Detection and Classification of Neurodegenerative Diseases: A Spatially Informed Bayesian Deep Learning Approach.

Dissertation submitted in partial fulfillment of the requirements for the Degree of Master of Science in Geospatial Technologies

February 24, 2021

David Enrique Payares García
dpayaresg@uni-muenster.de
https://github.com/DavidPayares

Supervised by:
Wiebke Schick
Institute of Geoinformatics
University of Münster

Supervised by:
Jorge Mateu
Department of Mathematics
Universitat Jaume I

Co-supervised by:
Ana Cristina Marinho da Costa
NOVA Information Management School
Universidade Nova de Lisboa

UNIVERSITAT JAUME I

NOVA IMS Information Management School

ifgi Institute for Geoinformatics University of Münster

Acknowledgements

I would like to express my profound gratitude, admiration, and appreciation to my supervisor Jorge Mateu. His excellent support and guidance, his ability to inspire, his vast expertise in statistics, together with plenty of patience and kindness helped me in numberless ways over the course of my Master's thesis, and essentially, to start my journey as a researcher.

I am exceptionally grateful to my supervisor Wiebke Schick for many interesting conversations and for sharing her broad knowledge in neuroscience. Her valuable contributions, constant motivation, and enthusiasm were crucial for conducting this research.

Many thanks and respect are due to my dear friends and colleagues, especially Poshan and David. Your friendship and advice inside and outside academia are invaluable to me. I had the joy and privilege to share this journey with talented and wonderful people who today I consider family, my unexpected family.

I also want to thank the Erasmus Mundus Program for allowing me to cultivate myself in a fascinating, supportive and cheerful environment. Words cannot describe how much this experience has taught me and helped me to grow academically and personally.

Finally, my inmost gratitude goes to my mother, Damary, for all her unconditional love, her support, and her encouragement. I would not be who I am today without her. *Mami, esto es por ti, todo gracias a ti!*

Contents

List of Tables

List of Figures

List of Acronyms

AZ	Alzheimer's Disease
BNN	Bayesian neural Network
CNN	Convolutional Neural Network
CT	Computed Tomography
MCI	Mild Cognitive Impairment
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
PCA	Principal Component Analysis
PD	Parkinson's Disease
PET	Positron Emission Tomography
ROI	Regions of Interest
SBNN	Spatially Informed Bayesian Neural Network

Abstract

Neurodegenerative diseases comprise a group of chronic and irreversible conditions characterized by the progressive degeneration of the structure and function of the central nervous system. The detection and classification of patients according to the underlying disease are crucial for developing oriented treatments and enriching prognosis. In this context, Magnetic resonance imaging (MRI) data can provide meaningful insights into neurodegeneration by detecting the physiological manifestations in the brain caused by the disease processes. One field of extensive clinical use of MRI is the accurate and automated classification of neurodegenerative disorders. Most studies distinguish patients from healthy subjects or stages within the same disease. Such distinction does not mirror clinical practice, as a patient may not show all symptoms, especially if the disease is in an early stage, or show, due to comorbidities, other symptoms as well. Likewise, automated classifiers are partly suited for medical diagnosis since they cannot produce probabilistic predictions nor account for uncertainty. Also, existent studies ignore the spatial heterogeneity of the brain alterations caused by neurodegenerative processes. The spatial configuration of the neuronal loss is a characteristic hallmark for each disorder. To fill these gaps, this thesis aims to develop a classification technique that incorporates uncertainty and spatial information for distinguishing four neurodegenerative diseases, Alzheimer's disease, Mild cognitive impairment, Parkinson's disease and Multiple Sclerosis, and healthy subjects. This technique will produce automated, contingent, and accurate predictions to support clinical diagnosis.

To quantify prediction uncertainty and improve classification accuracy, this study introduces a Bayesian neural network with a spatially informed input. A convolutional neural network (CNN) is developed to identify a neurodegenerative condition based on T_1-weighted MRI scans from patients and healthy controls. Bayesian inference is incorporated into the CNN to measure uncertainty and produce probabilistic predictions. Also, a spatially informed MRI scan is added to the CNN to improve feature detection and classification accuracy.

The Spatially informed Bayesian Neural Network (SBNN) proposed in this work demonstrates that classification accuracy can be increased up to 25% by including the spatially informed MRI scan. Furthermore, the SBNN provides robust probabilistic diagnosis that resembles clinical decision-making and accounts for atypical, numerous, and early presentations of neurodegenerative disorders.

Keywords: Bayesian deep learning, neurodegenerative diseases, classification, magnetic resonance imaging, spatially informed input, uncertainty.

Chapter 1

Introduction

1.1 Context

Neurodegenerative diseases are one of the most concerning group of diseases in the clinical context since they are unpredictable and incurable (Dugger & Dickson, 2017). The World Health Organization (WHO) predicts that by 2040, as many developed countries' aging population keeps growing, neurodegenerative diseases will become the second leading cause of death after cardiovascular disease (Gammon, 2014). Neurodegenerative diseases, also known as neurodegenerative disorders, comprise a heterogeneous group of chronic and irreversible conditions characterized by the progressive degeneration of the structure and function of the central nervous system (Mascalchi, 2005). These disorders attack the brain neurons by causing abnormal functioning and degeneration which eventually leads to their demise. Depending on the amount and severity of the neuronal damage, common symptoms include movement disorders (ataxias) and cognitive impairments (dementias). Thus, identifying and classifying neurodegenerative diseases is not only crucial in clinical practice but also beneficial for developing appropriate treatments and enriching clinical research.

One of the main tools to characterize neurodegenerative diseases is symptomatology. Conventional diagnostic approaches are based on the clinical evaluation of the symptoms; according to their range and frequency, a potential disease can be inferred. However, studies have proved that symptoms reflect only partially the functional changes in affected brain regions caused by the neural loss and the changes in signal transduction and that they can be linked to multiple pathologies (Dugger & Dickson, 2017). For instance, confusion and episodic memory impairment are common manifestations in the early stages of Alzheimer's, Parkinson's and Mild Cognitive Impairment. Likewise, in early disease phases or atypical variants, the differential diagnosis between disorders only based on symptomatology is inaccurate and unreliable (Chetelat & Baron, 2003; Gelfand, 2014; Morris et al., 2001). Against this background, techniques to measure alterations in brain physiology were proposed to support neurodegenerative disease classification.

An outstanding technique, magnetic resonance imaging (MRI), has transformed the study of the human brain, and especially brain diseases since its inception in 1977. MRI is highly effective at identifying anatomical and structural changes in

brain physiology by using the magnetic response of brain tissue molecules (Pool, 2002). Relatives changes in tissue water content indicate atrophy or degeneration of brain structures, and usually, MRI is used to capture these changes. Measures of abnormalities (also called biomarkers) derived from structural MRI have been markedly useful regarding diagnosis and assessment of neurodegenerative diseases (Stoessl, 2012). Clinical trials have exploited MRI potential in the characterization of neurodegenerative chronic diseases such as Parkinson's (Heim et al., 2017), Alzheimer's (Pantano et al., 1999) and Multiple Sclerosis (Filippi & Rocca, 2011) and have established itself as a reliable tool for both the diagnostic process and the monitoring of disease progression.

In MRI, neuronal damage caused by neurodegeneration is represented as atrophies or lesions whose intensities in MRIs differ from the surrounding, apparently healthy tissue. A neurological specialist will study the images and will quantitatively and qualitatively recognize the patterns. Both this information and patient symptoms help to diagnose and assess the diseases. However, MRI contributions are merely limited to visual inspection (Stoessl, 2012). An example of this is the widely used McDonald criteria to identify Multiple Sclerosis. It only considers the number of lesions (visible in MRI) and the number of clinical attacks. Although MRI provides a straightforward way to recognize neurodegenerative disorders, spatial and temporal dissemination of the structural brain alterations in early stages hinders disease characterization through human eye checkup (Heim et al., 2017; A. J. Thompson et al., 2018).

1.2 Motivation

Due to the availability of MRI, there is an increasing interest to apply analytical methods to provide insights into neurodegenerative disorders. Volumetric measures of regions of interest (ROI) (Fox et al., 2005; Lee et al., 2011; Serrano-Pozo et al., 2011; Smeets et al., 2016; Tapiola et al., 2008), voxel-based morphometry (Prinster et al., 2010; Schmitter et al., 2015) and comparison of cortical thickness (Burton et al., 2004; Sailer et al., 2003; P. M. Thompson et al., 2001), have explored MRI data to establish differential diagnostics. However, these methods measure group differences and they are not extendable to individual diagnosis.

Another prominent approach is imaging feature extraction for classification and prediction using deep learning. As opposed to the aforementioned methods, deep learning is a prime approach for pattern recognition. Its potential in MRI lies in feature extraction regardless of scale (Angermueller et al., 2016). Deep learning techniques can extract distinctive features from voxel, vertex, and ROI levels. Multiple deep learning studies have been successful in classifying and predicting neurodegenerative diseases. Rabeh et al. (2016) and Kruthika et al. (2019) successfully distinguish between normal controls and Alzheimer's disease (AZ) patients using supported vector machine (SVM), Aslani et al. (2019) employed two-dimensional convolutional networks to classify MS patients into major clinical stages. Another approach has focused on segmenting the brain changes and relate them to a neurodegenerative disorder (Brosch et al., 2015; Chitradevi & Prabha, 2020; Folle et al., 2019; Jain et al., 2015).

These approaches have achieved high accuracy (over 93%) in binary and (over 79%) multi-class diseases classifications. Nonetheless, most of the studies focus on distinguishing stages of a specific disease or discriminating healthy individuals from patients. Studies of the classification of multiple neurodegenerative diseases are scarce since this entails combining and standardizing diverse MRI sources. Furthermore, accuracy results in deep learning methods such as SVM and convolutional networks are partly suited for clinical prognosis since they cannot account for uncertainty in diagnosis. Brain abnormalities that define neurodegenerative diseases can be present before the onset of symptoms and more than one neurodegeneration process can be found in an individual (Dugger et al., 2014; Dugger & Dickson, 2017). Methods that provide contingent outcomes are needed for diagnosing patients in early stages or with numerous pathologies. Finally, spatial heterogeneity of the changes in the brain metabolism is ignored. Since the brain follows a spatial configuration, so does the neuronal loss that characterizes neurodogerenerative processes. Regional damage obeys a distinct neurodegenerative process, that is, it occurs in specific regions depending on the underlying condition. Xing et al. (2018) found that deep learning approaches for image recognition display higher accuracies and prevent overfitting when spatial heterogeneity is introduced into the data. Overfitting describes features that arise from noise or variance in the data, rather than its underlying distribution leading to loss of accuracy. Meyer et al. (2018) proved that adding simple spatial components such as geographic coordinates to machine learning techniques for spatial interpolation can enhance both the stability and predictions of the model. In MRI, Taschler et al. (2014) incorporated spatial locations of MS lesions to differentiate between MS stages. Also, Illan et al. (2014) demonstrated that MRI data coupled with cerebral spatial information - functional region labeling - increases Alzheimer's disease detection in early stages significantly.

These limitations are the main motivation for this research topic: the application of deep learning techniques to MRI data to identify neurodegenerative disorders. A particular focus of this work lies in the use of deep learning methods together with Bayesian inference to supply probabilistic predictions. This study is focused on the analysis of brain MRI scans and proposes a Bayesian neural network approach that uses spatially informed imaging data to diagnose Alzheimer's, Parkinson's, Multiple Sclerosis, and Mild cognitive impairment. Likewise, this thesis indirectly attempts to enhance the potential use of MRI and to highlight the importance of considering spatially informed data in MRI clinical analysis.

1.3 Related work

In the last 40 years, research on neurodegenerative diseases has blossomed as a result of the introduction of neuroimaging techniques to automatically discover disease biomarkers (Pool, 2002). One of the first researches on the detection of neurodegenerative diseases through imaging data was published thirty years ago, by de Leon et al. (1983). The study used Positron emission tomography (PET) to detect physiological biomarkers in the brain tissues of normally aging subjects and senile dementia patients. The conclusions of de Leon et al. (1983) formed an important basis for further works and stimulated the inclusion of imaging

techniques in the diagnosis of neurodegeneration disorders. In the upcoming years, with the emergence of more sophisticated medical imaging modalities such as MRI, disease detection and classification studies became prominent in major medical journals.

Jack et al. (2003) compiled Alzheimer's disease biomarkers based merely on MRI data. Hotter et al. (2009) and Ge (2006) followed their example and identified structural brain changes for Parkinson's and Multiple Sclerosis. In each study, a group of highly trained neurologists participated to support and ensure their findings. Many studies made notable contributions in expanding neuroimaging biomarkers for the diagnosis of neurodegenerative diseases: Frisoni et al. (2010) in his research work attributed medial temporal lobe atrophy visible in MRI to Alzheimer's disease; Mahlknecht et al. (2010) found that found that atrophy in the temporal, occipital and right frontal lobes is also evident in Parkinson's patients, however, volume and shape differ from the atrophic changes caused by AZ. Meijer et al. (2017) established sound MRI criteria to separate Multiple Sclerosis from Parkinson's and Huntington Disease. Series of studies compiled and analyzed MRI biomarkers to create reliable guidelines to expose neurodegenerative diseases with overlapping symptomatology and atypical behaviors (Agosta et al., 2017; Blamire, 2018; Martin-Macintosh et al., 2016a, 2016b).

Consequently, several semi-supervised techniques emerged in studies to support diagnosis from unexperienced clinicians. Starting from volumetric measurements of regions of interest (ROI), voxel-based morphometry, cortical thickness, and principal component analysis (PCA) have been applied to classify neurodegenerative diseases. In this regard, Lee et al. (2011) highlighted the region of interest approaches while conducting his research in identifying PD with 200 healthy individuals and PD patients. Similarly, Prinster et al. (2010) in their work on identifying gray matter loss in MS patients mentioned how to classify the disease following the different patterns of the brain lesions' regional distribution. Tapiola et al. (2008) distinguished between AZ and MCI by analyzing the volume and spatial distribution of the cortical thickness.

In most of the cases, classification studies were limited to distinctions within the same disorder. Development in MRI preprocessing suggested that many of these methods should be capable of integrating multiple conditions Kuperman, 2004. Studies of more than one disease were gradually introduced. Research was primarily focused on discriminating between neurodegeneration and other pathologies like brain tumors (Kilic et al., 2013) and psychiatric disorders (Lin et al., 2013). Subsequently, Burton et al. (2004) segregated AZ and PD considering voxel-based morphometry in white matter. Similarly, Blazhenets et al. (2019) used PCA to predict the development of AZ and to recognize dementia variations such as MCI. These methods' contributions grounded in brain imaging research but were bounded by the modest number of patients and the inaccuracy of the medical diagnosis.

In the last years, advances in machine learning and pattern recognition methods capable of handling high dimensional data have enabled the development of new diagnostic tools derived from morphological analysis in MRI. From a clinical point of view, the main challenge resides in recognizing signatures of a disease

in the structural images that allow discriminating pathological from healthy subjects. Since important features can be automatically learned by these techniques, information extraction from MRI in advance is not necessary. Research work by Díaz et al. (2010) and Sood and Khandnor (2019) have accentuated the potential of deep learning in learning characteristic hallmarks for disease designation. Several deep learning approaches have been proposed to automatically classify patients with neurodegenerative disorders. Magnin et al. (2009) used SVM to characterize patients with AZ from elderly controls. Aslani et al. (2019), Karaca et al. (2016), Khan et al. (2020) and Sweeney et al. (2014) developed convolutional neural networks to separate normal control from AZ, PD, and MS. Other approaches have targeted differential diagnosis within dementia (AZ and MCI) (Rabeh et al., 2016) and MS types (Narayana et al., 2020).

In deep learning methods, measuring uncertainty associated with their predictions is challenging (Valentin Jospin et al., 2020). This is particularly alarming in medical diagnostic applications, where silent failure can lead to dramatic results. Tishby et al. (1989) proposed Bayesian inference to represent probabilities in layered networks. In 1996, Neal (1996) in his doctoral thesis elaborated on Bayesian learning and prediction uncertainties for complex neural networks. Currently, Bayesian research in deep learning is infrequent and its application in many fields remains unexplored. Consequently, few studies have been conducted in medical imaging. McClure et al. (2018) performed brain segmentation and measured its accuracy for quality control purposes. A novel method for skull stripping in nonhuman primates using bayesian convolutional networks was presented by Zhao et al. (2018). Both Herzog et al. (2020) and Khairnar (2020) used Bayesian neural networks for detection of strokes and breast cancer respectively.

Similarly, a few attempts of the inclusion of spatial information in neuroimaging analysis, outside deep learning theory, have been conducted. Marschallinger et al.'s (2017) research diagnosed MS from MRI data with geostatistics by adding the coordinates of the lesion. Taschler et al. (2014) modeled the spatial density of MS lesions to distinguish MS stages using spatial point patterns. For AZ diagnosis, Illan et al. (2014) subdivided the spatial components of the brain (regions) and modeled their dependencies using a Bayesian network. All studies mentioned the correlation between high diagnosis accuracy and spatial information derived from the brain morphology.

1.4 Aim and Objectives

This study aims to detect and classify neurodegenerative diseases using spatially informed MRI data through deep learning models coupled with Bayesian inference. Additionally, the thesis indirectly seeks to enhance the potential use of multiple MRI sources and to highlight the importance of considering spatially informed data in MRI clinical analysis.

This study will attempt to answer the following questions:

- To what extent can the characterization of neurodegenerative diseases be improved with the integration of spatial information in MRI data?
- How can uncertainty be included in deep learning models to represent probabilistic predictions?
- How does spatially informed data improve the classification accuracy of a deep learning model?

1.5 Outline

The thesis is organized as follows: Chapter 2 provides a theoretical background of the medical and methodological concepts used in this thesis, which are all related to MRI data analysis in neurodegenerative diseases. In Chapter 3, the methodology of the proposed classifier is described in detail. Chapter 4 presents the data on which the model is tested, and describes the experimental setup. In Chapter 5, the results of the experiments are interpreted and discussed. Finally, Chapter 6 concludes the results of this thesis.

Chapter 2

Background

This chapter presents the background material relevant to the subsequent chapters of the thesis. Section 2.1 covers the technical details of magnetic resonance imaging, including image acquisition and processing methods. Section 2.2 provides medical background on neurodegenerative diseases. Section 2.3 summarises the methodology underlying the convolutional neural network models and methods presented later in Chapter 3. Section 2.4 reviews some aspects of classification including validation and, introduces a Bayesian approach for prediction uncertainty.

2.1 Neuroimaging: Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) has become an important noninvasive imaging technique to produce three-dimensional images of the anatomy and physiological processes of the body. MRI employs the body's natural magnetic properties to locate and map the amount of water in the tissues resulting in a detailed image. MRI has found several applications in the field of medical science since it provides a unique contrast between organic soft tissues and high spatial resolution (Kuperman, 2004).

In magnetic resonance imaging (MRI) of the brain, structural MRI (sMRI) is used to visualize details of the brain anatomy which are unchanging for short periods. It has become a predominant tool to detect and represent abnormalities in the physical appearance of the brain and to track changes over time. Structural MRI is highly effective at identifying lesions and cerebral structural changes and therefore is widely used in the diagnosis and assessment of neurological diseases such as Multiple Sclerosis, Alzheimer's, Parkinson's, and schizophrenia (Stoessl, 2012). Although other modalities like Computed Tomography (CT), Positron Emission Tomography (PET), and Electroencephalography (EEG) are also commonly used in brain anatomy and brain activity imaging, they present poor spatial resolution and, exempt for the EEG, dangerous exposure to ionizing radiations (Afaq et al., 2017; Noachtar & Rémi, 2009). sMRI is considered the backbone for detecting neurological disorders in the clinical context.

2.1.1 Image acquisition

Magnetic resonance imaging relies on the nuclei spins of the hydrogen nuclei present in high quantities in the body's tissue. MRI forces the alignment of the natural random nuclei spins' orientation - a phenomenon known as resonance - and its subsequent return to equilibrium - relaxation phenomenon - by using a large pulsed external magnetic field. An MRI sequence is characterized by two parameters. The Echo Time (TE) defines the time between the Radiofrequency (RF) pulse and the signal measurement, and the Repetition Time (TR) represents the time between two RF pulses. When the TR between consecutive pulses of the external magnetic field is short, a the T_1-weighted imaging sequence is produced. T_1-weighted MRI uses the spins relaxation time after which the longitudinal component of the nuclear spins have fully returned to their equilibrium orientation. In contrast, T_2-weighted scans are obtained by using a long TE. T_2-weighted MRI employs the spin to spin relaxation time which is the time needed for the transverse component of the signal to decay exponentially from its initial value (Currie et al., 2013). Depending on the tissues, relaxation times vary. In T_1 modality, only fatty tissue exhibits bright responses given that the relaxation time for water is about five times larger than for fat. Both fatty and water-based tissue (mainly protons) are bright in the T_2 modality (Stanisz et al., 2005).

