Deep learning of brain asymmetry digital biomarkers to support early diagnosis of cognitive decline and dementia

https://eprints.bbk.ac.uk/id/eprint/52554/

Version: Full Version

Citation: Herzog, Nitsa J (2023) Deep learning of brain asymmetry digital biomarkers to support early diagnosis of cognitive decline and dementia. [Thesis] (Unpublished)

Deep Learning of Brain Asymmetry Digital Biomarkers to Support Early Diagnosis of Cognitive Decline and Dementia

Nitsa J Herzog

Department of Computer Science,
Birkbeck College, University of London

A thesis submitted in partial fulfilment of the requirements
for the degree of *Doctor of Philosophy*

May, 2023

Abstract

Early identification of degenerative processes in the human brain is essential for proper care and treatment. This may involve different instrumental diagnostic methods, including computer tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET) scans. These technologies provide detailed information about the shape, size, and function of the human brain. Structural and functional cerebral changes can be detected by computational algorithms and used to diagnose dementia and its stages (amnestic early mild cognitive impairment - EMCI, Alzheimer's Disease - AD). They can help monitor the progress of the disease. Transformation shifts in the degree of asymmetry between the left and right hemispheres illustrate the initialization or development of a pathological process in the brain. In this vein, the thesis proposes a new digital biomarker for the diagnosis of early dementia based on the detection of image asymmetries and cross-sectional comparison of NC (normal cognitively), EMCI and AD subjects. Brain asymmetry images and features of brain asymmetries extracted from MRI of the ADNI and OASIS databases are used to analyze structural brain changes and apply machine learning classification of the pathology. The experimental part of the thesis investigates the potential of conventional machine learning algorithms, deep transfer learning and specially designed architectures of convolutional neural networks for distinguishing between cognitively normal subjects and patients with early or progressive dementia. The proposed pipeline offers a low-cost imaging biomarker for the classification of dementia. It can be potentially helpful to other brain degenerative disorders accompanied by changes in brain asymmetries.

Acknowledgements

I would like to acknowledge the Computer Science department of Birkbeck College University of London for awarding me with a 3-year BEI studentship, which helped me concentrate on my research topic fully.

A special grace to Professor George Magoulas for supervising my work in the critical last year of the study. His academic advice and feedback facilitated me in organizing publications and the thesis writing process.

I am grateful to all department administrative and academic staff who assisted me during my research and helped to run it smoothly.

Separately, I would like to thank my family for their supportive environment and understanding of my efforts.

Publications of the Thesis

1. Herzog, N.J. and Magoulas, G.D., 2021. Brain asymmetry detection and machine learning classification for diagnosis of early Dementia. *Sensors*, *21*(3), p.778.
   Material from this publication has been included in Chapters 3.3, 3.4, 3.5, 3.6, 4.3, 4.4, 4.5, 4.6. 5.2, 5.4, and 5.5.
2. Herzog, N.J. and Magoulas, G.D., 2021, June. Deep Learning of Brain Asymmetry Images and Transfer Learning for Early Diagnosis of Dementia. In *International Conference on Engineering Applications of Neural Networks* (pp. 57-70). Springer, Cham.
   Material from this publication has been included in Chapters 5.6, 5.7 and 5.9.
3. Herzog, N.J. and Magoulas, G.D., 2022. Machine Learning-Supported MRI Analysis of Brain Asymmetry for Early Diagnosis of Dementia. In *Medical Informatics and Bioimaging Using Artificial Intelligence* (pp. 29-52). Springer, Cham.
   Material from this publication has been included in Chapters 3.3, 3.4, 3.5, 3.6, 4.3, 4.4, 4.5, 4.6. 5.2, 5.4 and 5.5.
4. Herzog, N.J. and Magoulas, G.D., 2022. Convolutional Neural Networks-Based Framework for Early Identification of Dementia Using MRI of Brain Asymmetry. *International Journal of Neural Systems*, pp.2250053-2250053.
   Material from this publication has been included in Chapters 5.6, 5.7, 5.8, 5.10 and 5.11.

Table of contents

List of figures

List of tables

Chapter 1: Introduction

1.1 Background of the research

The human brain is examined with the help of advanced modern technology, which provides detailed scans of the brain tissues and demonstrates the functional activity of the brain regions associated with a specific mental or behavioural task.

Brain scanning is divided into two large categories: structural imaging and functional imaging. The most common structural neuroimaging methods are X-ray, structural Magnetic Resonance Imaging (sMRI), Diffusion Tensor Imaging (DTI), modification of MRI, and Computerized Tomography. Well-known functional methods are Electroencephalography (EEG), functional Magnetic Resonance Imaging (fMRI), and Positron Emission Tomography. MRI covers around 50% of imaging data used to diagnose brain diseases (Segato et al., 2020). A significant advantage of MRI over popular CT and X-ray scans is the absence of ionising radiation during the MRI session. The MRI contrasting agent is less allergic than iodine-based substances of CT scans and X-rays. Another advantage of MRI is the possibility to provide high contrast resolution of soft tissue compared to a CT scan, which is superior at imaging dense anatomical structures. All these factors make MRI the method of choice for regular health checks in the population older than 60. High-resolution images make a significant impact on the computer-aided diagnosis of brain-related disorders.

Early dementia, or amnestic Mild Cognitive Impairment (aMCI), belongs to the group of neurocognitive disorders and its symptoms include short-term short-time memory loss, language difficulties, lack of reasoning and judgment, hardship in coping with daily routines. Approximately 10% of the world population aged between 70 and 79, and 25% of the population older than 80 are diagnosed with MCI. It is acknowledged that 80% of the patients with aMCI develop severe dementia, in the form of Alzheimer’s disease, within seven years. The proportion of dementia in the general population is 7.1 %, roughly 46.8 million people. Figure 1 illustrates the growing number of people with dementia worldwide.

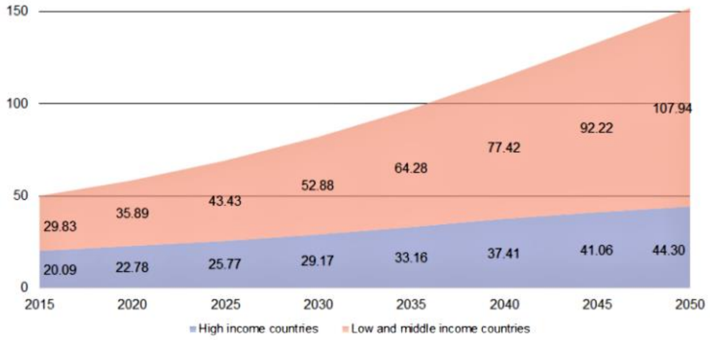


Figure 1: A number of people with dementia(millions) in low and middle-income countries compared to high-income countries (Source: https://www.alzint.org/about/dementia-facts-figures/dementia-statistics. World Alzheimer Report 2015 by Alzheimer's Disease International)

Neurogenerative disorders, such as Alzheimer's disease, are the most common, followed by vascular dementia, Lewy body dementia, Frontotemporal dementia, Parkinson's disease and Huntington's disease, which severely affect memory and other mental tasks. Amnestic MCI often becomes a prodrome of Alzheimer's disease. From this prospect, it is essential to identify this form of dementia in the early stage when proper care and treatment can stop or slow down the progression of the disease.

The diagnosis of MCI is based on neuropsychological testing, blood testing, and neuroimaging (ICD-11 2018; DSM-V 2013). Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), and Geriatric Mental State Examination (GMS) are the most common cognitive screening assessments. The tests usually include questions assessing orientation in place and time, short-time memory, attention, recall, and language ability of coherent speaking and understanding. For clinical judgment between MCI and AD, certain types of biomarkers, measured in the cerebrospinal fluid (CSF), are used. Amyloid-beta 42 (Ab42), total tau (T-tau), and phosphorylated tau (P-tau) globulins are identified in the early stage of Alzheimer's disease, whilst Hippocampal volume and rate of brain atrophy finalise the diagnosis.

Clinically, the progression or severity of dementia is measured according to the clinical dementia rating score. The rating scale varies from cognitive tests. Figure 2 demonstrates the progression of dementia.

Figure 2: Clinical stages of dementia (Source: mccare.com/education/mcidementia.html)

The research presented in this thesis is focused on the early detection and classification of dementia, aMCI using sMRI; when changes in the brain are not apparent to radiologists or clinical practitioners, the amyloid-beta deposition may be present or not, and the tau globulin is absent. This work involves computational algorithms used for the segmentation and evaluation of asymmetries in the brain's cortex.

The anatomical and functional brain asymmetries drew the attention of neuroscientists and other medical specialists since 19th century. Structural brain asymmetry is associated with the lateralisation of brain functions. The revealed lateralisation originates from genetic and epigenetic factors in the evolutionary development of the human brain (Bitar and Barry, 2018; Isles, 2018). Exposure to pathological factors during a human life also might cause changes in the lateralisation of the brain.

The evolutionary expansion of the left-hemispheric area is closely connected to speech production, perception, and motor dominance. The earliest observations of brain asymmetry were reported by the French physician, anatomist and anthropologist Pierre Paul Broca in the 19th century and then, ten years later, by German neurologist Carl Wernicke. They found that the patient's language was severely impaired when a stroke or tumour affected the left-brain hemisphere. Broca localised the afflicted area in the anterior left hemisphere, including some parts of the inferior frontal gyrus ("Broca's area"). The pathological process in that brain area significantly changed the patients' language production and syntactic processing. Changes in language comprehension, such as understanding spoken words, were primarily discovered by Wernicke in the posterior temporal-parietal region ("Wernicke's area"). It has been confirmed that differences in the brain's anatomical structure correlate with functional lateralisation. The left hemisphere is primarily responsible for language processing and logical thinking. The right hemisphere specialises in spatial recognition, attention, musical and artistic abilities. Emotions and manifestation are also connected to the right hemisphere (Gainotti 2019 – 2 articles).

The symmetry of the brain tends to change under the destructive processes accompanying a neuro-degenerative disease. These changes are significant in the most affected anatomical areas. The regions of the brain that are responsible for memory, language process and cognitive tasks show visible signs of tissue reduction. One of the first affected areas leading to the development of dementia is the hippocampus, which is followed by the amygdala and cortex. The hippocampus is needed for memory retrieval, and simultaneously the amygdala supports the emotional response to factual events. The anatomical areas of the cortex might be affected in the later stage of dementia in case of Alzheimer's Disease and be initially damaged if the subject belongs to an atypical form of Alzheimer's Disease, frontotemporal dementia, vascular dementia, or Lewi body dementia. Clinical symptoms of cortex atrophy vary in the anatomical areas. Changes in human language and behaviour depend on the right-left location of the pathological zone. The left temporal zone controls the semantic memory and perception of the words' meaning. The right zone affects the ability to recognise faces and objects. Damages in the frontal zone will modify the different patterns of human behaviour. Posterior cortical atrophy with a locus in the occipital lobes leads to the problem with visual processing information and spatial orientation.

Changes in brain asymmetry can be investigated through the brain scans such as MRI, PET, etc., which have become a popular diagnostic tool for brain-related disorders, including MCI and Alzheimer's Disease. A number of neuroscientists, Kim et al. (2012), Yang et al. (2017), Wachinger et al. (2015), and Liu et al. (2018), described the changes in the morphology and functionality and emphasized the asymmetrical development of the transformations in specific brain areas due to the development of cognitive decline and dementia. However, their studies were limited since their works did not discuss the possibility of using the changes of cortical and subcortical asymmetries as an additional classification biomarker. Machine learning approaches, including modern deep learning algorithms for the detection and classification of medical pathology, are widely represented in the literature and have practical implications in the diagnosis of AD and MCI. Imaging data collected from brain scans and used for ML modelling demonstrate promising prediction performance of dementia. Nevertheless, early changes in the brain structure due to the development of the degenerative processes often remain silent before their clinical or symptomatic manifestation. Hence, introducing diagnostic biomarkers that can predict the early onset of dementia has potential to contribute to the effective treatment of the disease.

1.2 Aim and objectives of the thesis and research questions

Aim

The research aims to propose a new brain asymmetry digital biomarker and investigate its potential to support automatic and robust detection of early stages of cognitive decline and dementia using structural magnetic resonance images and deep learning architectures.

Research questions and objectives

To achieve the above aim, the following research questions have to be addressed**.**

Q1. Can changes in structural brain asymmetry be detected by computer vision algorithms?

Q2. Can statistically engineered features from brain asymmetries distinguish between early and progressive stages of dementia?

Q3. Is it possible to use segmented images of structural brain asymmetry as a diagnostic source (biomarker) to predict early dementia (amnestic MCI) and Alzheimer's Disease?

Q4. What machine learning models are the most successful in the diagnosis of early cognitive impairment using images of brain asymmetries?

To answer these questions the following objectives should be met:

1. Derive a computer vision approach that helps to detect asymmetry between the left and right hemispheres of the human brain and segment them from MRIs.
2. Analyse and compare brain asymmetry imaging biomarkers obtained from three groups of subjects, namely Alzheimer’s Disease, Early Mild Cognitive Impairment and Healthy Controls, and use the transformation of asymmetry as a predictive factor or digital biomarker of the early degenerative processes in the human brain.
3. Investigate the potential of classic and deep machine learning algorithms to learn brain asymmetry biomarkers to support a medical diagnosis of generative brain disorders.

4. Analyse the functionality and effectiveness of classic machine learning algorithms in the classification of Alzheimer’s Disease, Early Mild Cognitive Impairment and Cognitively Healthy controls.
5. Analyse the functionality and effectiveness of Convolutional Neural Networks (CNN) and transfer learning models based on CNN for classifying Alzheimer’s Disease, Early Mild Cognitive Impairment and Cognitively Healthy control.

1.3 Thesis structure and contribution

The thesis is organised as follows.

Chapter 2, “Literature Review”, presents the literature related to this study and includes an overview of brain asymmetry in diagnosing brain-related disorders and the classification of early mild cognitive impairment and Alzheimer's Disease using one of the most extensive database MRI repositories, the Alzheimer's Disease Neuroimaging Initiative (ADNI), and the Open Affective Standardized Image Set (OASIS). The brain asymmetry section includes the description of the evolutionary changes in brain anatomy and the impact of these changes on brain functionality. The asymmetries are described concerning the person's age, gender and handedness. Recent findings are discussed based on the state-of-the-art literature. Computer-aided diagnosis of MCI and AD follows the antecedent part of the chapter. It provides an overview of the current state of the problem, including the proposed methodologies and the classification performance of the state-of-the-art methods. The chapter also provides an overview of computer vision and machine learning technologies used to analyse and classify medical data. It discusses computer vision applications in medical practice, focusing on diagnosing degenerative brain diseases, and gives an overview of image processing and segmentation techniques applied to imaging data. The machine learning part of this chapter discusses supervised machine learning algorithms and deep learning networks. Supervised learning algorithms, such as Support Vector Machines-SVM, k-nearest neighbour (KNN) and Linear Discriminant-LD, and CNN are described in detail, as they directly relate to the research presented in the thesis. This chapter is linked to the first objective of the research and indirectly contributes to research question one (Q1), two (Q2), and three (Q3) by providing a theoretical background for the work conducted to address these questions. Gaps in the literature are discussed and summarized at the end of the chapter. Generally, the chapter helps to understand the contemporary state of the diagnosis of dementia, identifies the gap in the diagnosis of MCI and raises questions that require further research.

Chapter 3, "Computational Framework for Brain Asymmetry Digital Biomarker Representation and Classification", introduces the computational framework of the thesis. This includes approaches used for the image retrieval, preprocessing and segmentation. First, a computational framework for the entire diagnostic process is described. Then the chapter presents the databases used in the research, provides an overview of magnetic resonance scan technologies and medical image formats used for saving the imaging data. The chapter discusses in detail the proposed methodology and provides the mathematical background for the detection object's symmetry. The chapter explains the image transformation and segmentation stages towards obtaining the brain asymmetry biomarker. This chapter directly contribute to the conducted research and establishes link to the first objective of the study and answers the research question one (Q1).

Chapter 4, "Features of Segmented Brain Asymmetry: Analysis and Classification" discusses the feature engineering methods. It explains the approach used for the generation of statistical features from MRI asymmetries. Then, it provides an analysis of the statistical features of three image classes, AD, EMCI and NC. The comparative feature analysis is done for five datasets which vary in size, symmetry axes, and gender. The analysis is supported by visual representation of the statistical features for each patients' group. The chapter explains how the collected features are represented among the patient groups and demonstrates the differences between early and progressive dementia. The chapter contribution is a feature analysis that helps to answer questions one (Q1) and two (Q2) and meet the research's first and second objectives.

Chapter 5, "Learning Brain Asymmetry Biomarkers" describes two classification machine learning workflows: conventional supervised machine learning and deep neural networks-based methods. The first workflow includes tests with classic machine learning algorithms and demonstrates the robustness of the proposed biomarker for classification. The chapter also describes deep learning methods in diagnosis of dementia, including transfer learning architectures proposed for the classification of early dementia. Transfer learning of AlexNet and VGG16 adapted for the MRI data is presented in detail, and the proposed CNN architecture, which incorporates different classification layers, is also included in this chapter. Later, the chapter provides experiments with machine learning classification algorithms and demonstrates the test results. The initial experiments were conducted with conventional supervised machine learning algorithms—the further tests were run with pretrained convolutional neural networks, and new CNN architectures created from scratch. Training, evaluation, and testing of the pretrained network were done using a transfer learning approach by connecting different classification modules to the output layer. The test results are combined in tables and supported by the ROC/AUC curves and other types of graphical representation. The statistical significance of the results was evaluated using Wilcoxon-Mann-Whitney rank-sum test. The chapter contributes to the diagnosis of EMCI and AD and demonstrates the robustness of brain MRI asymmetries in the early disease prediction by using machine learning models. The chapter meets objectives three, four and five and answers third (Q3) and fourth (Q4) research questions.

Chapter 6, "Conclusions and Future Work", includes detailed answers to research questions and outlines the study's significant findings. The chapter summarizes the information about features of brain asymmetry and interprets their changes in relation to age, gender, handedness, and brain size. It explains the classification results and suggests the ML models, which are more effective in the diagnosis of EMCI and AD using asymmetry biomarkers. The section also includes a comparison of the proposed models with those described in the literature. In the summary of the study, the chapter restates the research hypothesis and reiterates the key points of the work. Next, it explains the relevance and significance of the research and highlights its contribution to existing knowledge. The chapter discusses the study's limitations and gives an overview of future directions in diagnosing medical pathologies using computer vision and machine learning techniques.

Chapter 2: Literature Review

2.1 Introduction

Anatomy and functional features of the human brain are some of the most studied scientific areas, where in recent years, joint projects of medical and non-medical specialists have become a regular practice for the development of new strategies for the diagnosis and treatment of brain pathologies.

The two brain hemispheres have slightly different anatomy and function, and a detailed examination of their structure shows a variety of asymmetrical areas. Investigating the human brain's anatomical properties and functional ability is an intensively developing research area. The human brain has an overall leftward posterior and rightward anterior asymmetry (known as “Petalia and Yakovlevian torque”) (Toga and Thompson 2003). The right cerebral hemisphere protrudes forward, and the left hemisphere protrudes backwards compared to the right one. This type of asymmetry is mainly found in right-handed individuals - around 90% of the human population is right-handed (McManus 2019), while the opposite pattern in some cases is observed in left-handed individuals. Figure 3 illustrates the anatomical differences between the left and the right hemisphere of the brain of a cognitively normal person.

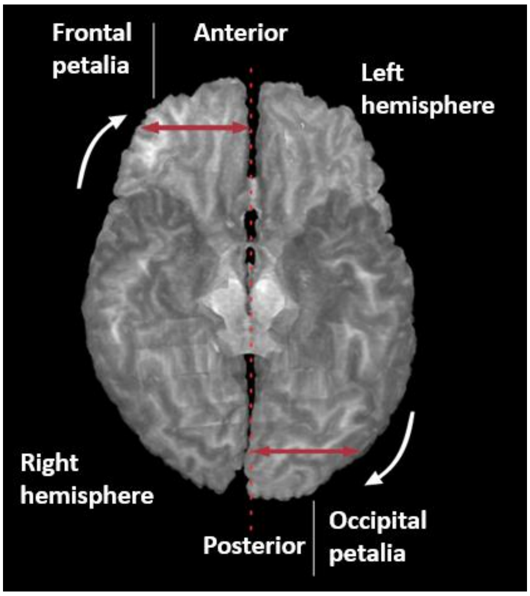


Figure 3: Normal brain lateralization (the Yakovlevian anticlockwise torque). (Source: CDI, Oswaldo Cruz German Hospital - Fleury Group / São Paulo 2015)

Brain asymmetry is closely related to human handedness. An interesting fact is that the foetal orientation during the pregnancy is correlated with the handedness of a newborn child. These asymmetries are first observed in the 29-31 weeks of gestational age. Almost

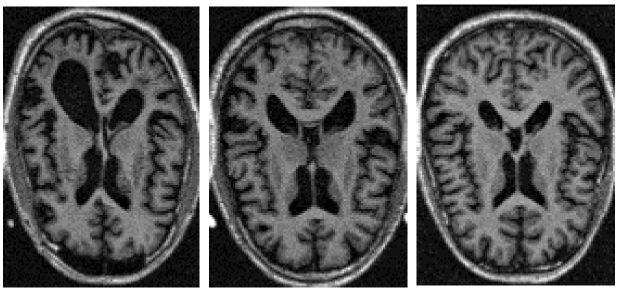
90% of the human population is right-handed (McManus, 2019). “Petalia and Yakovlevian torque” (Segato et al., 2020) is a term that describes an overall leftward posterior and rightward anterior asymmetry usually presented in right-handed individuals. Around 95% of right-handed persons have their speech and language zones in the left hemisphere, while only 5% show the language zone representation in the right hemisphere or bilateral. Compared to right-handed people, left-handed demonstrate a higher ratio of hemispheric lateralisation. A strongly dominant right hemisphere lateralisation is presented only in 7% of the left-handers. This proportion can vary from age. Up to 85% of left-handed children have language area dominance in the left hemisphere (Szaflarski et al., 2012).

Some studies highlight the differences in hemispheric lateralisation between males and females (Tomasi and Volkow 2012). The distinctions can be noticeable in linguistic performance, visuospatial or motor skills. The female brains show more symmetries in both cerebral hemispheres.

The level of asymmetry also depends on the age of the person. The brain functional hemispheric asymmetry in the frontal lobes of young adults is more lateralised than in elderly healthy persons. The activity reduction of the frontal cortex leads to age-related cognitive decline. It is registered by functional neuroimaging as changes in the domains of semantic, episodic, or working memory, perception, and inhibitory control. Elderly people demonstrate compensatory processes in the brain that transform brain lateralisation. Sometimes it looks like bilateral hemispheric activity (Cabeza 2002, Cabeza et al. 2004).

2.2 Brain asymmetry for the diagnosis of brain-related disorders

Alzheimer’s Disease and amnestic MCI, which in some cases become a prodrome stage of AD, are the most common types of dementia. Detecting dementia in the early stage benefits timely treatment and helps to optimize the care and management of patients. The diagnosis of Alzheimer’s Disease using MRI scans gives a clear picture of the grey and white matter brain atrophy. At the same time, early mild cognitive impairment does not have such obvious signs of brain tissues atrophy. Appendix C presents details about brain anatomy and the diagnosis of MCI. Figure 4 displays a comparison of MRI scans of brains transformed by neurodegenerative disorders and an intact brain.



(a) (b) (c)

Figure 4: MRI scans of the brain affected by Alzheimer's Disease (a), the brain of a person with early mild cognitive impairment (b), and the brain of the cognitively healthy person (c).

The white matter brain regions show a progressive decrease in the degree of asymmetry in patients with Mild Cognitive Impairment and an increase in asymmetry in patients with Alzheimer's disease (Yang et al., 2017). For concept approval, Yang et al. used diffusion tensor image tractography to construct the hemispheric brain white matter networks. The researchers came to the conclusion that the brain white matter (WM) networks show the rightward topological asymmetry when the right cerebral hemisphere becomes dominant in AD patients but not in the early phase of the MCI. Left-hemisphere regions are affected earlier and more severely. The abnormal hemispheric asymmetries of AD and MCI patients significantly correlated with memory performance. The research on image tractography left the investigation of the grey matter (GM) of the brain intact. The question "Are the cortical parts of the brain affected in the same way as white matter tracts?" has remained open.

Neuroscientists actively investigate changes in bilateral brain functionality in cases of diagnosed dementia. The functional cortical asymmetry progressively decreases in patients with MCI (Liu et al., 2018). Liu et al.'s research was based on whole-brain imaging. They registered and compared the spontaneous brain activity in patients with MCI, AD and NC using functional MRI. They discovered that patients with MCI and AD have abnormal rightward laterality in the brain compared to healthy controls with observed leftward lateralisation. At the same time, alterations in the brain lateralisation between patients with MCI and normal controls were different from alterations between patients with AD and normal controls. The rightward lateralisation in the patients with MCI and AD may be reflected as a relative increase in brain activation within the right hemisphere or a relative decrease in brain activation within the left hemisphere. Patients with MCI showed an increase in the activation of several brain regions in the right hemisphere during the processing of word memory tasks. Those areas were compensatory activated compared to the activation zones in the left hemisphere of the healthy controls. Liu et al. (2018) suppose that the reason for the abnormal right-lateralized pattern in patients with AD might be more complex than in patients with MCI. They think that structural differences between the groups potentially influence functional results. Still, their research did not investigate the relationship between brain structural asymmetry and brain functional lateralisation arch. All participants in the study were right-handed. The researchers did not determine whether right brain lateralisation occurs in left-handed persons. The researchers found a significant difference in brain functionality between MCI and AD patients. The patients with MCI had normal leftward lateralisation with some elements of abnormal rightward activity. In patients with AD, the typical pattern of left lateralisation disappeared, and some abnormal right-lateralized pattern was detected.

