adjuvar a previsão de diagnóstico de pacientes com AAC provável ou possível, bem como a identificação de biomarcadores inerentes ao desenvolvimento da doença ou de outras doenças neurodegenerativas.

Dada a metodologia do estudo, o modelo de consentimento informado é dispensado.

O estudo não é objecto de financiamento e a sua realização não acarretará encargos adicionais ao HB. Não são referidas situações de conflito de interesses.

A aptidão dos investigadores está demonstrada.

Em suma, o estudo cumpre os princípios da Bioética, pelo que a CE HB nada tem a opor à sua realização.

Documentos anexados.-

1- Requerimento ao CA do HB.

2- Requerimento de apreciação de projeto pela CESHB.

3- Informação do Responsável pela Unidade/Diretor de Serviço.

4- Protocolo do estudo, incluindo:

4.1- instrumentos de recolha de dados

4.2- informação para o participante

4.3.- modelo de Consentimento Informado

5- Curriculum Vitae abreviado dos Investigadores

6- Declaração de Compromisso de Confidencialidade dos investigadores envolvidos

7- Informação sobre financiamento do projeto

8- Informação do EPD do Hospital.

O presidente da Comissão

(Dr. Juan R. Garcia)

HOSPITAL DE BRAGA, EPE
DIRETOR CLÍNICO

(JORGE MARQUES)