Visualization of the affected and normal brain tissue only on T_1 or T_2 images is far from optimal. The anatomical changes appear hypo-intense or normal on T_1-weighted MRI and are therefore hard to detect. A common standard procedure for neurodegenerative diseases patients involves the use of a contrast agent, usually Gadolinium, to enhance the contrast in the brain tissues (Xiao et al., 2016). Gadolinium coupled with T_1-weighted MRI improve the visibility of lesions, inflammations, abnormal structure, and changes in the brain tissues.

2.1.2 Preprocessing

After the acquisition of an MRI scan, due to the nature of the data, it needs to be processed before any statistical analysis, especially if the study involves multiple sources, multiple scans, and/or multiple subjects. The collection of transformations from the data is called imaging preprocessing. There are numerous steps in imaging preprocessing commonly used to reduce noise, adjust and standardize the data. The steps' order and relevance depend on the study aim and the neurological criteria.

Basic prepossessing steps comprise image harmonization corrections to errors derived from images acquired by multiple scans. Noise and underemphasized features are amended using image contrast and inhomogeneity correction. Spatial registration addresses errors that occur when scans from different subjects are used and that stem from the different sizes and shapes of the brains. Errors occurring due to head motion during the scanning can be corrected in that way, too. It aligns the scans from each subject to a common brain template, usually, the brain atlas templates provided by Montreal Neurological Institute (MNI). Additionally, intensity normalization is a recommended technique to standardize the voxel intensities from every scan to be comparable among subjects and scanners.

Preprocessing techniques allow comparing and analyzing brain MRI of different diseases regardless of patient and scanner particularities.

2.2 Medical background: Neurodegenerative Diseases

In neuroanatomy, three main physiological components of the brain are distinguished: white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). WM consists of myelinated axons that transmit electrical impulses between various regions of the brain. GM contains mostly neuronal cell bodies and unmyelinated neurons that are involved in brain functions like sensory and motor control. The CSF derives from blood plasma and provides basic mechanical and immunological protection to the brain (Mtui et al., 2016).

Neurodegenerative diseases comprise a variety of sporadic as well as inherited conditions characterized by progressive loss and/or death of defined populations of neurons (Mascalchi, 2005). Hallmark pathology of neurodegenerative diseases is the damage of the neurons and their connections located in the GM and WM tissue. Since neurons do not reproduce or replace themselves, their degeneration results in incurable and debilitating conditions. The neurological damages manifest in form of lesions, inflammations, and atrophy that can occur across the brain and spinal cord (Gitler et al., 2017).

Neurodegenerative diseases can be broadly classified by their clinical presentations based on the anatomic vulnerable regions (brain areas of the affected nerves) and their resulting clinical presentations such as movement disorders and cognitive or behavioral disorders (Dugger & Dickson, 2017). Examples of neurodegenerative diseases are Alzheimer's disease, Parkinson's disease, and Multiple Sclerosis. These diseases are diverse in their pathophysiology – with some causing memory and cognitive impairments and others affecting motor and sensory functions (Dickson, 2018; Duyckaerts et al., 2009; Gelfand, 2014). Depending on the disease symptoms can include memory loss, muscle function loss, anxiety, cognitive impairment, digestive and sexual dysfunction, mood disorders, and chronic pain.

Currently, no neurodegenerative disease is curable. Most treatments available manage the symptoms or halt the progression of the disease to slow the cognitive and functional deterioration.

2.2.1 Diseases

Alzheimer's Disease

Alzheimer's Disease (AZ) is a specific neurodegenerative disease and is the most common cause of dementia in old people contributing to 60-70% of cases. AZ is one of the major causes of disability and dependency among elderly people worldwide. AZ is characterized by reducing the hippocampal volume and causing progressive cerebral atrophy. The cause for AZ is not yet fully understood. However, studies show that it is associated with the beta-amyloid deposits and

neurofibrillary tangles in the cerebral cortex and subcortical gray matter, leading to neuronal and synaptic loss (Wang et al., 2017). AZ clinical hallmarks include - but are not limited - to loss of memory, spatial perception deficit, aphasia, apraxia and personality and mood changes (Duyckaerts et al., 2009). AZ is inexorably progressive and fatal within 5 to 10 years.

Parkinson's Disease

Parkinson's Disease (PD) is a common neurodegenerative disorder that causes progressive damage in different regions of the brain. The pathologic feature that correlates with signs and symptoms of PD is the neuronal loss in the substantia nigra, more specifically, a portion of this nucleus called the pars compacta. Neuronal degeneration in the substantia nigra leads to a reduction of dopamine levels. This limits functionality in the regulation of major brain structures involved in the control of movement. PD pathological processes manifest as multiple system atrophy, progressive supranuclear palsy, and cerebrovascular affectations (Dickson, 2018). The triggering event in PD is unknown, but recent studies suggest a role for loss of nuclear membrane integrity. PD's main symptoms are bradykinesia, postural instability, resting tremor, and rigidity. Other symptoms include emotional disorders and cognitive deterioration (Antony et al., 2013).

Multiple Sclerosis

Multiple Sclerosis (MS) is one of the most concerning neurodegenerative diseases in the clinical context since it is unpredictable and it affects not only the elder population - as other chronic diseases like Alzheimer's or Parkinson's - but also young adults (Cohen & Rae-Grant, 2012). MS is a chronic inflammatory and demyelinating disease of the central nervous system that manifests as lesions in affected regions of the brain and the spinal cord. MS lesions occur in the brain's white matter which connects neurons in the gray matter leading to limited neuronal conductivity along neurons across the brain and between the brain and other organs of the body. MS clinical manifestations are not exclusively neurological (motor weakness, visual impairment, diplopia, dysarthria, ataxia, etc.). Recent studies have shown that MS is also related to cognitive deficits such as attention and concentration disorders, slow information processing, loss of memory, and executive functions (Noseworthy et al., 2000).

Mild Cognitive Impairment

Mild cognitive impairment (MCI) is rapidly becoming one of the most common clinical manifestations affecting the elderly population. MCI is the stage between the expected cognitive decline of normal aging and the more serious decline of dementia, Alzheimer's disease (Morris et al., 2001). MCI generally reflects an intermediate state that shares to a certain degree the pathological characterization of AZ in vulnerable cortical regions including memory-related regions like the hippocampus, and the visual association cortex (Riley et al., 2002; Schneider et al., 2009). Symptomatology of MCI is strongly related to AZ clinical manifestations such as memory loss and language deficits. However, MCI usually does not present mood and personality changes. Symptoms are not severe enough to

significantly interfere with MCI patients' daily life and usual activities.

2.2.2 Diagnosis

The diagnostic tool commonly used to assess neurodegenerative diseases is based on the clinical evaluation of the symptoms. Also, traditional structural neuroimaging techniques, such as magnetic resonance imaging (MRI), help diagnose neurodegenerative diseases (Stoessl, 2012); their clinical usage provides promise in the diagnosis and monitoring of disease progression. Other diagnostic methods comprise markers detectable in readily accessible tissues (blood or saliva) (Gelfand, 2014). Genetic markers are as well employed for diagnosing familiar forms of neurodegenerative diseases and they include amyloid precursor protein in AZ, gene mutations in α-synuclein protein in PD or IgG, and IgM antibodies in CSF in MS (Dickson, 2018; Dugger & Dickson, 2017; Noseworthy et al., 2000). The combination of multiple techniques seems to improve disease diagnosis.

The main classification for the diagnosis of AZ has been the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS–ADRDA) criteria (McKhann et al., 1984). These criteria combine clinical and neuropathological patterns and assign diagnoses based on the stage of symptoms. Since atrophy caused by AZ seems to be an inevitable, inexorably progressive concomitant of neurodegeneration, MRI-based measures of atrophy are regarded as valid markers of AZ state and progression for several reasons (Chetelat & Baron, 2003). MCI is considered an intermediate state between normal cognition and dementia, and therefore AZ biomarkers also help to diagnose it.

In PD, clinical criteria are key; there is no definitive test for diagnosis. Genetic mutations or variants, neuroimaging, and cardinal motor test features are potential biomarkers that improve diagnosis and allow the identification of persons at risk (Jankovic, 2008). Symptomatology is generally considered the cardinal signs of PD: PD is mainly diagnosed based on the symptoms. MRI helps to detect PD due to the reduction of the size of nigral regions caused by the disease.

There are no markers specific for MS diagnosis. Diagnosis mainly depends on medical history and neurological examination. MS is characterized by attacks which are defined as new neurological deficits associated with an anatomical brain localization that persist for more than 24 hours (Gelfand, 2014). MS patients present attacks only visible in magnetic resonance imaging.

2.2.3 MRI criteria

Neuronal degeneration develops into anatomical changes visible in magnetic resonance imaging. Although MRI is one of the main sources for the diagnosis of neurodegenerative diseases, the correlation between physical brain abnormalities and clinical symptoms is merely approximate. The reason for this are the neural plasticity and regeneration mechanisms that atone damage (Horner & Gage, 2000). MRI can help to detect a neurodegenerative disease, but it can not associate brain-damaged regions to any specific disease symptom (Stoessl, 2012).

MRI sequences show different aspects of each disease activity and therefore different MRI criteria are adopted. T_1-weighted and T_2-weighted scans are the most common sequence to identify the damage caused by the diseases. T_1-weighted sequences display most of the biomarkers associated with AZ, PD, MS, and MCI. T_2-weighted scans are used to assess the overall disease burden. However, T_2-weighted sequences suffer from high visual variability in regions with atrophies and/or inflammations. This can cause equivocal interpretations in neurodegenerative diseases MRI (Frisoni et al., 2010; Mangia et al., 2013; Sahraian et al., 2010).

Table 2.1: Summary of biomarkers in T_1-weighted sequences.

Disease	Appearance on MRI	Description in diagnostic criteria
AZ	hypointense (dark)	Atrophy in temporal lobes and medial parietal cortex
PD	hyperintense (bright)	Atrophy in nigral regions
MS	hyperintense (bright)	Lesions across the brain and spinal cord
MCI	hypointense (dark)	Mild atrophy in temporal lobes and medial parietal cortex

The presence of atrophy in the medial temporal lobes is an established and validated biomarker for the diagnosis of AZ, PD, and MCI. The simplest way to assess atrophy is by visual inspection of coronal T_1-weighted MRI. T_1-weighted scans for PD highlight vascular, space-occupying or demyelinating lesions within the basal ganglia or brainstem. In MS, T_1-weighted sequences display chronic or persistent hypo-intense lesions indicating permanent axonal loss and neuronal damage.

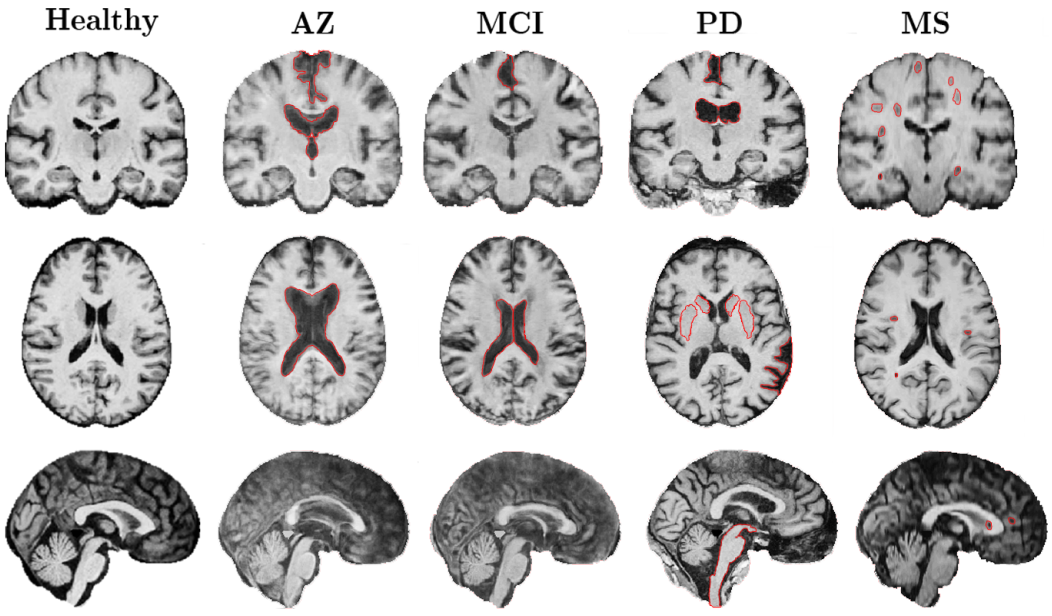


Figure 2.1: MRI most common biomarkers for degenerative patients and healthy controls.

2.3 Methodological background: Deep Learning

Deep Learning (DL) is a sub-field of Machine Learning Research, introduced as the intersection of research areas of neural networks, artificial intelligence, graphical modeling, optimization, and pattern recognition. DL involves learning multiple levels of representation and abstraction that help to make sense of data such as images, sound, and text (Arel et al., 2010). DL techniques are inspired by the functioning of human brains and are based on the principle of parallel processing. They consist of multiple layers or stages of nonlinear information processing and methods for supervised or unsupervised learning of feature representation at successively higher, more abstract layers (Deng & Yu, 2014). DL enables exploiting complex labeled or unlabeled data to learn distributed and hierarchical feature representations and to make effective predictions about them. Many DL networks such as convolutional neural networks, deep neural networks, recurrent neural networks and recursive neural networks have been successfully applied to diverse fields including computer vision, natural language processing, medical image analysis. DL architectures have proven high recognition and classification accuracy and have outperformed traditional machine learning and statistical techniques (Schmidhuber, 2015).

2.3.1 Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a special kind of neural network usually applied to analyzing visual imagery. CNNs are models inspired by biological neural networks that replace dense matrix multiplications with convolutions and filters in at least one layer. This allows representing the input data in hierarchical structures (Schmidhuber, 2015). CNNs are specialized neural networks that exploit data topologies, e.g., images as a grid of pixels. Due to its properties, CNN is arguably one of the most important current image classification methods. Traditional neural networks are characterized by one-dimensional architectures composed of Fully Connected (FC) layers, whereas CNN is composed of convolutional layers where hidden activation is calculated by multiplying small local inputs against weights, and spatially, local association is exploited by applying a local pattern of connectivity between adjacent neural layers (Lopez Pinaya et al., 2019). The resulting weights are then shared across the entire input locations and generate a feature map as an output. Finally, the collection of outputs of each convolutional layer is shared with the next layer. Sharing the output from neuron to neuron allows high level feature extraction and uncovers hidden latent relations in the data.

2.3.2 Architecture of a CNN

A CNN typically consists of a set of alternatively stacked layers that transform input volumes into outputs using differentiable functions. A convolutional block - also called convolutional layer - is composed by convolution, activation, and pooling layers (Figure 2.2). A convolutional layer is followed by one or more fully connected layers for high-level reasoning. Fully connected layers assign feed-forwarded scores to the n-dimensional inputs. At last, a loss layer measures

the deviation between the predicted and true labels. A detailed explanation is presented below.

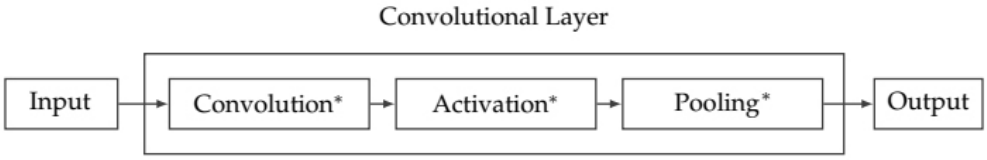


Figure 2.2: A CNN is viewed as a stack of convolutional layers formed by convolution, activation and pooling layers.

Convolution

The convolution layer represents the core concept of a CNN: to consider the spatial structure of the n-dimensional input. A convolution layer consists of filters - namely kernels - which extract features learned by the network over the images. A convolution layer produces different feature activation values at each location in the image to generate a feature map as output (Deng & Yu, 2014). From a mathematical point of view,

$$y(t) = (x * w)(t) = \int x(\tau)w(t-\tau)d\tau \tag{2.1}$$

where $t \in \mathcal{R}, \tau \in \mathcal{R}, x : \mathcal{R} \to \mathcal{R}$, and $w : \mathcal{R} \to \mathcal{R}$. The resulting function $y : \mathcal{R} \to \mathcal{R}$ after applying the convolution operator, denoted with an asterisk , to the functions x and w, is defined as the integral of the product of both functions after one is reversed and shifted (τ). The function x is usually referred to as the input and w is a weighting function known as kernel. The output y is named feature map.

In a computer the inputs are discrete. The implementation of a convolution layer in a discrete space can be written as

$$y(t) = (x * w)(t) = \sum_{\tau=-\infty}^{\tau=\infty} x(\tau)w(t-\tau) \tag{2.2}$$

where t can only take integer values.

In practice, the input x and the kernel w are not functions but multidimensional discrete arrays - also called tensors - of size n-dimension.

Finally, a discrete convolution can be redefined as a finite summation over tensor elements. For instance, a 3D convolution is defined as

$$Y(i,j,k) = (X * W)(i,j,k) = \sum^{m}\sum^{n}\sum^{l} X(m,n,l)W(i-m, j-n, k-l) \tag{2.3}$$

where Y, X, and W are three-dimensional tensors denoting the output feature map, the input three-dimensional data and the kernel function respectively.

Activation

The purpose of the activation layers is to introduce non-linearities into the network to enable non-linear representations. Activation functions increase the non-linear properties of the decision function learned by the network, without affecting the output of the previous convolution layer. Activation layers are often placed right after a convolution occurs, but many combinations are possible.

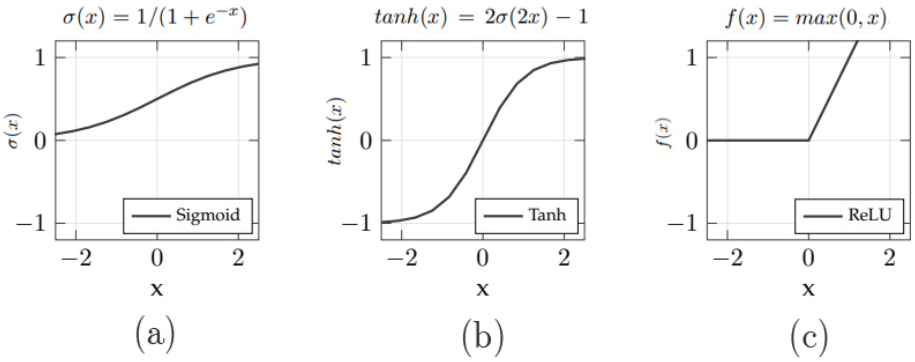


Figure 2.3: Activation functions: Sigmoid(a), Tanh (b), and ReLU (c)

The most common activation functions are shown. in Figure 2.3. Sigmoid and Tanh are similar functions, they squash the inputs into a non-linear space of range [0,1] and [-1,1] respectively. However, they are not commonly used because they saturate gradients and their outputs are zero-centered. In contrast, Rectified Linear Unit (ReLU) is the most common activation function for CNNs. It thresholds the negative inputs to zero (Csáji, 2001). ReLU has non-saturating properties and low computationally cost.

Pooling

Pooling layers are used to simplify the feature maps produced by the previous convolution and activation layers. In pooling, predefined functions are applied to summarize the signal and spatially preserve discriminant information (Schmidhuber, 2015). There are many pooling operations, the most common being the max pooling operation. Max pooling selects the maximum value of the neurons for a specific region.

Other functions such as average pooling or L2-norm pooling are used as well. These are characterized by extracting the average (or weighted average) value of the neurons. A pooling layer parameter comprises the size of the window and stride (number of steps between the consecutive convolutions).

Fully Connected

After a convolutional layer composed by convolution, activation, and pooling operations extracts low-level features from the data, the high-level reasoning of the network is interpreted by fully connected layers. A Fully Connected (FC) layer takes the output of the previous layers, flattens them and turns them into an n-dimensional vector where n is the number of classes or labels to classify. The

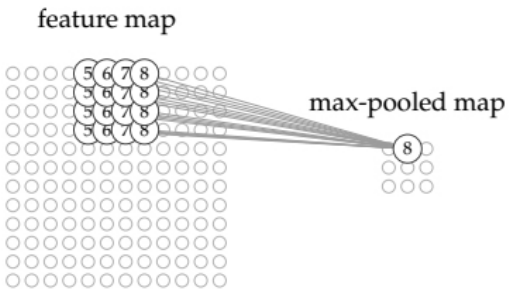


Figure 2.4: 4 x 4 Max pooling operation

output produced by this layer is the probabilistic score that a feature belongs to one of the classes. In a FC layer, neurons are connected to every neuron in the previous layers. FC layers are analogous to the layers in a typical multilayer perceptron (MLP) (Lopez Pinaya et al., 2019).

Classifier

The classifier layer calculates the output class probabilities for the input FC layers. It computes the deviation between the true labels of the input and the label predicted by the network (Deng & Yu, 2014). This is known as cross-entropy loss. The most common classifier for multi-class classification is the softmax function which transform the logits into probabilities.