The degree of asymmetry is not the same in the different parts of the brain (Kim et al., 2012). Kim et al. tested the hypothesis that individuals with amnestic Mild Cognitive Impairment and various stages of AD have reductions in asymmetries in the heteromodal neocortex. They found significant changes in the degree of asymmetry in right-handed adults' inferior parietal lobe of the brain. The cortical asymmetry was investigated using surface-based morphometry (SBM) to measure the cortical thickness. Their results show that the neocortical thickness asymmetries of the medial and lateral sides of the right and left parts of the brain were different from each other. The decrease in asymmetry was registered in the lateral parts of the frontal and parietal lobes and an increase in the temporal lobe. The left Perisylvian areas responsible for language functions, except Broca’s speech area, demonstrated leftward asymmetry. Other brain areas, which specialised in spatial perception, facial recognition, and memory processing, showed rightward asymmetry. Kim et al. assumed that the cortical asymmetry shown in healthy controls generally decreases in AD. But their study has some limitations, which the investigators highlighted. They did not directly examine the changes in cortical asymmetry observed during the AD progression, which gives a clear picture of an increase in asymmetry in case of severe AD. Also, it is unclear whether other degenerative diseases can cause similar changes in cortical asymmetry.

Wachinger et al. (2016) investigated the neurodegenerative processes in the subcortical brain structures of patients with Alzheimer’s disease. They proposed a measure of brain asymmetry based on spectral shape descriptors from the BrainPrint. BrainPrint is an ensemble of shape descriptors representing brain morphology and captures shape information of cortical and subcortical structures (Wachinger et al., 2015). Progressive dementia is associated with a significant increase in the neuroanatomical asymmetry in the hippocampus and amygdala. The research findings (see Table 1) prove that shape analysis can detect the progression of dementia earlier than volumetric measures. Shape asymmetry, based on longitudinal asymmetry measures in the hippocampus, amygdala, caudate and cortex, can be a powerful imaging biomarker for the early presymptomatic prediction of dementia.

Table 1: State-of-the-art neuroscientific methods of registration of brain asymmetry

Reference	Method	Conclusion
Yang et al. (2017)	Diffusion tensor image (DTI) tractography to construct the hemispheric brain white matter networks	Decrease of asymmetry in patients with MCI and an increase of asymmetry in patients with AD
Liu et al. (2018)	Registration of the spontaneous brain activity in patients with MCI, AD and NC using functional MRI (fMRI)	The functional cortical asymmetry progressively decreases in patients with MCI
Kim et al. (2012)	Surface-based morphometry (SBM) to measure cortical thickness	The degree of asymmetry is not the same in the different parts of the brain
Wachinger et al. (2016)	Measurement of brain asymmetry based on spectral shape descriptors using BrainPrint	Progressive dementia is associated with an increase in asymmetry in the hippocampus and amygdala; shape analysis can detect the progression of dementia earlier than volumetric measures

2.2.1 Summary of imaging technologies for dementia progression in neuroscience research

This part of the literature review focuses on the detailed study of the anatomy and functionality of the human brain. It uses the idea that the symmetry of the brain structure changes during the development of the neurodegenerative disorder. This section gives an overview of findings obtained with several medical imaging technologies that investigated the differences between the left and right parts of the brain due to the progression of dementia. The research methods of those studies include diffusion tensor image tractography, functional MRI, surface-based morphometry, and BrainPrint software, which belongs to the shape descriptors of brain morphology.

Diffusion tensor image tractography investigated the white matter tracts. The findings demonstrated the rightward topological asymmetry of the brain white matter in AD patients (Yang et al., 2017). The study confirmed the fact of asymmetrical changes in cases with developed dementia. However, the levels of the left-right symmetry shifts were not observed and described for the different stages of degenerative processes. Moreover, the study does not examine the grey matter property, which is also affected by the destructive process. Exploration of the grey matter thickness described by Kim et al. (2012). The investigators calculated the degree of asymmetry of the heteromodal neocortex in the multiple areas of the brain's frontal, temporal and parietal lobes. The measurements were done for the group of individuals with amnestic MCI and AD using surface-based morphometry. The obtained results clearly indicated the differences in the level of asymmetries in some parts of the neocortex between groups of people. The researchers found a partial decrease in cortical asymmetry in AD patients compared to the healthy controls. Nevertheless, the progressive changes in the asymmetry were not monitored and analyzed, and the pattern of the left to right symmetry conversion remains unclear. Wachinger et al. (2015) provided additional information about a structural peculiarity of the brain. Brain Print descriptor detected the shape changes of the subcortical areas such as the hippocampus and amygdala and indicated increased asymmetry in patients with AD. The research did not explore the shape features of MCI patients and the transformation of subcortical areas due to the progression of cognitive decline. The discovery of the brain functionality added some important information about the progression of dementia. Liu et al. (2018) found that patients with MCI and AD have abnormal rightward laterality in the brain compared to healthy controls. However, patients with MCI demonstrated the rightward activation only in a few brain regions responsible for word memory tasks. At the same time, AD patients had a significant increase in the entire right hemisphere activity. The researchers have assumed the connection between functional and structural brain changes, but they did not provide proof for this hypothesis.

All the provided findings point to the changes in brain symmetry due to the development and progression of dementia, but some questions remain unanswered.

In line with the above literature, the work presented in the thesis uses data collected in neuroimaging studies, as it will be discussed in the next section. The first part of the thesis describes the implementation of computer vision methods for the detection and segmentation of a structural image asymmetry from MRIs of NC, EMCI and AD subjects. The proposed computational framework includes methods for the visualisation of the shape of asymmetry and the recognition of the pattern of asymmetrical changes in an image. To the best of our knowledge, it is the first time this approach is proposed for generating digital biomarkers of brain asymmetry.

The second part of the thesis is focused on machine learning methods for the diagnosis of early dementia using the generated biomarkers of image asymmetries. The critical points of this part of the computational framework are to show the robustness of the proposed biomarkers in the early diagnosis of hidden dementia using MRI data.

The following literature review section describes the classification of AD and MCI. All these studies used the same ADNI database, which lies at the base of this thesis.

2.3 Classification of Alzheimer's Disease and early mild cognitive impairment using the ADNI database

The Alzheimer's Disease Neuroimaging Initiative database (adni.loni.usc.edu) was launched in 2003 as a public-private partnership led by Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease (for up-to-date information, see www.adni-info.org). Multiple research studies have benefited from MRI data from the ADNI database, the same used in the thesis. Some of them are represented below. These studies focused on machine learning approaches for the modelling and classification of neurogenerative disease, including stable and progressive forms of mild cognitive impairment and Alzheimer's Disease. However, these studies worked with whole brain images but not segmented asymmetries.

A large part of research in this field includes machine learning-based diagnostic approaches that use features engineering, which has been demonstrated to contribute to successful modelling. For instance, Beheshti et al. (2017) implemented feature-ranking and a genetic algorithm to analyze structural magnetic resonance imaging data of 458 subjects. The researchers state that the proposed system can distinguish between stable and progressive MCI and predict the conversion of MCI to Alzheimer's Disease from one to three years ahead it is clinically diagnosed. Beheshti et al. (2017) identified atrophic grey matter regions using voxel-based morphometry (VBM). The features were extracted after applying a 3D mask, and they were ranked according to their t-test scores. Features with t-test values higher than 70% were combined into new subsets. A genetic algorithm with the Fisher criterion function (Welling 2005) evaluated the separation between the two data groups and helped select the most discriminative feature subsets for the classification. The classification process was finalized with linear SVM (Evgeniou and Pontil 1999). The classification performance was evaluated with a 10-fold cross-validation procedure. The calculated accuracy shows 93.01% for stable MCI and 75% for progressive MCI. The feature selection process raised one of the accuracy from 78.94% to 94.73%.

Another group of scientists (Moradi et al., 2015) investigated the conversion of MCI to AD. Their algorithm identifies AD in a time between one to three years prior to the development of clinical symptoms. The proposed algorithm is based on aggregated biomarkers and a random forest (RF) classifier (Breiman, 2001). The MRI data were preprocessed by removing images with age-related changes in the brain’s anatomical structure using a linear regression model. Feature selection was implemented on AD and NC images by a regularized logistic regression (RLG) algorithm (Tripepi, 2008). The classification stage is performed using a semi-supervised low-density separation (LDS) method (the LDS is a two-step algorithm that relies on the graph-distance kernel and the Transductive Support Vector Machine-TSVM learning). At the beginning of this stage, the classifier is trained with labelled AD and NC data. Then, unlabeled MCI images are fed into the classifier. It helps to separate the stable and progressive MCI and get them labelled. In the final stage, the output of the LDS classifier, as an input feature, is combined with age and cognitive measurements feature vectors into the RF classifier. The aggregated biomarker distinguishes between stable and progressive MCI and approximates the probability of conversion of MCI to AD. The image sequences of 825 subjects were evaluated with a 10-fold cross-validation method. The results showed that the predictive performance of the aggregated biomarker is higher than the performance of single biomarkers. MRI data with a combination of cognitive and age measures improves the classification accuracy by 5.5% (from 76.5% to 82%).

Another approach (Zhu et al., 2017) proposed an algorithm for the Joint Regression and Classification (JRC) problem in the diagnosis of MCI and AD. This regularization-based method aims to consider the related similarity of features, samples, and their responses. The features are associated with each other if their respective weight coefficients are similar. The weight coefficients are linked to the response variables via feature vectors and demonstrate the resembling type of relation. The same rule is applied to a matching pair of samples and their respective response values. The regularization method was assessed with MRI and PET image sequences of 202 subjects (Wong et al., 2003). The images were separated into 93 regions of interest (ROI) using a volumetric measure of the grey matter of the brain. Structural MRI scans were aligned with functional PET images using affine registration. The average intensity values were calculated for each ROI. Structurally and functionally related, features were extracted from each ROI and sent to the feature selection process using the regularization algorithm. Extracted features were expected to predict jointly one clinical label and two clinical scores. Imaging data for clinical labelling were classified with SVM. Other types of data obtained from cognitive tests were used for training two more Support Vector Regression (SVR) models for the prediction of clinical scores of AD Assessment Scale-Cognitive Subscale (ADAS-Cog) and Mini-Mental State Examination (Nogueira et al. 2018). The results were obtained using binary classification methods and 10-fold cross-validation. In the initial stage of the experiment, the classification and regression tasks were performed without feature selection. The results of this stage were considered as a baseline. In the next run, the baseline was compared to single-task results, when the selected features are classified independently, and multi-task results, when the features are classified jointly for the classification and regression models. The proposed joint approach shows the superiority of the single-task approach by 5.6%. Compared to the baseline, the average accuracy for a single task increases by 6%, and for multi-task by 8.8%. The proposed models were compared with two state-of-the-art methods: High-Order Graph Matching (HOGM) (Duchenne et al. 2011) and Multi-Modal Multi-Task (M3T) (Zhang and Shen 2011). The Joint Regression and Classification model outperforms their competitors by improving classification accuracy by 5% (vs HOGM) and 4.7% (vs M3T) for MRI, and 4.6% (vs HOGM) and 4.2% (vs M3T) for PET. The highest archived accuracy for the classification AD vs NC is 95.7%, MCI vs NC is 79.9%.

Many studies describe changes in the cortical and subcortical brain areas during the development of Alzheimer’s Disease. Special attention requires the transformation in the architecture of white matter. The properties of white-matter tracts can be measured using

Diffusion Tensor Imaging (Glozman and Le, 2014). Features collected from white-matter tracts of 509 subjects and processed with the SVM and Logic Regression classifiers showed the volumetric loss in fibres during the progression of the disease. Classification performance of DTI data showed 92% accuracy for detecting Alzheimer’s Disease.

A novel ensemble-based machine learning algorithm able to predict the conversion of Mild Cognitive Impairment to Alzheimer's Disease was proposed in the study by Grassi et al. (2019). The information was obtained from socio-demographic characteristics, clinical records, and neuropsychological test scores. The study did not use imaging data. Data of 550 subjects with MCI and follow-up assessments for at least three years were included in the study. Collected features were processed with 13 supervised ML algorithms. The final result was obtained by combining a weighted rank average of the predictions generated by each individual algorithm. Estimated performance was received with a 10-fold cross-validation procedure, repeated ten times. The classifier demonstrated an AUROC (Area Under the Receiving Operating Curve) of 0.88, a sensitivity of 77.7%, and a specificity of 79.9% on unseen data.

Zhou et al. (2018) applied the transfer learning Adaboost (TrAdaboost) algorithm for the classification of AD, EMCI and NC. The MRI data for the research was collected from 507 subjects in the ADNI database and 36 subjects from the local hospital. The images were segmented using the spm-Dartel toolbox (Ashburner et al., 2014) and registered with the Automated Anatomical Labeling (AAL) atlas (Rolls et al., 2020) to allocate the regions of interest. Grey matter tissue volumes calculated from 90 anatomical areas became the main parameters used for classification. The gained information was passed to the feature selection process based on the Shannon mathematical theory of communication (Shannon, 1948) and the C4.5 decision tree statistical classifier (Arellano et al., 2018). As a result, the number of features presented in the three classes was reduced to seven. The TrAdaboost algorithm was used to classify AD, MCI, and normal controls data. The training and testing data was drawn from different distributions for transfer learning. If AD data was used for training, then the MCI data was applied for testing. A similar rule was applied to AD vs NC and MCI vs NC groups. Obtained features were examined on their impact on the accuracy using iterative optimisation. Finally, the training data from the ADNI subjects were used for testing the local hospital data. The classifier was trained with the Adaboost algorithm. The highest achieved accuracy was 85.4% for the ADNI database and 93.7% for the local hospital data. The accuracy of the proposed algorithm was compared with SVM (support vector machine), KNN (K-nearest neighbour), ITML (information theory metric learning), and Linear-MSVM (linear metric-based support vector machine). The TrAdaboost algorithm significantly overperformed all the competitors.

Another stream of research has taken advantage of machine learning methods that generate features as part of the training process. These methods employ Artificial Neural Networks and Deep Learning and have recently attracted a lot of attention in medical image analysis and classification. They can process a large amount of data and learn in a supervised (labelled) or unsupervised (unlabelled) mode. Particularly, diagnostic approaches that use Deep Learning in most cases do not require complicated, time-consuming image preprocessing and feature engineering techniques and produce state-of-the-art results.

In this context, the Convolutional Neural Network (CNN) is one of the models successfully adapted to classify imaging data (Yamashita et al., 2018). Basaia et al. (2019) have built and evaluated a CNN algorithm that predicts AD, progressive cognitive mild impairment, and stable cognitive impairment. The researchers used T1-waited structural MRIs of 1409 subjects from the ADNI database. The image data was split into training, validation, and testing sets in the proportion of 90% for the first two and 10% for the last one. 10-fold cross-validation was applied. They used transfer learning, which exploits pre-trained initial weights, and were able to reduce the training time and increase the network performance. High predictive accuracy was achieved in both databases with no significant difference. The highest percentage for AD vs HC (healthy control) classification accuracy was: 99% for ADNI. For C-MCI vs HC and s-MCI vs HC, accuracy was 87% and 76%, respectively, while for AD vs c-MCI and AD vs s-MCI, performance was 75% and 86%, and for c-MCI vs s-MCI, 75%.

Multi-Layer Perceptron (MLP) and a Convolutional Bidirectional Long Short-Term Memory (ConvBLSTM) model were proposed by Stamate et al. for the diagnosis of dementia (Stamate et al. 2020). Different clinical sources and protocols of 1851 participants of the ADNI database were combined. The collected biomarkers consist of 51 input attributes. They include baselines demographics data, functional activity questionnaire, MMSE, CSF biomarkers, neuropsychological tests, and measurements received from MRI, PET and genetic data. The ReliefF method (Robnik-Šikonja and Kononenko, 2003) and permutation test (Pesarin and Salmaso, 2010), including 500 permutations of labels, were combined for feature selection and ranking. The top-10 ranked features have been sent to classification models. 75 % of the data were used for training and the rest for testing. The predictive results were obtained using Monte Carlo simulations (Johansen et al., 2010). All models were able to predict dementia and mild cognitive impairment accurately. The highest accuracy of 86% was achieved with the Multi-Layer Perceptron model.

Another study (Lama et al., 2017) proposed an unsupervised deep learning method for the classification of AD, MCI, and NC. The algorithm extracts the features with PCA (Abdi and Williams 2010) and processes them with a Regularized Extreme Learning Machine (RELM) (Ding et al. 2014). RELM is based on a single hidden-layer feedforward neural network. The investigators chose high-level features using the softmax function (a function that takes a vector of real numbers as input and normalizes it into a probability distribution). The results of RELM are compared with multiple kernel SVM and import vector machine (IVM) (IVM classifier based on Kernel Logic Regression uses a few data points to define the decision hyperplane and has a probabilistic output). The researchers have done 100 tests of imaging data collected from 214 subjects using 10-fold cross-validation and ten trials with the leave-one-out method. They separated training and testing images with a ratio of 70/30 for the 10-fold cross-validation and 90/10 for the leave-one-out validation. The study confirmed that RELM improves the classification accuracy of AD and MCI from 75.33% to 80.32% for binary classification and 76.61% for multiclass classification.

Lu et al. (2018) describes the multiscale neural network-based analysis for early diagnosis of Alzheimer's Disease. Imaging modalities of fluorodeoxyglucose positron emission tomography (FDG-PET) of the ADNI database are used and analyzed as the main imaging data. FDG-PET is one of the imaging techniques that allows receiving quantitative measures of the metabolic activities of the human brain. Regional abnormalities of the metabolic brain activities are considered to be emphasizing the cognitive and functional decline noticeable in patients with AD. The data of 1051 subjects were analyzed using a multiscale deep neural network that distinguished metabolic patterns of the NC, MCI and AD patients. FDG-PET images were preprocessed before the classification task. In this vein, the manually segmented and subdivided T1 MRIs into multiple small ROI were co-registered with FDG-PET images. This procedure allowed it to extract metabolic measures from ROI and train them with a multiscale deep neural network (MDNN). The MDNN is a composition of four deep neural networks. Three of them were trained independently with different scaled metabolic features. The joined output of these networks was sent to the input layer of the final DNN classifier. Each single DNN was initially pretrained with a stacked-autoencoder (SAE) (Vincent et al., 2010). SAE represents a type of artificial neural network based on stacking layers of autoencoders that denoise corrupted versions of unlabelled data in the hidden layers and reconstruct the data in the output. The pretrained networks were fine-tuned as MLP using transfer learning architecture by keeping only three encoded layers and adding to them a Softmax layer. All the data were divided into ten separate subsets with the intention to obtain robust and stable results. Nine subsets were used for the training of the classifier and one subset for testing it. The final result was obtained as an average of ten repeated runs from ten different multiscale deep neural networks. Presymptomatic AD was distinguished from non-progressive to the AD of MCI with an accuracy of 82.51 %.

3D neural network models have been recently developed for the diagnosis of dementia. One of these models was applied to structural MRIs for the detection of Alzheimer's Disease (Yagis et al., 2020). The researchers implemented a 3D VGG model for the brain MRI datasets downloaded from the ADNI and OASIS databases. VGG stands for the Visual Geometry Group based at Oxford University, UK. The VGG model is a variant of convolutional neural net researchers adapted to the 3D images. The data was collected from 200 subjects from the ADNI database and 200 subjects from the OASIS database that equally represent HC and patients with AD. Data preprocessing and normalization atlas-based co-registration with a standardized brain template and skull stripping. The proposed 3D VGG16-based model contains four blocks with convolutional and pooling layers. The image maps are constructed with an increased number of features in each subsequent layer. The dropout layer finalizes all convolutional blocks with a selective probability of 0.5 and three fully connected layers with Softmax function in the output. The architecture was built using the open-source TensorFlow machine learning platform created by Google (Abadi et al., 2016). The model is evaluated using a 5-fold cross-validation procedure, and the averaging accuracy is obtained after five repeated runs. The testing is done on two different sets based on imaging data of 40 subjects. The received accuracy in the diagnosis of AD is 73.4%± 0.04 for the ADNI dataset and 69.9%± 0.06 for the OASIS dataset.

The fusion of deep learning models was proposed by Qiu et al. (2018) for the diagnosis of cognitive decline. They investigate the impact of the Logical Memory (LM) test and the Mini-Mental State Examination on enhancing the diagnostic accuracy received through MRI data. The data of 368 subjects with normal cognition and MCI were acquired from the National Alzheimer Coordinating Center (NACC) database. In these datasets, 303 individuals had normal cognition and 83 individuals were diagnosed with MCI. Of the MCI patients, 82 had amnestic MCI, and one had the form of non-amnestic MCI. The imaging data sets consist of 2D slices of structural MRIs in the axial plane. At least 20 images were collected from each MRI sequence. The selected slices covered the previously reported areas as regions of interest in the diagnosis of AD and MCI. Memory and mental score results were included as the numeric data values. Two MLP models were trained with the results of the LM and MMSE tests and combined with the fused MRI model using the majority voting schema. The training of MRI data was done using the CNN VGG-11 architecture adapted to perform the binary classification task using the transfer learning strategy. Two fully connected (FC) layers were added to the configuration with the dropout layer after the first FC and Softmax layer after the second FC. The datasets were split into three parts with 50% images for training, 25% for validation, and 25% for testing. Each split had the same proportion of MCI and NC subjects. Five iterations were performed to compute the tests' averaging value and standard deviation. F1 score was calculated after a single iteration on the validation set. The neural network with the highest F1 score was used for further testing. For memory testing, MLP models included the combination of linear and nonlinear activation functions. This makes the MLP models capable of exploring complex relationships between selected features and the detection of MCI. The MLP model has three layers: input, hidden, and output, with 3 or 4 nodes in the first level, 20 nodes in the second level and two nodes in the decision level, respectively. Sigmoid Softmax and nonlinear Softmax functions were attached to the hidden and output layers. The specific weights for MCI and NC classes were assigned for models analyzing MMSE and LM. The voting approach and model fusions are performed in the following way. The voting approaches, such as majority voting, max voting, and mean voting, were applied to three independent transfer learning VGG-11 models. Then, the second level of majority voting originated in an MRI-based fusion model. The final level of majority voting combined the features collected from the MRI fusion model and two MLP cognitive test result models for the prediction of MCI. Averaging results across five runs evaluated the performance of the final multimodal model. The researchers concluded that the multimodal fusion model that incorporates features from MRI scans, LM data and

MMSE tests could predict MCI better than single-based models and models with the majority voting approach. The accuracy of the final multimodal fusion model reached 90.9%.

Classification differences between dementia with Lewy bodies and Alzheimer's Disease were investigated using a deep neural network (Iizuka et al., 2019). The 480 brain perfusion single-photon emission computer tomography (SPECT) images were obtained from 300 individuals and equally divided between DLB, AD and cognitively normal (NL) groups. The additional feature, used for differential diagnosis among the pathologies and known as the cingulate island sign (CIS), was collected from the posterior cingular cortex of the brain. The pick of this feature exists in the stage of mild dementia and gradually disappears in the progression to DLB. CNN was applied for the diagnosis of dementia. 80 % of images were used for training and 20 % for testing. CNN architecture included four convolutional, four max-pooling layers with ReLU (rectified linear activation unit) functions and was finalized with a fully connected and Softmax layer for the classification task. The optimal number of the discriminative features was obtained after 100 epochs. CNN features from the output layer were processed with the Gradient-weighted class activation mapping (GradCAM) to visualize those areas of the brain that were emphasized by the classifier. The predictive accuracy was evaluated with additional statistical measures such as receiver operating characteristic (ROC) curve and area under the curve (AUC). The computed AUC results for the DLB vs NL, DLB vs AD, and AD vs NL classes are 0.954, 0.935 and 0.943, respectively. CIS correlations between DLB vs AD and DLB vs NL were estimated using Pearson's correlation coefficient (Schober et al., 2018). Correlations between CIS ratios and clinical scores for DLB vs AD and DLB vs NL were evaluated using Spearman rank correlation coefficients (Schober et al., 2018). The obtained diagnostic accuracy was 93.1%, 89.3% and 92.4% between DLB vs NL, DLB vs AD and AD vs NL classes accordingly.

Biomarkers of the prefrontal cortex of the brain related to the diagnosis of MCI are analyzed (Yang et al., 2019). Three groups of biomarkers, such as mental task performance, digital and imaging biomarkers, were investigated by the linear discriminant analysis (LDA), statistical analysis and CNN. All data were collected from 15 MCI patients and nine healthy controls of the same age and similar educational backgrounds. Mental task performance was evaluated using functional near-infrared spectroscopy (fNIRS) (Bunce et al., 2006). Digital biomarkers included statistical measures of the mean value change between the rest and task periods, slopes of the peak time in the hemodynamic response, and skewness and kurtosis of the above values. Imaging biomarkers were collected from a correlation map and t-map. Analysis of cognitive performance included evaluating working memory, mental reaction to the dominant responses, and the multiple words in a given semantic category. T (topographic) or activation map was generated after registering hemodynamic cortical activities during the mental tasks. The correlation map was built based on the temporal functional correlation between connected channels in the prefrontal lobe. For the classification task, digital biomarkers were analyzed using the LDA. The imaging biomarkers were classified with CNN. Four main layers presented CNN: two convolutional and two fully connected. Max-pooling and dropout layers were added to the CNN architecture to improve the classification performance. The obtained results showed the advantage of CNN over the LDA and statistical analysis. The highest classification result of 90.62% accuracy in the detection of MCI was received using imaging biomarkers. LDA and statistical analyses demonstrated 76.67% and 60% accuracy, respectively.

A deep learning approach based on convolutional autoencoders is proposed for the detection of AD (Martinez-Murcia et al., 2019). Imaging and non-imaging data were used for the study. Imaging data (T1-weighted MRI) consists of 2182 images collected from 479 subjects. Non-imaging data includes the age of patients, presence of the Apolipoprotein E4 in chromosome 19, the concentration of Tau protein in cerebrospinal fluid (Zou et al., 2020), results of MMSE and Alzheimer’s Disease Assessment Scale (ADAS) cognitive tests and Clinical Dementia Ratio (CDR) (McDougall et al., 2021). Before classification, all images were normalized using integral normalization when the spector of image intensities is divided by the average intensity value (Martinez-Murcia et al., 2018). Encoder-decoder CNN architecture has been chosen for the classification task (Baldi, 2012). This type of CNN is able to minimize reconstruction errors. The encoder of the architecture consists of 5 convolutional layers with a global average-pooling layer at the end. The decoder has an additional five convolutional layers with dense and reshaped layers at the beginning of the configuration. The encoder is connected to the decoder via Z-layer. Z-layer is a fully-connected layer with Z-neurons and low-dimensional representation. It was proposed to use the Mean Squared Error (MSE) as a reconstruction error between input and output images. This type of convolutional network has position-invariance, can work with volumetric images and requires smaller memory than the dense-connected network. In the proposed architecture, most pooling layers were replaced by convolutions with strides that prevent the loss of information without losing efficiency. The architecture was trained with 20 epochs using the RMSprop optimization algorithm (Ruder, 2016), with a 0.001 learning rate. 10-fold cross-validation was used to estimate the methodology. Performance metrics such as mean, standard deviation, sensitivity, specificity, the total balanced accuracy and F1-score were calculated. During the classification tasks, the optimal number of Z-manifolds was equal to 20. The Z-manifold demonstrated a high ability to predict classes and clinical variables and indicated the atrophic brain regions. Regression analysis helped establish the correlation between imaging characteristics and other clinical data. Alzheimer’s Disease was diagnosed with an accuracy of 84%.

Duc et al. (2020) proposed to join 3D CNN and MMSE features for the diagnosis of Alzheimer’s Disease. The researchers created special feature maps for the regression and classification tasks using resting-state functional Magnetic Resonance Images (rs-fMRI) obtained from 331 participants. The fMRIs were co-registered with their corresponding structural T1 MRIs. The regression approach for calculating MMSE score included tests with linear least square regression (LLSR), tree regression with group independent component analysis (gICA) features, bagging-based ensemble regression and support vector regression. Some optimization techniques were applied to the regression algorithms. They include MVPA (multivariate pattern analysis)-based approaches. One of them is Least Absolute Shrinkage and Selection Operator (LASSO), and another one is the support vector machine recursive feature eliminator (SVM-RFE) (Duc et al., 2020). The SVM-RFE is a feature optimization technique that looks for reciprocal relations between multiple features. Both these techniques, LASSO and SVM-RFE, take an active part in feature ranking and selection processes. For the classification task, it was developed a modified version of the VGG network. The changes were done by adding the batch normalization layers to the existing convolution layers. Stratified 10-fold cross-validation was performed on a 3D CNN model. 90 % of imaging data was used for training and 10 % for testing the model. The classifier’s performance was evaluated using statistical measures such as accuracy, sensitivity, specificity, and balanced accuracy. The performance of regression models was examined with R-squared (R^2), and root mean square error. The mean accuracy of 85.27% was received for the classification of AD versus cognitively normal subjects. MMSE results were received with different optimized regression models. The best-optimized model was the linear least square regression with the combined gICA and

RFE features. Its RMSE value is 3.27 ± 0.58 and R² value is 0.63 ± 0.02.

The prognosis of MCI and the risk of AD development is one of the main lines of research (Li et al., 2019). Longitudinal time-to-event study of clinical and imaging changes accompanying the development of MCI involved the observation of 2046 individuals. The research method was based on the analysis of the structural hippocampal MRI using a deep learning framework. The imaging data were obtained from the ADNI and the Australian Imaging Biomarkers and Lifestyle Study of Aging (AIBL) databases. The images from the first database were used for the model training, and the second group of images was used for evaluating the results. Bilateral hippocampal regions were received from MRIs using the atlas-based segmentation technique. The right and left hippocampal images were processed in separate streams using CNN. Deep neural network architecture consists of a convolutional layer, two max-pooling layers and three residual blocks finalized with a global averaging pooling layer. Two streaming processes were concatenated in the dropout layer and sent to the fully-connected layer for another labelling. Each residual block includes two convolutional layers with batch normalization function (BN) and rectified linear units (ReLUs) (LeCun et al., 2015). The time-to-event prognosis of the progression of MCI to AD was conducted using LASSO regularized Cox regression model (Cox, 1972; Du and Tibshirani, 2018). This model generates an overall risk score for the disease progression. These results were combined with non-imaging data such as age, gender, education, and cognitive tests measures. It was evaluated how the hippocampal deep imaging features are correlated with clinical measures and how the parts of the hippocampus contribute to the prognosis and classification of dementia. The accuracy of the results was evaluated using time-dependent ROC curves and concordance index (C-index) (Brentnall and Cuzick, 2018). C-index is defined as a ratio of concordant pairs to the total number of evaluation pairs and, as a ROC curve, has a range between 0 to 1. The deep learning classification accuracy for classes AD vs NC was 0.900 for the ADNI and 0.929 for the AIBLwith AUC values equal to 0.956 and 0.958 accordingly. The CNN model predicted the progression of MCI to AD with a C-index of 0.762 for the ADNI and 0.781 for the AIBL. The model's performance improved after combining the imaging features with the cognitive measure. Obtained C-index is 0.864 and 0.848 for both databases, respectively.

Another approach that predicts the conversion of MCI to AD was proposed by (Spasov et al., 2019). The research exploited the imaging data received from MRI scans, demographic information, neuropsychological cognitive assessment tests results, and the details from APOe4 genotyping. All analyzed data were collected from 785 participants in the ADNI database. Imaging data were preprocessed in two ways providing MRI and JD sets used later for the classification. The first way includes an image normalization by bias field correction. The second approach involves an image brain segmentation procedure using the BET tool of the FSL software package (Jenkinson et al., 2012). The data classification was performed using deep neural network architecture. The leading architecture includes the convolutional sub-network inspired by efficient well-known AlexNet (Krizhevsky et al., 2012) and Xception (Chollet, 2017) CNNs. The main idea was to share the layers with the extracted features between two tasks, MCI vs AD and AD vs HC, and find the underlying factors for the conversion of MCI to AD. The features of two separated fully-connected layers were processed by applying a sigmoid activation function and a binary cross-entropy loss function to the output. The output assigned results in the range between 0 to 1. All results that are closed to 1 correspond to the progressive MCI or detect AD. The classification task started from two lines of two convolutional blocks processing images from MRI and JD sets that were concatenated in the next level and sent to a three-layer sub-network. Additional convolutional blocks continue to work with collected features creating the first processing line. The second processing line was established by adding MRI-set features obtained from the first two convolutional blocks. Two more convolutional blocks were concatenated again and sent by the fully-connected layer for prediction. Additional experiments concatenated features collected from the imaging data and those obtained from the clinical data and cognitive tests. It gave the possibility to compare the performance of CNN using different feature combinations. The best results were obtained by grouping MRI with clinical features. The evaluation metrics demonstrated 86% accuracy, 87.5% sensitivity, 84% specificity, and 0.925 AUC.

The study (Chitradevi and Prabha, 2020) evaluated several types of image segmentation methods and their impact on the diagnosis of Alzheimer's Disease. The segmentation was focused on the brain parts affected by neuronal loss due to AD development. GM, WM, Corpus Callosum (CC), and Hippocampus (HC) were separated from the surrounding areas using various segmentation optimization algorithms, including Genetic Algorithm (GA) (Coley, 1999), Particle Swarm Optimization algorithm (PSO) (Sengupta et al., 2019), Grey Wolf Optimization (GWO) (Mirjalili et al., 2014), and Cuckoo Search (CS) (Joshi et al., 2017). For the study, 200 brain images of patients with Alzheimer's Disease and cognitively healthy subjects were obtained from Chettinad Health City, Chennai. The segmentation quality was validated using ground truth images by computing Feature Similarity Index Metrics (FSIM) and Structure Similarity Index Metrics (SSIM). The image analysis and classification were performed using the AlexNet convolutional neural network. The best results were obtained with images processed with the GWO optimization algorithm, cable to segment brain sub-region with a validation accuracy of 98%. The data collected from the hippocampus provided accuracy of 95%, sensitivity of 95% and specificity of 94% and demonstrated the highest reliability among other segmented sub-regions investigated in the study.

A combined strategy of Random Forest feature selection with deep neural network classification is proposed for the diagnosis of early onset of dementia (Amoroso et al., 2018). T1-weighted MRIs collected from 240 subjects of the ADNI database were divided into four classes AD, MCI, cMCI (c states for "converted" to AD) and healthy controls. All imaging categories were balanced and analyzed using FreeSurfer v.5.3. Additional non-imaging features such as age, gender, diagnosis, baseline score from MMSE were added for each participant. FreeSurfer analysis provided supplementary information about cortical and subcortical volumes, cortical thickness, surface area and curvature, the volume of hippocampal subfields and several structural measures. The overall number of representative features was 431. The number of features was reduced to 242 after removing features with zero variances and interdependent features with a correlation coefficient over 0.9. The feature selection process was done using the Random Forest classifier. One hundred five-fold cross-validation rounds of the classifier helped choose the twenty most essential features. Among them were hippocampal volume, entorhinal cortex thickness, cerebrospinal fluid volume, lateral ventricle volume and baseline MMSE score. The final classification task was done using a feedforward DNN. The optimal neural network configuration of 11 layers and 2056 input units was found after several cross-validation tests. The prediction of the four final classes was performed using the Softmax function. The additional classification, based on the hippocampal volume, was performed using the fuzzy logic approach when each subject has a membership score in each set. The classification score was obtained with the RF classifier, and the diagnosis was made using the Bayesian approach. The performance of the fuzzy logic model was evaluated after 100 rounds of 10-fold cross-validation. The testing set included 160 real and 340 simulated samples. The DNN classifier demonstrated the highest recall for all classes (87.5% for AD, 52.5 % for HC and 27.5% for MCI), except cMCI, compared to the fuzzy model. The fuzzy model had a recall of 57.5% for the diagnosis of cMCI. At the same time, the DNN was more precise concerning MCI and cMCI, unlike the fuzzy model that showed higher precision for AD and HC.

Deep learning-based prediction of AD using fluorine-18 (18F) fluorodeoxyglucose (FDG) PET of the brain was investigated (Ding et al., 2019). For model training, they used 2109 images of 1002 individuals from the ADNI database and 40 images of 40 patients for testing the deep learning model. Preprocessed images were sent to the ImageNet Inception V3 CNN model (Szegedy et al., 2016). The model was fine-tuned by adding the drop-out layer with the rate of 0.6 before the fully-connected layer of the architecture. 90 % of images were used for the training and 10% for testing. The model was also tested with 40 images from an independent data set.

Alzheimer’s Disease was predicted with a specificity of 82% and sensitivity of 100% on average of 75.8 months before the final diagnosis. Comparative analysis of diagnostic performance was done for the DL model and human radiology reader. DL outperformed the reader in terms of sensitivity (100% vs 57% for the AD diagnosis and 43% vs 14% for the MCI) and terms of precision (18% vs 11% for the diagnosis of MCI). However, in terms of specificity, the reader overcame the DL model by 9% in the prediction of AD and by 18% in the detection of MCI. Regarding the precision, the reader also demonstrated the advantage in the diagnostic accuracy by 3%. Evaluating all diagnostical differences and computing the statistical significance of the results, the researcher concluded that the deep learning model generally performs better than human experts in the diagnosis of AD and prediction of early onset of dementia.

Lastly, Oh et al. (2019) proposed a volumetric convolutional neural network model to diagnose neurodegenerative diseases. The study used six hundred ninety-four structural MRIs from the ADNI database. Before the classification, the T1 MRI scans were normalized for bias-field homogeneities. No segmentation was applied to the imaging data. After that, the researchers implemented a convolutional autoencoder (CAE) with an inception module from GoogLeNet InceptionV2 for unsupervised learning, dimensionality reduction and differentiation between AD and NC. However, the autoencoder (AE) is unable to analyze the spatial structure of the imaging data. Spatial feature relationships were seen as an essential condition for the reliable explanation of classification decisions. A gradient backpropagation-based visualization technique was applied to the model to learn the model and understand the positioning of the important biomarkers. Supervised transfer learning architecture was proposed for the binary classification of pMCI (p states for "progressive") and sMCI (s states for "stable"). For an approximation of the CNN prediction, a gradient-based visualization method was applied to the group of biomarkers specified the progressive MCI and AD. During the study, the temporal and parietal lobes of the brain were identified as the main regions significantly affecting the classification outcome. The proposed approach demonstrated accuracies of 86.60% for the detection of AD and 73.95% for the prediction of pMCI.

A summary of the characteristics of the proposed above research methods and their classification findings are provided in Table 2.

Table 2: State-of-the-art methods of the diagnosis of Mild Cognitive Impairment and Alzheimer's Disease

Authors	Methods	Results
Lama et al. (2017)	PCA Features + Regularized Extreme Learning Machine (unsupervised classification learning algorithm based on single hidden-layer feedforward neural networks) of MRI (Classification of AD, MCI, NC).	Accuracy is 80.32% (for binary classification), 76.61% (for multiclass.)
Zhou et al. (2018)	Feature selection using C4.5 classifier + Transfer Learning Method (includes Transfer AdaBoost algorithm) of MRI (Classification of AD, MCI, NC.)	Accuracy is 85.4% (improved with optimised feature selection).
Beheshti et al. (2017)	Feature-ranking + genetic algorithm + SVM classifier of MRI (Classification of AD, MCI).	Accuracy is 93.01% (for stable MCI), 75% (for progressive MCI); without feature selection is 78.94%, with feature selection is 94.73%.
Moradi et al. (2015)	Logic regression + MRI biomarker (based on low-density separation) + SVM + neuropsychological test results + random forest classifier of MRI (Classification of AD, MCI, NC).	MRI + cognitive test improves the accuracy by 5.5% (from 76.5% to 82%).
Glozman and Le (2014)	Feature ranking of the white matter (WM) + SVM (with Linear and RBF Kernels) and Logic Regression of DTI (Classification of AD vs NC).	The average accuracy is 92%.

Grassi et al. (2019)	Ensemble algorithm using sociodemographic information, clinical characteristics, neuropsychological measures; supervised ML. (Conversion from MCI to AD).	AUROC (area under a receiver operating characteristic curve) is 0.88, and sensitivity is 77.7%, specificity is 79.9%. The range of AUROC for proposed models is between 0.83 and 0.90.
Basaia et al. (2019)	CNNs. Did not use feature engineering. (Classification of AD, stable MCI and converted MCI)	The accuracy of AD vs CN is 98%, stable MCI vs converted MCI is 75%.
Stamate et al. (2020)	Deep Learning models: two Multi-Layer Perceptron (MLP1 and MLP2) models and a Convolutional Bidirectional Long Short-Term Memory (ConvBLSTM) model. The features were collected from clinical and genetic data, MRI data, PET data, and additional biospecimen. (Classification of Dem, MCI, NC).	The best models (MLP1 and MLP2) show an accuracy of 0.86 for Dem, MCI and NC classes.
Zhu et al. (2017)	Selection of structural and functional features with a regularisation algorithm according to the similarity of the features. A Joint Regression and Classification (JRC) algorithm was applied to imaging (MRI, PET) and non-imaging (cognitive tests) data. (Classification of AD, MCI, NC)	The highest accuracy for the classification of AD vs NC is 95.7%, MCI vs NC is 79.9%.
Lu et al. (2018)	FDG-PET image analysed with multiscale deep neural network (MDNN). Each DNN was pretrained with a stacked-autoencoder (SAE); pretrained networks were fine-tuned as Multilayer Perceptron (MLP). (Classification of AD, stable MCI, progressive MCI and normal controls)	The accuracy of stable MCI vs progressive MCI is 82.51%
Yagis et al. (2020)	Applied 3D VGG neural network model; MRI data is used (Classification of AD, NC)	Accuracy in the diagnosis of AD vs NC is 73.4%± 0.04 for the ADNI dataset and 69.9%± 0.06 for the OASIS dataset
Qiu et al. (2018)	Combination of deep learning models: two multilayer perceptron (MLP) models + fusion of CNN VGG-11 networks; Investigated MRI + cognitive test scores (Classification of MCI, NC)	The accuracy of the final multimodal fusion model with MRI and cognitive score features is 90.9%

Iizuka et al. (2019)	CNN; SPECT images (Classification AD, dementia with Lewy bodies (DLB) and normal)	The obtained accuracy is 93.1%, 89.3% and 92.4% between DLB vs NL, DLB vs AD and AD vs NL classes, respectively
Yang et al. (2019)	Three groups of biomarkers are analysed: mental task performance, digital (Stat) and imaging biomarkers (fNIRS) with the linear discriminant analysis (LDA), statistical analysis and CNN (Classification of MCI, normal)	The highest accuracy is 90.62% with imaging biomarkers. LDA and statistical analyses show an accuracy of 76.67% and 60%, respectively
Martinez-Murcia et al. (2019)	DL based on convolutional autoencoders with Z-manifold block; Used imaging (MRI) and non-imaging (age, ApoE4, TAU, cognitive test results) data; Regression analysis investigated the correlation between imaging and other clinical data (Classification of AD, normal)	The accuracy of AD vs normal is 84%.
Duc et al. (2020)	Applied 3D CNN VGG for images, linear least square regression (LLSR) tree regression, bagging-based ensemble regression and support vector regression for MMSE; used resting-state functional MRI + MMSE measures (Classification of AD, normal)	The mean accuracy is 85.27% for AD vs normal
Li et al. (2019)	A longitudinal study of clinical and imaging changes for the development of MCI; Used MRI data from ADNI and AIBL databases for the classification with CNN; LASSO regularized Cox regression model predicted the progression of the disease (Classification of AD, MCI, normal)	Accuracy for classes AD vs NC is 90% for the ADNI and 093% for the AIBL. C-index of the progression of MCI to AD is 0.762 (ADNI) and 0.781 (AIBL)
Spasov et al. (2019)	Progression of MCI explored with AlexNet and Xception CNNs; concatenated features from different data types; used imaging (MRI), non-imaging (demographic, cognitive tests, and APOe4) data; (Classification of AD, MCI, HC)	Best performance from grouping imaging and clinical data is: accuracy 86%, sensitivity 87.5%, specificity 84%, AUC 0.925
Chitradevi and Prabha (2020)	Investigated the impact of the various segmentation techniques on the classification performance using the AlexNet CNN; used MRI data (Classification of AD, normal)	The highest accuracy is 98% (whole brain). The data from the hippocampus: accuracy

		95%, sensitivity 95%, specificity 94%
Amoroso et al. (2018)	Used Fuzzy logic model (Random Forest (RF) + Bayesian) and feedforward DNN; investigated MRI data + age, gender, diagnosis, MMSE score (Classification of AD, MCI, converted MCI, HC)	The highest recall was received with DNN for AD (87.5%), for HC (52.5 %), for MCI (27.5%) and with a fuzzy model for cMCI (57.5%)
Ding et al. (2019)	Used Inception V3 CNN model with FDG-PET images (Classification of AD, normal)	AD prediction 75.8 months before the final diagnosis with a specificity of 82% and sensitivity of 100%
Oh et al. (2019)	Used volumetric CNN: convolutional autoencoder (CAE) with an inception module for unsupervised learning (from GoogLeNet inceptionV2), supervised transfer learning architecture for the classification of progressive MCI vs stable MCI; used MRI data (Classification of AD, sMCI, pMCI, normal)	Accuracies for the AD is 86.60% and for pMCI is 73.95%

2.3.1 Discussion

Various feature engineering approaches and classification techniques have been applied to diagnose degenerative brain disorders. The structure of some of these methods is briefly described and summarised in this section. Most of these methods include feature detection, extraction and selection stages. Some of them have a high level of complexity and are time-consuming.

The methodologies described in the literature differ in image dimensionalities, data preprocessing stages, feature engineering techniques and classification approaches. Most methods rely on the 2D imaging data obtained from MRI slices, but some experiments process the 3D image sequences for classification tasks. The data preprocessing stages include mostly image normalization. In some approaches, image filtering and resizing can be applied. The feature engineering varies in length and complexities. It is essential to mention that the method's performance depends not only on the feature engineering technique itself but also on brain areas where features are collected. Atrophic regions of the cortex and subcortical zones are considered standard places for feature collection. Also, volumetric measures of the grey and white matter areas might assign waits to the extracted features. The following paragraphs will give more details about feature extraction and selection techniques. The diagnostic or classification approaches in the literature are presented mainly by the traditional machine learning algorithms, ensembled groups or deep learning networks. Their peculiarity is highlighted in the incoming text.

The feature collection process is one of the essential steps in computer vision technologies. In the study about the prediction of conversion of MCI to Alzheimer's Disease (Beheshti et al. 2017), the features were collected from atrophic grey matter areas of the brain by applying a 3D mask. Then, features passed two levels of the selection process, including t-test score and genetic algorithm. The data was classifier using linear SVM. According to the article, the accuracy score increased by around 17% due to the feature selection process. However, the complexity of the method makes it time-consuming. In the paper (Moradi et al., 2015), the features were collected from two image classes, NC and AD, using the regularized logistic regression algorithm. A high complexity semi-supervised low-density separation (LDS) method was proposed for feature selection. The classification was performed with labelled data of two classes, AD and NC. Unlabelled data of the third class of MCI was added in the next stage, followed by the LDS classification process. The output of the LDS classifier was combined with additional measurements. The final result was obtained using a random forest classifier. In this way, the proposed classification approach integrates many stages, including feature integration mechanism and semi-supervised and supervised learning, making the algorithm cost consuming. The paper described an attempt to align the structural and functional features according to their similarity (Zhu et al., 2017). The feature selection process was done with the same, as in the previous study, regularisation algorithms., The alignment technique was performed using two types of scans, MRI and PET. The extracted features were combined with clinical score measures and selected using a regression model. Classification of dementia was performed using an SVM classifier. Combined regression and classification models improved the diagnosis by 4-5% but demonstrated similar to other papers' prediction scores (see Table 2).

Some research techniques prefer to aggregate together with imaging features non-imaging biomarkers for diagnosing neurodegenerative diseases. For example, Moradi et al. (2015) combined imaging data with age and cognitive rating scores. Stamate et al. (2020) used demographics, genetic, cognitive, CSF biomarkers, and MRI and PET measurements. Zhu et al. (2017) combined MRI and PET data with cognitive test scores.

To enhance the diagnostic prediction of the MCI, Qiu et al. (2018) added the Logical Memory test score and the Mini-Mental State Examination records to the MRI data.

Diagnostic techniques also have multiple varieties. Among classical machine learning algorithms, the SVM is found as one of the most popular classifiers with a high level of precision. Beheshti et al. (2017) performed a classification task using linear SVM. Zhu et al. (2017) applied a joint the SVM and SVR model when the images were classified using the SVM, and the features received from two cognitive tests were trained and classified using the SVR. However, deep neural networks had more preferences over the other machine learning techniques in the last decades. They can deal with imaging and non-imaging information. In addition, they can perform classification tasks on labelled and unlabelled data. Most deep learning networks do not require or have less complex image preprocessing and feature engineering stages. The ability to process and analyse tens of thousands of patients' records increases their popularity in medical diagnosis, including the detection of dementia.

One of the neural network models that can process a large number of images with high precision is a Convolutional Neural Network. For example, Basaia et al. (2019) applied this model for MRI data distinguishing between stable and progressive MCI. Iizuka et al. (2019) proposed CNN for the detection of AD using SPECT images. Yang et al. (2019) applied a convolutional network using mental task performance measurements and digital and imaging biomarkers to predict AD. The convolutional autoencoders were proposed by Martinez-Murcia et al. (2019) and Baldi (2012). This type of CNN can minimize the reconstruction error between input and output images. This type of network can process volumetric images and requires smaller memory than dense-connected ones, but the prediction AD is not high enough. It gives 84% accuracy.

A wide range of other classification models includes, for example, a Multilayer Perceptron with transfer learning architecture (Vincent et al., 2010), a Convolutional Bidirectional Long Short-Term Memory model (Stamate et al., 2020), a Regularized Extreme Learning Machine model in combination with multiple kernel SVM and import vector machine (Moradi et al. 2015). Feature selection techniques can also be combined with a deep neural network (Amoroso et al., 2018).