The softmax is a normalized exponential function that transforms the k-dimensional CNN output vector Y of values to another k-dimensional vector $S(Y)$ of values in the range [0, 1] that add up to 1. $S(Y)$ represents a categorical probabilistic distribution. The softmax function is defined as:

$$S\left(Y_i\right) = e^{Y_i} / \sum_{j=0}^{k} e^{Y_j} \tag{2.4}$$

2.3.3 Training a CNN

The training of a CNN is the optimization process to find the set of weights that minimize the differences between the predicted classes and the ground truth classes, i.e, the actual classes of the data.

Loss

Loss is the function that minimizes the deviation between the predicted and the real classes. The loss function determines the optimal set of parameters that reduce model error (Lopez Pinaya et al., 2019). These parameters are known as the weights and bias learned by the network. In multi-class classification problems, the SoftMax loss is computed using the cross-entropy, a natural way

to measure the difference between two probability vectors.

$$L(S(Y), T) = -\sum_{i=0}^{k} T_i \log(S_i) \tag{2.5}$$

L is the loss function of a categorical probabilistic distribution $S(Y)$ and true class labels T defined through the ground truth classes. The loss function denotes the negative log-likelihood of the i given instances.

Regularization

One of the most important aspects of CNNs is how well they generalize to unseen data. When the loss in the training set is much higher than the one in the test set, the model is overfitting and thus achieving poor generalization.

CNNs are expressive models particularly prone to overfitting given their non-linear hidden nature and a large number of parameters. In this context, a well-defined CNN relies on the selection of architectures that maximize generalization and the development of techniques to prevent overfitting. The process of avoiding or preventing overfitting is called regularization and some of the most successful techniques include L1 and L2 regularization and dropouts.

Dropout Dropout is the most applied technique to avoid overfitting in convolutional neural networks. Dropout operations are usually performed after fully connected layers (Srivastava et al., 2014). The concept is to randomly drop units and connections between fully connected layers with a determined probability (Figure 2.5). Dropout prevents units from co-adapting and reduces overfitting significantly.

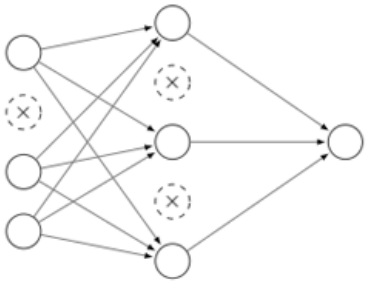


Figure 2.5: Dropout technique applied to a fully connected network.

Optimization

Optimization algorithms exist to force fast convergence in the training process. Gradient descent is an iterative optimization technique to minimize the loss by updating the weights of neurons in every iteration through a backpropagation mechanism. A CNN models a function that processes a training data set using a set of weights for the network layers and computes the loss for that input. This process is called the $forwardpass$. The loss is then used to compute the

gradients, and the gradients are used to update the weights. This process is called *backwardpass*.

Famous algorithms include the Stochastic Gradient Descent with momentum (SGD) and the Adaptive Moment Estimation (Adam).

Stochastic Gradient Descent Stochastic Gradient Descent (SGD) is a specific case of gradient descent that estimates the loss and gradient by using a mini-batch of n training samples per iteration (epochs). In SGD, a learning rate α determines the step size in the direction of the gradient of the loss function. SGD with momentum is computed by adding a fraction γ of the update vector of the last step M_{t1} to the current update vector.

$$M_t = \gamma M_{t-1} + \alpha \, \nabla f\left(W_t, X, T\right) \tag{2.6}$$

Adam Adaptive Moment Estimation (Adam) proposed by Kingma and Ba (2015) is a optimizer based on mini-batch gradient descent that computes adaptive learning rates for each parameter. It keeps both the exponentially decaying average of past gradients M'_t and the past squared gradients M_t

$$\begin{aligned} M'_t &= \gamma' M'_{t-1} + (1-\gamma') \, \nabla f\left(W_t, X, T\right) \\ M_t &= \gamma M_{t-1} + (1-\gamma) \nabla f^2\left(W_t, X, T\right) \end{aligned} \tag{2.7}$$

where M_t is the momentum in iteration t, γ is a fraction of the update vector of the last step M_{t-1}. ∇ the gradient and $f(*)$ the loss function depending on the weights W, the training set X and the true labels T.

The estimates of the first and second moment of the gradients are defined as:

$$\hat{M}'_t = \frac{M'_t}{1-\gamma'_t}, \quad \hat{M}_t = \frac{M_t}{1-\gamma_t} \tag{2.8}$$

Finally, these biased corrected moment estimates are used to update parameters using the Adam update rule:

$$W_{t+1} = W_t - \frac{\alpha}{\sqrt{\hat{M}_t + \epsilon}} \hat{M}'_t \tag{2.9}$$

where α represents the learning rate.

2.3.4 Bayesian Neural Networks

Bayesian neural networks (BNN) combine neural network and stochastic mo-dels with the latter forming the core of this integration (Neal, 1996). Traditional Neural Networks focus on representing data through arbitrary functions. In contrast, statistical models (also called probabilistic models) represent the data based on direct specifications over the model parameters. BNN introduces stochastic

components in a neural network to simulate multiple possible models with an associated probability distribution. This allows explicit uncertainty inference in the underlying processes of the neural network (Valentin Jospin et al., 2020). BNN can produce uncertainties on the predictions and draw the distributions of the parameters learned by the model (Bishop, 1997).

Let θ denote the models which represent the weights W and the biases b of a neural network NN. Be $p(\boldsymbol{\theta})$ the probability distribution of all possible $\boldsymbol{\theta}$ models. $\boldsymbol{\epsilon}$ represents the random noise derived from a NN being an approximation. The more similar the models are, the lower is the uncertainty.

$$\begin{aligned} \boldsymbol{\theta} &\sim p(\boldsymbol{\theta}) \\ \boldsymbol{y} &= NN_{\theta} + \boldsymbol{\epsilon} \end{aligned} \tag{2.10}$$

BNN is an stochastic artificial neural network trained using Bayesian inference (MacKay, 1992).

Bayes' Theorem

Bayes' theorem describes the probability as a measure of belief in the occurrence of events and states that the prior beliefs influence the posterior ones. The Bayes theorem can be stated as

$$P(H \mid D) = \frac{P(D \mid H)P(H)}{P(D)} \tag{2.11}$$

where H is the hypothesis and D is data. $P(D \mid H)$ is the likelihood, $P(H)$ the prior belief and $P(D)$ the evidence. $P(H \mid D)$ is considered the posterior belif.

A BNN understands the neural network architecture as a functional model y that depends on data x and the stochastic model as the prior distribution over $p(\boldsymbol{\theta})$ and prior confidence in the predictions of the model $p(y \mid x, \boldsymbol{\theta})$. $p(\boldsymbol{\theta})$ can be considered to be the hypothesis H and the training set is the data D. Applying independence between the model parameter and the inputs,

$$P(\boldsymbol{\theta} \mid D) = \frac{P(D \mid \boldsymbol{\theta})P(\boldsymbol{\theta})}{P(D)} = \frac{P(D \mid \boldsymbol{\theta})P(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} P(D, \boldsymbol{\theta}')P(\boldsymbol{\theta}')d\boldsymbol{\theta}'} \tag{2.12}$$

A marginal probability distribution that measures the uncertainty of the model can be calculated once the Bayesian posterior is computed. The marginal probability distribution $p(y \mid x, D)$ is indirectly sampled from equation 2.10. $\boldsymbol{\theta}$ is sampled from the variational distribution $q_{\Phi}(\boldsymbol{\theta})$ defined in the section below.

$$p(\boldsymbol{y} \mid \boldsymbol{x}, D) = \int_{\theta} p\left(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}'\right) p\left(\boldsymbol{\theta}' \mid D\right) d\boldsymbol{\theta}' \tag{2.13}$$

BNN gives the relative probabilities of each class as an average prediction. This can also considered as a measure of uncertainty (MacKay, 1992).

$$\hat{\boldsymbol{p}} = \frac{1}{|q_\Phi(\theta)|} \sum_{\boldsymbol{\theta}_i \in q_\Phi(\theta)} NN_{\boldsymbol{\theta}_i}(\boldsymbol{x}) \tag{2.14}$$

Priors

In Bayesian inference, a model must be fed with prior information before it observes the data (Fan et al., 2008). In BNN setting prior distribution is a difficult task due to neural network's complexity and a large number of parameters (Bishop, 1997).

BNN assumes a prior distribution on the model parameters, e.g, a Gaussian prior distribution over the model weights. Assuming a normal prior on the coefficient of the network is known as a good practice in over-parametrized neural networks (Kingma & Welling, 2014). According to the central limit theorem, regardless of the true underlying distribution, samples taken from data will steadily approximate a normal distribution. Likewise, as the number of hidden units increases, the prior over functions implied by such priors converges to a Gaussian process(Neal, 1996).

$$p(\boldsymbol{\theta}) = \boldsymbol{\mathcal{N}}(\mathbf{0}, \sigma\mathbf{I}) \tag{2.15}$$

Priors are closely related to deep learning regularization techniques (Valentin Jospin et al., 2020). Optimizing model weights to minimize the loss function through a regularization function is equivalent to finding the most likely weights under a posterior distribution given a prior distribution.

Variational Inference

Variational inference is a method for learning an approximation of the posterior distribution (Kingma & Welling, 2014). Variational inference is very popular in BNN since it is a scalable method that performs particularly well in large neural network architectures (Valentin Jospin et al., 2020).

Variational inference as opposed to its counterpart Markov Chain Monte Carlo (Hoffman et al., 2013), is not an exact method. Instead of sampling from the exact posterior distribution, a variational distribution $q_\Phi(\boldsymbol{\theta})$ parametrized by parameters Φ is defined. The neural network learns the values of Φ such that $q_\Phi(\boldsymbol{\theta})$ resembles closely the exact posterior; Kullback-Leibler divergence can measure this resemblance.

In terms of Φ, Kullback-Leibler divergence is defined as,

$$D_{KL}\left(q_\phi \| P\right) = \int_{\boldsymbol{\theta}} q_\phi\left(\boldsymbol{\theta}'\right) \log\left(\frac{q_\phi\left(\boldsymbol{\theta}'\right)}{P\left(\boldsymbol{\theta}' \mid D\right)}\right) d\boldsymbol{\theta}' \tag{2.16}$$

In order to compute $D_{KL}(q_\phi \| P)$, $P(\boldsymbol{\theta} \mid D)$ also has to be computed. To avoid this problem, the evidence lower bound (ELBO) is defined:

$$ELBO = \int_{\boldsymbol{\theta}} q_\phi(\boldsymbol{\theta}') \log\left(\frac{P(\boldsymbol{\theta}', D)}{q_\phi(\boldsymbol{\theta}')}\right) d\boldsymbol{\theta}' = \log(P(D)) - D_{KL}(q_\phi \| P) \tag{2.17}$$

Minimizing $D_{KL}(q_\phi \| P)$ is equivalent to maximizing the ELBO given that $\log(P(D))$ only depends on the prior. The ELBO, as any other loss function in CNN, can be optimized applying stochastic gradient descent to variantional inference (Hoffman et al., 2013).

Deep learning, as in statistics, constructs $q_\Phi(\boldsymbol{\theta})$ from sample conjugate distributions in the exponential family such as multivariate normal, betas and gammas (Valentin Jospin et al., 2020).

2.4 Methodological background: Hidden Markov Random Fields

A hidden Markov random field (HMRF) model is a particular case of hidden Markov models (HMM). HMM are stochastic processes derived by a Markov chain whose state sequence cannot be observed directly, only through a sequence of observations. HMM are very convenient to model context-dependent entities by characterizing their mutual influence through conditional Markov random field (MRF) distributions (Held et al., 1997). In instances with spatial information (2D and 3D images), HMM constrains this information using the entities (pixels or voxels) of the neighboring context. This allows forming mutually exclusive classes that depend on the spatial similarities of the entities (Zhang et al., 2001). Since the HMRF is spatially dependent, it can encode both the spatial and statistical properties of an instance.

In MRF, the set of spatial locations $\mathcal{S}$ are associated among them following a neighborhood structure $\mathcal{N} = \{\mathcal{N}_i, i \in \mathcal{S}\}$. The joint probability of a pair of random fields (X_i, Y_i) is defined as

$$P(y_i, x_i \mid x_{\mathcal{N}_i}) = P(y_i \mid x_i) P(x_i \mid x_{\mathcal{N}_i}) \tag{2.18}$$

where $\forall i \in \mathcal{S}$, and X_i's neighborhood configuration is $X_{\mathcal{N}_i} \in \mathcal{L}$, $\mathcal{L} = \{1, 2, ..., l\}$ being a set of possible classes or labels.

The HMRF model is defined as the marginal probability distribution of Y_i depending on the random variable parameter θ and $X_{\mathcal{N}_i}$,

$$\begin{aligned} p(y_i \mid x_{\mathcal{N}_i}, \theta) &= \sum_{\ell \in \mathcal{L}} p(y_i, \ell \mid x_{\mathcal{N}_i}, \theta) \\ &= \sum_{\ell \in \mathcal{L}} f(y_i; \theta_\ell) p(\ell \mid x_{\mathcal{N}_i}) \end{aligned} \tag{2.19}$$

where $= \{\theta_\ell, \ell \in \mathcal{L}\}$.

Assuming a Gaussian emission distribution, the HMRF model can be re-written as

$$p\left(y_i \mid x_{\mathcal{N}_i}, \theta\right) = \sum_{\ell \in \mathcal{L}} g\left(y_i; \theta_\ell\right) p\left(\ell \mid X_{\mathcal{N}_i}\right) \tag{2.20}$$

where $\theta_\ell = \left(\mu_\ell, \sigma_\ell\right)^T$ and

$$g\left(y; \theta_\ell\right) = \frac{1}{\sqrt{2\pi\sigma_\ell^2}} \exp\left(-\frac{\left(y - \mu_\ell\right)^2}{2\sigma_\ell^2}\right). \tag{2.21}$$

2.4.1 Spatial Brain Segmentation

The Hidden Markov Random Field (HMRF) has been widely adopted in image segmentation due to its classification nature (Nie et al., 2009). In the brain imaging context, HMRF models are used to segment 3D images of the brain into different classes (usually tissue) whilst also correcting for spatial intensity variations. It is a robust and reliable method insensible to noise with probabilistic volume tissue segmentation.

The segmentation of an MRI using an HMRF model involves assigning to each MRI voxel a class label value taken from the set of classes, typically WM, GM, and CSF tissues. The voxels are indexed using a 3D rectangular lattice structure and each voxel is characterized by an intensity value. The labeling of each voxel depends on the probability of that voxel's intensity belonging to a spatial neighborhood of similar voxel intensities.

Chapter 3

Spatially Informed Bayesian Neural Network

This chapter describes the methodology of the Spatially informed Bayesian Neural Network (SBNN). It is built on the theory discussed in Chapter 2. This chapter is structured as follows: In the first section, a general workflow to design, train, and use the SBNN is explained. The second section presents the model's input data emphasizing the spatially informed input. The third section covers the neural network architecture and stochastic model configuration, including the prior and variational posterior family. The fourth discusses the output final predictions based on the summary of the marginal distributions given the output posterior distribution. The final section summarizes the software and tools to build the SBNN model.

3.1 Model Design

The Spatially Informed Bayesian Neural Network aims to identify a possible neurodegenerative disease of a suspected patient based on a MRI scan. The models analyze the MRI data together with a spatially informed layer to classify a patient MRI scan into one of the five classes: Healthy, AZ, MCI, PD, and MS. The output classification is a potential disease with an associated probability and prediction error. SBNN combines a convolutional neural network, a stochastic variational model, and a spatially informed input to produce predictions.

Figure 3.1 displays the workflow to build the SBNN. First, preprocessed MRI scans (x_λ) are used to generate the spatially informed layers (x_γ) using the HMRF model. The resulting MRI exemplifies a brain segmented in ℓ mutually exclusive classes. Both the preprocessed MRIs and the segmentation MRIs are split into training, validation, and test datasets. The training dataset (D) is the input of the SBNN model. Then a functional model ($\mathcal{NN}_{\boldsymbol{\theta}}$) is chosen, i.e., a convolutional neural network (CNN). This thesis selected four predefined model architectures and proposed a new one. The five models were tested to detect the architecture that best represents and generalizes the MRI data and to evaluate the effect of the spatially informed input. The next step is the choice of a stochastic model, that is, the prior distribution over the CNN kernel and bias parameters $p(\boldsymbol{\theta})$ and the

prior confidence in the model predictions $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$. Also, a variational posterior family $q_\Phi(\boldsymbol{\theta})$ is defined; as the prior follows a normal distribution and so does the variational posterior. The CNN is then trained with the input training dataset and the stochastic parameters. During the training, the KL-divergence function adjusts the posterior based on the priors and the ELBO function is maximized to learn the variational parameters (CNN weights and bias). The training and inference output is a posterior distribution over the model coefficients $p(\boldsymbol{\theta}|D)$ that is then marginalized $p(y|\boldsymbol{x}, D)$ to get the distribution of possible predictions for a particular MRI scan for each disease. The predictions are summarized to find the final prediction and uncertainty.

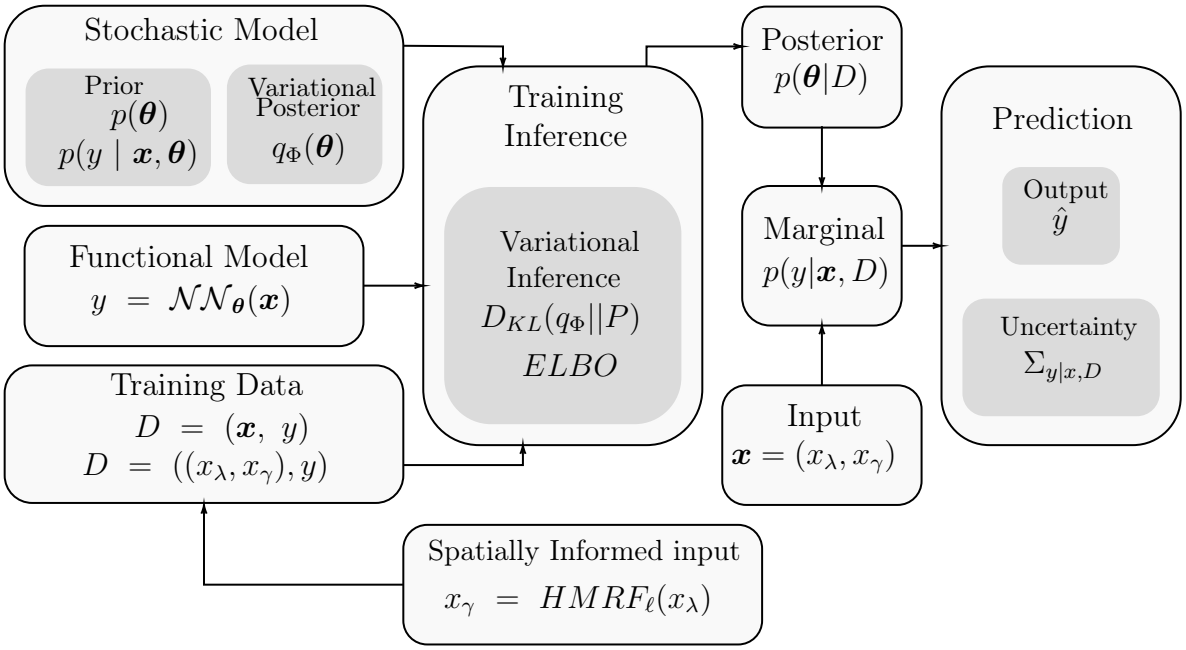


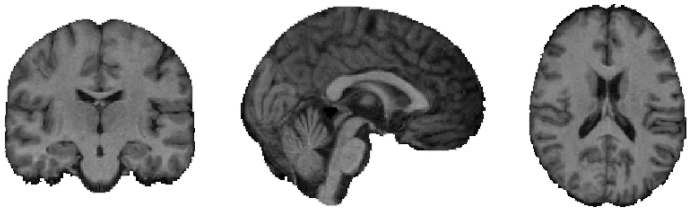
Figure 3.1: Design to train and use the spatially informed Bayesian neural network.

3.2 Model Input

One of the main goals of the SBNN is to prove that the inclusion of an explicit spatial input in a CNN model improves classification accuracy. Consequently, the SBNN model is fed not only by standard data but also by a spatially informed input. The standard data corresponds to the preprocessed MRI scans (see section 4.3) and the spatially informed data to the brain segmentation using the HMRF model.

The brain segmentation MRI scans (x_γ) are obtained by applying the HMRF model to the preprocessed MRIs (x_λ). Three classes $\ell = 3$ corresponding to the WM, GM, and CSF are defined to the HMRF model for segmenting the brain. These structures are selected since their spatial distribution over the brain and volumes are highly correlated to the brain damage caused by neurodegenerative diseases (Gitler et al., 2017). The output MRI scans summarize implicitly the neighborhood structures between the voxels intensities and evidence abnormal changes in those structures. Figure 3.2 displays the coronal, transverse, and

sagittal views of a healthy subject MRI scan and its corresponding segmentation. It is evident how similar neighboring voxel intensities are being assigned to the same class or classified with the same label.