Pretrained and more complex deep learning models were proposed in recent studies for the detection of brain pathology. For example, Yagis et al. (2020) examined a 3D VGG model in AD diagnosis and received an accuracy of 74%. The fusion of deep learning models that combines two Multilayer Perceptron models and the CNN VGG-11 architecture (Qiu et al., 2018) performed the binary classification task with the prediction of MCI of around 90%. Modifier and optimized 3D CNN VGG network demonstrated 85% accuracy in Duc et al. (2020) paper. Another deep learning configuration composed of the Xception CNNs and convolutional AlexNet-like sub-networks investigated the progressive MCI and demonstrated a performance between 84 and 88% (Spasov et al., 2019). Another transfer learning model is discussed in the paper by Oh et al. (2019). It aggregates a convolutional autoencoder with an inception module from GoogLeNet InceptionV2. However, the autoencoder was unable to analyze the spatial structure of the imaging data, which is vital for understanding the classification decision. Additional visualization techniques had to be applied to detect the positioning of the essential biomarkers. The temporal and parietal lobes of the brain were identified as the main regions affected by the degenerative process. With the proposed methodology, AD was detected with an accuracy of 86.60% and the progressive MCI with an accuracy of 73.95%. Ding et al. (2019) investigated the classification performance ImageNet Inception V3 CNN model for diagnosing Alzheimer's Disease using PET scans. The archived diagnostic performance of this method demonstrated a specificity of 82% and a sensitivity of 100%.

The combined approach of image segmentation methods and CNNs demonstrated the increase in the prediction and diagnosis of brain pathologies. The study (Chitradevi and Prabha, 2020) compared several image segmentation methods and their classification performance in diagnosing Alzheimer's Disease. Images segmented brain areas such as grey matter, white matter, Corpus Callosum, and Hippocampus were tested using the AlexNet convolutional neural network. Accuracy above 95% was obtained for some sub-regions of the brain.

3 Summary and contribution of the chapter

This chapter presented an overview of anatomical brain asymmetries related to the lateralization of brain functionality. The chapter highlighted factors leading to the differences in brain structure that affect the performance of cortical and subcortical brain areas. The main factors influencing morphological changes in the grey and white brain matter belong to the group of neurodegenerative disorders. The most common among those is Alzheimer’s Disease. The prevalence of AD is between 60 to 80 % of all dementia cases. Mild cognitive impairment becomes a prodrome of Alzheimer's disease on 80% of occasions.

The literature review explained the changes in brain asymmetries that point to the development of the degenerative process in the brain. Modern imaging computerized technologies can detect the early transformation of structural and functional symmetry. The research methods in this chapter focused on investigating functional changes in hemispheric asymmetry in the process of the development of MCI and AD. MRIs of a healthy person and persons with MCI and AD visualize the local differences in the structural asymmetry between some brain areas in these three groups of individuals. Several studies confirm the gradual decrease of the left-sided brain asymmetry in cognitively healthy persons in the early stages of mild cognitive decline. These changes are developed due to atrophic processes in the brain areas responsible for memory and language processing. If the degenerative processes progress further, the asymmetry in the cortical and subcortical brain areas becomes noticeably right-shifted.

Based on magnetic resonance imaging, a computer-aided diagnosis (Yanase and Triantaphyllou 2019, Lazli et al. 2020) of anatomical changes in the brain gives an accurate result for the early prediction of brain disorders. The research methods analysed in the literature review apply multiple feature engineering and classification algorithms for the diagnosis of dementia. The performance of the research techniques described above demonstrates quite good results. However, a diagnosis of the early stages of cognitive decline still requires new, more precise methods.

Overall, the literature review helped to identify some gaps in this research area that are described below.

Researchers who discovered brain asymmetries were mostly focused on the brain functionalities or explored structural differences between subcortical brain areas for specific anatomical structures. These investigators did not apply segmentation algorithm to extract the asymmetry from entire brain and also, they did not use machine learning techniques for the diagnostic purpose. *In this way, the development of imaging biomarker that outline the asymmetries between right and the left hemisphere of the brain is the main contribution of the thesis to this scientific area. The proposed method allows the detection of early brain changes during routine MRI scan.*

Classification of brain pathology, which is represented in the second part of literature review, was done using multiple methodological approaches that integrated feature engineering and ML methods. The researchers test their models using original MRIs, but not asymmetries. They did not run and describe experiments when pretrained network incorporate different from Softmax classification modules. *Thus, the additional contribution of the thesis focuses on the classification of dementia using asymmetry biomarker and transfer learning approaches which incorporate modification of classification module. Created machine learning models help identify affected parts of the brain with high precision and distinguish early and more severe stages of dementia.*

The next chapter gives detailed representation of computer vision technologies that contribute to the processing and analysis of medical imaging data. The reader has to keep in mind the stages of preprocessing of MRIs and brain segmentation approaches that help to understand the reason behind the choice of the thesis's computational framework. At the same time, the thesis research proposes and describes an additional brain segmentation technique that detects asymmetries in brain hemispheres. This new technique is yet to be described in the state-of-the-art literature. As an additional imaging biomarker, the imaging dataset of brain asymmetries will be tested later in Chapter 5 for its robustness in diagnosing of Early Mild Cognitive Impairment and Alzheimer's Disease, whose data is compared with cognitively healthy subjects. The tests will be done with classification algorithms, including different kinds of transfer learning approaches. For this reason, well-known architectures of pretrained CNNs will be introduced to the reader.

Chapter 3: Computational Framework for Brain Asymmetry Digital Biomarkers Representation and Classification

3.1 Introduction

This chapter of the thesis introduces the proposed brain asymmetry computer vision methodology and answer the questions raised earlier in Chapter 2 about detecting changes in the brain asymmetries during the progression of the disease. The thesis investigates the structural property of the brain in the initial stage of cognitive decline and severe changes in Alzheimer's Disease. Common patterns of these changes are visualized after processing the MRIs with a number of segmentation algorithms. The information obtained from brain asymmetries is checked for robustness and classified with machine learning algorithms. The correlation between affected cortical parts is explored using MRI sets of subjects with normal cognitive function, mild cognitive impairment and Alzheimer's Disease using deep networks either via a transfer learning approach applied to pretrained convolutional neural networks or networks tailored to this task.

Computer vision methods lie at the base of the suggested computational framework. CV part starts from image retrieval, goes through multiple image transformation stages, and ends with the visualization of brain asymmetries. A detailed discussion of the proposed methodology is given in Section 3.8. Brain imaging data used in the project are obtained from MRI scans, normalised and processed with segmentation algorithms and sent to the ML modelling for the prediction of the disease.

Multiple machine learning approaches for diagnosing cognitive decline and dementia were described in the literature review chapter. The choice of the algorithm depends on the image type, size of the dataset, characteristics of the available hardware environment, and skills of the researcher or IT specialist.

The classification of medical pathologies using Machine Learning and Artificial Neural Networks (ANN) has attracted a lot of attention in recent years. In this context, the Convolutional Neural Network is one of the models successfully adapted to classify imaging data in numerous applications (Yamashita et al., 2018). Hybrid architectures based on CNN models were successfully implemented in various medical areas. For example, a combination of the Fast Region-Based Convolutional Neural Networks (R-CNN) and SVM demonstrated high performance for detecting and categorising brain tumors (Khairandish et al., 2020). Singh and Majumder (2020) applied a decision fusion approach using CNN for feature extraction and SVM for the final decision-making to diagnose electrocardiogram (ECG) abnormalities. Transfer learning methods adapted to two CNN models, the VGG16 and the InceptionV3 network, demonstrated performance above 90% for pneumonia classification (Yadav and Jadhav, 2019).

Compared to the classical machine learning algorithms, deep neural networks can provide an end-to-end solution, automating the image preprocessing and feature engineering stages by considering those as part of the training process and achieving a high prediction rate of brain pathology. Advances in deep learning together with the availability of hardware, such as GPUs and cloud computing infrastructures, have allowed the training of deep neural networks on large volumes of images.

Deep learning can be used as a single classifier or in ensemble architectures for the diagnosis of degenerative brain diseases (Nanni et al., 2020). Deep networks can handle 2D and 3D data in order to distinguish between healthy and dement subjects (Yagis et al., 2020). At the same time, the advantages of deep learning models can be used for limited datasets by applying a layer-wise transfer learning approach (Mehmood et al., 2021) and image augmentation techniques (Mikołajczyk and Grochowski, 2018). Deep transfer learning models propose an effective way of image segmentation and can automatically classify brain scans focusing only on small brain regions (Aderghal et al., 2020).

This chapter introduces the classification methods applied to the imaging sets of brain asymmetries using different types of machine learning algorithms and transfer learning models of deep neural networks.

This chapter is organized as follows.

Section 3.2 presents an overview of the proposed computational framework for processing and classification of brain asymmetry digital biomarkers. The computational framework consists of two stages. The first stage involves computer vision algorithms that focus on processing and analysis of imaging data based on segmented brain asymmetries. The second stage exploits machine learning algorithms in the diagnosis of early cognitive decline using the images of brain asymmetries. It investigates machine learning algorithms, transfer learning methods of Deep Neural Networks and proposes new deep network architectures for classification of brain asymmetry biomarkers.

Section 3.3 describes the MRI data repositories used in thesis. This section describes two imaging sources. It provides qualitative and quantitative characteristics of the imaging data supported by demographic information.

Section 3.4 explains the physical principals of magnetic resonance imaging techniques. It discusses medical image formats that were obtained from the databases and converted later to .PNG for data modelling.

Section 3.5 describes the image preprocessing stages of the computational framework. Two levels of data preprocessing, including image resizing and normalization, are given in this part. Image resizing was applied for the image datasets to reduce the computational power.

Section 3.6 overviews the image segmentation approaches used in neuroscience including methods applied to the MRI set for the detection of brain asymmetries.

Section 3.7 represents the mathematical operations discussed in the literature and used for symmetry detection.

Section 3.8 introduces brain segmentation methods and explains the segmentation of hemispheric asymmetries. The brain segmentation approach based on thresholding and a region-growing algorithm is included in the first part of the section. The second part proposes a new approach for the detection and segmentation of hemispheric asymmetries that become the main data source for the analysis of brain changes due to the development of dementia.

The chapter ends with a summary of the work presented and contribution to the research area.

3.2 Overview of the Computational Framework for Processing and Classification of Brain Asymmetries

The machine learning workflow for early diagnosis of dementia (see Figure 5) includes image preprocessing, segmentation of image asymmetries, extraction of statistical and non-statistical features and image analysis (see Chapter 4 for relevant review). Machine learning algorithms (see Chapter 5 for relevant review) are implemented in the classification stage of the processed images. Images illustrating differences between the right and the left hemispheres of the MRI slices (2D images) of the brain are used for features extraction. This simplifies the feature engineering stage because the collected features are selected from brain regions affected by degenerative processes. The images of segmented asymmetries require less storage than original MRIs. This speed up the classification processing of large datasets using images as input.

In the last stage of the workflow, different kinds of machine learning algorithms can be applied (Herzog and Magoulas, 2021). This can include two potential pathways: one that exploits image asymmetry features and another one that uses images of segmented asymmetry. Machine learning classifiers, such as Naïve Bayes (NB), Linear Discriminant (LD), Support Vector Machines (SVMs) and K-Nearest Neighbour (KNN), typically operate on the basis of feature vectors, such as image asymmetry features that can be used for training and testing. In contrast, a Deep Network (DN) classifier receives images of segmented asymmetry and generates its own features through training. At this stage, various Convolutional Neural Network (CNN) models are investigated by applying transfer learning to stablished pretrained architectures and distinctive architectural configurations of CNNs are proposed (Herzog and Magoulas, 2022).

The data processing pipeline, including image processing and machine learning classification, has been implemented in Matlab using affordable and easy to obtain commodity hardware: Windows 10 Enterprise, processor – Intel (R) Core (TM), i7-7700 CPU@ 3.60GHz, 16 GB RAM.

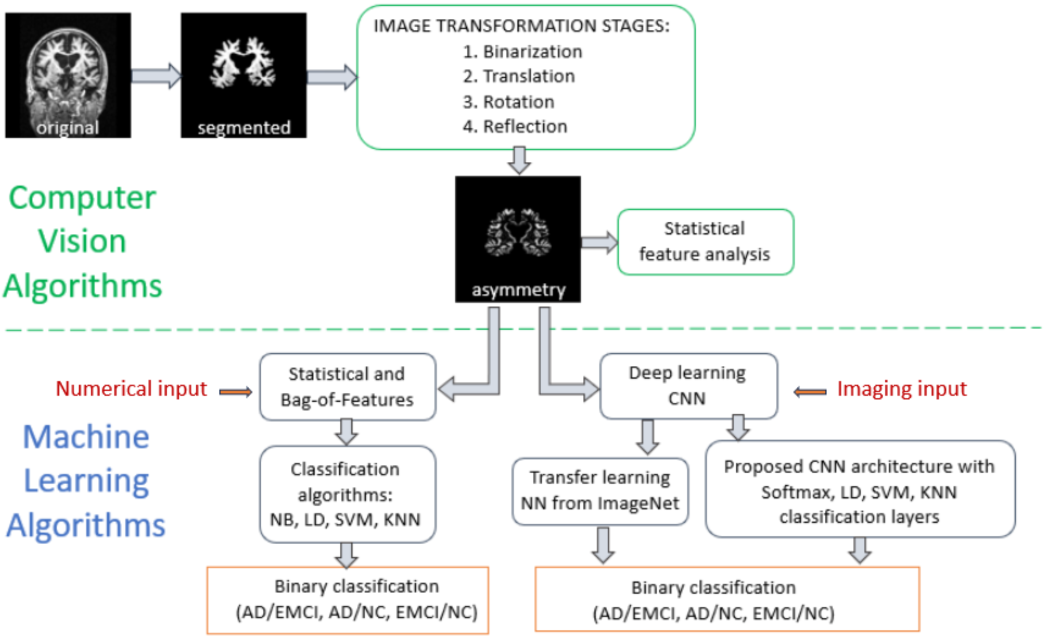


Figure 5: Computational framework including image transformation stages, asymmetry features generation and machine learning classification algorithms

3.3 MRI data repositories

3.3.1 ADNI

MRI data used in the thesis were obtained from the Alzheimer’s Disease Neuroimaging Initiative database (adni.loni.usc.edu), which was launched in 2003 as a public-private partnership led by Michael W. Weiner, MD. The primary goal of ADNI has been to assess whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of Mild Cognitive Impairment and early Alzheimer’s disease. More details about the project and up-to-date information are available at www.adni-info.org. Since 2004 the ADNI has launched three different databases, namely ADNI 1, ADNI GO/2, and now ADNI 3.

The created datasets include T1-weighted images of structural MRI data of 150 subjects (100 male, 50 female) from the ADNI-3 database. Among them are 50 patients with Mild Cognitive Impairment at the age between 55 and 65 years old, 50 patients with Normal Cognition at the age between 55 and 65 years old, and 50 patients with Alzheimer’s Disease at the age between 65 and 90 years old. The age range between 55 and 65 years for the EMCI and NC groups was chosen to eliminate the ageing effect on the MRI data. Magnetization Prepared Rapid Gradient-Echo (MPRAGE) T1-weighted MRI images were taken from the same type of 3T scanners (sequence at 1.2 Tesla), Siemens Medical Solutions (http://adni.loni.usc.edu/methods/mri-tool/mri-acquisition/).

3.3.2 OASIS

The second MRI data repository is the Open Access Series of Imaging Studies (see www.oasis-brain.org). The OASIS brain project was created by Washington University in 2007. The OASIS is an open-access database of neuroimaging and processed medical imaging data supported by non-imaging information, including demographic, cognitive, and genetic. The platform is focused on providing information about normal ageing and cognitive decline for neuroimaging, cognitive and clinical research. The project was released in three phases. OASIS-1 includes cross-sectional data (Marcus et al., 2007); OASIS-2 is a longitudinal study (Marcus et al., 2010) that was created for hypothesis-driven data analyses, development of segmentation algorithms and design of neuroanatomical atlases. OASIS-3 is composed of longitudinal neuroimaging, cognitive, clinical, and biomarker datasets for normal ageing and Alzheimer's Disease.

The OASIS-2 longitudinal collection of T1-waited MRIs of 100 (50 male, 50 female) very mildly demented (VM-D) and non-demented (non-D) right-handed subjects aged between 60 and 80 was used for the experiments.

3.4 Types of magnetic resonance images

Magnetic resonance brain images are the primary source of data used in the thesis. Initially, they were represented in DICOM and NIFTI formats. This section added to the thesis to explain how the original brain images are represented in the ADNI and OASIS data directories, and what type of conversion was applied to them to create readable for machine learning algorithms datasets for use in the thesis. This section also provides a background of magnetic resonance technology and explains the structure of image matrices used by computer vision algorithms for image processing.

Medical images are represented in standardized file formats (Larobina and Murino, 2014). The file format represents how the imaging data are organized in terms of pixels or voxels inside the image. In most cases, the file formats need to be adapted to the system or software requirements for data loading, interpretation, and visualization.

The minimal structuring element representing image intensity values is a pixel. Each pixel has a depth, a number of bits used to encode the pixel information. Photometric interpretation of the image specifies the image colour. A monochrome image has only one sample of pixels and does not contain information about the image colour. A binary black-and-white image has pixel information stored in bits equal to zero or one. Radiological images, like X-ray, MRI and CT, have a greyscale photometric interpretation matching 8 or 16 bits. Nuclear images, like PET and SPECT, are represented with a colour map with 24 bits per pixel. Images employing the RGB colour palette are called “true colour” images.

Any file format is associated with information beyond the pixel data. This information is called metadata. It includes the matrix dimensions, pixel depth, spatial resolution and image photometric interpretation. The most popular formats that are present in the publicly available databases are Digital Imaging and Communications in Medicine (Dicom), Analyze, Neuroimaging Informatics Technology Initiative (Nifti), and Medical Imaging NetCDF (Minc). The metadata can be stored together with other medical imaging data at the beginning of a single file. This file format is used by DICOM, NIFTI and MINC formats. Another configuration stores the metadata and image data in separate files. The Analyze format is an example of it. This file format is preferable for complex imaging structures when modifying information is required. Table 3 represents the standard medical image formats and their main characteristics. More information on the formats can be found in Appendix D.

Table 3 represents the standard medical image formats and their main characteristics.

Table 3: File formats characteristics

Format	Extension	Header	Data type		
			Integer	Float	Complex
Analyze	.hdr/.img	Fixed length, 348 bites, binary format	Unsigned (8 bit), signed (16, 32 bit)	32, 64 bit	64 bit
NIFTI	.nia, .nii, .hdr/.img	Fixed length: 352 bytes binary format (348 bytes if data stored as .img/.hdr)	Signed and unsigned (from 8 to 64 bit)	From 32 to 128 bit	From 64 to 256 bit

MINC	.mnc	Extensible binary format	Signed and unsigned from 8 to 32 bit	32, 64 bit	32, 64 bit
DICOM	.dcm, .dicom	Variable length binary format	Signed and unsigned integer, (8, 16 bit; The 32-bit dose is allowed for radiotherapy only)	Not supported	Not supported

Magnetic resonance images can be defined as a function $I(i, j, k)$ in 3D space where $i = 0,\dots, M-1, j = 0,\dots, N-1, k = 0,\dots, D-1$ denote coordinates. The values of $I(i, j, k)$ are voxel (pixel in 2D space) values represented by grey intensities $\{0,\dots, 255\}$. Each image element is uniquely described by its coordinates (i, j, k) and its intensity value for voxels ((i, j) for pixels), where i is the row number, j is the column number, and k is the slice number of volumetric image sequences. Voxel (pixel) values depend on magnetic resonance characteristics of the corresponding element in the biological tissue, the strength of a magnet, time of acquisition and some other factors.

In the thesis, T1-weighted MRI brain sequences have been chosen as a source of medical imaging data.

MRI brain scans are used more frequently than other scanning technologies because of the radiation safety and high detection level of soft tissue structure (Sprawls, 2000). Two types of weighted, T1 and T2, images depend on the relaxation time for protons and their density in the tissue. T1- and T2-weighted images are differentiated by the brightness of cerebrospinal fluid. CSF looks dark on T1-weighted imaging and bright on T2-weighted imaging. The white colour component of T1-weighted scans is determined by the presence of fatty substances in the tissue, such as white matter. The dark colour demonstrates the tissues filled with water, such as CSF. T2-weighted images have the opposite colouring pattern. In the case of dementia, the brain's anatomical structure is subject to change. So, both T1- and T2-weighted image sequences can demonstrate these transformations.

Diffusion tensor imaging was proposed for use in 1994 by Peter Basser (Alexander et al., 2007). It is a modification of magnetic resonance imaging technology developed from diffusion MRI and based on the nuclear magnetic resonance field. The diffusion tensor model is characterised by a rotational invariance of the shape of water diffusion. It makes the DTI method preferable in investigating the complex anatomy of the white matter tracts of the human brain. This technology is sensitive to the cellular structure of the tissue and works by measuring the diffusion coefficient of water molecules. The diffusion measured in tissue is anisotropic or varies with direction. DTI is taking part in the visualization of the white matter tracts called tractography. DTI offers some specific measures such as diffusion magnitude (e.g., mean diffusivity) and diffusion anisotropy (e.g., fractional anisotropy) (O'Donnell and Westin, 2011).

The quality and size of image sequences of observed patients might differ from scan to scan. It depends on the types and makes of image scanners. Processed, analysed, and compared images must have the same type. In other words, they have to be preprocessed.

3.5 Image preprocessing

3.5.1 Image normalization

Image preprocessing starts from image normalization. The normalization process brings the pixel intensity values of all imaging data to one standard. Often, this process leads to general image enhancement (Image enhancement techniques are described in Chapter 3.6) if normalization is based on contrast or histogram stretching as in the current research. Image normalization method can be divided ruffly into two groups: *histogram-based* methods and *statistic-based* methods. Histogram-based methods has been used in the research as image normalization techniques, so they are briefly described next.

Each given greyscale image can be represented as a histogram of its grey levels. Grey values of a poorly contrasted image are clustered mainly in the center of the histogram (see Figure 6). The histogram normalization function spreads the pixel values in a way that fills the entire available pixel's intensity range between 0 and 255.

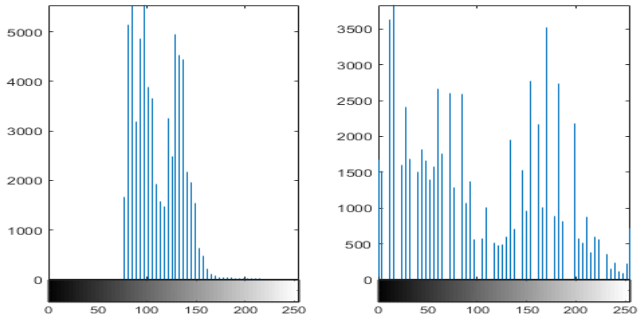


Figure 6: The histogram of the original image (left) and normalized image (right)

The histogram normalization function of the greyscale image between two points with coordinates (a, c) and (b, d), and initial (x) and final (y) pixel intensity values is performed according to the formula:

$$y = \left(\frac{x - a}{b - a}\right)^{\gamma} (d - c) + c \tag{1}$$

The gamma value describes the shape of the function and can perform linear and non-linear normalization. If gamma is equal to 1(in the default option), the mapping of the pixels will be linear. The gamma values which are less or greater than one lead to non-linear pixel distribution (McAndrew, 2004).

Examples of histogram normalization techniques used for image preprocessing are given in (Štruc and Pavešić, 2017).

3.5.2 Image resizing

Image resizing is a routine procedure in image processing. Usually, all images have to be resized to one standard fixed size, for example 256-by-256-by-3 pixels (the image size is used in the current research and explained later). The purpose of resizing is fitting images in a particular space. Very often, the size of big imaging data must be reduced to speed up a machine learning process and minimize the storage space on the device.

Resized imaging data can be stored in different file formats. The popular medical data have four major file formats: Analyze, Nifti, Minc, and Dicom (Larobina and Murino, 2014). These file formats have a specific structure such as pixel depth, photometric interpretation, metadata, and pixel data. Most contemporary computer vision techniques are adapted to work with such sort of data (Willemink et al., 2020), but machine learning methods can work with conventional (non-medical) file formats only.

3.5.3 Conversion imaging data

Conversion of imaging data to other file formats is a common approach (for example, Dicom to PNG). These formats might be lossy and lossless. Lossless data are preferable for efficient machine learning training. Popular lossless formats are TIFF, PNG, GIF (Graphics Interchange Format), PDF (Portable Document Format), EPS (Encapsulated PostScript), AI (Adobe Illustrator), RAW (contains the unprocessed data captured by a digital camera or scanner's sensor). EPS and AI are used in photoshop and graphic design.

JPEG/JPG format is lossless but widely used online because of the flexibility of compression.

3.5.4 Image preprocessing summary

For the research, all image sequences were downloaded in 3D NIFTI and DICOM formats from ADNI and OASIS databases. From each image sequence, 3-4 central slices were converted to 2D images in PNG format using Dicom Converter (https://dicomapps.com/dicom-converter/index.html). The created datasets were sent to the preprocessing stages that included image normalization and image resizing procedures.

The image normalization brings the pixel intensity values of all imaging data to one standard. Each given greyscale image can be represented as a histogram of its grey levels. Grey values of a poorly contrasted image are clustered mainly in the centre of the histogram. A histogram normalization method spreads the pixel values in a way that fills the entire available pixel's intensity range between 0 and 255. In the current study, MRI brain images are normalized using the histogram stretching technique.