(a) Original MRI

(b) Segmentation

Figure 3.2: (a) Original MRI scan from a healthy subject; (b) Segmentation image produced by the HMRF model. White color denotes the white matter tissue, light gray the gray matter tissue and dark gray the cerebrospinal fluid.

Both the preprocessed and segmented MRI scans are split into datasets to train, validate, and test the SBNN. For training and validation purposes, each pair of MRI scans $(x_{\lambda i}, x_{\gamma i})$ is accompanied by its corresponding true ground disease (y_i). In the case of the model testing, true classes y are only used to evaluate the predicted classes $\hat{y}$ of the test dataset. Data augmentation for the input $\boldsymbol{x}$ was performed by rotating, translating, and flipping the MRI scans. However, it did not affect the models' performance and prediction.

3.3 Functional and Stochastic models

The functional model $\mathcal{NN}_{\boldsymbol{\theta}}$ consists of a 3D CNN with two inputs and two branches. The first branch is designed to extract high-level features from the preprocessed MRI scans. The second branch extracts information from the spatially informed MR images. Both branches have identical convolutional layer architectures to maximize the features detection and to guarantee output common dimensions. A fusion layer combines the output semantic features of both CNN branches which are passed on to a set of fully connected layers. The FC layers and a SoftMax activation at the last layer are trained to assign a predefined class to the input 3D images, i.e., to predict the neurodegenerative disease for each patient. Figure 3.3 summarizes the $\mathcal{NN}_{\boldsymbol{\theta}}$ architecture.

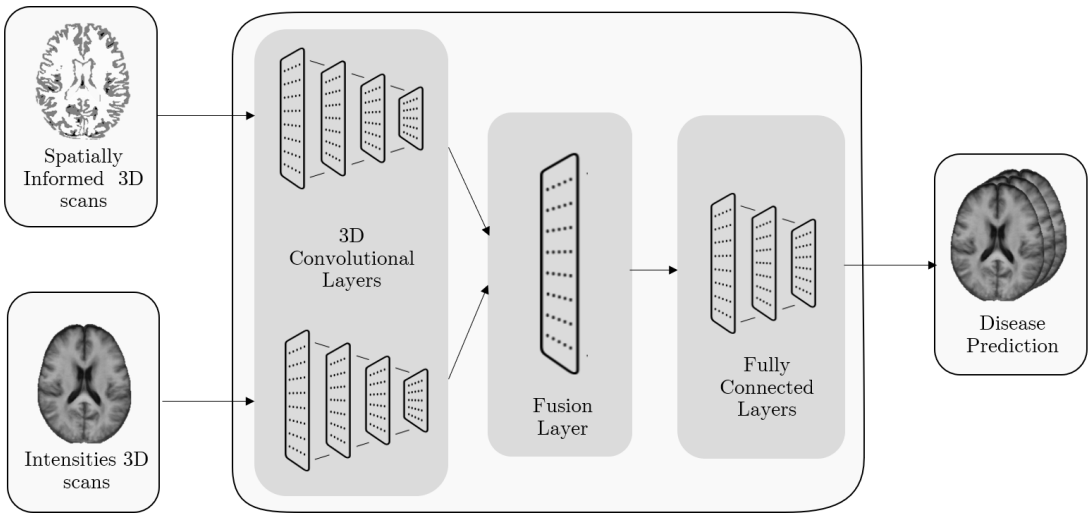


Figure 3.3: Schematic description of the proposed functional model architecture for neurodegenerative diseases patients prediction.

Five 3D CNN architectures are used for the functional model branches. These architectures consist of several 3D convolutional layers (convolution, activation, normalization, and pooling) with different parameters (filter, kernel, strides, etc). This study selects four 3D CNN architectures based on related state-of-the-art brain multi-class classifications and proposes a new one.

The first 3D CNN architecture uses three simple convolutional layers and two Fully Connected (FC) layers to classify computed tomography (CT) brain scans into healthy scans (H) and abnormal scans containing subarachnoid hemorrhage (SAH), intraparenchymal hemorrhage (IPH), acute subdural hemorrhage (ASDH) and brain polytrauma hemorrhage (BPH). The second architecture, proposed by Payan and Montana (2015), consists of three convolutional layers and three FC layers. This model differentiates between Alzheimer's disease (AZ), Mild cognitive impairment (MCI), and healthy (H) subjects. Mzoughi et al. (2020) proposes a more complex CNN architecture to distinguish between low-grade (LG) and high-grade (HG) gliomas: eight convolutional layers with three fully connected layers. Another multi-class classification approach is suggested by Parmar et al. (2020). In their study, they classify four Alzheimer's disease stages using fMRI scans and a set of five convolutional layers and three FC ones. Table 3.1 shows the details of the used state-of-the-art 3D CNN architectures for brain imaging classification. Further details of the models can be found in the studies mentioned.

Table 3.1: State-of-the-art CNN-based methods for neuroimaging multi-class classification.

Author	Disease	Classes	MRI scans	Testing Accuracy (%)
Ker et al. (2019)	CT Brain Hemorrhages	H, SAH, IPH, ASDH, BPH	399	72.68
Payan and Montana (2015)	Alzheimer's stages	H, MCI, AZ	2,275	89.47
Mzoughi et al. (2020)	Brain glioma tumors	HG, LC	284	87.2
Parmar et al. (2020)	Alzheimer's stages (fMRI)	H, AZ LMCI, EMCI	120	93.10

This study proposed architecture, inspired by the AlexNet CNN (Krizhevsky et al., 2017), has four convolutional layer groups and three fully-connected layers. The detailed configuration of the networks is shown in figure 3.4. The input to the CNN is a 3D preprocessed/segmentation MRI scan with a 91x109x91 size. The convolutional layers compute their outputs by applying the convolutional operations (and a ReLU activation) with 3D filters of the size 3x3x3, 5x5x5, 7x7x7, 9x9x9. The results are followed by max-pooling of size 2 and 1x1x1, 2x2x2, 3x3x3, 4x4x4 strides. The last three layers in the CNN are fully connected. These FC layers include the neurons connected to all outputs of their precedent layers. The FC last layer with a softmax activation has 5 neurons, which represent the probabilities of a patient to belong to one of the five classes (healthy, AZ, MCI, PD, or MS).

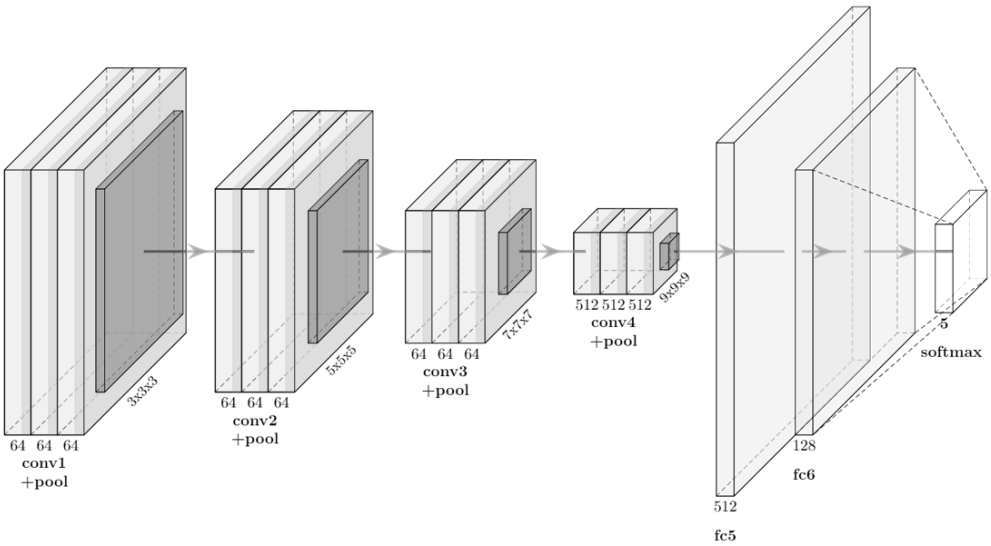


Figure 3.4: Proposed CNN architecture for each model branch feature extraction from 3D MRI scans.

Batch normalization layers are placed after each convolutional layer to increase network stability. FC layers are followed by dropouts to regularize and account for overfitting. A detailed architecture implementation is mentioned in the Appendix.

The second step, after defining a functional model, is to build the stochastic one. The stochastic model assumes that the convolution and FC layers kernels and bias ($\boldsymbol{\theta}$) are drawn from distributions. Using Bayesian inference, the prior $p(\boldsymbol{\theta})$, surrogate posterior $q_\Phi(\boldsymbol{\theta})$, and divergence for both the kernel and bias distributions are specified.

As variational inference method, the reparameterization estimator proposed by Kingma and Welling (2014) is used. This method performs a Monte Carlo approximation of the distribution integrating over the kernel and bias. It also regularizes these parameters by minimizing the Kullback-Leibler divergence up to a constant, also known as the negative $ELBO$ on the marginal likelihood.

The prior $p(\boldsymbol{\theta})$ and variational posterior $q_\Phi(\boldsymbol{\theta})$ are taken as Gaussian distributions with initialized mean $\mu = 0$ and random small variance σ^2:

$$p(\boldsymbol{\theta}) = \prod \mathcal{N}\left(\boldsymbol{\theta} \mid 0, \sigma^2\right) \tag{3.1}$$

$$q_\Phi(\boldsymbol{\theta}) = \prod \mathcal{N}\left(\boldsymbol{\theta} \mid \mu, \sigma^2\right) \tag{3.2}$$

The prior confidence in the model predictions $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ is computed using the softmax function in the last FC layer.

3.4 Training and inference

The CNN is trained using variational inference. The bias's and kernel's prior and variational posterior are defined as Gaussian distributed. The KL-Divergence is weighted by the number of training samples (see section 4.4.1) and the $ELBO$ functions measure the loss of the model.

The training constructs the posterior $p(\boldsymbol{\theta}|D)$, in other words, the kernel and bias distributions (marginalized distributions) given the training data. Ideally, these distributions should be Gaussian distributed around the most probable values of the CNN parameters.

3.5 Prediction

For the prediction, the variational Monte Carlo inference is applied to each pair of MRI scan in the testing dataset to get its corresponding probability distribution $p(y|x, D)$ depending on the output posterior $p(\boldsymbol{\theta}|D)$ (CNN output weights). This can be done by executing the SBNN n times (or iterations) for an input $\boldsymbol{x}$.

A vector of relative probabilities (uncertainties) per disease is obtained by averaging the n predictions (the mean value of the probability distribution) of the SBNN. The final predicted disease $\hat{y}$ is taken as the most likely disease, i.e., the disease with the highest averaged probability.

Chapter 4

Data and experimental design

This chapter describes the set of material employed in this work, in terms of data and software. The chapter is structured as follows. Section 3.1 introduces the characteristics of the MRI images and clinical data used in the experiment. Section 3.2 explains the data retrieval process. Section 3.3 presents the set of tools to standardize MRI coming from different sources. At last, Section 3.4 describes the experimental design.

4.1 Data sources

This study employed 330 3D T_1-weighted MRI scans from three different studies.

The first study corresponds to the Alzheimer's Disease Neuroimaging Initiative* (ADNI). The ADNI data set is part of a large multicenter study initiated in 2004 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI is to test whether serial magnetic resonance imaging (MRI) and other biological markers and clinical and neuropsychological assessment combined can measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AZ). The full study has over 10,000 MRIs from 2294 cases and 500 control subjects. The data set available for this thesis comprises a subset from the ADNI 3 cohort of the ADNI study. The subset consists of 60 AZ, 30 MCI, and 75 Healthy (normal controls) subjects scanned on a Siemens Prisma 3T scanner. 3D MPRAGE images were acquired since they are considered the best in the quality ratings (Jack et al., 2008). The scans were obtained with the following acquisition protocol: echo time 5 ms, time-to-repetition 20.0 ms, flip angle 40, matrix size 256×256, slices 160, voxel dimensions mm^3). The subset also includes patient-specific covariates such as sex and age.

The second dataset was obtained from the Parkinson's Progression Markers Initiative (PPMI) study (Marek et al., 2011). The PPMI is a 5-year longitudinal study

*Data used in the preparation of this thesis were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`

funded by The Michael J. Fox Foundation for Parkinson's Research (MJFF) and industry partners to identify biomarkers of PD progression. The PPMI counts with brain imaging, clinical evaluation, and biospecimen collections for more than 2500 Parkinson's disease patients and 1500 healthy controls subjects. The dataset selected for this thesis is conformed by 60 PD patients and 75 Healthy subjects' MRI scans and clinical data. The 135 T_1-weighted scans were acquired following the MPRAGE protocol with an echo time of 4 ms, 15.0 ms time-to-repetition, a 60 flip angle, matrix size of 256x256, 170 slices, and voxel dimensions of 1x1x1.2 mm^3. The clinical data refers to the sex and age of the patients.

Table 4.1: Number of subjects and their mean age.

Disease	Subjects			Mean Age		
	Female	Male	Total	Female	Male	Total
AZ	25	35	60	72.8	76.2	74.8
MCI	10	20	30	60.5	70.5	65.1
PD	23	37	60	60.8	63.1	64.25
MS	23	7	30	40.8	34.1	39.2
Healthy	91	59	150	63.7	66.3	64.7
Total	172	158	**330**	62	67	64.5

In 2017, the University Medical Center Ljubljana (UMCL) disseminated a public MS dataset of 30 MS patients including MR images and biological data. The publicly available dataset was created to encourage further use and research in MS lesions segmentation. A 3T Siemens Magnetom Trio MR system imaged the cohort of 30 patients. 3D T_1-weighted scans have the following specifications: echo time 20 ms, time-to-repetition 200.0 ms, flip angle 120, matrix size of 408x512, 144 slices and voxel dimensions of 0.42x0.42x3.3 mm^3. The study conducted by the UMCL provided a novel protocol for creating reference white-matter lesion segmentation based on multi-rater consensus. For more details see Lesjak et al. (2018). This thesis uses the full dataset from 30 patients comprising 3D T_1-weighted scans and biological data such as sex and age of the study participants.

Table 4.1 provides an overview of the main characteristics of the data set. The average age is relatively similar across diseases. The largest discrepancy can be observed for MS, which affects the young population. Subjects in this group on average tend to be younger than subjects in groups of the other diseases.

4.2 Data retrieval

All 330 MRI scans were downloaded in the newer standard NIfTI format for brain imaging processing. Patients' sex and age variables from each study were merged into a single data-set in CSV format.

For downloading data of the ADNI and PPMI studies, an access request was sent to the Image Data Archive (IDA) center through its website (`https://ida.loni.usc.edu/`). IDA is in charge of providing systems and resources to help collect, manage, and share data from brain clinical trials and research studies.

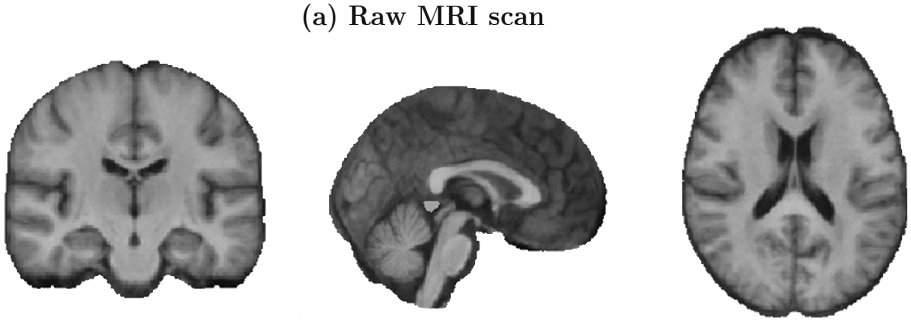
IDA granted access to the data from both studies within two to three weeks after sending the request application. Data from the UMCL is publicly available and was downloaded from the university's laboratory of imaging technologies (`http://lit.fe.uni-lj.si/tools`).

4.3 Data preprocessing

Preprocessing of the dataset involves numerous steps to clean and standardize the raw data before its analysis. Preprocessing is a mainstay to increase the sensitivity of analysis and to certify the validity of any model that involves MRI scans.

Figure 4.1 shows a raw and preprocessed MRI scan from a healthy person. In the processed MR image, the voxel intensities and locations were corrected. Also, subjects and scanner specific effects were removed. The processed scan shares the spatial architecture and dimensions of the Montreal Neurological Institute (MNI) template brain atlas (91x192x91 $2mm^3$). This noise-free MRI scan can be used in further analysis, i.e. classification analysis.

(a) Raw MRI scan



(b) Preprocessed MRI scan

Figure 4.1: (a) Original MRI scan from a healthy subject; (b) Preprocessed MRI scan of the same healthy subject.

Classifying different neurodegenerative diseases requires accurate preprocessing and careful feature learning due to the fact that similar changes can be caused by different diseases. Preprocessing prevents ambiguous recognition during image analysis and inequity in the voxels' intensities. The preprocessing of MRI data has a major influence on the high and consistent accuracy of MRI further analysis (Parmar et al., 2020).

In the sub-sections below, a detailed preprocessing workflow for this study dataset is explained.

4.3.1 Inhomogeneity correction

Inhomogeneity intensity correction (also known as bias field correction) is a technique to adjust low frequency smooth undesirable signals that corrupts MRI images because of the inhomogeneities in the magnetic fields of the MRI scanners. A bias field blurs images reducing their contents and changing the intensity voxel values so the same tissue has different gray level distribution across the image (Juntu et al., 2008). Inhomogeneity correction regulates voxels intensity values based on a particular tissue and allows further analysis such as segmentation and classification which assume spatial invariance of the MRI scans. For this thesis dataset, an N4 (Improved N3 Bias Correction (Tustison et al., 2010)) was applied.

4.3.2 Brain extraction

Brain extraction removes non-brain tissue such as the skull and neck from an MRI scan of the whole head. Since this project focuses on the brain tissue, non-brain voxel areas are deleted from the MRI data. The brain extraction was performed using the Fast Robust Automated Brain Extraction algorithm (BET) (Smith, 2002). As the raw MRI scans contain a lot of neck tissue (non-brain matter), the centre-of-gravity is computed to find the brain true centre in order to improve the brain tissue extraction.

4.3.3 Spatial Registration

Spatial registration refers to the process of aligning two images so that their common features overlap and their differences are emphasized and readily visible (Penny et al., 2007). Spatial registration performs spatial transformations to one or multiple images to make locations (voxels) have a similar interpretation. Specific voxels in particular locations can be only constrained in brains with equivalent spatial domains.

During a MRI session, a particular high-resolution T_1-weighting structural scan is obtained according to the subject's brain shape and layout and the particular parameters of the scanner. To conduct a population-level analysis every image must have the same conditions in resolution, size, and spatial distribution. This type of registration is called registration to a template.

Since the data of this thesis come from different sources and patients, a non-linear symmetric diffeomorphic image registration (Avants et al., 2008) method was used to spatially transform the 330 T_1-weighting scanner into a reference template. Literature suggested the brain atlas of Montreal Neurological Institute (MNI) due to its high-spatial-resolution and unbiased properties. The MNI template is a standard magnetic resonance imaging template brain volume for normal population (Penny et al., 2007). This template defines a representative brain of the population derived of averaging the dimensions, size and locations of brain MRI's from 152 healthy individuals.

4.3.4 Intensity Normalization

MRI intensities are acquired in arbitrary units making them incomparable across sites and between subjects. Even MRI scans acquired with the same protocol can not be compared. This affects the performance, prediction, and inference of further MRI analysis. Intensity normalization is a key step before performing between-subject or between-time intensity comparisons at the voxel level. It ensures comparability across images by bringing the intensities to a common scale across patients (Penny et al., 2007).

The voxel intensity values of this thesis dataset were normalized using the Removal of Artificial Voxel Effect by Linear regression (RAVEL) (Fortin et al., 2016). The method removes present unwanted variation after a white stripe intensity normalization, a robust method based on parameters obtained from a sample of normal-appearing white matter. The RAVEL algorithm normalizes the voxel intensity values by decomposing the variation of the intensities into a biological component (clinical covariates) and technical variation (scan effects). RAVEL algorithm used patients' age and sex as clinical variables to produce the normalized voxel intensities for each MRI in this study.

4.3.5 Image Harmonization

Differences in MRI scanners and acquisition protocols generate technical between-scanner variations that hinder the detection of imaging features associated with clinical covariates. Image normalization eliminates scanner effects and allows comparisons of images across imaging sites and scanners. Like intensity normalization, image harmonization algorithms model the biological variables and the scanner effects to remove unwanted variation. However, image harmonization methods include explicitly scanner parameters.

In this study, the ComBat image harmonization algorithm (Fortin et al., 2018) was used to harmonize cortical thicknesses of the 330 MRI scans. ComBat proved to be a powerful method to combine and study multi-source brain imaging data while increasing the power and reproducibility of subsequent statistical analyses.

4.4 Experimental design

4.4.1 Dataset Splitting

To train, validate, and test the SBNN model, the 330 MRI scans were split using the simple random sampling technique. It is an intuitive and efficient technique to randomly assign observations to subset datasets. The images were randomly shuffled and assigned to training, validation, or test dataset.

Table 4.2: Number of MRI scans per training, validation and test dataset.