Another routine procedure in image processing applied on MRI data is image resizing, where all images are resized to one standard fixed size. The purpose of resizing is to fit the images into a certain dimensionality space. Very often, the size of big imaging data has to be reduced to speed up a machine learning process and minimize local storage requirements. In the current research, the images are resized to 256×256 pixels with RGB colour channels for further use in the machine learning classification stage.

3.6 Image segmentation

Image segmentation methods are at the base of the thesis methodology. Object segmentation is the most challenging part of Computer Vision and constantly searches for new, more accurate algorithms. One main reason for this difficulty is that human biology, including the visual system, is very complicated, and there is no technology that can identically reproduce it. A human can recognize objects under all kinds of variations in scale, illumination, and positioning. There is no limit to objects that can be memorized by the human brain and recognized in the future. Thus, the list of computer vision challenges is prominent. The first main challenge is image retrieval. The image should be identified and processed under changeability in viewpoint and scale, under impermanence of illumination conditions, presence of background noises, image occlusion and image deformation. Variations in image parameters inside the class can also be an issue. The second major challenge is the storage and processing of imaging data which require a vast amount of space and computational time. Many solutions are already proposed to overcome these sorts of problems, but still, there is plenty of space for extended research and improvements.

Segmentation is the process of image partitioning into sets of pixels (segments) united under a particular set of rules. The main goal of the segmentation is to locate objects or image boundaries such as corners, curves, lines, etc. The process of estimating the difficulties of image segmentation before it starts can help with a choice of the segmentation algorithm to be applied for. Liu et al. proposed to use a new feature that combines information from the log histogram of the log gradient and the local binary pattern histogram to predict the success of the segmentation (Liu et al., 2011). Each tracking object needs to be separated from the background. This task is achieved by the ***segmentation*** algorithms, which have a central part in computer vision. Depending on the application or project requirements, the segmentation procedure can be applied to single image or image sequences. *Local segmentation* isolates specific regions in an image. In comparison, *global segmentation* extracts the whole image from the scene of multiple objects. Referring to the scene of segmentation, it can be semantic and instance. *Semantic segmentation* is based on understanding and labelling each pixel in the image (pixel-wise predictions) (Liu et al., 2020). The row of pixels with maximum similarity are classified as the same group of objects. For example, if there are four dogs in an image, semantic segmentation labels all dogs' pixels the same. The object will be coloured identically. The boundaries of each object are outlined. Beyond semantic segmentation, *instance segmentation* takes into account the different instances of classes (Watanabe and Wolf, 2019). Thus, each sample of the same class is coloured differently from the other. In tests with dogs, all of them will be assigned different colours and individual labels. Instance segmentation carries out the more complex task of identification of relations, boundaries, and differences of the objects. In some cases, the segmentation has complicated sights of multiple overlapping objects and backgrounds. Methods used in medical image segmentation are discussed in detail in Appendix E.

MRI segmentation can be performed on 2D images separated from an image sequence or on 3D series. If 2D images are segmented slice-by-slice, they can be connected to a 3D volume or a continuous surface. In brain MRI, there are three main tissue types: grey matter, white matter, and cerebrospinal fluid. But typically, the scanned image also incorporates the skull and a nonbrain area. Therefore, the standard approach of MRI processing involves the extraction of nonbrain tissues before the brain segmentation methods can be used. The applied segmentation technique is represented in more detail in Section 3.8.

3.7 Mathematical operations for symmetry detection

In computer vision, symmetry can be described as a balanced similarity between two parts of the object. The imaginary axis along the figure that separates it into equal parts that can be mirrored from one to another is called the axis of symmetry.

Humans in nature can easily detect the symmetries in surrounding objects. Artificial intelligence tries to reproduce the human perception of symmetries, but the technical ability of computer vision algorithms is still far from the simulation of sophisticated human perception. Image symmetry detection relies on the basic concepts of 2D Euclidian space (Liu et al., 2010). Euclidian space has four types of primitive symmetries: reflection, translation, rotation, and glide reflection. 3D Euclidian space has an additional two types of symmetry, they are roto-reflection and helical symmetry. The mathematical expressions for the first four types of symmetries are provided below.

Reflection (the axis of reflection remains invariant):

$$f(x, y) = f(-x, y) \tag{2}$$

Rotation (the centre point or axis of rotation remains invariant, and n is an integer demonstrating the number of equal parts in the rotation circle):

$$f(x, y) = f\left(rcos\left(\frac{2\pi}{n}\right), rsin\left(\frac{2\pi}{n}\right)\right), \text{ where } r = \sqrt{(x^2 + y^2)} \tag{3}$$

Translation (there are no invariant points for some $\Delta x, \Delta y \in R$):

$$f(x, y) = f(x + \Delta x, y + \Delta y) \tag{4}$$

Glide reflection (demonstrate the translation along the reflection axis; there are no invariants point for some $\Delta x \in R$):

$$f(x, y) = f(x + \Delta x, -y) \tag{5}$$

There are multiple symmetries and symmetries groups beyond those presented above, but they are not considered in the current research.

The mirror symmetry has a fixed structure. The fixed structure means that the axis of symmetry has a line of fixed points. Every pair of symmetric points is also connected by lines which are intersected with this axis. The intersection of all these lines helps to find the object's centre. The automatic detection of symmetry in digital images is a constant topic in CV. Despite the long computational effort, computer vision technology is still far from a robust "symmetry detector". The state-of-the-art detection methods can be described as feature-based, correlation-based, energy-based and graphic models. More advanced methods include statistical learning models such as spectral-based and MRF-based models.

Numerous approaches have been suggested to detect bidirectional or mirror symmetry in digital images. *The direct approach* is determined when the symmetry, such as rotation or reflection, is applied to the transformed image which is later compared to the original one. This approach assumes that the reflected object is either perfectly symmetric or not symmetric at all. However, this method is sensitive to noise. Another *voting approach* is based on the statement that the axis of symmetry is uniquely identified by the two corresponding points and similar to the centre of symmetry by the two corresponding pairs of points. The voting schema is applied to determine the best axis of symmetry with the highest voting score. This schema is less sensitive to noise than the direct approach but has a high level of complexity.

The symmetry is divided into *global* and *local*. The first type is more efficient but sensitive to noise. The determination of local asymmetry can be more accurate but takes a lot of time to be processed. The current research focuses on the detection of global symmetry, and the local symmetry techniques are not in the scope of the study. In addition to the direct and voting approaches, two additional approaches are established on basic function and moment-based methods. An example of basic function method is the Walsh function (Yodogawa, 1982), which evaluates the bidirectional symmetry. Walsh function assesses four subtypes of symmetry in the object, which are horizontal, vertical and doubly mirrors and rotational symmetry. Welsh coefficients are calculated from the vectors of values obtained from the corresponding symmetries. The symmetry evaluation is done according to the entropy of these four values. The moment-based method was developed to differentiate the mirrored images, which are rotationally symmetric. The specific and more complicated cases of mirror symmetries are skewed symmetries (Tuytelaars et al., 2003). These symmetries are found through the search of affine-invariant matching points. Hough transform is used to detect these points.

The human brain has, on average, bilateral symmetry. One of the methods used for evaluating this symmetry is described by Liu et al. (2001) and based on the finding of the midsagittal plane (MSP). The authors developed an algorithm which is able to identify a 3D MSP. The algorithm demonstrated robustness when it was applied to normal brains and brains with large areas of the lesion. The method's robustness was tested on MRI and CT images with significant gaps between slices. The purpose of MSP extraction was to find out where the bilateral symmetry was before the onset of the disease. The algorithm was applied to the sets of original, reflected and rotated images to determine the resulting 3D plane that stands at the minimal distance from all 2D middle lines. The method did not show any statistical difference in the determination of MSP between its results and the judgments of human experts. The MSP techniques were applied to the AD and Schizophrenia diagnoses (Liu et al., 2004; Liu et al., 2007). Similar quantified or statistical brain asymmetry methods described in (Teverovskiy et al., 2008) were successfully used for age estimation.

As mentioned at the beginning of this chapter, there are multiple tangible research directions related to image symmetries—for instance, the symmetry-based registration and symmetry-based recognition of the biomedical objects. There is a need for the development of additional strategies able to deal with real-world complexity, hidden patterns in image structures, as well as symmetrical and distorted objects under noisy conditions. In contrast to mathematical methods, the proposed approach uses a natural anatomical brain fissure as a symmetry line. The choice is based on previous finding indicating no differences between mathematical allocated MSP and anatomical fissures in patients without brain tumours.

3.8 Segmentation of brain’s hemispheric asymmetries

In this stage, normalized and resized images are processed using brain segmentation algorithms, which aim to localize an object of interest or the image boundaries. The process concerns the partitioning of the image into sets of pixels, or segments, united under a set of rules. In the case of MRI, segmentation can be performed on 2D images separated from an image sequence or on 3D series. If 2D images are segmented slice-by-slice, they can be used to construct a 3D model. Typically, a scanned image incorporates the skull and a nonbrain area as well as WM brain tissue, GM brain tissue, and CSF. Therefore, a common approach in MRI processing involves the extraction of nonbrain tissues before the brain segmentation methods can be used.

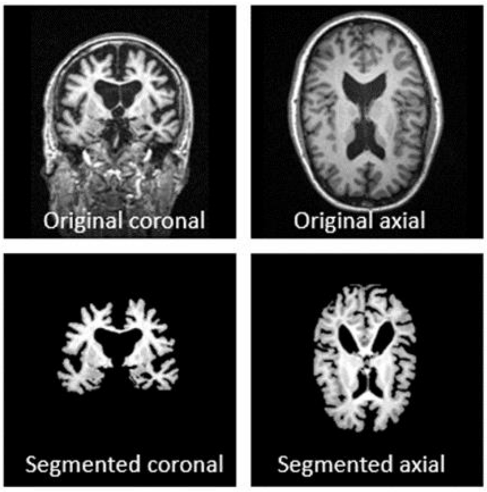


Figure 7: Segmented brain from MRIs in coronal and axial planes

Although there is no single method appropriate for all images because of image diversity, presence of noise and artefacts, a segmentation method that has been developed for one imaging context can be adapted to another class of images.

The segmentation method that is used in the research belongs to the group of intensity-based methods, which includes thresholding and region-growing or region-based types. The three main MRI brain tissue types - WM, GM and CSF - can be easily distinguished due to their differences in pixel intensity levels. However, the presence of noise, artefacts, overlapped objects, and inhomogeneity of the tissues is an objective factor that can require the incorporation of additional tools and the implementation of advanced techniques. Figure 7 demonstrates the performance of the segmentation algorithm on an MRI slice.

Thresholding is a well-known image segmentation method that separates an object from its background. The method divides pixels of the image according to their intensity levels and can be used to create a binary image. Double or multiple thresholding is based on several threshold values like T0 and T1. Segmentation of the original image $I(i, j)$ can be defined as:

$$I'(i,j) = \begin{cases} m, \ if \ I(i,j) > T1, \\ n, \ if \ I(i,j) \leq T1, \\ o, \ if \ I(i,j) \leq T0, \end{cases} \tag{6}$$

where I' (i, j) is an output image, pixels labelled with the number 1 correspond to the segmented object, and pixels labelled with the number 0 correspond to the background.

Thresholding methods are fast and computationally efficient, but they are sensitive to noise and image textural inhomogeneity. The thresholding has many variations that roughly can be divided into global (single), locally adaptive, and multiple thresholding. In our framework, a double thresholding method is applied for brain segmentation from the skull and other surrounding tissues; that is accompanied by a region growing method.

Region-growing, or region-based method, highlights areas of the image according to predefined uniformity or homogeneity criteria of pixel intensity values. The region merging starts from a “seed point” (a single pixel or group of pixels) that belongs to the target object. Either manual or automatic initialization of the seed allows additional neighbouring pixels, which satisfy similarity criteria of the tissue, to be added to the growing area.

Stages of the algorithm include: (a) partitioning the image into seed regions, (b) fitting planar or biquadratic models to the seed regions, (c) finding all neighbouring pixels that are compatible with those regions, (d) increasing the model’s order if no compatible points are found (stop growing the region if the model order reached the maximal number; otherwise, continue region growing by returning to the previous step), compare the seeds regions by calculation the goodness of fit.

Compatibility of neighbouring points is defined as:

$$C_i^{(k)} = \{((x,y)|d^2(x,y,a,m)) \leq \epsilon \ and \ (x,y) is \ a \ 4-neighbour \ of \ R_i^{(k)} \cup C_i^{(k)}\} (7)$$

where ϵ is the compatibility threshold and (x, y, a, m) are 4-neighbouring points of the region R.

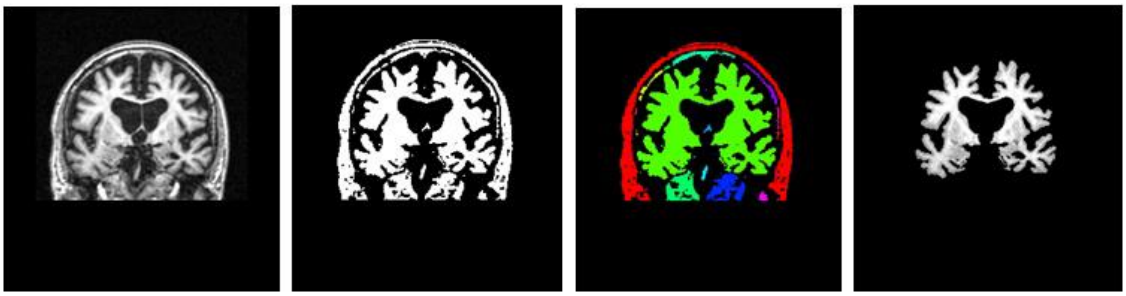
Differences between old and new regions of the model are identified by computing the goodness of fit (P):

$$P^{(k+1)} = x^2\left(M, a^{(k+1)}, R_i^{(k+1)}\right) - x^2(M, a^k, R_i^k) \tag{8}$$

where M is the Model.

The region-growing procedure is repeated until no more matching pixels are left. A combination of thresholding and region-based methods is applied for each preprocessed 2D brain image.

Figure 8 demonstrates the brain segmentation approach that process image through binarization thresholding and region detection stages.



(a) (b) (c) (d)

Figure 8: Brain segmentation stages: original image (a), binary image (b), image after application of region detection algorithm, (d) final image of the segmented brain

The next level of segmentation helps to detect differences between the right and left hemispheres defining a brain asymmetry image. The method is based on the finding that there is a loss of the grey and white matter at the initial stage and along with the development of the neurodegenerative disorder, which leads to variations in the symmetry of the brain's structure. Initially, the symmetry between the left and right hemispheres of the brain increases. However, the progression of the disease increases the degree of asymmetry as the left-sided hemispheric lateralization of a healthy individual gradually becomes right-sided with the development of severe dementia such as Alzheimer's Disease. For a cognitively normal person, Figure 9 demonstrates the variations in the anatomy of the left and right hemispheres.

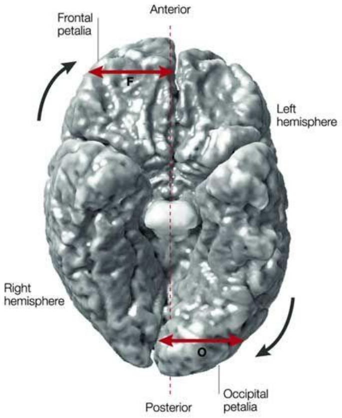


Figure 9: Lateralization of a healthy brain (Source: Toga & Tompson, 2003, Mapping Brain Asymmetry, Nature Reviews Neuroscience)

Differences in the symmetry of the two hemispheres can be detected through segmentation.

There are many computer vision techniques proposed for the segmentation of specific brain areas in accordance with the anatomical atlas (Despotović et al., 2015). The thesis presents an algorithm for the segmentation of the hemispheric asymmetries whose key point is the detection of the vertical axis of symmetry between the left and right hemispheres of the brain. The hypothesis being tested in this part of the work is that there is an axis of reflective symmetry running through the centre of the brain (Liu et al., 2001). The centre point of the brain is allocated using an image binarization technique and calculating the image centroid (Teverovskiy and Li, 2006). In the context of image processing and computer vision, the centroid is the weighted average of all the pixels in an image. The "weighted" centroid, or centre of mass, is always at the exact centre and depends on the grey levels in the image.

The binarization algorithm (Michalak and Okarma, 2019) converts a 256-shaded greyscale image to a binary (black and white-coloured). The binarization is done according to the adjusted level of a threshold. All pixels in the image above the threshold level are replaced by the value 1 (white) and other pixels that are below that level by the value 0 (black).

The brain centre might differ from the centre of the whole image, including the background. If such a case occurs, the brain needs to be translated into the centre of the image and rotated to the correct angle via the vertical axis. As soon as the brain centralization, translation, and rotation techniques are performed, the image can be flipped or reversed from left to right across the vertical axis (Ruppert et al., 2011). The mirroring process is finalized by the segmentation of image asymmetries (Herzog and Magoulas, 2021).

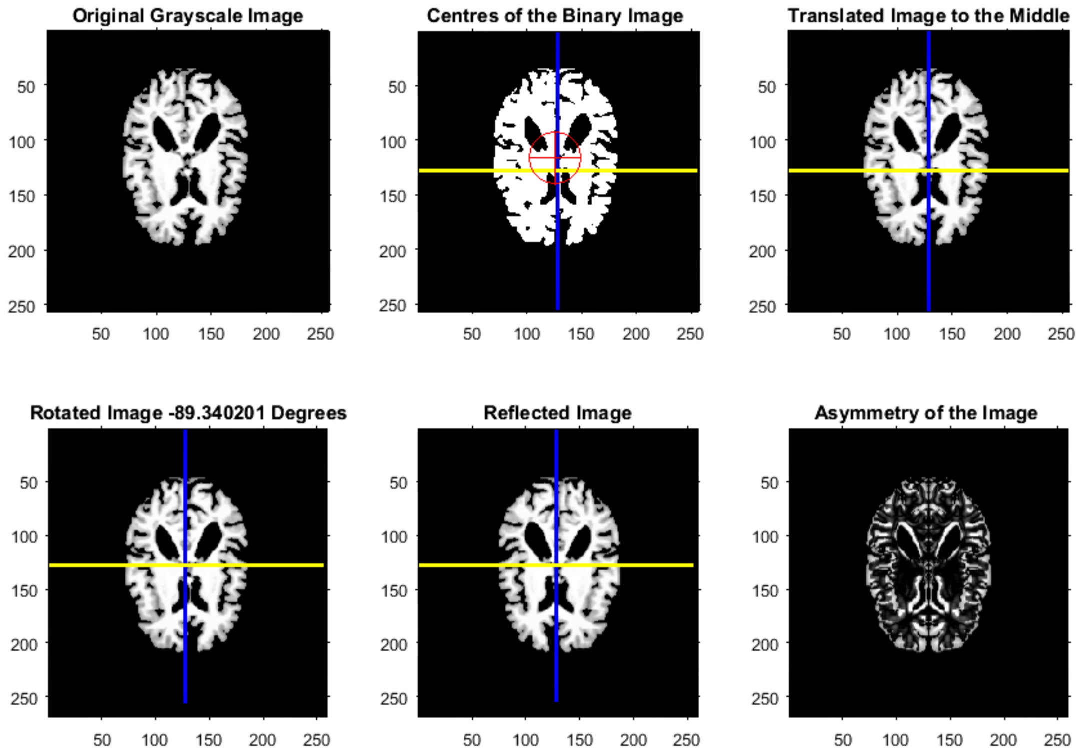
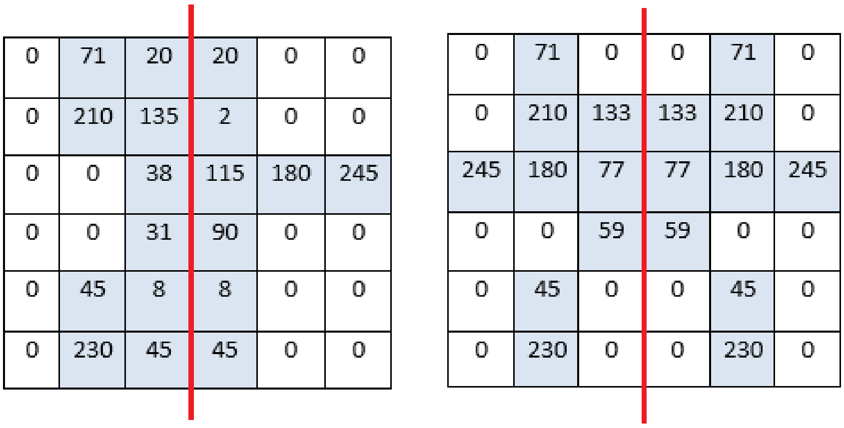

Original Grayscale Image
Centres of the Binary Image
Translated Image to the Middle
Rotated Image -89.340201 Degrees
Reflected Image
Asymmetry of the Image

Figure 10: The image transformation stages for detection and segmentation of image asymmetries

The last image of this stage (see Figure 10) is obtained as a result of mirroring the left-brain hemisphere to the right and of the right-brain hemisphere to the left, which is followed by subtraction of the hemispheres from each other (the norm representation for pseudocode is not require as we assume only positive values):

$$D = (L - R) + (R - L) \tag{9}$$

where D is an image asymmetry, L is an image matrix of the left hemisphere, R is an image matrix of the right hemisphere.



0	71	20	20	0	0
0	210	135	2	0	0
0	0	38	115	180	245
0	0	31	90	0	0
0	45	8	8	0	0
0	230	45	45	0	0

0	71	0	0	71	0
0	210	133	133	210	0
245	180	77	77	180	245
0	0	59	59	0	0
0	45	0	0	45	0
0	230	0	0	230	0

Figure 11: An illustrative example of matrix transformation values of a greyscale image of size 6-by-6: initial matrix (*left*) and matrix of segmented asymmetry (*right*), mirrored via the vertical axis. The numbers in the cells correspond to the grey level of the pixel values

The symmetrical image areas (see Figure 11) get a value of 0 due to matrix subtraction. They are visualized as black areas in the image. The asymmetrical parts of the image are represented as different intensity grey levels from 1 to 255.

As a result of matrix operations, asymmetrical regions of the brain, whose pixels have different intensity levels according to the level of asymmetry, become segmented from the image. The proposed algorithm highlights shape and pattern of asymmetry in the image (Herzog and Magoulas, 2021).

The algorithm was tested on single slices of the brain, but the same idea can be extended and applied to the whole 3D brain image.

3.9 Summary and contribution of the chapter

This chapter introduced the computational framework of the thesis. The first part of the methodology includes the image preprocessing and segmentation stages. The second part is focused on the machine learning tasks (This part is represented in detail in Chapter 5).

Medical imaging can be obtained due to a number of imaging techniques that include X-rays, computed tomography (CT) scans, Positron Emission tomography (RET) scans and Magnetic Resonance Imaging (MRI). MRI technology is the most popular method for medical diagnosis. They are frequently used as they are radiation-free and have high efficiency. MRI scans are pretty good for visualising the biological tissues and showing the different levels of their intensity. This type of scan has been chosen as a source of data for the research on the detection of brain pathology. MRI scans can be obtained in certain types and formats. Image types differ by the physical characteristics of the biological tissues and can be represented as T1 and T2. The pixel data property and metadata specify MRI formats. NIFTI and DICOM are some of the formats used for the study. The imaging data often needs to be adduced to one standard for the machine diagnosis. It might include converting the medical formats to the traditional digital image formats such as '.jpg', '.bmp', or '.tif', image resizing and image normalisation procedures. Histogram-based and statistical-based methods can be applied for image normalisation.

The key role of the diagnostic cycle of the current research play image segmentation. Group of intensity-based methods, such as thresholding and region-growing techniques, are used in the study for the segmentation of region of interest and the detection of brain pathologies.

Neuroimaging data used in the research were obtained from the ADNI and OASIS databases. Magnetic resonance 3D image sequences were downloaded in NIFTI and DICOM formats and converted to PNG format composed of 2D MRIs. The neuroimaging data investigated in the research include MRIs of patients with Early Mild Cognitive Impairment, Alzheimer's Disease and control images of cognitively healthy subjects.

All the obtained images passed through the resizing and normalization stages. Two types of segmentation algorithms have been applied at the image processing level. The first, the brain segmentation algorithm, separated the brain from the skull and non-brain tissues using double thresholding and region-based methods. The second segmentation algorithm detected and visualized the brain asymmetries. The last approach was based on binarization, translation and reflection techniques.

Segmented images of brain asymmetries become a source of features used to analyse and classify neuroimaging pathologies, as it will be discussed in the next chapter. For example, the feature engineering process was based on transforming image pixels into wavelets by applying DWT. It facilitates generating the first and second-order statistical features responsible for the distribution and positioning of the grey-level values. Later, the Bag-of-Feature approach was also adapted for the neuroimaging data to obtain a detailed image signature. The features collected using the Bag-of-Feature algorithm were added to the statistical ones to improve the classification performance of the diagnostically sophisticated medical images (see Chapter 5.2 for details).

This chapter’s main contribution to the early diagnosis of cognitive decline is a proposal for a new digital biomarker and a computational framework for obtaining the biomarker from MRIs of the brain.