Disease	Training	Validation	Test
Percentage	70%	15.4%	14.6%
AZ	42	9	9
MCI	21	5	3
PD	42	9	9
MS	21	5	3
Healthy	105	23	22
Total	231	51	48

The training dataset, used by the SBNN model to learn each disease features, consists of 231 (70%) T_1-weighted preprocessed scans. The validation dataset and test dataset are composed of 51 (15.4%) and 49 (14.6%) preprocessed images respectively as shown in Table 4.2. The full dataset was split per disease to guarantee the same distribution of each disease in each subset dataset.

4.4.2 Experimentation

The five CNNs are trained for a maximum of 100 iterations using the Adam optimizer with an exponential decay function of initial rate $\delta = 1\text{x}10^{-5}$. The batch size is set to a default value of 32 training samples. The input training and validation sets are shuffled upon generation.

The 3D CNN with the highest accuracy over the test dataset (data that the model never has seen) is chosen and transformed into a SBNN using variational inference. The bias and kernel prior and variational posterior are defined as Gaussian distributed. The mean μ and σ^2 variance are initially set to 0 and 0.1 respectively. The KL-Divergence is weighted by 231, the number of training samples.

The Monte Carlo number of iterations is set to $n = 1000$. Probability distributions per disease are obtained for each of the 49 test samples. For one sample, the mean value of every disease probabilities is chosen to obtain the predicted relative probabilities. The disease with the highest relative probability is taken as the final prediction.

4.4.3 Evaluation

The predicted disease is constrained with the ground truth one to generate an evaluation metric. This metric is usually summarized by a Confusion Matrix (Table 4.3) which consists of four possible combinations: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN).

Table 4.3: Confusion Matrix.

		Actual Class	
		True	False
Predicted Class	True	True Positive	False Negative
	False	False Positive	True Negative

A True statement refers to a correctly classified class (positive or negative), and a False statement refers to an incorrect classified class (positive or negative). Based on the Confusion Matrix configuration, four performance metrics can be derived:

$$Precision = \frac{TP}{TP + FP} \tag{4.1}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.2}$$

$$F_1 - Score = \frac{2 * Precision * Recall}{Precision * Recall} \tag{4.3}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.4}$$

Precision refers to the ratio of positive MRI scans correctly classified within the correct class. Recall describes the portion of positive MRI scans correctly classified in all classes. F1-Score is the harmonic mean of precision and recall. Accuracy measures the correctly identified MRI scans over the total classification.

Chapter 5

Results and Discussion

This chapter presents and discusses the results of the experiment described in Chapter 4. It is structured as follows. The first section shows the incidence of the spatially informed layer on the models and a neural network model is chosen as a functional model of the BNN. Section two presents a performance evaluation of the SBNN and its parameters. Then, the third section focuses on the accuracies of the predictions of the test data.

5.1 Spatially informed Input

The five CNN architectures were trained with the same parameters and the same training, validation, and test datasets (see section 4.4.2). Table 5.1 shows the accuracy results of the five networks with and without the spatially informed input.

For the five architectures, without including the spatially informed input, the training, validation, and test accuracy oscillate between 0.49 to 0.75%. The architecture proposed by Mzoughi et al. (2020) performs particularly well in the training and validation dataset but poorly in the test dataset; this means that the CNN is overfitting, in other words, it is learning the data high-level features but it is unable to generalize them to unseen data. Among the CNN architectures, the lowest classification accuracy (49%) was obtained from the model of Ker et al. (2019). This can be possibly explained by its simple architecture: three convolutional layers and two FC layers. The few semantic features learned by the network are not enough to classify or detect a specific class. The highest test accuracy for this group corresponds to this thesis proposed architecture with a value of 0.64%. Although it represents a relatively high value compared to other architectures, it is considered a low value for classification purposes.

Differentially, the test accuracy for the five CNNs architectures with the spatially informed input range from 0.68 to 0.83%. This study proposed architecture outperforms the state-of-the-art model in terms of the validation (85%) and test (83%) accuracies. The Mzoughi et al. (2020) architecture presents the second-highest test accuracy among the five models; however, as it happened in the architecture without the spatially informed input, it suffers from overfitting. The

CNN configuration from Parmar et al. (2020) as well as the one from Payan and Montana (2015) exhibit the lowest test accuracies of the group with the same value of 68%. However, Payan and Montana (2015) architecture performs better in the training and validation datasets.

Table 5.1: Training, validation and test accuracy for the four state-of-the-art and the one proposed in this thesis with one branch (only MRI scans) and two branches (MRI scans plus the spatially informed MRI scans).

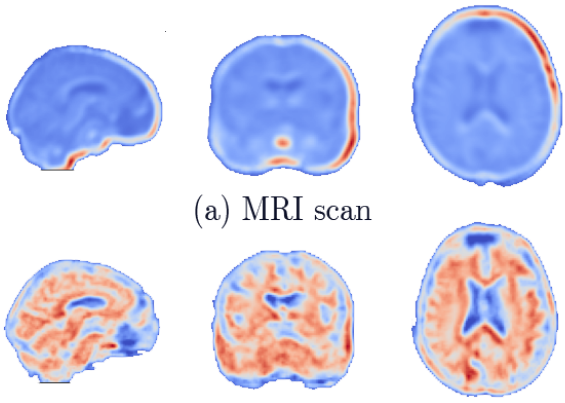
		Accuracy		
CNN Architecture	Inputs	Training	Validation	Test
Ker et al. (2019)	MRI	0.62	0.66	0.49
	MRI, SiMRI	0.74	0.70	**0.74**
Payan and Montana (2015)	MRI	0.54	0.56	0.60
	MRI, SiMRI	0.61	0.62	**0.68**
Mzoughi et al. (2020)	MRI	0.92	0.72	0.61
	MRI, SiMRI	0.99	0.77	**0.76**
Parmar et al. (2020)	MRI	0.75	0.56	0.56
	MRI, SiMRI	0.92	0.75	**0.68**
Proposed	MRI	0.75	0.66	0.64
	MRI, SiMRI	0.96	0.85	**0.83**

MRI: Preprocessed MRI scans; **SiMRI:** Spatially informed MRI scans.

Taking the results into account, two conclusions can be drawn. First, Adding the spatially informed input improves significantly the training, validation, and test accuracy for all the architectures. Overall, the accuracies increased between 8% and 25%.

Figure 5.1 displays output from the first convolutional layer of the proposed architecture for a patient with AZ. This output, also called a feature map or activation map contains the most important features identified by a convolutional layer. For the MRI scan (a), the activation features are mainly located in the outer parts of the front and parietal lobes. In other words, CNN is detecting scarce features for characterizing AZ. On the other hand, for the spatially informed MRI scans (b), the identified features are clustered around the brain ventricles and emphasize cerebral atrophy. These regions are known biomarkers to detect AZ through MRI (Frisoni et al., 2010). The combination of both inputs boots the CNN feature detection for the diseases.

In summary, the CNNs with a spatially informed input (MRI scans with explicit spatial relations) can locate more and better features associated with a neurodegenerative disease. The spatially informed input provides richer information to the classifier and thus it can detect neurodegeneration hallmarks for a disease.



(a) MRI scan

(b) Spatially Informed scan

Figure 5.1: (a) Sagittal, coronal and axial slices from the feature map of the first convolutional layer for a AZ patient's MRI scan. (b) Spatially informed MRI scan.

Second, this work proposed CNN architecture achieves better accuracy results than the state-of-the-art ones. In all cases, the proposed architecture excels in correctly classifying most of the MRI scans unknown to the model. This means that it understands and generalizes the most relevant features of the diseases of this study. A detailed evaluation of the CNN is described in the next section.

5.2 Model evaluation

As mentioned in the section before, the proposed CNN architecture overperformed the four state-of-the-art models. It learns key spatial features disease-wise and can generalize them.

Initially, the model showed an overall (testing) accuracy of 83%: 83 out of 100 MRI scans will be classified correctly. When the stochastic model, i.e., the prior and variational posterior over the parameters were incorporated in the CNN, the accuracy decreased to 81%. A loss of 2% occurred by transitioning from the functional model to the stochastic model. Although the CNN accuracy is reduced by adding the stochastic model, this latter allows confidence in the predictions. It is fair trade-off between accuracy and inclusion of uncertainty. As defined in Chapter 3, the combination of the spatially informed CNN and the stochastic model is called spatially informed Bayesian neural network (SBNN).

Since there are distributions associated with the model parameters, the accuracy obtained after testing the SBNN is just one value of a distribution of possible values (probability distribution). After evaluating the model with 1000 trials (iterations) an average accuracy of 81.52% was achieved. Overall, an input MRI scan will have an 81.52% probability of being classified with the correct disease. Table 5.2 shows the accuracy evaluation for the entire model and for each class.

The SBNN evaluation was performed using the 48 MRI scans of the test dataset. Precision, recall, and F1-scores for each disease were calculated based on the averaged multi-class confusion matrix for the 1000 trials. Macro average and

weighted average metrics are also provided to measure the model accuracy with consideration to the imbalance in the number of samples per class.

As observed in table 5.2, the highest precisions with a perfect score of 1.00 were obtained for the classification of the diseases MS and PD. This means that there is a 100% probability of classifying correctly an MRI scan coming from one of these two diseases. Healthy and AZ precisions are highly accurate with values of 86% and 76% respectively. MCI presents the lowest precision (25%) of the group; only one out of four MCI MRI scans was properly classified. Since MCI is considered an early stage of AZ and usually a subtle transition between normally aging old patients and AZ (Morris et al., 2001), MCI precision was expected. Most of the MCI MRI scans were classified into AZ and healthy classes (see confusion matrix in Appendix). Recall for the five diseases varies from 50% to 100%. In MS, the classification is perfect and so is the recall. In WM tissue, MS biomarkers present the most distinctive features among neurodegenerative diseases (Mascalchi, 2005). The lowest recall belongs to MCI. This means that if a MRI scan of a MCI patient (Actual Positive) goes through the SBNN, it will be predicted in another category (Predicted Negative) with a 50% probability. In the case of a healthy individual, there is only a 14% (1-0.86) chance of a MRI scan to be assigned to another disease. F1-score might be considered as an alternative precision measure since there is an uneven class distribution in the test dataset. In terms of the F1-score, most of the diseases perform well except for MCI. MS exhibits a perfect score and both PD and Healthy scores are considerably high. AZ F1-score is lower than their counterparts' precision and recall due to the wrong classification of MCI MRI scans into the category. The support column shows the number of MRI scans classified per class regardless if they belong to the correct class or not.

Table 5.2: Evaluation metrics for the SBNN model and the studied classes.

Disease	Precision	Recall	F1-score	support
AZ	0.76	0.75	0.71	8
MCI	0.25	0.50	0.33	2
PD	1.00	0.75	0.86	12
MS	1.00	1.00	1.00	4
Healthy	0.86	0.86	0.86	22
Accuracy			0.81	48
Macro avg	0.77	0.77	0.75	48
Weighted avg	0.85	0.81	0.82	48

Macro average resumes the precision, recall, and F1-score for all diseases assuming that the classes are balanced. Considering the unbalanced test dataset, the weighted average better summarizes these metrics. The SBNN has a precision of 85%, a recall of 81%, and an F1-score of 82%. These metrics are consistent with the SBNN averaged accuracy and support the potential of the model for classifying neurodegenerative diseases.

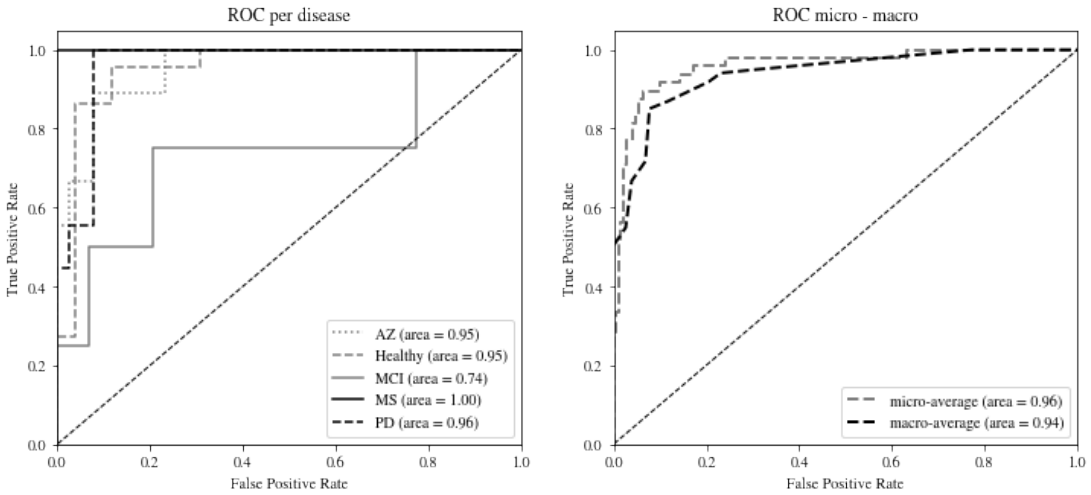


Figure 5.2: (a) ROC curves per disease; (b) micro and macro curves for SBNN.

Figure 5.2 illustrates the ROC (Receiver Operator Characteristic) curves for each disease. The ROC curve shows the trade-off between the proportion of observations that were correctly predicted to be positive (true positive rate) and the proportion of observations that are incorrectly predicted to be positive (false positive rate). Classes whose curves are closer to the top-left corner indicate better performance. Curves closer to the 45-degree diagonal of the ROC space expose less accurate classes. AZ, PD, MS, and Healthy classes tend towards ideal classifications. Contrarily, the SBNN exhibits less accurate MRI scan classification for MCI patients. This suggests that MRI scans without distinctive features - associated with other diseases - could be randomly assigned to the MCI class. The micro and macro average ROC curves also highlight the high accurate classification performance of the SBNN.

5.3 Model predictions

The SBNN was executed 1000 times in a particular MRI scan to obtain the probability distributions per disease. The mean and standard deviation values were extracted from the probability distributions to represent class-wise the predicted probability and error (two times the standard deviation) associated with it. The class with the highest probability is considered the predicted class for that MRI scan. If the predicted class corresponds to the ground truth class, the classification is correct. If the predicted class differs from the ground truth class, the classification is incorrect. If the mean probability for every disease does not exceed a threshold probability of 0.5 (the minimum acceptable probability), the MRI scan is not assigned to any of the five classes, i.e., the classification is uncertain.

This procedure was performed in the 48 MRI scans of the test dataset. Three predictions for three different AZ MRI scans are presented in Figure 5.3. The probability distributions are log-transformed to improve visualization; the closer the values are to 1, the higher the probability is. The first MRI scan (a) was

correctly classified; the mean probability is 0.83 and error 0.13. The mean probabilities for the other diseases are very low and, in most cases, it tends to 0. There is an 83% chance that the MRI scan belongs to an AZ patient. The second MRI scan (b) is not classified into any disease even though AZ exhibits the highest probability among the group. The mean probability (0.45) is not high enough to decide to which class the MRI scan resides. In this case, other clinical assessments such as symptomatology or cognition tests should be included to diagnose a disease. Uncertain classification can inform about early disease stages, wrong ground truth class, or diseases ignored in the model. The final analyzed MRI scan (c) demonstrates an incorrect classification. Healthy is the most likely predicted class with a probability of 0.55 and an error of 0.26. Three main aspects affect the erroneous classification: i) the mean probability surpassed the decision threshold only by 0.5, ii) the error is half of the predicted probability, and iii) the mean probability for AZ is 0.43. This scan could come from an AZ patient in an early stage even though the evidence states it belongs to the healthy class. Uncertainties give contextual information to decide over a disease or another. Depending on the predicted probabilities, their errors, and the level of uncertainty, a clinician can decide whether to integrate or not more information to support a diagnosis.

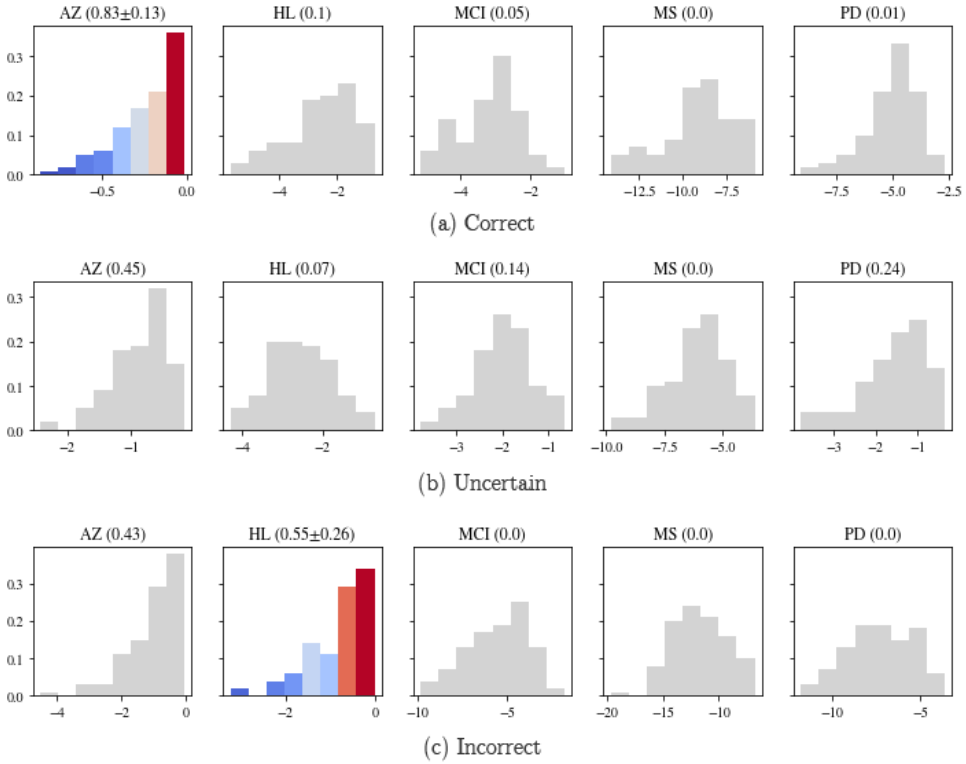


Figure 5.3: (a) AZ MRI scan correctly classified; (b) AZ MRI scan not classified; (c) AZ MRI scan incorrectly classified.

Figure 5.4 displays the mean prediction and error bars for the test dataset and Figure 5.5 shows the boxplots summarizing the mean prediction of the MRI scans

per each disease. AZ probabilities vary from 0.68 to 0.80 approximately. The maximum error (second MRI scan) is about 10% and the minimum error (ninth scan) is 3% approximately. AZ expected probability is 76%, in other words, the chance of an AZ MRI scan to be correctly assigned to the AZ class. For Healthy MRI scans, the probabilities go from 0.85 up to 1.0. The error bars show that the classification of some scans is less precise. The expected probability for healthy scans is around 92%.

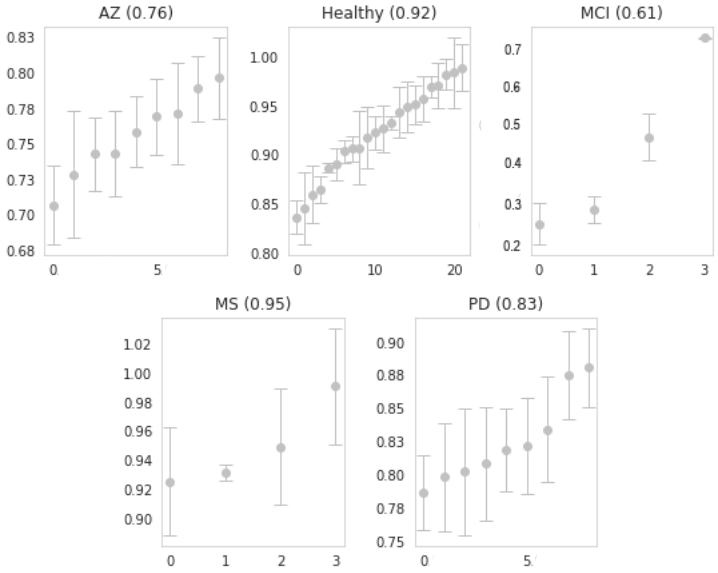


Figure 5.4: Predicted probabilitY errorbar plots of the test dataset per disease.

The probability predictions for MCI are mainly concentrated between 0.3 and 0.5. Only one scan is correctly classified with 70% confidence and low error. The expected MCI probability is 61%, however, this value is biased by the only corrected classified MRI scan. MS presents the highest probabilities ranging from 0.90 to 1.0. For most of the scans, the error does not exceed 2%. This class presents the highest expected accuracy (95%). PD predictions oscillate between 0.8 and 0.9. The mean error for the class is approximately 4.5% and 8.3 out of 10 PD scans are expected to be classified accurately.

The dispersion and accuracy of the mean probabilities are linked to the level of alikeness between the training and test data. Since the SBNN learned the characteristic features of each disease, a MRI scan whose features disagree is a potential cause for an error in the classification.