As mentioned above, the next chapter performs the analysis of statistical features extracted from the hemispheric asymmetries of three image classes, EMCI, AD and NC, and explains their differences. The obtained features are processed and analysed for each class of imaging data. The analytical tests are repeated for multiple datasets of different lengths and dimensions. Also, the following chapter demonstrates the robustness of image asymmetry features in the diagnoses of brain pathology using a variety of classification algorithms.

Chapter 4: Features of Segmented Brain Asymmetry: Analysis and Classification

4.1 Introduction

In this chapter, the robustness of brain asymmetry features as biomarkers for the early diagnosis of dementia within a traditional machine learning workflow is examined. In this context, statistical analysis plays an important role in understanding the nature of medical imaging data. Statistical interpretation of the data gives a possibility to evaluate the brain structure from different perspectives. The analytical part of the study is based on the assessment of the statistical properties of segmented asymmetries. The purpose of this stage of research is to understand the changes in the brain structure leading to the progression of the disease.

The statistical information is obtained from 5 image sets of segmented asymmetries: a 150-image set of male MRIs in the coronal plane, a 150-image set of female MRIs in the coronal plane, a 300-image set of male subjects in the coronal plane, a 300-image set of males in the axial plane, and a 300-image set of males in the coronal plane after processing the statistical data with principal component analysis (PCA) algorithm. The statistical interpretation of the asymmetrical brain changes involves ten statistical features described in Sections 4.3 – 4.6. In the first stage of analysis, the average value of results is computed from the statistical features collected from each class of images of the dataset. After that, the averaging results are normalized. An additional analytical test was performed with 13 statistical features after processing them with Principal Component Analysis (PCA).

4.2 Feature engineering overview (feature generating processes)

Feature detection and description (Salahat and Qasaimeh, 2017) are important steps of image evaluation. The concept of feature detection and description refers to the procedure of identifying interest points in an image or any object that can be used to describe and analyze the image (object) contents.

The chosen features should have the following essential qualities (Tuytelaars and Mikolajczyk, 2008; Shalev-Shwartz and Ben-David, 2014):

1. Repeatability, which refers to the possibility of the algorithm detecting the same points independently in the original and transforming them from authentic images;
2. Invariance to the affine transformation such as translation, rotation, and scale;
3. Invariance to the presence of noise, blur etc.;
4. Locality, which characterizes the robustness to occlusion, clutter and illumination change;
5. Distinctiveness, which is means that the structure of the region should be rich in variations and intensity patterns;
6. Quantity represents the sufficient number of points characterising the image;
7. Time efficiency reflects the shortest time required to identify the features.

The first five characteristics should be as high as possible to consider a feature detection algorithm successful. At the same time, features' quantity and time efficiency have to be minimized as much as possible without getting a negative impact on image representation.

4.2.1 Feature detection

Feature detection algorithms can be divided into groups depending on the localization of the detected features - points, regions, corners, and edges of the image. These algorithms detect the differences in the pixel's grey levels gradient of the neighbouring pixels. For instance, the algorithm localises the edges by changing pixel values in a specific vertical or horizontal direction. If the algorithms are constructed to operate simultaneously in both directions, they can detect corners and regions.

The first kind of feature is *key-point features* or interest points. They are described by the group of neighbouring pixels surrounding the point of interest. One kind of these types of features is 'corners'. Key-point features can align different images. The advantages of key-point features are the ability to match objects, even if they have occlusions, different scales or orientation changes. These sorts of features have often been used for object instance and object category recognition. There are several famous feature detection algorithms. The classic "Harris" detector uses a specific filter for image convolution and perfectly detects corners of the image (Harris and Stephens 1988). More modern variants of detectors filter the image in horizontal and vertical directions with derivatives of a Gaussian function (Schmid, Mohr, and Bauckhage 2000; Triggs 2004). The feature description process often accompanies feature detection. Local features or key-point descriptors help identify and describe the unique characteristics of the image. The list of those detectors and descriptors can be started from Scale-invariant feature transform (SIFT) (Lowe, 2004), histogram of oriented gradients (HOG) (Işık, 2014), and speeded up robust features (SURF) (Bay et al., 2008). In general, the algorithms mentioned above analyse the pixel information within a local neighbourhood area and indicate the gradient of the pixel's orientation around the point of interest. Thus, each feature receives orientation according to the dominant gradient of all pixels in the local area.

Another category of features indicating the object boundaries is '*edges*'. They can be 'curves' or 'straight-line segments'. The information obtained from the edges complements both key-point and region-based descriptors. The most popular edge detector and descriptors, such as Canny, Sobel, Prewitt, Robert, and LoG are based on detecting discontinuities in brightness (Muthukrishnan and Radha, 2011). These and alternative descriptors are widely used for image segmentation and analysis.

Feature detection is followed by feature extraction (García et al., 2016). Sometimes before the extraction step, the features are selected from the initial feature set to simplify the training model.

4.2.2 Feature selection

Feature selection includes multiple techniques. They help to reduce the complexity of a model. This benefits its efficiency, reduces training time, and makes the model easier to interpret. Feature selection techniques roughly fall into three categories: filtering, embedding and wrapping.

Filtering techniques preprocess features and remove those that will be of no use for the modelling process. Most of these methods allocate some sort of weight to features and delete the ones that fall below the allocated level of a threshold (Sánchez-Marono et al., 2007; Jović et al., 2015). In other words, the features are ranked, and highly ranked features are left for analysis and classification. The filtering techniques are not expensive for implementation, but they do not take into account the model being employed. This limitation does not guarantee that the right features will be selected for the model. For this reason, conservative prefiltering strategies are recommended to eliminate irreversible mistakes of removal of useful features.

Wrapper methods allow for the processing of subsets of features and checking their efficacy (Shardlow, 2016). Each time, the features between subsets are reshuffled, and the efficacy of the new sets are evaluated. The reshuffling process continues until a notable increase in efficacy is reached. It prevents accidental deleting of those features that are non-informative by themselves but useful when taken in a combination. This method treats the model as a black box that provides a quality score for every proposed subset of features. Subsets are refined iteratively while the model is processed. The wrapper methods are pretty expensive and time-consuming.

Embedded methods incorporate feature selection as part of the model training process (Guyon and Elisseeff, 2003). Examples of this method can be mentioned as a decision tree algorithm that selects features in every training step or split of the tree (Quinlan, 1986) and a multinomial logistic regression algorithm that predicts the probabilities of the different possible outcomes of a dependant variable (Böhning, 1992). Embedded methods select the specific model features and balance between computational expenses and quality of results.

4.2.3 Feature extraction

Feature extraction is a dimension reduction process when unneeded, irrelevant and redundant attributes are removed from the data. This process reduces the complexity of a predictive model and contributes to its accuracy. Usually, feature extraction algorithms are focused on specific patterns of the image. These patterns can be retrieved from shapes, colours, and textures. Feature descriptors identify and extract features from the object.

Shape descriptors roughly can be divided into two groups depending on the geometrical area of the object: region-based and contour-based (Amanatiadis et al., 2011). The region-based descriptor extract features from the whole area of a shape, and the contour-based descriptor process features from the contour of a shape only. Some other descriptors have a combination and variation of the main two.

The representatives of *region-based descriptors* are Image Moments (IM) and Angular Radial Transform (ART) descriptors. IM descriptors use a statistical approach for shape recognition and are widely applied in object classification tasks) (Kotoulas, and Andreadis, 2005). Image Moments shows an invariant to translation, rotation, and scale. ART descriptor is a moment-based method that demonstrates pixel distribution within an object region and is able to describe both connected and disconnected shapes (Ricard et al., 2004). The important characteristics of this algorithm are the robustness to different kinds of rotations and perspective deformations.

Contour-based descriptors calculate values around an object's outline in either clockwise or anti-clockwise directions. Fourier descriptors (FD), Curvature Scale Space (CSS) descriptors are examples of this group. The FD is a classical method for shape recognition (Dalitz et al., 2013). This method has multiple derivatives and is applied to encode various shape signatures. The FD can quickly be normalized to represent shapes independently of their orientation, size, and location. A significant advantage of this descriptor is the robustness to noise. The CSS method describes shape boundary as a one-dimensional signal and analyses this signal in scale space (Frejlichowski, 2012). The algorithm is invariant to the scale, rotation, location of an object within the image and robust to noise.

Colour descriptors define the colour property of the image within a particular colour space. The colour descriptors consist of several histogram descriptors, a dominant colour descriptor, and a colour layout descriptor (CLD) (Manjunath et al., 2001). Examples of colour spaces applied across studies are red-green-blue (RGB), hue-saturation-value (HSV), hue-max-min-diff (HMMD), hue-saturation-intensity (HSI). The colour descriptors are actively used for image recognition.

Texture descriptors characterize the surface of the object. In computer vision, textural image analysis is one of the main methods used for object recognition and image classification. Compared to the colour descriptors, which use single-pixel properties, texture descriptors evaluate and interpret the group of pixels. Greyscale texture analysis can be grouped into four categories: statistical methods, structural methods, model-based methods, and transform-based methods.

Statistical methods explore the spatial distribution of pixels when a set of local statistical features is obtained from the distribution of the grey values at each point of the image (Nanni et al., 2016). Statistical methods include first-, second-order statistics and run-length matrix (RLM) analysis. First-order features, such as mean, variance, skewness, kurtosis, etc., estimate pixels properties eliminating the spatial relations between neighbouring pixels. Second-order statistics named contrast, correlation, energy, homogeneity, and entropy, summarize the relationships between the group of pixels according to their orientation and distances in the grey level co-occurrence matrix (GLCM). The GLCM approach has a high level of matrix dimensionality. RLM refers to higher-order statistics. The RLM texture analysis describes a distribution of pixels in constant grey level runs and captures coarse textures in a grey-scale image (Bharati et al., 2004). Besides traditional statistical texture analysis, a combination of all these methods can be applied.

Structural methods represent texture by well-defined texture elements (microtexture) and rules describing their features and placement (macrotexture) (Materka and Strzelecki, 1998). Texture elements can be different shapes (lines, circus, etc.), and the structural approach provides a good symbolic description of the image. It can be more beneficial for image synthesis than for image analysis. Various proposed approaches are based on textual structures. However, these methods appear to be limited since they only describe regular textures.

Model-based methods generate empirical models of the image texture based on the variations of pixel parameters. One of the parameters, for instance, can be defined as a weighted average of the pixel intensities in the neighbourhoods. This principle is used in the autoregressive (AR) model (Materka and Strzelecki, 1998). The conditional probability of the intensity of a particular pixel depending on the values of the neighbouring pixels underlies the Markov random fields (MRF) and fractal model descriptors (Al-Kadi, 2008).

Transform-based or signal processing methods are based on spectral transform. These methods convert the image into a new form using the spatial frequency properties of the pixel intensity variations. Representatives of transform-based methods are Fourier descriptors, Gabor filters (GF), and Wavelet Transform (WT) descriptors. Implementation of the Fourier descriptor is limited due to its lack of spatial localisation. Gabor filters are more effective in spatial localisation. However, they struggle to localise a spatial structure in natural textures (Armi and Fekri-Ershad, 2019). The Wavelet Transform descriptors propose a variety of special resolutions that helps to represent a texture at a suitable scale, and the range of wavelet functions allows finding the best option for texture analysis. GF and WT descriptors are effectively used in texture analysis for image segmentation and classification tasks.

In image processing, a discrete wavelet transform is a technique to transform image pixels into wavelets (Kalaiselvi and Nagaraja 2016; Usman and Rajpoot 2017). It is used for time-frequency image analysis, which selects the appropriate frequency bands based on the characteristics of the signal. The DWT can be applied for lossless image compression and image denoising. The method of signal processing decomposes an image into different subbands (sub-images) with separable low- (L) and high-pass (H) filters. The filters are applied in two directions, horizontal and vertical. They divide the image into four sub-images with approximated (LL) and detailed (HL, LH, HH) components. The sub-image LL is obtained from the low-pass filtering in both directions, horizontal and vertical. The HL sub-image is received from high-pass filtering along the horizontal axis and low-pass filtering along the vertical axis. The LH sub-image includes filters in the opposite, compared to the HL sub-image, directions. The remaining HH sub-image is obtained from high-pass filtering in both directions. The third and higher levels of decomposition are generated in a similar way. Figure 12 demonstrate how the DWT operates at the imaging level.

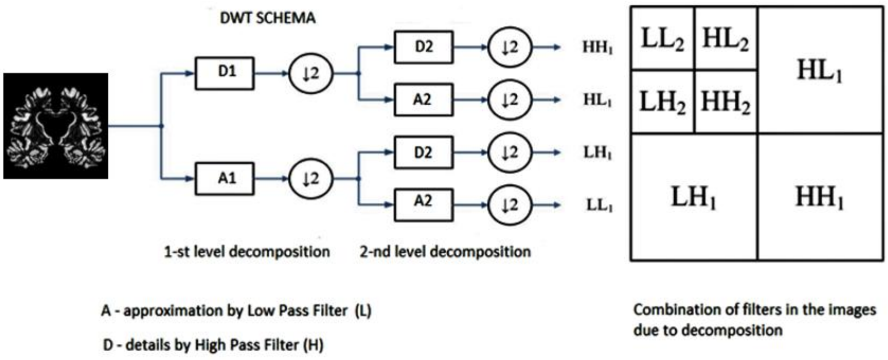


Figure 12: The DWT schema of the 1-st and the 2-nd level of the image decomposition after applying high- and low-pass filters in the horizontal and vertical directions

4.3 Feature engineering for brain asymmetry images

Feature engineering gives knowledge about the most remarkable characteristics of the image (Zheng and Casari, 2018). Features detection and description refer to the procedure of identifying points of interest in an image (or object) that can be used to describe and analyse the image (object) contents providing valuable data for image analysis (García et al., 2016). The first machine learning pathway of the computational framework for brain asymmetry includes the analytics part that is based on the evaluation of the statistical properties of imaging data. The statistical descriptors of the image texture can generate a number of relevant and distinguishable features, which is crucial for the interpretation of the research findings.

Imaging property can be analysed using structural, statistical or hybrid approaches. Multivariate statistical analyses help to discriminate between different types of brain tissue that are represented by their own textural parameters. The texture analysis approach is also relevant for those tissues and organs that appear to be less detectable.

4.3.1 Statistical feature engineering

Approaches that are based on statistical features for representing image properties are well-established in image processing (Di Ruberto and Fodde, 2013). The statistical descriptors of the image texture, colour or morphological properties generate a limited number of relevant and distinguishable features. Often, the first and second-order statistics operators significantly impact the texture analysis of the established medical domains (Avola et al., 2013). The first-order statistical features provide information about the distribution of the grey-level values in the image. However, these types of features do not demonstrate the position of relative grey-level values within the image. This limitation can be overcome with second-order statistical features. Matrices of relative pixels can represent the configuration of grey-level occurrences I_1 and I_2. The matrix describes the frequency of grey level I_1 and I_2 pixels that are separated by a distance d and have a specific direction θ within the same window. In such a way, the second-order statistics provide the information extracted from the co-occurrence matrix about the relative distance d and orientation θ of the pair of pixels. The orientation θ has four directions: horizontal, vertical, diagonal and anti-diagonal, with angles of 0°, 45°, 90° and 135° accordingly. The distance d is calculated as an average value of the four directions. The feature will not change its value if the image is rotated, i.e., they are rotationally invariant.

All statistical features together have a high possibility to determine the differences between two kinds of images.

The proposed machine learning workflow uses ten statistical features to represent the image asymmetries: MSE (Mean Squared Error), Mean, Std (Standard deviation), Entropy, RMS (Root Mean Square), Variance, Smoothness, Kurtosis, Skewness, IDM (Inverse difference moment). These features are commonly selected and widely used for statistical analysis. Their number was chosen randomly to compare the property of three classes of MRIs processed with segmentation algorithms. The first feature on the list, MSE, has been calculated directly from the original image and its mirrored version, while the rest of the features are generated using discrete wavelet transform (DWT).

As an example, the features vector, extracted from the image asymmetry of a patient's 2D MRI, is provided below:

[849.477703 10.47024065 90.10031233 0.764676239 51.73779963 7260.005288 0.999995013 52.78224277 5.661092548 9052.865381]

A brief description of each statistical feature used in this study is provided below.

4.3.2 Statistical feature description

The features calculated from images or image asymmetries give information about the likelihood of grey pixel values in a random position in an image, their orientation, and interaction with other surrounding pixels. They are defined as follows:

Mean Squared Error (1st order statistics)

The MSE is defined as the average squared intensity difference in the pixel values between the corresponding pixels of two images (Wang et al. 2004). Many applications use this measure as an image quality metric (Thung and Raveendran, 2009).

General steps to calculate the MSE of a set of X and Y values are: 1) finding the regression line; 2) inserting the X values into the linear regression equation to find the new Y values (Y'); 3) subtracting the new Y' value from the original to get the error; 4) calculating the square errors; 5) summarizing the errors; 6) finding the mean of the errors.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Yi - Yi')^2, \tag{10}$$

The regression line describes a relationship between two variables X and Y through an equation of a straight line. This line is called "the line of best fit". The calculation of the regression line can be done with a formula:

$$Y = a + bX\ , \tag{11}$$

where

$$b = \frac{\sum_{i=1}^{N} X_i Y_i - NX'Y'}{\sum_{i=1}^{N} X_i^2 - NX'^2}, \tag{12}$$

and

$$a = Y' - bX', \tag{13}$$

In the thesis, MSE was calculated between the original image and its mirrored version. The formula for MSE calculation is provided below:

$$MSE = \frac{1}{M*N}\sum_{i=0,j=0}^{M-1,N-1}[I(i,j) - K(i,j)]^2 \tag{14}$$

In the formula, M is the width, and N is the height of the images where *I (i, j)* is a reflected image and *K (i, j)* is the original image. The *i* and *j* are the pixel values in the rows and columns of the original (*K*) and reflected (*I*) images. The MSE is equal to zero when both images are identical.

Mean (1st order statistics)

The mean in computer vision is the texture feature that measures the average value of the intensity pixel values (Kumar and Gupta, 2012). The mean is one of the parameters that represent the brightness of the image. The image is bright when the mean is high, and the image is dark when the mean is low. In the formula below Ng is a number of grey level pixel values and *p(i)* is a probability of pixels.

$$\mu = \sum_{i=0}^{Ng-1} ip(i) \tag{15}$$

Standard deviation (1st order statistics)

The standard deviation (STD) in terms of image processing indicates how much deviation or dispersion exists from the mean or average (Esmael et al., 2015). A low standard deviation indicates that the data points tend to be very close to the mean, while a high standard deviation demonstrates that the data points are spread out over an averaging value. The standard deviation indicates the contrast of grey level intensities. The low value of the standard deviation shows the low contrast, and the high value presents the high contrast of the image. In the formula, *Ng* indicates the number of the grey pixel values, and μ is a mean intensity value:

$$STD = \sqrt{\frac{\sum_{i=0}^{Ng-1}(x_{i-}\mu)^2}{Ng-1}} \tag{16}$$

Entropy (2nd order statistics)

The entropy in computer vision characterizes the image texture and measures the randomness of the pixel intensity distribution (Yang et al. 2012). The entropy is the highest when all the pixel probabilities $p(i,j)$ are unequal. The entropy has smaller values when the entries in pixels $p(i,j)$ are equal. As a result, the homogeneous image has a lower entropy value and heterogeneous regions have a higher entropy value.

$$ENT = -\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} p(i,j)log\big(p(i,j)\big)\,, \tag{17}$$

where $p(i,j)$ is a probability of grey levels in a pair of pixels and defined as:

$$p(i,j) = \frac{C(i,j)}{\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} C(i,j)}, \tag{18}$$

where $C(i,j)$ is a matrix of frequencies of the grey level pairs (i,j) and Ng is a number of grey level pairs (i,j).

Root Mean Square (1st order statistics)

The RMS is a measure of the magnitude of a set of values (Lee et al., 2017). It shows how far these values are from the line of best fit. RMS gives a sense of the typical size of the values. It works for positive and negative numbers. To calculate the RMS of a set of values: 1) square all the values in the set; 2) find the average (arithmetic mean) of the squares; 3) take the square root of the result.

$$RMS = \sqrt{\frac{\sum_i x_i^2}{N}} \tag{19}$$

where N is the number of measures, and x_i is each value.

Variance (1st order statistics)

The variance in a digital image is a measure of image heterogeneity (Yang et al., 2012). It refers to the grey level variability of the pixel pairs. The variance is calculated as the sum of squared differences between the intensity values of the central pixel and its neighbours. It shows how the greyscale values differ from their mean.

$$VAR = \sum_{i=0}^{Ng-1}(i-\mu)^2 p(i) \tag{20}$$

or for pair of pixels:

$$VAR = \sum_{i=o}^{Ng-1}\sum_{j=0}^{Ng-1}(i - \mu_x)^2\, p(i,j) + \sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1}\big(j - \mu_y\big)^2\, p(i,j), \tag{21}$$

where pixels value in pairs *p(I, j)* and *Ng* is a number of grey level values.

Inverse difference moment (2nd order statistics)

Inverse difference moment (IDM) indicates the local homogeneity of an image (Yang et al., 2012). The IDM increases when pixel pairs are close in their greyscale values. So, a low IDM value characterizes heterogeneous images, and a relatively high IDM value indicates homogeneous images.

$$IDM = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{1}{1+(i-- j)^2}\, p(i,j), \tag{22}$$

Smoothness (1st order statistics)

Smoothness is one of the parameters of homogeneity. It measures the relative smoothness of intensity in an image (Malik and Baharudin, 2013). It is high for an image region of constant intensity and low for regions with large deviations in their intensity values. The smoothness is measured by using the standard deviation value. A larger number of STD indicates a smooth image texture.

$$SM = 1-- \frac{1}{1+(STD)^2} \tag{23}$$

Kurtosis (1st order statistics)

In digital image processing, kurtosis is used to measure the peak of the distribution of the intensity values around the mean (Ho and Yu, 2015). A high value of the kurtosis indicates that the peak of the distribution is sharp-edged, and the tail is long and heavy. A low value of the kurtosis demonstrates that the peak of the distribution is rounded, and the tail is short and thin. Kurtosis values are often interpreted in combination with noise and resolution of the image. A high kurtosis value is accompanied by low noise and low resolution.

$$KURT = \sigma^{-4} \sum_{i=0}^{Ng-1} ((i-\mu)^4 p(i)) - 3, \tag{24}$$

where σ is a standard deviation, μ is a mean value, and i is a pixel value.

Skewness (1st order statistics)

In digital imaging, skewness shows the asymmetry of the probability distribution of the pixel intensity values about the mean value (Esmael et al., 2015). The skewness can be positive or negative. Any symmetric data have skewness near zero. Negative values of skewness indicate data that are skewed left, and positive values indicate data that are skewed right. The skewness to the left means that the left data tail is relatively long compared to the right one. In image processing, data skewness can give information about image surfaces. Darker and glossier surfaces tend to be more positively skewed than lighter and matte surfaces.

$$SKEW = \frac{1}{(stddev)^3} \sum_{i=1}^{N} (x_i - - mean)^3 , \tag{25}$$

$$SKEW = \sigma^{-3} \sum_{i=0}^{Ng-1} (i - \mu)^3 \, p(i) \tag{26}$$

Contrast (2nd order statistics)

Contrast represents the amount of colour or greyscale differentiation or ratio of pixel values between various image features. Images with higher contrast levels display a greater degree of colour or greyscale pixels' variation than images with lower contrast. The contrast demonstrates the density of tissues in the image.

$$CON = \sum_{n=0}^{Ng-1} n^2 \left\{ \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i,j) \right\} \tag{27}$$

$$|i - j| = n$$

Correlation (2nd order statistics)

Correlation is a measure of grey level linear dependences of neighbouring pixels. 0 speaks about uncorrelated values, 1 shows a high level of correlation.

$$COR = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i,j) \frac{(i-\mu_x)(j-\mu_y)}{\sigma_x \sigma_y} , \tag{28}$$

where $\mu_x, \mu_{y,} \sigma_x, and\ \sigma_y$ are the means and standard deviations of *p(i, j)*.

Energy (2nd order statistics)

Energy or angular second moment (ASM) speaks about uniformity of distribution. Energy has a high value when the frequency of repeated pixels in the matrix is high. This is related to the homogeneous image. It is calculated by summarizing the squared elements in the Grey Level Co-occurrence Matrix:

$$E = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i,j)^2 \tag{29}$$

4.4 Statistical feature analysis of 150 images set

150 images set in the coronal plane (male)

The first dataset is represented by 150 images of the male subjects in the coronal (frontal) plane. All images are equally distributed between three groups of AD, EMCI and NC participants. Table 4 presents the averaging feature results based on normalized features data extracted from each group's segmented image asymmetry set. The following Figure 13 provides a visual representation and comparison of the data.

Table 4: Mean values of 10 statistical features extracted from 150 2D images of male participants

	MSE	Mean	Std	Entropy	RMS	Variance
AD	1.610606	1.65855	1.232508	0.835095	0.706434	1.364122
EMCI	1.185587	1.39993	1.041964	0.844349	0.592445	0.964807
NC	1.522815	1.650167	1.215422	0.84552	0.68953	1.28672

Smooth	Kurtosis	Skewness	IDM
0.999997	0.46855	0.539209	1.459882
0.999996	0.482657	0.548148	1.184178
0.999997	0.505423	0.561646	1.414765

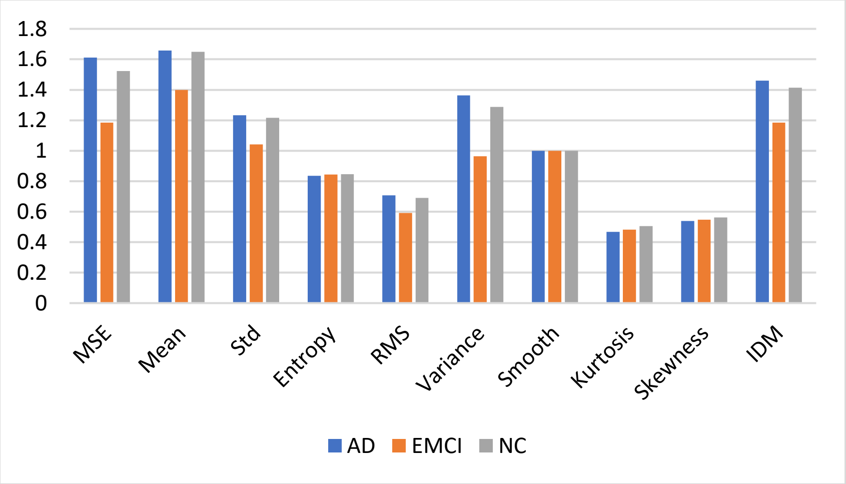


Figure 13*:* Statistical mean of each statistical image feature for normal cognition (NC), early mild cognitive impairment (EMCI), and Alzheimer's Disease (AD) patients (150 images set in the coronal plane for male subjects)

The comparison of statistical features of AD, EMCI, and NC classes demonstrates differences in their statistical characteristics. The features belonging to the EMCI class have the lowest values compared to the features of the NC and AD classes, with the AD class demonstrating the highest averaging feature values. The first feature, MSE, which calculates the difference between original and inverted image matrices, clearly indicates that EMCI data show more symmetry than NC and AD imaging data. The other statistical features are also supporting this hypothesis.

150 images set in the coronal plane (female)

The next created dataset belongs to the female participants. It consists of the same number of images as the previous set of male subjects. Initially, image sets were separated according to gender. This condition was applied due to the differences in the brain size and write-left lateralisation of the anatomical structures of the two genders. Statistical data extracted from the segmented asymmetries were processed and combined in the feature Table 5 and illustrated in Figure 14.

Table 5: Averaging feature values received from 150 2D images of female participants

	MSE	Mean	Std	Entropy	RMS	Variance
AD	1.596137	1.496493	1.21072	0.792835	0.676323	1.3241
EMCI	1.008233	1.190574	0.936756	0.785103	0.506458	0.820428
NC	1.077443	1.234181	0.985403	0.791182	0.541784	0.867183

Smooth	Kurtosis	Skewness	IDM
0.999996	0.56743	0.590364	1.334345
0.999995	0.58882	0.603903	1.05987
0.999995	0.539956	0.577644	1.09192

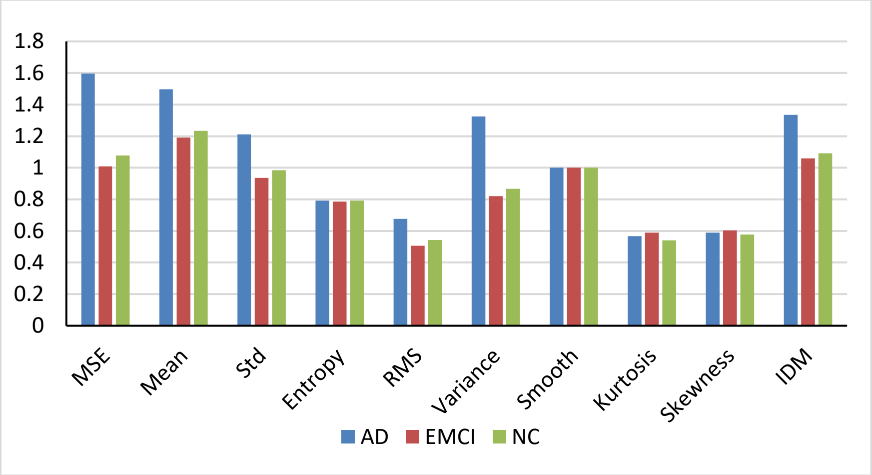


Figure 14: Statistical mean of each statistical image feature for normal cognition (NC), early mild cognitive impairment (EMCI), and Alzheimer's Disease (AD) patients (150 images set in the coronal plane for female subjects)

From the tables and figures above (see Tables 4, 5 and Figures 13, 14), the statistical values highlight the differences between features collected from AD, EMCI and NC groups. Features represent EMCI class with lower values than AD and NC classes for both genders.

MSE value analysis of 150 image sets of male and female subjects

The MSE feature analysis with a Pareto chart has been performed for male and female participants. The MSE value for each class has been calculated from the differences between the original image and its mirrored version for all images in classes of AD, EMCI, and NC representatives and indicates the impact of asymmetry for each class. Figure 15 provides an MSE feature analysis with a Pareto chart. The cumulative line on the secondary axis shows the contribution of each bar (image class) in the total value as a percentage.

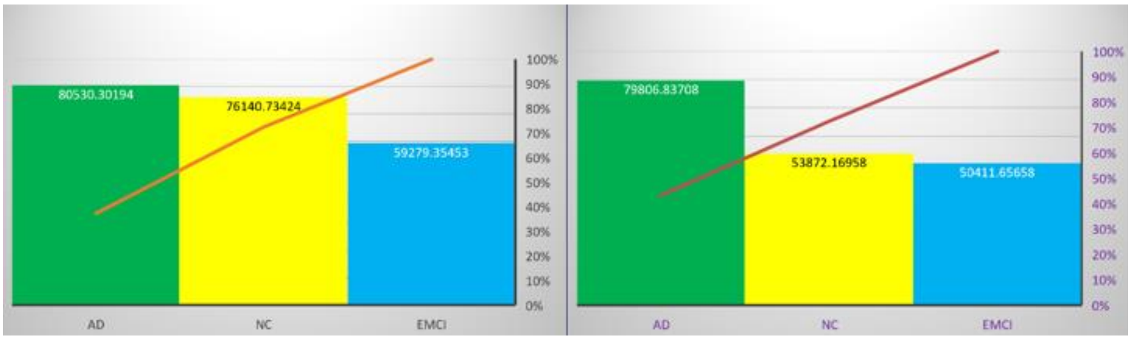


Figure 15: Pareto charts of MSE feature analysis for MRI data of the male (*left*) and female (*right*) subjects for three classes. The total MSE feature value for both bar charts is placed in the coloured bars. The axis on the right indicates the cumulative percentage of the total value for each class

The Pareto bar chart indicates the impact of asymmetry for each class. The highest MSE bar is associated with patients of the AD class. The highest MSE bar for this class confirms that changes in symmetry in the MRIs of this image group are substantial compared to changes in symmetry in the MRIs of the other groups. At the same time, the EMCI image group has smaller values and looks more “symmetrical” than others. The pattern of changes in male and female subjects does not show significant differences.

4.5 Statistical feature analysis of 300 image set

300 images set in the coronal plane (male)

A new male 300 image set was created from 150 different participants to check the influence of the double-size data on the evaluation performance. The feature engineering was performed in a similar way for these imaging data. All the data are equally divided into three groups of Alzheimer’s Disease-AD, Early Mild Cognitive Impairment-EMCI and Normal Cognitively-NC participants. Table 6 and Figure 16 show averaging normalized statistical data of each extracted feature collected from 300 images of 150 subjects with segmented asymmetries.

Table 6: Normalized mean values of 10 statistical features obtained from 300 images of segmented asymmetries of 150 male participants

	MSE	Mean	Std	Entropy	RMS	Variance
AD	1.80849	1.719051	1.288539	0.820447	0.719896	1.49946
EMCI	1.399871	1.484234	1.12843	0.822393	0.625747	1.119476
NC	1.594207	1.677185	1.209318	0.876431	0.702067	1.286044

Smooth	Kurtosis	Skewness	IDM
0.999997	0.510943	0.565706	1.577456
0.999996	0.502798	0.557935	1.347272
0.999997	0.456303	0.527473	1.420371

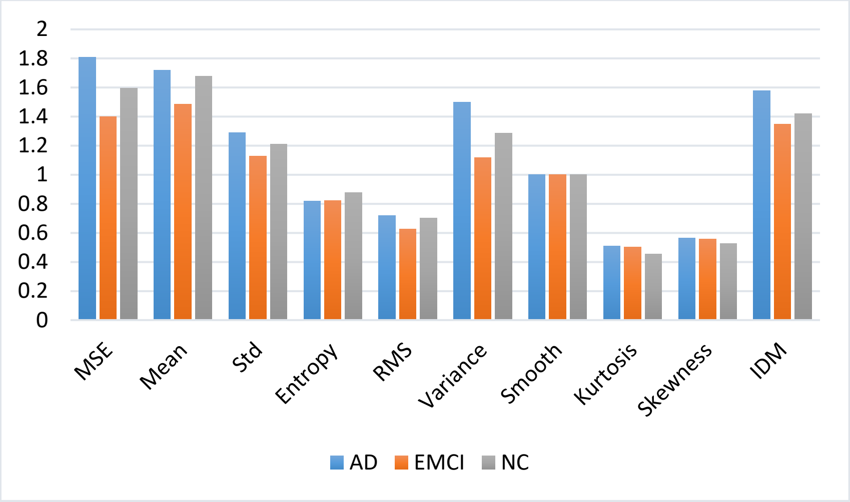


Figure 16*:* Statistical mean of each image asymmetry feature for normal cognition (NC), early mild cognitive impairment (EMCI), and Alzheimer's disease (AD) patients calculated from 300 images of male subjects in the coronal plane

The evaluation of the results demonstrates, similar to the previous datasets, changes in the asymmetries of AD, EMCI and NC subjects. EMCI group looks more symmetrical compared to the AD and NC groups. The highest asymmetry shows the group of patients with Alzheimer's Disease.

300 images set in the axial plane (male)

An additional experiment was conducted with another 300 images set that was created from the image asymmetries, not in the coronal (frontal) but the axial (horizontal) plane. The horizontal axis was chosen to investigate the contribution of the topologically different features to the classification process. These features are processed similarly to the previous ones and provided in Table 7. Figure 17 gives a graphical representation of the statistical values.

Table 7: Normalized mean values of 10 statistical features obtained from 300 images (axial plane) of segmented asymmetries of 150 male participants

	MSE	Mean	Std	Entropy	RMS	Variance
AD	0.261327	0.25081	0.157751	0.97021	0.876396	0.1997
EMCI	0.210589	0.232144	0.139754	0.997116	0.781286	0.153199

NC	0.266186	0.266342	0.159873	1.012704	0.918746	0.203346

Smooth	Kurtosis	Skewness	IDM
0.999998	0.34638	0.461496	0.180701
0.999998	0.320396	0.441634	0.172252
0.999998	0.327003	0.45007	0.188292

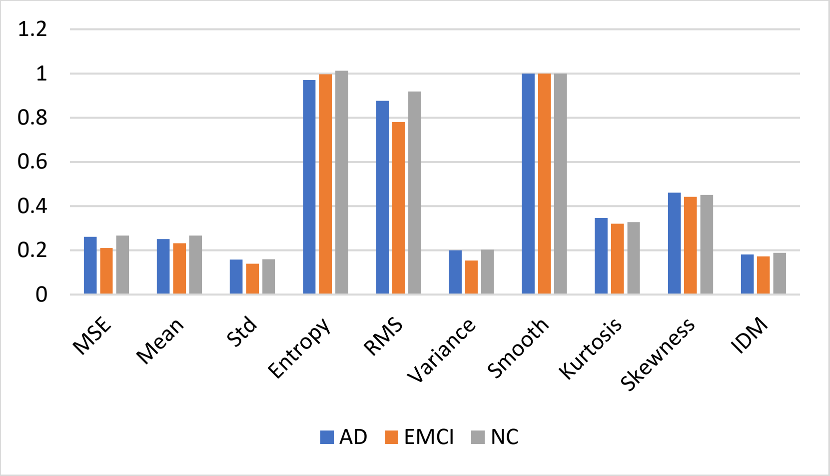


Figure 17: Statistical mean of each image statistical feature for normal cognition (NC), early mild cognitive impairment (EMCI), and Alzheimer's Disease (AD) patients (300 images set in the axial plane for male subjects)

The EMCI group shows a higher tendency to symmetry than a class of AD and a class of NC subjects from the data above.

300 images set in the coronal plane (male)

The features sets were extended with three more 2nd order GLCM statistical features, named contrast, correlation, and energy. GLCM features are well known for texture analysis, but due to the high level of dimensionality, they often require the implementation of the dimensionality reduction technique. In this case, PCA (principal component analysis) was added to the feature extraction algorithm. Table 8 and Figure 18 demonstrate the varieties among AD, EMCI and AD classes after applying PCA on features extracted with DWT. Eight out of thirteen features such as MSE, 1st order statistics (mean, kurtosis, skewness), and 2nd order statistics (IDM, entropy, contrast and correlation) show visual differences between image classes. Early stages of dementia, that might be related to the amnestic MCI, lead to a decrease in asymmetry between the right and the left hemisphere of the brain.

Table 8: Normalized mean values of 13 statistical features obtained from 300 images (coronal plane) of segmented asymmetries of 150 male participants.

	MSE	Mean	Std	Entropy	RMS	Variance	Smooth	Kurtosis	Skewness	IDM	Contrast	Correlation	Energy
AD	0.151633	0.397616	0.809645	0.219148	0.8105414	0.6560376	0.9583988	0.3514284	0.2948674	0.2958834	0.3032587	0.1256219	0.8458561
EMCI	0.119791	0.412472	0.810086	0.224008	0.8111071	0.6567431	0.9587062	0.3230652	0.2721116	0.266333	0.2921182	0.1351724	0.8412450
NC	0.150042	0.394934	0.809335	0.22391	0.8102608	0.6554778	0.9582721	0.3406915	0.2841802	0.3112698	0.2974115	0.1301043	0.8460503

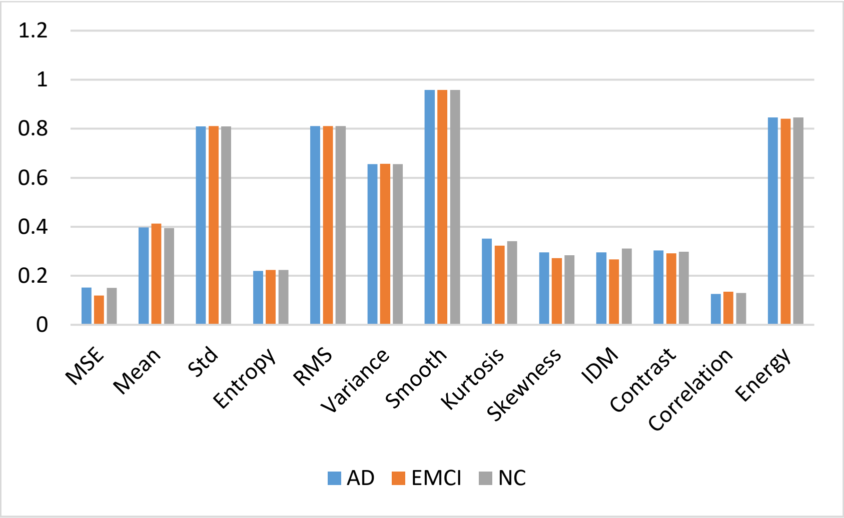


Figure 18: 13 statistical features (MSE, 1st and 2nd line of statistics) after applying PCA on 150 image male datas

4.6 Comparative analysis of feature sets

Comparison of MSE and other statistical values of the four datasets

This subsection compares the statistical properties of four datasets with ten statistical features for the data analysis. Figure 19 illustrates the changes in MSE values between the three image classes, AD, EMCI and NC, in four completely different image sets.

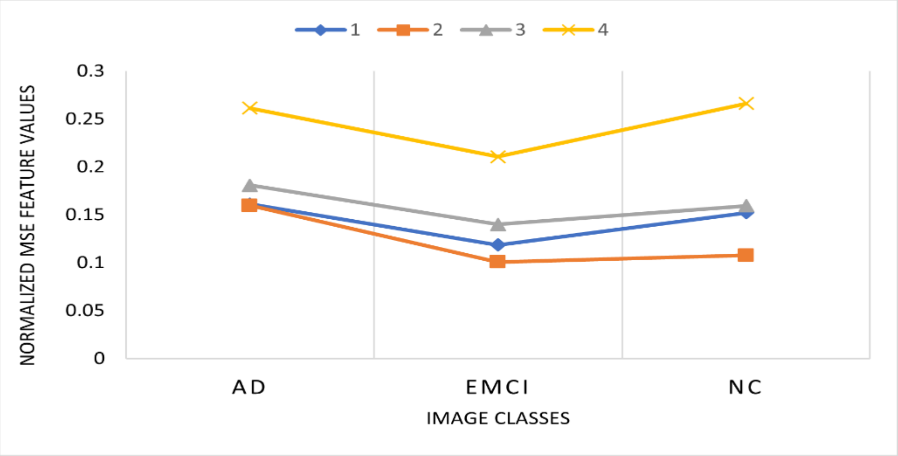


Figure 19: Comparison of MSE feature values between the three classes of MRIs. Numbers, 1,2…4, indicate the investigated datasets: #1 refers to a 150-image set of male MRIs in the coronal plane, #2 is a 150-image set of female MRIs in the coronal plane, #3 corresponds to a 300-image set of male subjects in the coronal plane, and #4 represents a 300-image set of males in the axial plane

Images of the EMCI class obtain the lowest MSE value for all image sets. It indicates more symmetry for the EMCI class than for AD and NC classes. The differences between AD and NC classes are less noticeable.

Further analysis of brain asymmetries was performed using the ratio of statistical values of binary classes such as AD vs EMCI, EMCI vs NC and AD vs NC. Interclasses ratios are displayed in Table 9.

Table 9: Ratios of statistical values for binary image groups combined from four different datasets (*Var states for Variance, **Kurt for Kurtosis, ***Skew for Skewness)

Datasets	MSE	Mean	Std	Entropy	RMS	Var	Smooth	Kurt	Skew	IDM
AD/EMCI	1.36	1.19	1.18	1.00	1.20	1.42	1.00	0.98	0.98	1.24
EMCI/NC	0.78	0.85	0.86	0.99	0.84	0.74	1.00	0.94	0.98	0.84
AD/NC	1.10	1.00	1.02	0.99	1.01	1.05	1.00	0.92	0.96	1.04
AD/EMCI	1.60	1.26	1.29	1.00	1.31	1.61	1.00	0.97	0.98	1.25
EMCI/NC	0.99	0.97	0.95	1.00	0.95	0.94	1.00	1.09	1.03	0.97
AD/NC	1.48	1.22	1.22	1.00	1.26	1.52	1.00	1.06	1.03	1.22
AD/EMCI	1.29	1.16	1.14	1.00	1.14	1.34	1.00	1.02	1.00	1.17
EMCI/NC	0.88	0.88	0.93	0.93	0.90	0.87	1.00	1.09	1.08	0.95
AD/NC	1.13	1.02	1.07	0.93	1.03	1.16	1.00	1.11	1.08	1.11
AD/EMCI	1.24	1.09	1.14	0.97	1.13	1.33	1.00	1.09	1.05	1.06
EMCI/NC	0.77	0.85	0.88	0.99	0.85	0.75	1.00	0.97	0.98	0.89

AD/NC	0.96	0.93	1.00	0.96	0.96	1.00	1.00	1.06	1.02	0.95

From the calculations above, we can see that the highest ratio for the features MSE, Mean, Std, RMS, Variance, and IDM is received for the AD and EMCI subjects' group for all datasets. AD/EMCI ratio for MSE feature for all datasets is between 1.24 and 1.60, for Mean is between 1.09 and 1.26, for Std is between 1.14 and 1.29, for RMS is between 1.13 and 1.31, for Variance is between 1.33 and 1.61 and for IDM is between 1.06 and 1.25. It means that the differences in the symmetry between AD and EMCI classes are substantial. The same features demonstrate the smallest and most stable ratio for the EMCI vs NC groups. EMCI/NC ratio for MSE feature for all datasets is between 0.77 and 0.99, for Mean is between 0.85 and 0.97, for Std is between 0.86 and 0.95, for RMS is between 0.84 and 0.95, for Variance is between 0.74 and 0.94 and for IDM is between 0.84 and 0.97. This provides evidence for the presence of initial degenerative changes in the brain that can be visible by dedicated software and diagnosed. It aligns with previous neuroscience study by Liu et al. (2018). The statistical value ratio for AD and NC classes for most of the features in datasets demonstrates the differences in symmetries but on a smaller scale than for AD and EMCI classes. AD/NC ratio is provided below, and for the MSE feature in the four datasets it is between 0.96 and 1.48; for Mean it is between 0.93 and 1.22; for Std it is between 1.00 and 1.22; for RMS it is between 0.96 and 1.26; for Variance it is between 1.00 and 1.52 and for IDM it is between 0.95 and 1.22. It is possible to see that the ratio for this group indicates the different types of symmetry. In some cases, images of the AD class are more symmetrical than the NC class, and in other cases, they are less symmetrical. The constant changes in symmetry due to the progression of the degenerative processes in the brain can explain this fact. If dementia progresses, the left-right asymmetry also grows. Thus, the level of asymmetry depends on the stage of dementia, which was also verified by neuroscience studies such as Yang et al. (2017) and Wachinger et al. (2018).

Entropy, Kurtosis and Skewness features are not very informative for these types of imaging data. Their ratios for binary classes have a small difference that cannot help with diagnosis. Also, the smoothness values tend to be unity and are not suitable for the predictions.

Additional analysis of the statistical values provides the following information about the structure of MRI asymmetries. From Tables 6, 7, 8, and 9, the lowest average values

(MSE, Mean, Std, RMS, Variance and IDM) compared to AD, and NC data for all datasets has EMCI. The MSE provides information about the differences between the original and reflected images. The Mean represents the brightness of the image, while the Std indicates the contrast of grey level intensities. RMS measures the magnitude of a set of values and the Variance measures the image homogeneity, which shows how the greyscale values differ from their mean. IDM also indicates the local homogeneity of the image (low values are specific for heterogeneous images). In this vein, images of EMCI can be characterised as images with small differences between original and reflected ones. The images have low contrast and are not homogeneous. The pixel values of the EMCI group have a low magnitude. Compared to EMCI and NC classes, AD has the opposite picture and the highest average values for the same statistical features.

The statistical feature analysis supports the view that image asymmetry decreases in the initial stage of the generative process in the brain (Early Mild Cognitive Impairment) and grows when the person develops moderate and severe dementia (Alzheimer's disease).

4.7 Summary and contribution of the chapter

The feature engineering process is an important part of the diagnostic cycle. It plays significant role in understanding the image characteristics and in their analysis. Correctly identified and selected features make the diagnostic process relevant and efficient. Multiple feature generation algorithms give the opportunity to receive detailed image descriptions and highlight an image exclusiveness. This is vital for image analysis and classification tasks. The collected features must have several essential characteristics such as locality, distinctiveness, repeatability, invariance to affine transformation, and presence of noise. Therefore, the chosen correctly feature detection algorithm helps to identify the area of interest in the image. Numerous textures, colour, and shape descriptors are proposed to find and extract the most important features.

The statistical information on the image asymmetries was used to understand the structural changes in the brain due to the progression of cognitive decline. It contributes to the early diagnosis of dementia and helps to predict the development of the neurodegenerative process. The study compared and analysed the statistical features of five different multiclass MRI sets. Each set represents three image classes of EMCI patients, AD patients and NC subjects. In all tests, the EMCI class demonstrated the highest level of symmetry between the left and right brain hemispheres. This fact can be used as a diagnostic biomarker of early cognitive decline. The highest asymmetry level is detected due to the development of the advanced stages of Alzheimer’s Disease.

Thus, the proposed approach contributes to the diagnosis of early dementia by introducing segmented asymmetries as an additional imaging biomarker. The exploration of the statistical features collected from hemispheric asymmetries and the following machine learning tests demonstrate the robustness of the algorithms.

The next chapter gives a detailed description of the machine learning experiments with classic and deep neural networks and their results in the diagnosis of Mild Cognitive Impairment and Alzheimer’s Disease. The inputted imaging data for the tests is represented by MRIs of segmented brain asymmetries.

Chapter 5: Learning Brain Asymmetry Biomarkers

5.1 Introduction

This chapter investigates the potential of brain asymmetry biomarkers to support a diagnosis of cognitive decline and early dementia by exploiting classic machine learning algorithms and convolutional neural network architectures. It contains three studies. The ***first study*** checks the brain asymmetry features for their robustness in the detection of brain pathologies. The ***second study*** explores the potential of deep transfer learning for image asymmetries with fine-tuning of the pretrained CNN, AlexNet, incorporating an SVM module. The experiments are conducted using a gradual increase in training epochs. The ***third study*** is focused on the diagnostic potential of the specialised CNN architectures based on the 5CLNN model, and comparisons with the AlexNet and VGG16. The 5CLNN model is tested with Softmax, SVM, KNN and LD classification modules.