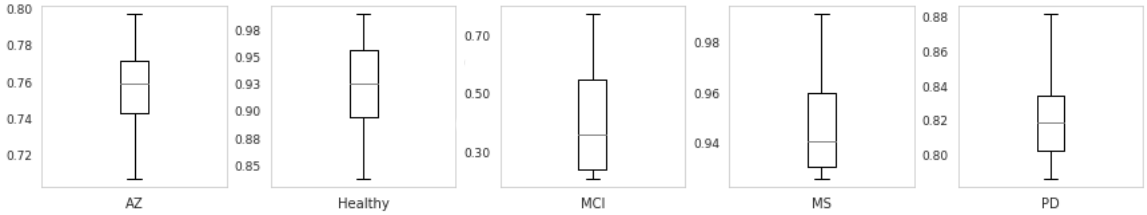


Figure 5.5: Predicted probabilities boxplot per disease.

Table 5.3 introduces the 95% confidence intervals for each disease. A confidence interval indicates a likely range of values that encompasses the expected value with a certain probability. MCI probabilities are lower compared to other diseases; this can be explained by the number of samples in the test data and by the overlapping features it presents concerning other categories such as AZ and Healthy. MS MRI range of probabilities is almost perfect which means the SBNN can easily categorize it. PD and Healthy classification accuracy can arise to 90%, a relatively high accuracy value considering the false positives coming from other classes (see Table 5.2). AZ predictions are precise considering the short range of its confidence interval.

Table 5.3: 95 % Confidence intervals for the mean of the predicted probabilities.

Disease	Interval
AZ	(0.734, 0.778)
MCI	(0.236, 0.657)
PD	(0.820, 0.917)
MS	(0.952, 0.999)
Healthy	(0.834, 0.902)

Chapter 6

Conclusions

This thesis presented a Bayesian convolutional network for the classification of neurodegenerative disease - Alzheimer's Disease, Mild Cognitive Impairment, Parkinson's Disease and Multiple Sclerosis - using magnetic resonance imaging data. To improve accuracy, the model took advantage of spatially informed MRI data (introduced as a regional brain segmentation) and used a Bayesian approach in which the model predictions have associated an associated level of confidence.

The proposed model was tested with MRI scans from four neuronal degenerative disorders and healthy controls. The results showed that a Bayesian 3D neural network, informed only by MRI data, can produce accuracy that outperforms state-of-the-art methods. However, they also highlighted the limited separability of some of the diseases, especially in disorders with few samples.

The results show a substantial increment (up to 25%) in the classification accuracy of 3D convolutional neural networks by incorporating a spatially informed MRI scan. Such an increment is a direct consequence of enhancing the CNN feature detection through brain spatial information. Including the explicit spatial scan results in detailed differentiation and separability between diseases. Although the segmentation MRI scan improves the distinction of neurodegeneration processes, there is a limitation in this approach. In highly correlated diseases, such as Alzheimer's and Mild cognitive impairment, the captured spatial changes are insufficient to separate surely these conditions.

An advantage of the Spatially informed Bayesian neural network model lies in the fact that it accounts for uncertainty. The SBNN model formulation explicitly allows for probabilistic estimation of the CNN parameters which are used in the convolutional layers. Additionally, the Bayesian framework allows for inference on the predictions by summarizing the posterior predictive probabilities. Concerning the prediction of neurodegenerative disorder, the probabilistic formulation of the SBNN model naturally accounts for aleatoric uncertainty (inherent random noise) and epistemic uncertainty (ignored information), and therefore, better represents the uncertainty in clinical diagnosis. SBNN predictions can not only detect the most likely disease of a patient, but they can also inform about patients with atypical variants, numerous pathologies, early disease stages, or diseases ignored in this research.

The overall classification accuracy of the SBNN model shows that approximately 8 out of 10 MRI scans are classified correctly. However, there is room for improvement regarding individual disease accuracy. PD, MS HL accuracies surpassed the overall accuracy, but MCI and AZ could not. MCI's low accuracy is not surprising since many studies consider this condition as a stage of AZ. Most of the MCI MRI scans were wrongly classified by the model into the AZ class.

It is worth noting that data preprocessing is instrumental in comparing and analyzing MRI data coming from diverse data sources. This study indirectly proposes a data preprocessing pipeline to integrate MRI scans from neurodegenerative patients. Although preprocessing varies depending on the research purpose, most brain imaging classification studies need to standardize and clean data to guarantee accurate predictions. The authors of this work believe that the proposed preprocessing schema could be implemented in neuroimaging studies involving numerous MRI data sources.

In conclusion, this research shows how detecting neurodegenerative disorders can be improved, considering explicit spatial information in MRI data. One of the benefits of the proposed model is the combination of multiple diseases and the ability of generate contingent diagnosis.

In future studies concerned with the same problem, the prediction accuracy may be improved by including exogenous variables, related to clinical assessments, which account for uncaptured hallmarks of the diseases. For example, psychological and cognitive tests that describes the antecedent clinical history and behavioural aspects of the patient. Furthermore, larger datasets and original MRI scan dimensions need to be included to provide richer information.

The findings of this study are expected to be of direct clinical interest for neurologists, physicians, and medical researchers. In a comprehensive overview, the most important contribution of this thesis is that it is one of the first works that considers an explicit spatial input for diseases classification, and assists diagnosis with uncertainties. Essentially, it introduces a reliable automatic tool, that can provide a fast and accurate diagnosis, and support both the treatments and prognosis of neurodegenerative disorders.

6.1 Limitations and recommendations

Limited dataset size and class imbalance are some of the main limitations of this study. Given the learning nature of deep learning methods, larger datasets result in better performance. Neural networks with small data sets cannot capture all the heterogeneity representations of a disease, leading to less accurate predictions. On the other hand, MRI image acquisition processes are usually complex and expensive. Dataset sizes are bounded and restricted by specific study needs; in many cases, for privacy considerations, MRI-related studies data is confidential and only accessible to participating researchers and organizations. More publicly available MRI data is necessary to exploit the potential of data-driven methods as deep learning, and to develop robust classifiers insusceptible to other pathologies and disease variations.

Similarly, MRI data size and dimension are key features of computational requirements for deep learning algorithms. A MRI scan is a higher-dimensional image data composed of millions of voxels. Processing hundreds of MRI scan present challenges associated with memory and computation consumption when using CNNs. For instance, MRI scans considered in this study were down-sampled to a reasonable size (from 182x218x182 to 91x109x91) to be processed by the BNN.

Some studies have proved that covariates influence differential diagnosis classifiers (Narayana et al., 2020; Taschler et al., 2014). Having clinical and biological information from patients besides age and sex is needed to improve classification accuracies. Further covariates are omitted in most studies to anonymize patients and protect their identities.

6.2 Future Work

The use of automated techniques using brain imaging data is a promising area of medical research, especially in the field of neurology. In a clinical context, the development of methods that aid decision-making is decisive for the assessment, treatment, and prognosis of neurodegeneration processes such as Alzheimer's, Parkinson's, and Multiple Sclerosis. The methods presented in this thesis can serve as a basis for including spatial information in neuroimaging analysis. The use of spatially informed data offers an opportunity to enhance the detection of neurodegenerative disorders. Additionally, the spatially informed input is not only limited to neurodegenerative diseases. Given its nature, it can be extendable to other brain conditions such as tumors, gliomas, and strokes, in which brain damage is also spatially located.

A direct extension that could be considered is the addition of other spatial information, for example, relative coordinates of voxels, distances between brain regions, and voxel intensity densities. These alternative ways to encode spatial information could enrich disease detection and account for subtle differences between the voxel intensities from two similar disorders (like AZ and MCI).

The SBNN model considers four neurodegenerative disorders and healthy subjects. In future work, the model can be extended by adding more degenerative affections (Huntintung's disease, Amyotrophic Lateral Sclerosis) and introducing stratification within each disease (disease stages or types). This latter could be useful for early diagnosis and monitoring of the disease progression.

Another approach unexplored by this work but plentiful in clinical practice is the inclusion of other MRI modalities. Although in most of the cases T1-weighted scans display neurodegeneration biomarkers, other modalities such as T2-weighted and FLAIR scans can be used to support the findings of the former. In a more robust model, other modalities could be added to account for disease features that are undetectable in the T1-weighted sequence.

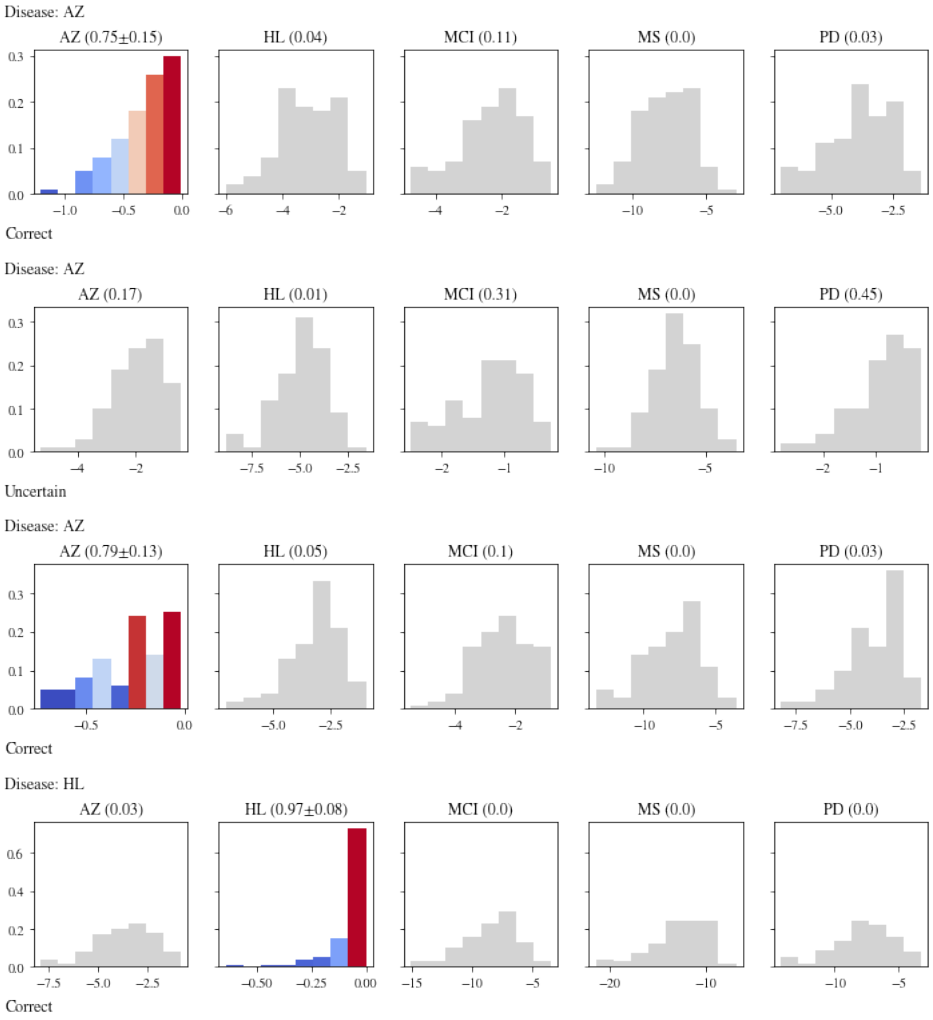
An enhancement of this study concerning some diseases' low accuracy and the problem of miss classified MRI scans would be the integration of clinical and biological information about patients. Currently, the only criterion for the SBNN to detect a patient's neurodegenerative condition is MRI data. However, clinical

trials suggest that MRI-based information together with information derived from clinical assessments and individual characteristics lead to accurate differential diagnosis.

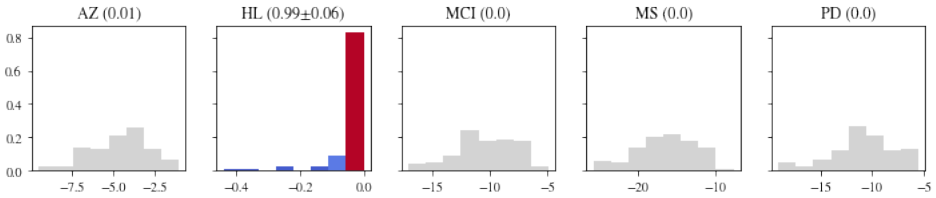
Appendix A

SBNN predictions

This appendix provides some predictions of the SBNN model that were generated after evaluating the 48 MRI scans of the test dataset. This information includes the prediction estimate, error variance, most likely disease, and classification status (correct, uncertain, incorrect). For those predictions that were uncertain, only the prediction estimate is provided. The predictions correspond to MRI scans from AZ, MCI, PD, and MS patients, and HL subjects.

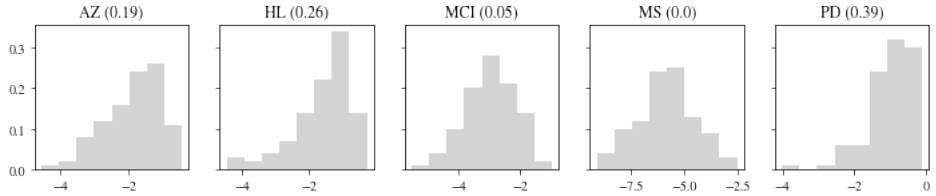


Disease: HL



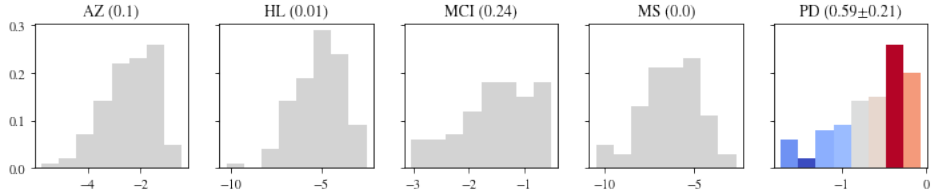
Correct

Disease: HL



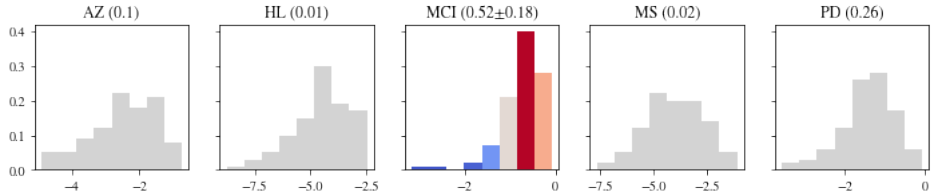
Uncertain

Disease: MCI



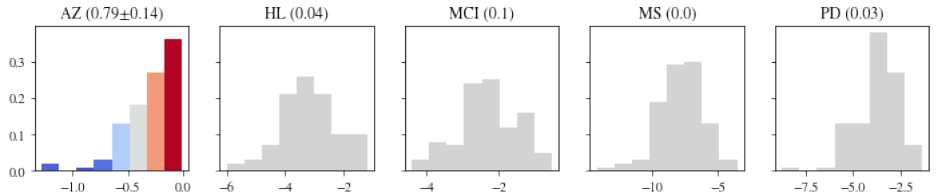
Incorrect

Disease: MCI

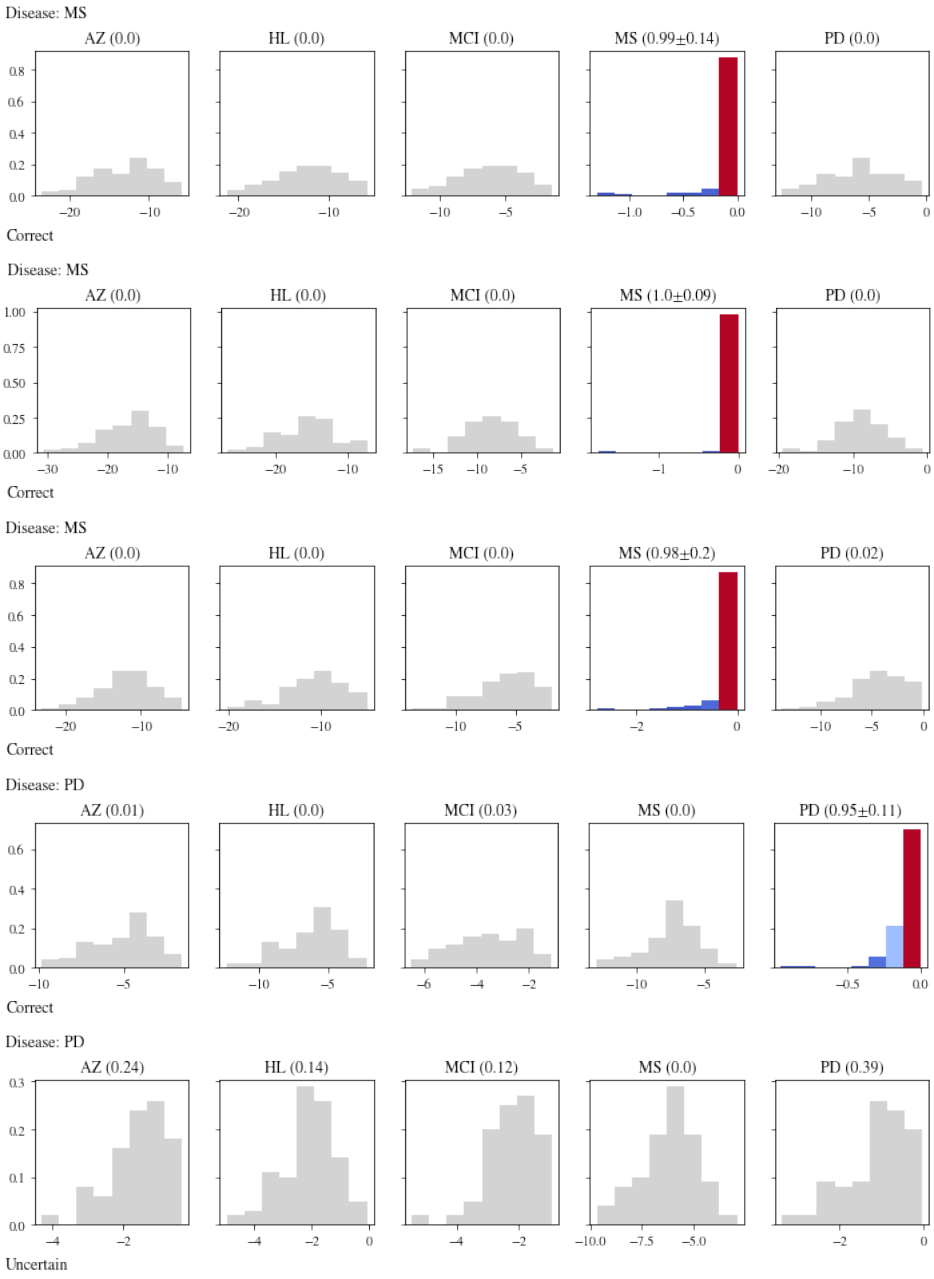


Correct

Disease: MCI



Incorrect

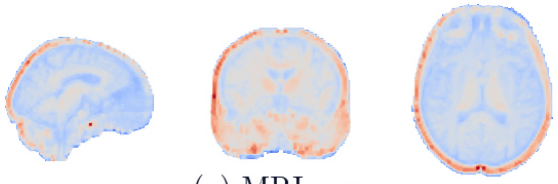

Disease: MS
AZ (0.0)
HL (0.0)
MCI (0.0)
MS (0.99±0.14)
PD (0.0)
Correct
Disease: MS
AZ (0.0)
HL (0.0)
MCI (0.0)
MS (1.0±0.09)
PD (0.0)
Correct
Disease: MS
AZ (0.0)
HL (0.0)
MCI (0.0)
MS (0.98±0.2)
PD (0.02)
Correct
Disease: PD
AZ (0.01)
HL (0.0)
MCI (0.03)
MS (0.0)
PD (0.95±0.11)
Correct
Disease: PD
AZ (0.24)
HL (0.14)
MCI (0.12)
MS (0.0)
PD (0.39)
Uncertain

Appendix B

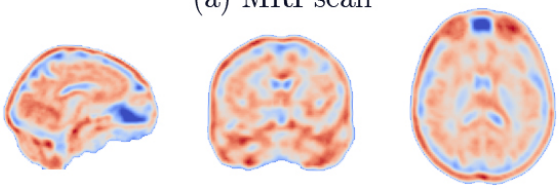
Feature Maps

The accuracy classification of the functional models improved after including the explicit spatial MRI scan. This appendix includes the activation maps of the first convolutional network for the architecture proposed in this thesis with and without the spatially informed input. The feature maps were computed for each neurodegenerative disease and healthy controls. The activation maps enhance the CNN feature recognition (Biomarkers) associated with each disorder.

B.1 Alzheimer's Disease



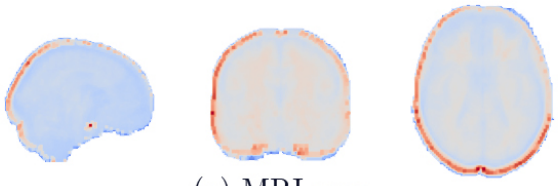
(a) MRI scan



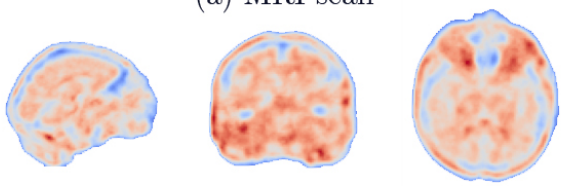
(b) Spatially Informed scan

For the MRI scan (a), the activation features are located in the outer parts of the *front* and *parietal lobes*, which are not characteristic features of AZ. Contrarily, the spatially informed MRI scans (b) identifies features around the brain ventricles and emphasize *the medial temporal lobe atrophy*; this is a known biomarker for the prediction of AZ. Likewise, high activation features occur in the *entorhinal cortex* and *hippocampal* regions.

B.2 Parkinson's Disease



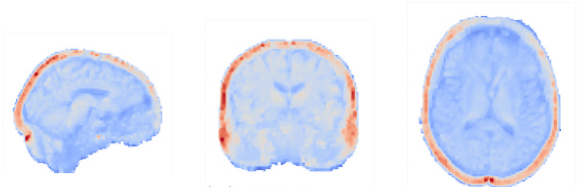
(a) MRI scan



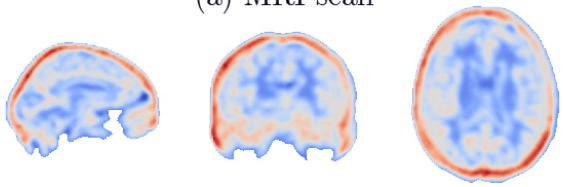
(b) Spatially Informed scan

The activation features produced by excluding the spatially informed scan (a) suggest that PD can be detected in the outer parts of the brain tissue. This is erroneous. The feature maps including the spatially informed MRI scan (b) are distributed all over the brain lobes. However, regions where the *substantia nigra* is located are highlighted. The sign of nigral degeneration is a known biomarker of PD.

B.3 Mild Cognitive Impairment



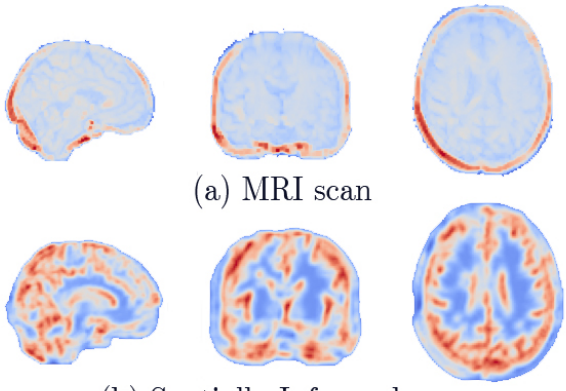
(a) MRI scan



(b) Spatially Informed scan

The MRI scan is locating features in the outer parts of the *front* and *parietal lobes*, which are not characteristic features of MCI. The spatially informed MRI scans are also erroneously detecting features outside the lobes. However, some soft activation regions are located near to the *entorhinal cortex* and *hippocampus*.

B.4 Multiple Sclerosis

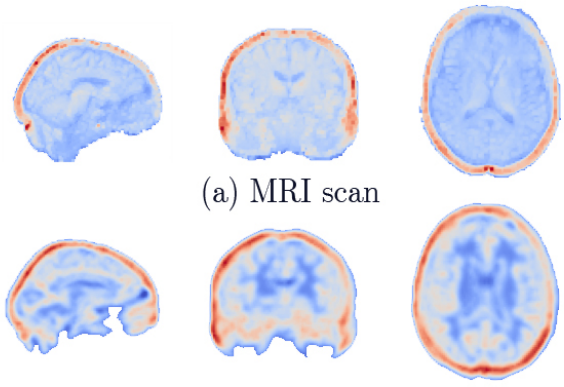


(a) MRI scan

(b) Spatially Informed scan

For multiple sclerosis it is evident how the spatially informed input of feature maps improves the recognition of the *white matter lesions* caused by the disease. Likewise, the activation map emphasizes the *corpus callosum*. One manifestation of MS is the slimming of this structure.

B.5 Healthy



(a) MRI scan

(b) Spatially Informed scan

For both the preprocessed (a) and spatially informed (b) MRI scan from the healthy subject, most of the activation features are located in the outer parts of the brain lobes. Also, some weak activations can be seen around the brain *ventricles, cerebellum, and temporal lobe*.

Appendix C

Confusion Matrix

A visualization format that produces insight into the behavior of the SBNN is the confusion matrix. The confusion matrix represents the validation results after performing 1000 trials for the test dataset. Correct, uncertain, and incorrect classified scans were assigned to the most likely disease.

It can be observed that there is confusion between MRI scans from AZ and MCI patients. Besides, some PD MRI scans get misclassified in AZ, MCI, and HL classes. Another notable fact is that MS shows a perfect classification with 4 out of 4 scans classified correctly. Some HL subject scans are classified wrongly into AZ and MCI.

		Actual Class				
		AZ	Healthy	MCI	MS	PD
Predicted Class	AZ	6	2	0	0	1
	Healthy	1	19	1	0	1
	MCI	1	1	1	0	1
	MS	0	0	0	4	0
	PD	0	0	0	0	9

Appendix D

Code

D.1 Preprocessing

All the R code used in this thesis is bundled in the NeuroNorm package (Payares, 2021). The source code, including documentation for each function, can be found on GitHub, through the following link: `https://github.com/DavidPayares/neuronorm`. The package, as many neuroimaging tools, only works on Linux OS. The package is optimized for images of neurodegenerative diseases such as Alzheimer's Disease, Mild Cognitive Impairment, Parkinson's Disease and Multiple Sclerosis, but can easily be adapted to other pathologies.

The `NeuroNorm` package can be installed from github with the following code. The `devtools` package should be installed in advance.

```
devtools::install_github("DavidPayares/neuronorm")
```

A reproducible preprocessing can be performed using the sample data in the package. Example scripts can be found at `https://github.com/DavidPayares/neuronorm/blob/main/inst/example.R`.

```
## ----------------------- Data Loading --------------------------

# Get general folder
folder <- system.file("extdata", package = "neuronorm")

# Get covariates
covariates <- system.file("covariates.txt", package = "neuronorm")

# Read covariates information
clinical_info <- read.csv(file = covariates, sep = ';')

## ----------------------- Preprocessing --------------------------

require('neuronorm')
# Preprocess MRI scans
paths_preprocess_patients <- preprocess_patients(folder, clinical_info)

# Outputs paths of the preprocessed MRI scans.
```

```
paths_preprocess_patients

## ----------------------- Image Visualization -------------------------

require('oro.nifti')
# visualize a preprocessed MRI scan for a patient.
img <- readNIfTI(file.path(paths_preprocess_patients$patient04$ravel))
orthographic(img)
```

D.2 SBNN

The python code used in this thesis for implementing the SBNN can be found at `https://github.com/DavidPayares/SBNN`. The SBNN python scripts comprise data loading, model training, model testing and predictions.

```
# Neurodegenerative Diseases Classification
# 3D Spatially Informed Bayesian CNN.

from tensorflow.keras.layers import Input, MaxPooling3D, AveragePooling3D, Flatten, BatchNormalization, Dropout,
     Concatenate
from tensorflow.keras.optimizers import Adam, schedules
from tensorflow.keras.callbacks import ModelCheckpoint
from tensorflow.keras.models import Model

import tensorflow as tf
import tensorflow_probability as tfp

class SBNNModels(object):

   def __init__(self,
            model_name = 'reparametrization',
            input_shape = (91,109,91,1),
            pooling = "max",
            drop_rate = 0.5,
            training_data_length = 10,
            classes = 5):

      """Spatially Informed Bayesian Neural Network.

        Please see link-of-thesis for more information.

         Parameters:
         ----------

         model_name: string, select one bayesian model among
                   "reparametrization", "flipout-kl" and "flipout-prior".
                   default model "reparametrization".
         input_shape: list or tuple of four ints, the shape of the input data.
                  (91, 109, 91, 1) by default.
         pooling: string, pooling methods. "max" for max pooling,
                 "avg" for average pooling. Default is "max".
         drop_rate: float, dropout rate, default is 0.5.
         training_data_length : int, numer of samples in the training data.
         classes: number of classes.


         Returns
         -------

         Model object.


         Raises
         ------
         ValueError if model is not an allowable value.

      """

      # Set model parameters
      self.model_name = model_name
      self.input_shape = input_shape
      self.pooling = pooling
      self.drop_rate = drop_rate
      self.training_data_length = training_data_length
      self.classes = classes

      # Set bayesian model parameters

      ### Gaussian priors
      # Layers weights and bias distribution priors
      self.kernel_prior = tfp.layers.default_multivariate_normal_fn
      self.bias_prior = tfp.layers.default_multivariate_normal_fn

      ### Normal posteriors
      # Layers weights and bias distribution posteriors
      self.kernel_posterior = tfp.layers.default_mean_field_normal_fn(is_singular = False)
      self.bias_posterior = tfp.layers.default_mean_field_normal_fn(is_singular = False)
```

```
    ### Kullback-Leibler divergence
    self.divergence_fn = lambda q,p,_: tfp.distributions.kl_divergence(q,p)/ self.training_data_length

    # Build SBNN models
    if model_name in {'reparametrization', 'flipout-kl', 'flipout-prior'}:
       self.model = self.bayesianModel(self.input_shape)
    else:
       raise ValueError("Unknown model name. Allowed models are 'reparametrization', 'flipout-kl' and 'flipout-
            prior'")

    return

def conv3D(self, inputs, filter_size, strides):

    """Build a 3D bayesian convolution layer .


       Parameters:
       ----------

       inputs: input tensor, it should be the original input,
              or the output from a previous layer.
       filter_size: int, the number of filters.
       strides: int tuple with length 3, the stride step in
               each dimension.


       Returns
       -------

       Tensor for convolutional layer.

    """

    # Decide convolutional layer architecture based on selected model
    if self.model_name == 'reparametrization':
       conv_layer = tfp.layers.Convolution3DReparameterization(filters = 64,
                                                kernel_size= filter_size,
                                                strides= strides,
                                                activation='relu',
                                                kernel_prior_fn = self.kernel_prior,
                                                kernel_posterior_fn = self.kernel_posterior,
                                                kernel_divergence_fn = self.divergence_fn,
                                                bias_prior_fn = self.bias_prior,
                                                bias_posterior_fn= self.bias_posterior,
                                                bias_divergence_fn = self.divergence_fn)(inputs)
    elif self.model_name == 'flipout-prior':
       conv_layer = tfp.layers.Convolution3DFlipout(filters=64,
                                        kernel_size= filter_size,
                                        strides= strides,
                                        activation= 'relu',
                                        kernel_prior_fn= self.kernel_prior)(inputs)
    elif self.model_name == 'flipout-kl':
       conv_layer = tfp.layers.Convolution3DFlipout(filters=64,
                                        kernel_size= filter_size ,
                                        strides= strides,
                                        activation= 'relu',
                                        kernel_divergence_fn= self.divergence_fn)(inputs)

    return conv_layer

def dense3D(self, inputs, units, activation = 'relu'):

    """Build a 3D bayesian convolution layer .


       Parameters:
       ----------

       inputs: input tensor, it should be the original input,
              or the output from a previous layer.
       units: int, Integer or Long, dimensionality of the output space.

       Returns
       -------

       Tensor for convolutional layer.

    """

    # Decide convolutional layer architecture based on selected model
    if self.model_name == 'reparametrization':
       dense_layer = tfp.layers.DenseReparameterization(units= units,
                                          activation= activation,
                                          kernel_prior_fn = self.kernel_prior,
                                          kernel_posterior_fn = self.kernel_posterior,
                                          kernel_divergence_fn = self.divergence_fn,
                                          bias_prior_fn = self.bias_prior,
                                          bias_posterior_fn= self.bias_posterior,
                                          bias_divergence_fn = self.divergence_fn)(inputs)
    elif self.model_name == 'flipout-prior':
       dense_layer = tfp.layers.DenseFlipout(units= units,
                                    activation= activation,
                                    kernel_prior_fn= self.kernel_prior)(inputs)
    elif self.model_name == 'flipout-kl':
       dense_layer = tfp.layers.DenseFlipout(units= units,
                                    activation= activation,
                                    kernel_divergence_fn= self.divergence_fn)(inputs)
```

```
    return dense_layer

def covLayer(self, inputs, filter_size, conv_strides, pool_strides):


    """Build a convolutional layer.


       Parameters:
       ----------

       inputs: input tensor, it should be the original input,
              or the output from a previous layer.

       Returns
       -------
       Tensor for convolutional layer.

    """

    # Decide type of pooling (max or avg)
    if self.pooling == "max":
       pool = MaxPooling3D
    elif self.pooling == "avg":
       pool = AveragePooling3D

    # Define the convolutional layer block
    conv_layer = self.conv3D(inputs, filter_size, conv_strides)
    conv_pool = pool(pool_size=2, strides= pool_strides)(conv_layer)
    conv_batch = BatchNormalization()(conv_pool)

    return conv_batch

def bayesianModel(self, inputs):

    """Build the bayesian convolutional network.


       Parameters:
       ----------

       inputs: input tensor, it should be the original input.

       Returns
       -------
       model: Keras Model instance, SBNN model.

    """

    #### Intensities Branch (preprocessed MRI scans)

    ## Input layer
    input_int = Input(shape= self.input_shape)

    # Convolutional Layers
    conv_int = self.covLayer(input_int, (3,3,3), (1,1,1), (1,1,1))
    conv_int = self.covLayer(conv_int, (5,5,5), (1,1,1), (2,2,2))
    conv_int = self.covLayer(conv_int, (7,7,7), (1,1,1), (3,3,3))
    conv_int = self.covLayer(conv_int, (9,9,9), (1,1,1), (4,4,4))

    # Fully Connected Layers
    conv_int = Flatten()(conv_int)
    conv_int = self.dense3D(conv_int, units = 512)
    conv_int = Dropout(self.drop_rate)(conv_int)
    conv_int = self.dense3D(conv_int, units = 128)
    conv_int = Dropout(self.drop_rate)(conv_int)

    #### Spatially Informed Branch (Segmented MRI scans)

    ## Input layers
    input_si = Input(shape= self.input_shape)

    # Convolutional Layers
    conv_si = self.covLayer(input_si, (3,3,3), (1,1,1), (1,1,1))
    conv_si = self.covLayer(conv_si, (5,5,5), (1,1,1), (2,2,2))
    conv_si = self.covLayer(conv_si, (7,7,7), (1,1,1), (3,3,3))
    conv_si = self.covLayer(conv_si, (9,9,9), (1,1,1), (4,4,4))

    # Fully Connected Layer
    conv_si = Flatten()(conv_si)
    conv_si = self.dense3D(conv_si, units = 512)
    conv_si = Dropout(self.drop_rate)(conv_si)
    conv_si = self.dense3D(conv_si, units = 128)
    conv_si = Dropout(self.drop_rate)(conv_si)

    # Fusion layer
    concat = Concatenate(axis = -1)([conv_int, conv_si])
    concat = self.dense3D(concat, units = 512)
    concat = Dropout(self.drop_rate)(concat)

    # Classification Layer
    if self.model_name == 'reparametrization':
       outputs = self.dense3D(concat, units = tfp.layers.OneHotCategorical.params_size(self.classes), activation
             = 'softmax')
       outputs = tfp.layers.OneHotCategorical(self.classes)(outputs)
    else:
       outputs = self.dense3D(concat, units = self.classes, activation = 'softmax')

    # Model definition
```

```
      model = Model(inputs = [input_int, input_si], outputs = outputs)

      ## Output Model
      return model

#### Get the SBNN model
if __name__ == "__main__":

   # A test to print model's architecture.
   model = SBNNModels().model

   # learning rate
   initial_learning_rate = 0.0001
   lr_schedule = schedules.ExponentialDecay(initial_learning_rate,
                                            decay_steps=100000,
                                            decay_rate=0.96,
                                            staircase=True)
   # Checkpoints
   checkpoint_cb = ModelCheckpoint("Spatially-informed-Bayesian-Neural-Network.h5",
                                   save_best_only=True)
   # Compilation
   model.compile(loss="categorical_crossentropy",
              optimizer = Adam(learning_rate = lr_schedule),
              metrics=["acc"]
              )
   # SBNN architecture
   model.summary()
```

References

Afaq, A., Fraioli, F., Sidhu, H., Wan, S., Punwani, S., Chen, S. H., Akin, O., Linch, D., Ardeshna, K., Lambert, J., Miles, K., Groves, A. & Kayani, I. (2017). Comparison of PET/MRI with PET/CT in the Evaluation of Disease Status in Lymphoma. *Clinical Nuclear Medicine*. https://doi.org/10.1097/RLU.0000000000001344

Agosta, F., Galantucci, S. & Filippi, M. (2017). Advanced magnetic resonance imaging of neurodegenerative diseases. https://doi.org/10.1007/s10072-016-2764-x

Angermueller, C., Pärnamaa, T., Parts, L. & Stegle, O. (2016). Deep learning for computational biology. *Molecular Systems Biology*. https://doi.org/10.15252/msb.20156651

Antony, P. M., Diederich, N. J., Krüger, R. & Balling, R. (2013). The hallmarks of Parkinson's disease. https://doi.org/10.1111/febs.12335

Arel, I., Rose, D. & Karnowski, T. (2010). Deep machine learning-A new frontier in artificial intelligence research. *IEEE Computational Intelligence Magazine*. https://doi.org/10.1109/MCI.2010.938364

Aslani, S., Dayan, M., Storelli, L., Filippi, M., Murino, V., Rocca, M. A. & Sona, D. (2019). Multi-branch convolutional neural network for multiple sclerosis lesion segmentation. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2019.03.068

Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Bishop, C. M. (1997). Bayesian Neural Networks. *Journal of the Brazilian Computer Society*, *4*. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-65001997000200006&nrm=iso

Blamire, A. M. (2018). MR approaches in neurodegenerative disorders. https://doi.org/10.1016/j.pnmrs.2018.11.001

Blazhenets, G., Ma, Y., Sörensen, A., Rücker, G., Schiller, F., Eidelberg, D., Frings, L. & Meyer, P. T. (2019). Principal components analysis of brain metabolism predicts development of Alzheimer dementia. *Journal of Nuclear Medicine*. https://doi.org/10.2967/jnumed.118.219097

Brosch, T., Yoo, Y., Tang, L. Y., Li, D. K., Traboulsee, A. & Tam, R. (2015). Deep convolutional encoder networks for multiple sclerosis lesion segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-24574-4_1

Burton, E. J., McKeith, I. G., Burn, D. J., Williams, E. D. & O'Brien, J. T. (2004). Cerebral atrophy in Parkinson's disease with and without dementia: A comparison with Alzheimer's disease, dementia with Lewy bodies and controls. *Brain*. https://doi.org/10.1093/brain/awh088

Chetelat, G. & Baron, J. C. (2003). Early diagnosis of Alzheimer's disease: Contribution of structural neuroimaging. https://doi.org/10.1016/S1053-8119(02)00026-5

Chitradevi, D. & Prabha, S. (2020). Analysis of brain sub regions using optimization techniques and deep learning method in Alzheimer disease. *Applied Soft Computing Journal*. https://doi.org/10.1016/j.asoc.2019.105857

Cohen, J. A. & Rae-Grant, A. (2012). *Handbook of multiple sclerosis: Second edition*. https://doi.org/10.1007/978-1-907673-50-4

Csáji, B. (2001). Approximation with artificial neural networks. *MSc. thesis*.

Currie, S., Hoggard, N., Craven, I. J., Hadjivassiliou, M. & Wilkinson, I. D. (2013). Understanding MRI: Basic MR physics for physicians. https://doi.org/10.1136/postgradmedj-2012-131342

de Leon, M. J., Ferris, S. H., George, A. E., Christman, D. R., Fowler, J. S., Gentes, C., Reisberg, B., Gee, B., Emmerich, M., Yonekura, Y., Brodie, J., Kricheff, I. I. & Wolf, A. P. (1983). Positron emission tomographic studies of aging and Alzheimer disease. *American Journal of Neuroradiology*.

Deng, L. & Yu, D. (2014). Deep learning: Methods and applications. *Found. Trends Signal Process.*, *7*(3–4), 197–387. https://doi.org/10.1561/2000000039

Díaz, G., Romero, E., Hernández-Tamames, J. A., Molina, V. & Malpica, N. (2010). Automatic classification of structural MRI for diagnosis of neurodegenerative diseases. *Acta Biologica Colombiana*, *15*(3), 165–180.

Dickson, D. W. (2018). Neuropathology of Parkinson disease. *Parkinsonism and Related Disorders*. https://doi.org/10.1016/j.parkreldis.2017.07.033

Dugger, B. N., Adler, C. H., Shill, H. A., Caviness, J., Jacobson, S., Driver-Dunckley, E. & Beach, T. G. (2014). Concomitant pathologies among a spectrum of parkinsonian disorders. *Parkinsonism and Related Disorders*. https://doi.org/10.1016/j.parkreldis.2014.02.012

Dugger, B. N. & Dickson, D. W. (2017). Pathology of neurodegenerative diseases. https://doi.org/10.1101/cshperspect.a028035

Duyckaerts, C., Delatour, B. & Potier, M. C. (2009). Classification and basic pathology of Alzheimer disease. https://doi.org/10.1007/s00401-009-0532-1

Fan, C., Gao, F., Sun, S. & Cui, F. (2008). Bayesian neural networks and its application. *2008 Fourth International Conference on Natural Computation*, *3*, 446–450. https://doi.org/10.1109/ICNC.2008.624

Filippi, M. & Rocca, M. A. (2011). MR imaging of multiple sclerosis. *Radiology*, *259*(3), 659–681. https://doi.org/10.1148/radiol.11101362

Folle, L., Vesal, S., Ravikumar, N. & Maier, A. (2019). Dilated deeply supervised networks for hippocampus segmentation in MRI. *Informatik aktuell*. https://doi.org/10.1007/978-3-658-25326-4_18

Fortin, J. P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., McInnis, M., Phillips, M. L., Trivedi, M. H., Weissman, M. M. & Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2017.11.024

Fortin, J. P., Sweeney, E. M., Muschelli, J., Crainiceanu, C. M. & Shinohara, R. T. (2016). Removing inter-subject technical variability in magnetic resonance imaging studies. *NeuroImage*, *132*, 198–212. https://doi.org/10.1016/j.neuroimage.2016.02.036

Fox, N. C., Black, R. S., Gilman, S., Rossor, M. N., Griffith, S. G., Jenkins, L. & Koller, M. (2005). Effects of Aβ immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. *Neurology*. https://doi.org/10.1212/01.WNL.0000159743.08996.99

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P. & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. https://doi.org/10.1038/nrneurol.2009.215

Gammon, K. (2014). Neurodegenerative disease: brain windfall. *Nature*, *515*(7526), 299–300. https://doi.org/10.1038/nj7526-299a

Ge, Y. (2006). Multiple sclerosis: The role of MR imaging.

Gelfand, J. M. (2014). Multiple sclerosis: Diagnosis, differential diagnosis, and clinical presentation. *Handbook of clinical neurology*. https://doi.org/10.1016/B978-0-444-52001-2.00011-X

Gitler, A. D., Dhillon, P. & Shorter, J. (2017). Neurodegenerative disease: Models, mechanisms, and a new hope. https://doi.org/10.1242/dmm.030205

Heim, B., Krismer, F., De Marzi, R. & Seppi, K. (2017). Magnetic resonance imaging for the diagnosis of Parkinson's disease. *Journal of Neural Transmission*, *124*(8), 915–964. https://doi.org/10.1007/s00702-017-1717-8

Held, K., Kops, E. R., Krause, B. J., Wells, W. M., Kikinis, R. & Müller-Gärtner, H. W. (1997). Markov random field segmentation of brain MR images.

IEEE Transactions on Medical Imaging. https://doi.org/10.1109/42.650883

Herzog, L., Murina, E., Dürr, O., Wegener, S. & Sick, B. (2020). Integrating uncertainty in deep neural networks for MRI based stroke analysis. *Medical Image Analysis*. https://doi.org/10.1016/j.media.2020.101790

Hoffman, M. D., Blei, D. M., Wang, C. & Paisley, J. (2013). Stochastic variational inference. *Journal of Machine Learning Research*.

Horner, P. J. & Gage, F. H. (2000). Regenerating the damaged central nervous system. https://doi.org/10.1038/35039559

Hotter, A., Esterhammer, R., Schocke, M. F. & Seppi, K. (2009). Potential of advanced MR imaging techniques in the differential diagnosis of Parkinsonism. https://doi.org/10.1002/mds.22648

Illan, I. A., Górriz, J. M., Ramírez, J. & Meyer-Base, A. (2014). Spatial component analysis of MRI data for Alzheimer's disease diagnosis: A Bayesian network approach. *Frontiers in Computational Neuroscience*. https://doi.org/10.3389/fncom.2014.00156

Jack, C. R., Slomkowski, M., Gracon, S., Hoover, T. M., Felmlee, J. P., Stewart, K., Xu, Y., Shiung, M., O'Brien, P. C., Cha, R., Knopman, D. & Petersen, R. C. (2003). MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD. *Neurology*. https://doi.org/10.1212/01.WNL.0000042480.86872.03

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., Whitwell, J. L., Ward, C., Dale, A. M., Felmlee, J. P., Gunter, J. L., Hill, D. L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., ... Weiner, M. W. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. https://doi.org/10.1002/jmri.21049

Jain, S., Sima, D. M., Ribbens, A., Cambron, M., Maertens, A., Van Hecke, W., De Mey, J., Barkhof, F., Steenwijk, M. D., Daams, M., Maes, F., Van Huffel, S., Vrenken, H. & Smeets, D. (2015). Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *NeuroImage: Clinical*. https://doi.org/10.1016/j.nicl.2015.05.003

Jankovic, J. (2008). Parkinson's disease: Clinical features and diagnosis. https://doi.org/10.1136/jnnp.2007.131045

Juntu, J., Sijbers, J., Dyck, D. & Gielen, J. (2008). Bias Field Correction for MRI Images. https://doi.org/10.1007/3-540-32390-2_64

Karaca, Y., Zhang, Y.-D., Cattani, C. & Ayan, U. (2016). The Differential Diagnosis of Multiple Sclerosis Using Convex Combination of Infinite Kernels. *CNS & Neurological Disorders - Drug Targets*. https://doi.org/10.2174/1871527315666161024142439

Ker, J., Singh, S. P., Bai, Y., Rao, J., Lim, T. & Wang, L. (2019). Image thresholding improves 3-dimensional convolutional neural network diagnosis of dif-

ferent acute brain hemorrhages on computed tomography scans. *Sensors (Switzerland)*. https://doi.org/10.3390/s19092167

Khairnar, P. (2020). *A bayesian convolutional neural network based classifier to detect breast cancer from histopathological images and uncertainty quantification* (Doctoral dissertation). Michigan Technological University. Houghton, Michigan. https://doi.org/10.37099/mtu.dc.etdr/1066

Khan, R. U., Tanveer, M. & Pachori, R. B. (2020). A novel method for the classification of Alzheimer's disease from normal controls using magnetic resonance imaging. *Expert Systems*, (March), 1–22. https://doi.org/10.1111/exsy.12566

Kilic, A. K., TuncerKurne, A., KarliOguz, K., Soylemezoglu, F. & Karabudak, R. (2013). Mass lesions in the brain: Tumor or multiple sclerosis? Clinical and imaging characteristics and course from a single reference center. *Turkish Neurosurgery*. https://doi.org/10.5137/1019-5149.JTN.7690-12.3

Kingma, D. P. & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Kingma, D. P. & Welling, M. (2014). Auto-encoding variational bayes. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Kruthika, K. R., Rajeswari & Maheshappa, H. D. (2019). Multistage classifier-based approach for Alzheimer's disease prediction and retrieval. *Informatics in Medicine Unlocked*. https://doi.org/10.1016/j.imu.2018.12.003

Kuperman, V. (2004). Magnetic Resonance Imaging Physical Principles and Applications. *Current opinion in nephrology and hypertension*.

Lee, S. H., Kim, S. S., Tae, W. S., Lee, S. Y., Choi, J. W., Koh, S. B. & Kwon, D. Y. (2011). Regional volume analysis of the Parkinson disease brain in early disease stage: Gray matter, white matter, striatum, and thalamus. *American Journal of Neuroradiology*. https://doi.org/10.3174/ajnr.A2372

Lesjak, Ž., Galimzianova, A., Koren, A., Lukin, M., Pernuš, F., Likar, B. & Špiclin, Ž. (2018). A Novel Public MR Image Dataset of Multiple Sclerosis Patients With Lesion Segmentations Based on Multi-rater Consensus. *Neuroinformatics*. https://doi.org/10.1007/s12021-017-9348-7

Lin, C. H., Chen, C. M., Lu, M. K., Tsai, C. H., Chiou, J. C., Liao, J. R. & Duann, J. R. (2013). VBM Reveals Brain Volume differences between Parkinson's disease and essential tremor patients. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2013.00247

Lopez Pinaya, W. H., Vieira, S., Garcia-Dias, R. & Mechelli, A. (2019). Convolutional neural networks. *Machine learning: Methods and applications to brain disorders*. https://doi.org/10.1016/B978-0-12-815739-8.00010-9

MacKay, D. J. C. (1992). A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation*. https://doi.org/10.1162/neco.1992.4.3.448

Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehéricy, S. & Benali, H. (2009). Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*. https://doi.org/10.1007/s00234-008-0463-x

Mahlknecht, P., Hotter, A., Hussl, A., Esterhammer, R., Schocke, M. & Seppi, K. (2010). Significance of MRI in diagnosis and differential diagnosis of Parkinson's disease. https://doi.org/10.1159/000314495

Mangia, S., Michaeli, S. & Tuite, P. (2013). Magnetic resonance imaging (mri) methods in parkinson's disease. In P. Tuite & A. Dagher (Eds.), *Magnetic resonance imaging in movement disorders: A guide for clinicians and scientists* (pp. 1–13). Cambridge University Press. https://doi.org/10.1017/CBO9781139207294.003

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., Poewe, W., Mollenhauer, B., Sherer, T., Frasier, M., Meunier, C., Rudolph, A., Casaceli, C., Seibyl, J., Mendick, S., ... Taylor, P. (2011). The Parkinson Progression Marker Initiative (PPMI). https://doi.org/10.1016/j.pneurobio.2011.09.005

Marschallinger, R., Mühlau, M., Schmidt, P., Atkinson, P., Trinka, E., Golaszewski, S. & Sellner, J. (2017). Using Classical Geostatistics to Quantify the Spatiotemporal Dynamics of a Neurodegenerative Disease from Brain MRI. https://doi.org/10.1007/978-3-319-46819-8_67

Martin-Macintosh, E. L., Broski, S. M., Johnson, G. B., Hunt, C. H., Cullen, E. L. & Peller, P. J. (2016a). Multimodality imaging of neurodegenerative processes: Part 1, the basics and common dementias. https://doi.org/10.2214/AJR.14.12842

Martin-Macintosh, E. L., Broski, S. M., Johnson, G. B., Hunt, C. H., Cullen, E. L. & Peller, P. J. (2016b). Multimodality imaging of neurodegenerative processes: Part 2, atypical dementias. https://doi.org/10.2214/AJR.14.12910

Mascalchi, M. (2005). Neurodegenerative diseases with associated white matter pathology. In M. Filippi, N. De Stefano, V. Dousset & J. C. McGowan (Eds.), *Mr imaging in white matter diseases of the brain and spinal cord* (pp. 377–388). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-27644-0_26

McClure, P., Rho, N., Lee, J. A., Kaczmarzyk, J. R., Zheng, C., Ghosh, S. S., Nielson, D. M., Thomas, A. G., Bandettini, P. & Pereira, F. (2018). Knowing what you know in brain segmentation using Bayesian deep neural networks. https://doi.org/10.3389/fninf.2019.00067

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D. & Stadlan, E. M. (1984). Clinical diagnosis of alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on alzheimer's disease. *Neurology*. https://doi.org/10.1212/wnl.34.7.939

Meijer, F. J. A., Goraj, B., Bloem, B. R. & Esselink, R. A. J. (2017). Clinical Application of Brain MRI in the Diagnostic Work-up of Parkinsonism. *Journal of Parkinson's Disease*, *7*, 211–217. https://doi.org/10.3233/JPD-150733

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M. & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling and Software*. https://doi.org/10.1016/j.envsoft.2017.12.001

Morris, J. C., Storandt, M., Miller, J. P., McKeel, D. W., Price, J. L., Rubin, E. H. & Berg, L. (2001). Mild cognitive impairment represents early-stage Alzheimer disease. *Archives of Neurology*, *58*(3), 397–405. https://doi.org/10.1001/archneur.58.3.397

Mtui, E., Gruener, G. & Dockery, P. (2016). *FITZGERALD'S Clinical Neuroanatomy and Neuroscience*.

Mzoughi, H., Njeh, I., Wali, A., Slima, M. B., BenHamida, A., Mhiri, C. & Mahfoudhe, K. B. (2020). Deep Multi-Scale 3D Convolutional Neural Network (CNN) for MRI Gliomas Brain Tumor Classification. *Journal of Digital Imaging*, *33*(4), 903–915. https://doi.org/10.1007/s10278-020-00347-9

Narayana, P. A., Coronado, I., Sujit, S. J., Wolinsky, J. S., Lublin, F. D. & Gabr, R. E. (2020). Deep learning for predicting enhancing lesions in multiple sclerosis from noncontrast MRI. *Radiology*. https://doi.org/10.1148/radiol.2019191061

Neal, R. M. (1996). Priors for infinite networks. *Bayesian learning for neural networks* (pp. 29–53). Springer New York. https://doi.org/10.1007/978-1-4612-0745-0_2

Nie, J., Xue, Z., Liu, T., Young, G. S., Setayesh, K., Guo, L. & Wong, S. T. (2009). Automated brain tumor segmentation using spatial accuracy-weighted hidden Markov Random Field. *Computerized Medical Imaging and Graphics*. https://doi.org/10.1016/j.compmedimag.2009.04.006

Noachtar, S. & Rémi, J. (2009). The role of EEG in epilepsy: A critical review. https://doi.org/10.1016/j.yebeh.2009.02.035

Noseworthy, J. H., Lucchinetti, C., Rodriguez, M. & Weinshenker, B. G. (2000). MS - Clinical course and diagnosis. *New England Journal of Medicine*.

Pantano, P., Caramia, F. & Pierallini, A. (1999). The role of MRI in dementia. *Italian Journal of Neurological Sciences*, *20*(5 SUPPL.), 250–253. https://doi.org/10.1007/s100729970006

Parmar, H., Nutter, B., Long, R., Antani, S. & Mitra, S. (2020). Spatiotemporal feature extraction and classification of Alzheimer's disease using deep learning 3D-CNN for fMRI data. *Journal of Medical Imaging*, *7*(05), 056001. https://doi.org/10.1117/1.jmi.7.5.056001

Payan, A. & Montana, G. (2015). Predicting Alzheimer's disease a neuroimaging study with 3D convolutional neural networks. *ICPRAM 2015 - 4th International Conference on Pattern Recognition Applications and Methods, Proceedings*.

Penny, W., Friston, K., Ashburner, J., Kiebel, S. & Nichols, T. (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. https://doi.org/10.1016/B978-0-12-372560-8.X5000-1

Pool, J. L. (2002). Magnetic resonance imaging. *Biomedical Instrumentation and Technology*, *36*(5), 341–346.

Prinster, A., Quarantelli, M., Lanzillo, R., Orefice, G., Vacca, G., Carotenuto, B., Alfano, B., Brunetti, A., Brescia Morra, V. & Salvatore, M. (2010). A voxel-based morphometry study of disease severity correlates in relapsing-remitting multiple sclerosis. *Multiple Sclerosis*. https://doi.org/10.1177/1352458509351896

Rabeh, A. B., Benzarti, F. & Amiri, H. (2016). Diagnosis of Alzheimer Diseases in Early Step Using SVM (Support Vector Machine). *Proceedings - Computer Graphics, Imaging and Visualization: New Techniques and Trends, CGiV 2016*. https://doi.org/10.1109/CGiV.2016.76

Riley, K. P., Snowdon, D. A. & Markesbery, W. R. (2002). Alzheimer's neurofibrillary pathology and the spectrum of cognitive function: Findings from the Nun Study. *Annals of Neurology*. https://doi.org/10.1002/ana.10161

Sahraian, M. A., Radue, E. W., Haller, S. & Kappos, L. (2010). Black holes in multiple sclerosis: Definition, evolution, and clinical correlations. https://doi.org/10.1111/j.1600-0404.2009.01221.x

Sailer, M., Fischl, B., Salat, D., Tempelmann, C., Schönfeld, M. A., Busa, E., Bodammer, N., Heinze, H. J. & Dale, A. (2003). Focal thinning of the cerebral cortex in multiple sclerosis. *Brain*. https://doi.org/10.1093/brain/awg175

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. https://doi.org/10.1016/j.neunet.2014.09.003

Schmitter, D., Roche, A., Maréchal, B., Ribes, D., Abdulkadir, A., Bach-Cuadra, M., Daducci, A., Granziera, C., Klöppel, S., Maeder, P., Meuli, R. & Krueger, G. (2015). An evaluation of volume-based morphometry for pre-

diction of mild cognitive impairment and Alzheimer's disease. *NeuroImage: Clinical*. https://doi.org/10.1016/j.nicl.2014.11.001

Schneider, J. A., Arvanitakis, Z., Leurgans, S. E. & Bennett, D. A. (2009). The neuropathology of probable Alzheimer disease and mild cognitive impairment. *Annals of Neurology*. https://doi.org/10.1002/ana.21706

Serrano-Pozo, A., Frosch, M. P., Masliah, E. & Hyman, B. T. (2011). Neuropathological alterations in Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*. https://doi.org/10.1101/cshperspect.a006189

Smeets, D., Ribbens, A., Sima, D. M., Cambron, M., Horakova, D., Jain, S., Maertens, A., Van Vlierberghe, E., Terzopoulos, V., Van Binst, A. M., Vaneckova, M., Krasensky, J., Uher, T., Seidl, Z., De Keyser, J., Nagels, G., De Mey, J., Havrdova, E. & Van Hecke, W. (2016). Reliable measurements of brain atrophy in individual patients with multiple sclerosis. *Brain and Behavior*. https://doi.org/10.1002/brb3.518

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*. https://doi.org/10.1002/hbm.10062

Sood, T. & Khandnor, P. (2019). *Classification of parkinson's disease using various machine learning techniques* (Vol. 1045). Springer Singapore. https://doi.org/10.1007/978-981-13-9939-8_27

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*.

Stanisz, G. J., Odrobina, E. E., Pun, J., Escaravage, M., Graham, S. J., Bronskill, M. J. & Henkelman, R. M. (2005). T1, T2 relaxation and magnetization transfer in tissue at 3T. *Magnetic Resonance in Medicine*. https://doi.org/10.1002/mrm.20605

Stoessl, A. J. (2012). Neuroimaging in the early diagnosis of neurodegenerative disease. *Translational Neurodegeneration*, *1*, 1–6. https://doi.org/10.1186/2047-9158-1-5

Sweeney, E. M., Vogelstein, J. T., Cuzzocreo, J. L., Calabresi, P. A., Reich, D. S., Crainiceanu, C. M. & Shinohara, R. T. (2014). A comparison of supervised machine learning algorithms and feature vectors for MS lesion segmentation using multimodal structural MRI. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0095753

Tapiola, T., Pennanen, C., Tapiola, M., Tervo, S., Kivipelto, M., Hänninen, T., Pihlajamäki, M., Laakso, M. P., Hallikainen, M., Hämäläinen, A., Vanhanen, M., Helkala, E. L., Vanninen, R., Nissinen, A., Rossi, R., Frisoni, G. B. & Soininen, H. (2008). MRI of hippocampus and entorhinal cortex in mild cognitive impairment: A follow-up study. *Neurobiology of Aging*. https://doi.org/10.1016/j.neurobiolaging.2006.09.007

Taschler, B., Ge, T., Bendfeldt, K., Müller-Lenke, N., Johnson, T. D. & Nichols, T. E. (2014). Spatial modeling of multiple sclerosis for disease subtype

prediction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-10470-6_99

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., ... Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, *17*(2), 162–173. https://doi.org/10.1016/S1474-4422(17)30470-2

Thompson, P. M., Mega, M. S., Woods, R. P., Zoumalan, C. I., Lindshield, C. J., Blanton, R. E., Moussai, J., Holmes, C. J., Cummings, J. L. & Toga, A. W. (2001). Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cerebral Cortex*. https://doi.org/10.1093/cercor/11.1.1

Tishby, N., Levin, E. & Solla, S. A. (1989). Consistent inference of probabilities in layered networks: Predictions and generalization. *IJCNN Int Jt Conf Neural Network*. https://doi.org/10.1109/ijcnn.1989.118274

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A. & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/TMI.2010.2046908

Valentin Jospin, L., Buntine, W., Boussaid, F., Laga, H. & Bennamoun, M. (2020). Hands-on Bayesian Neural Networks - a Tutorial for Deep Learning Users.

Wang, J., Gu, B. J., Masters, C. L. & Wang, Y. J. (2017). A systemic view of Alzheimer disease - Insights from amyloid-β metabolism beyond the brain. https://doi.org/10.1038/nrneurol.2017.111

Xiao, Y. D., Paudel, R., Liu, J., Ma, C., Zhang, Z. S. & Zhou, S. K. (2016). MRI contrast agents: Classification and application (Review). https://doi.org/10.3892/ijmm.2016.2744

Xing, H., Meng, Y., Wang, Z., Fan, K. & Hou, D. (2018). Exploring geo-tagged photos for land cover validation with deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/j.isprsjprs.2018.04.025

Zhang, Y., Brady, M. & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/42.906424

Zhao, G., Liu, F., Oler, J. A., Meyerand, M. E., Kalin, N. H. & Birn, R. M. (2018). Bayesian convolutional neural network based MRI brain extraction on nonhuman primates. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2018.03.065