5.2 Machine Learning Workflow for Classification of Brain Asymmetry Images

The first part of the computational framework based on computer vision algorithms has been described in detail in Chapter 3. The machine learning workflow for early diagnosis of dementia (see Figure 20) is presented in this chapter (Herzog and Magoulas, 2021). It includes different types of machine learning algorithms that can be applied to the images of segmented brain asymmetries. The first pathway, which consists of the classic ML algorithms such as SVM, LD, KNN and NB, exploits features collected from asymmetry images. The second one, based on the deep neuronal structure, uses asymmetry images directly. This research stage investigated various convolutional neural network models by applying transfer learning to the established pretrained architectures and newly proposed configurations.

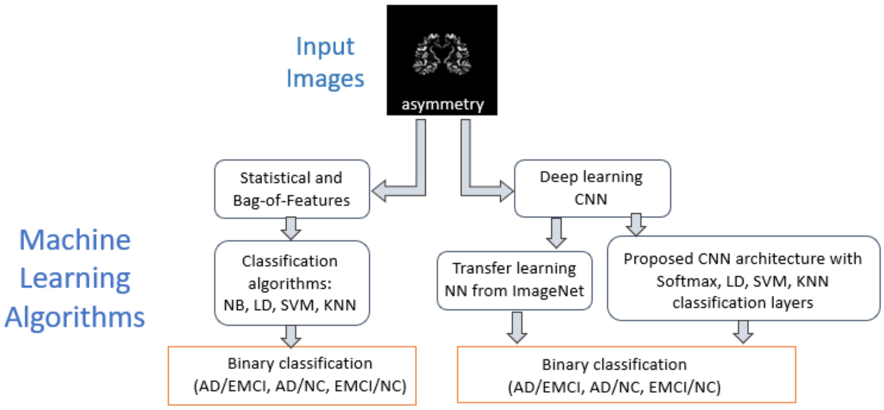


Figure 20: Computational framework including image transformation stages, asymmetry features generation and machine learning classification algorithms

Before the classification tasks, the MRI data passed through normalization, resizing and segmentation stages. The data preparation stages are described in Chapters 4.

5.2.1 Bag-of-Features approach

The Bag-of-Features (Rueda et al., 2012; Gabryel, 2018) is an adaptation of the Bag-of-Words algorithm that is used in natural language processing (NLP) and information retrieval for image analysis. In computer vision, the Bag-of-Features can be represented as a vector of occurrence or histogram of occurrences of the local image features (see Figure 21). The histogram is called the “visual signature” of an image.

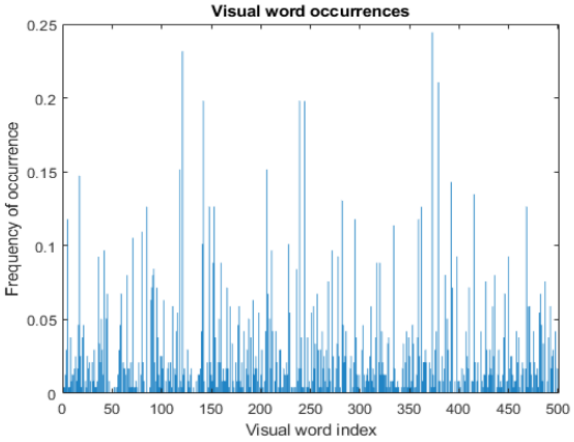


Figure 21: Example of the histogram of occurrences of an image

The BOF algorithm starts from the detection and representation of features in the image as numerical vectors. Feature descriptors are involved in these processes. There are several descriptors proposed in recent years. The Speeded Up Robust Features-SURF algorithm is used in the thesis.

The ***SURF*** is a region-based feature detection and extraction algorithm (Bay et al., 2008; Le and Gonzalez, 2009). SURF algorithm is several times faster than the well-known Scale Invariant Feature Transform (SIFT) (Karami et al., 2017) and outperforms it with respect to robustness, repeatability and distinctiveness. The speed of the SURF algorithm is achieved by relying on the Integral image (which represents a summed-area table that is performed for calculating the average intensity in the given image) (Ehsan et al., 2015) for image convolutions with a box filter (Haar-like filter) (Rezaei et al., 2013). The example of this filter is provided in Figure 48. A Hessian matrix-based measure is used for feature detection. A Hessian matrix is a squared matrix of second-order partial derivatives and has scalar-valued functions as entries (Carrazza et al., 2015). The SURF detector finds the localization and scale of interest points as extrema of the determinant of the Hessian matrix. Haar wavelet responses in the horizontal and vertical directions are applied for the feature description. Haar wavelet responses can indicate the presence or absence of specific characteristics in the image, such as edges or textural changes (Porwik and Lisowska, 2004).

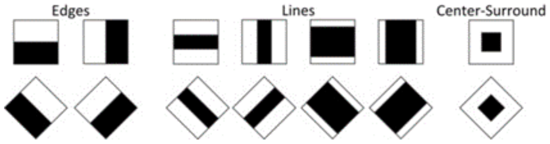


Figure 22: Haar-like filters examples (Source: Bianco et al., 2014***)***

For feature description, the SURF algorithm summarizes the pixel information within a local neighbourhood area and indicates their gradient orientation around the point of interest (see Figure 22). A local feature receives orientation according to the dominant gradient direction. A square sub-region with selected orientations (vectors) forms a descriptor. 4×4 sub-regions with four marginal statistics give a descriptor with a total of 64 dimensions (4×4×4).

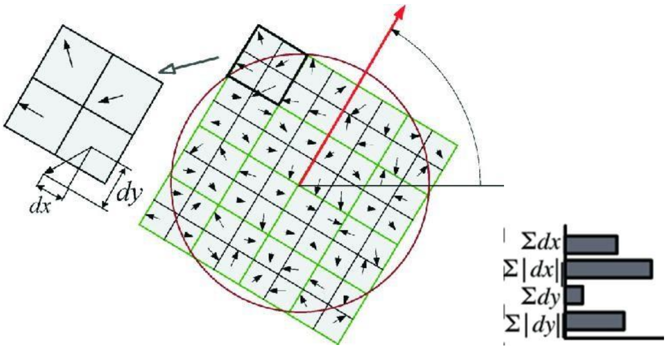


Figure 23: SURF descriptor (Source: Wang et al., 2011)

Generally, the SURF algorithm is characterized by rotation and scale invariance and works well with blurred images. But it is sensitive to viewpoint and illumination changes.

In the next stage of the BOF process, all described and extracted features are organized in clusters. This is done using the ***K-means*** clustering algorithm. The K-means is one of the well-known clustering algorithms (Bair, 2013). Clustering algorithms perform the grouping of data points (objects) with similar properties or features into clusters (see Figure 24). The data points are divided into clusters according to the distances between them in the data space. Defining an adequate distance measure is a critical process for successful clustering. Data points into clusters can belong completely to the assigned clusters (this is called “hard clustering”), or they can be addressed to one of them with some probabilities (this is applied to the “soft clustering”). There are more than 100 clustering algorithms that follow the different rules of defining the similarity among the data points. K-means belongs to the sub-class of centroid-based algorithms.

Centroid-based algorithms are repetitive algorithms where the concept of similarity is obtained by the proximity of a data point to the centroid of the clusters. For the K-means algorithm, the number of clusters should be defined in advance. In the first step of K-means clustering, any random data point is assigned as the initial mean of the cluster. Then, the centroid of the cluster is computed. The next computational step is re-assigning another data point to the closest cluster centroid. The re-calculations continue until the data convergence. The “data convergence” means that there is no further switching of data points between the clusters.

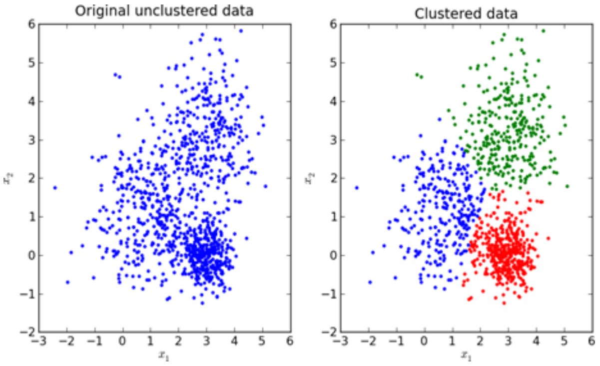


Figure 24: K-means clustering (Source: https://i.stack.imgur.com/cIDB3.png)

In the final stage of the BOF approach, each image is represented by a normalized histogram of occurrences of the image features. These features can be processed further with machine learning algorithms.

The BOF model is simple to understand. It gives a lot of flexibility for customization and can be easily implemented. The significant advantage of the BOF is that the model is suitable for both medium and large datasets. Unfortunately, there is a challenge to generate a common visual vocabulary that will be able to index any set of images. One of its limitations of it is the data sparsity when a lot of weights are equal to 0. It does not significantly impact the calculation, but it can lead to increased data space and affect the length of the processing time.

In the current research, the BOF algorithm was mainly used as an enriching algorithm for image classification to improve the precision of the detection of clinical pathology.

5.3 Model validation and evaluation

Every created model needs to be checked for its quality. The model evaluation process quantifies the quality of the predictions. For this reason, each created model should pass through the validation process and be able to make predictions on new independent variables or datasets. There are established metrics that are used to evaluate the model's performance. The better its current performance is, the harder to make changes for further improvements. At the same time, the very high performance on a training set is way to suspect the model overfitting.

5.3.1 Model validation

The validation procedure helps create an optimal model that can be applied to unseen data. Before validating the model, the data have to be split into the training, validation and testing parts. There are a number of ways to perform the validation process. Three classic evaluation approaches include hold-out, K-fold, and K-fold validation with shuffling.

For the ***hold-out validation***, the fraction of the data is separated from the training part. The validation runs on the testing part. This approach has a limitation when the dataset is small, and the testing set has only a few samples that cannot represent the data statistically. Cross-validation is preferable in this situation (Gonzalez et al., 2016)

The ***K-fold*** cross-validation approach split the data into *k* equal parts. For every partition *n*, a model is trained on the remaining partitions *k– 1* and evaluated on partition *n*. The final evaluation model's score is the *k* scores average. "***Leave-one-out***" is one of the cross-validation cases when the number of folds and instances in the dataset are equal. Every run on the algorithm is accompanied by separating one instance from the training set. Selected instances are different for every subsequent run. ***Iterated cross-validation with shaffling*** is another way of the validation procedure. It is the most precise method for small datasets and includes additional every time data shuffling before splitting the set (Gonzalez et al., 2016). Thus, cross-validation methods are helpful when the model performance demonstrates significant variance based on the data split for the training and testing parts.

5.3.2 Model performance

The confusion matrix or contingency table helps understand the model's results and, for example, answers the question of why the created system works worse or better than expected see Table 10). The confusion matrix of the binary classifier includes details about correctly and wrongly classified values. There are four possible outcomes on any given instance: True Positives (TP) when the classifier correctly labels positive values as positive; True Negatives (TN), when the negative representatives of the class are precisely determined; False Positives (FP) when the classifier mistakenly names a negative value as a positive, deriving in a "type one" classification error; and False Negatives (FN) when positive values incorrectly are called as negative, producing a "type two" classification error. A schematic view of the confusion matrix for the binary classifier is provided below.

Table 10: Binary confusion matrix

		Predicted classes	
		yes	no
Actual classes	yes	True Positive (TP)	False Negative (FN)
	no	False Positive (FP)	True Negative (TN)

Performance statistics

Performance statistics of the classification model can be computed from the confusion matrix discussed above. It is evaluated in terms of accuracy, sensitivity (recall), specificity, precision, F-score and area under the curve (Catal, 2012).

Accuracy shows how close a calculated value gets to the actual (true) value. The accuracy score is computed by multiplying calculated fractions by 100. The mathematical expression of accuracy is:

$$Accuracy = \frac{TP+TN}{P+N} \qquad (30)$$

where P stands for positive value (event), N for negative, F for false, and T for true value.

Accuracy is an efficient number that easy can be understood and explained. So, it can be easily used in any evaluation environment. The accuracy calculation's limitation can be met when the positive class is represented with much fewer numbers than the negative class. In this situation, another evaluation metric that can detect the positive class more accurately should be applied.

Precision is a more sensitive metric for detecting the positive class. Precision is a fraction of the correctly labelled positive events to all positive events.

$$Precision = \frac{TP}{TP+FP} \qquad (31)$$

Sensitivity (recall) is a fraction of true positives among all the correct events.

$$Sensitivity\ (Recall) = \frac{TP}{(TP+FN)} \qquad (32)$$

This measure is vitally important in medical diagnosis. The recall is different in cases of misdiagnosis or hyper diagnosis of the illness.

Specificity shows the opposite recall measure information. It calculates a fraction of negative events that are correctly identified.

$$Specificity = \frac{TN}{FP+TN} \tag{33}$$

F1-score (F-score, F-measure) balances the precision and recall statistics.

It is calculated as the harmonic mean or average of them.

$$F1\text{-}score = \frac{2(Recall * Precision)}{(Recall + Precision)} \tag{34}$$

The harmonic mean is constantly equal to or less than the arithmetic mean, and its lower value has a large incommensurable impact. The high value of the F1-score can be archived only in the situation when both recall and precision measures are high.

All the described above measures help to evaluate the performance of the classifier. However, it is important to remember that classifier works better in the most distinct situations when the sample is far from the threshold level. It prevents the sample from the assignment to it the alternative label. In this case, it is possible to speak about the confidence level that demonstrates how far the sample is from the given threshold. The confidence level can be assigned to the classifier with the borderline cases when the diagnosis is problematic.

Receiver Operator Characteristic curves (ROC)

The Receiver Operator Characteristic curve is a visual, graphical representation of the tested data, which illustrates how the number of correctly classified positive values (true positive rate) varies with the number of incorrectly classified negative values (false positive rate). ROC curve shows the relationship between sensitivity and specificity (see Figure 25).

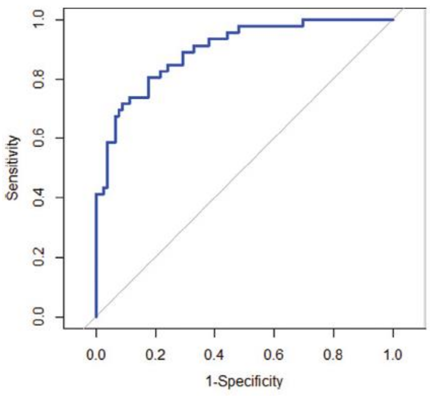
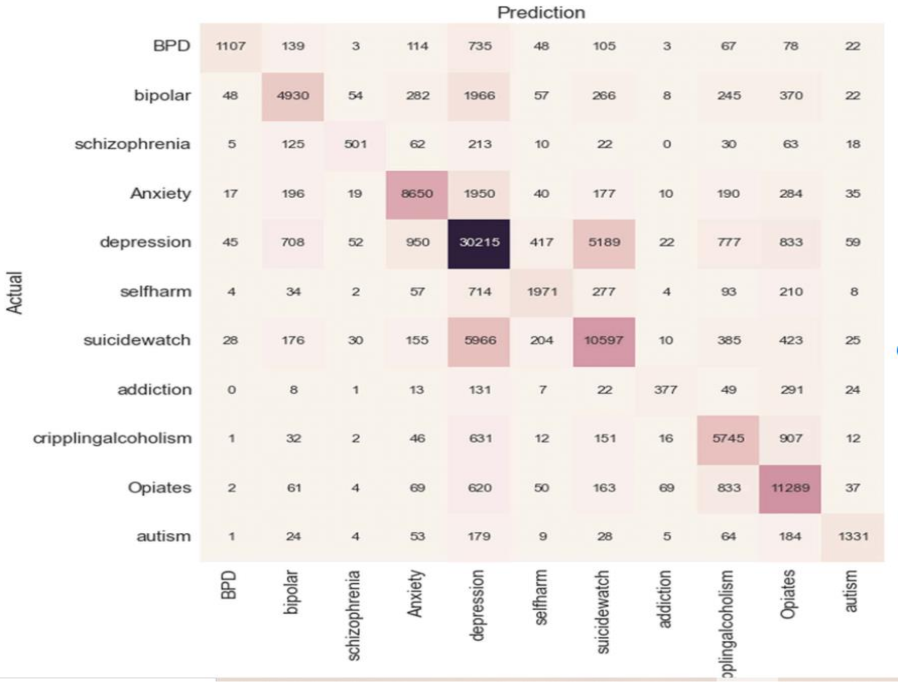


Figure 25: ROC curve example

The Area Under the ROC Curve demonstrate the distribution of positive and negative instances and measures the potential of a classifier to differentiate between classes. AUC can be used as a summary of the ROC curve. High values of the AUC indicate better performance of the classification model. The best possible 100% score equals one and looks like a square. The triangle shape of the AUC talks about 50% (or ½) of the classifier's efficiency.

Performance of multiclass systems.

The performance of multiclass systems is more challengeable than binary classifiers. The performance of the classifiers can drop with an increased level of complexity. More accurate results are reflected in the diagonal line of the confusion matrix (see Figure 26).



	BPD	bipolar	schizophrenia	Anxiety	depression	selfharm	suicidewatch	addiction	oplingalcoholism	Opiates	autism
BPD	1107	139	3	114	735	48	105	3	67	78	22
bipolar	48	4930	54	282	1966	57	266	8	245	370	22
schizophrenia	5	125	501	62	213	10	22	0	30	63	18
Anxiety	17	196	19	8650	1950	40	177	10	190	284	35
depression	45	708	52	950	30215	417	5189	22	777	833	59
selfharm	4	34	2	57	714	1971	277	4	93	210	8
suicidewatch	28	176	30	155	5966	204	10597	10	385	423	25
addiction	0	8	1	13	131	7	22	377	49	291	24
cripplingalcoholism	1	32	2	46	631	12	151	16	5745	907	12
Opiates	2	61	4	69	620	50	163	69	833	11289	37
autism	1	24	4	53	179	9	28	5	64	184	1331

Figure 26: Confusion matrix of multiclass classifier example (Source: https://www.researchgate.net/publication/315534048_Characterisation_of_mental_health_conditions_in_social_media_using_Informed_Deep_Learning)

The diagonal of the two-dimensional $d \times d$ confusion matrix M [i; j] holds the most important items from class *i* that were correctly classified as class i. The formula below expresses the precision (i) of the class *i* correctly identified.

$$Precision(i) = M[i,i] / \sum_{j=1}^{d} M[j,i] \tag{35}$$

Recall (i) demonstrates the fraction of all correctly identified class *i* members and is expressed by the following formula.

$$Recall(i) = M[i,i] / \sum_{j=1}^{d} M[i,j] \tag{36}$$

Error distribution

The created model can be evaluated by calculating the model's errors. Errors demonstrate the differences between the predicted $y' = f(x)$ and actual value y. The types of the error function include calculation of absolute error, relative error and squared error. *Absolute error* expresses the differences between predicted and true values as:

$$\Delta = y' - y \tag{37}$$

Relative or approximation error can be obtained by normalization the error magnitude and can be expressed as a fraction or percentage (when it is multiplied by 100%):

$$\epsilon = \frac{y - y'}{y} \tag{38}$$

Squared error calculation always gives positive results that can be easily interpreted.

$$\Delta^2 = (y' - y)^2 \tag{39}$$

However, large values contribute disproportionally to the total value. It makes the outliers dominate in error statistics for the large ensemble algorithms.

The commonly used error statistical measures used for evaluation of the model performance are Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). MSE formula was used to calculate the pixel errors between the original and mirrored versions of segmented image asymmetry. In this chapter, it is used for the calculation of the model's error.

Visualization of the error can be done using a plot of a histogram. It demonstrates the absolute error distribution for any predicted values (see Figure 27).

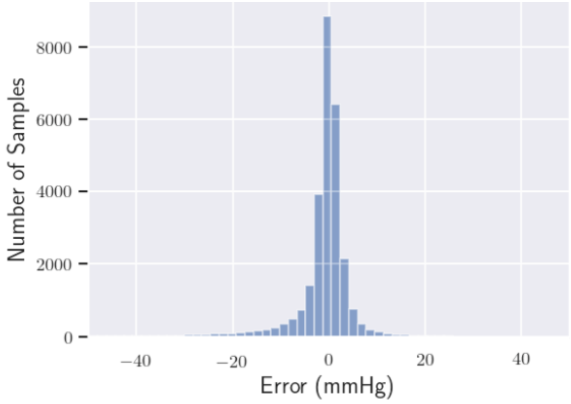


Figure 27: Error histogram example (Source: https://www.researchgate.net/figure/Mean-Error-Histogram-Here-we-present-how-the-mean-absolute-error-of-predicting-DBP-SBP_fig2_341148172)

The ideal distribution should be symmetrically centered around zero and be bell-shaped. This shape tells that small errors are more meaningful and common than large ones. Extreme outliers of such type of distribution are rare. If any of the conditions are not met, the error bias can be corrected using the model optimization algorithms.

5.4 Classification algorithms used in the thesis

Classification is relevant to a large group of supervised learning methods designed for predictive modelling problems. These methods can be performed using structural and unstructured data. They are efficient for binary and multiclassification problems.

The binary problem refers to the classification tasks with two class labelled outcomes and gives an answer if the current object belongs to one of the observed classes. Often in medical diagnosis, one class belongs to the normal state and another to the abnormal. For instance, “tumour detected” states the medical tests abnormalities and “tumour not detected” confirms a normal medical condition. Multiclass classification problem deals with a range of specified classes. It is not based on the principle of “normal” or “abnormal” outputs. It predicts only one possible class from the entire range. For instance, there are three brain tumours classes: astrocytoma, oligodendroglioma, and glioblastoma. While making a prediction, the multiclass classifier will name only one tumour from the above list.

The classification process workflow for supervised learning is provided in the Figure 28. It demonstrates the main stages of the data modelling and predictive labelling.

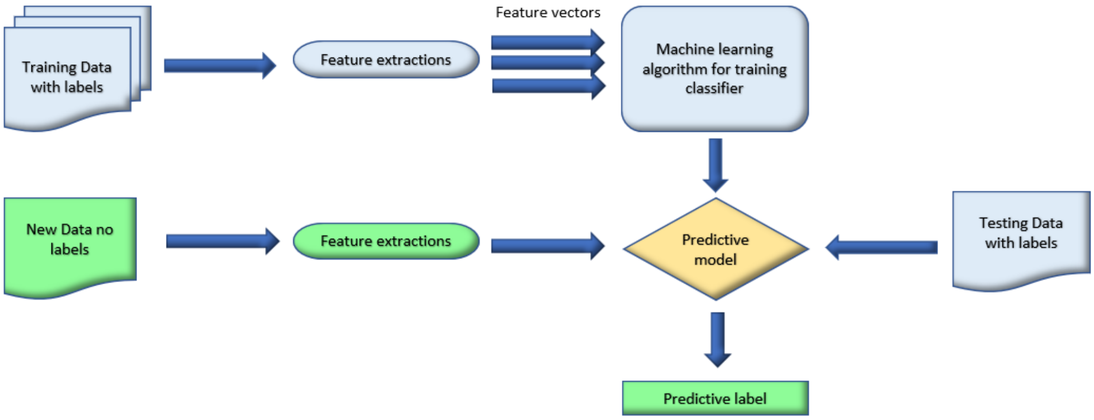


Figure 28: Classification process for the supervised learning workflow

The standard algorithms used for the classification tasks are described in this section.

The ***Naïve Bayes*** (NB) algorithm is based on Bayes' theorem. The algorithm is used for probabilistic statistical models that estimate the probability of each variable or class variables of data. The NB algorithm is suitable for binary and multiclass classification problems. A naive Bayes classifier (Schneider, 2005) is a statistical classifier that can predict class membership probabilities. The classifier assumes that there is strong independence between the features of the class. This type of assumption is called "conditional class independence". The representation of the Naïve Bayes is a list of probabilities. They are stored in the Naïve Bayes model. They include the probabilities of each class in the training dataset and the conditional probabilities of each input value (feature) in the class. It is essential to consider that the Naïve Bayes can learn the significance of the individual features but cannot determine their relationship. This type of classifier can be applied to both binary and multiclass datasets. The key benefit of the Naïve Bayes classifier is that it needs a small amount of training data to make a quick prediction. The popular representatives of the NB classifier are Gaussian, Multinomial, and Bernoulli.

The ***Linear Discriminant*** (LD) is a linear decision boundary classifier based on the data's Linear Discriminant Analysis (LDA) (Tharwat et al., 2017). This classifier is created by calculating data's class conditional densities following the Bayesian rules.

LDA or Fisher’s linear discriminant works through the projection of the dataset to the low dimensional space. It leads to the reduction of the data dimensionality. Usually, the LDA model associates each dataset class with a Gaussian density. Under this condition, all classes share the same covariance matrix. LDA estimates the between-class variance (sum of squared differences) and defines a distance between the mean of different classes. In the same way, it computes the within-class variance. After calculations of between-class and within-class variances, there is a possibility to construct the low-dimensional space with maximized between-class and minimized within-class variances. LDA has a close relationship with the regression analysis. They both express the dependent variables as a linear combination of different measurements.

The ***K-Nearest Neighbours*** (KNN) is a representative of the instance-based or memory-based learning algorithms (Aha et al., 1991). These algorithms are focused on the representation of the stored instances and measurements of the similarity between the instances. The instance-based algorithms look at data points and their characteristics around a single point to determine what group this data point belongs to. KNN is a straightforward supervised machine learning algorithm used to resolve classification and regression problems. It also can be applied to nonlinear data. KNN is considered a “non-parametric” and “lazy learning” algorithm. “Non-parametric” means that the algorithm does not rely on a fixed structure of a set of data parameters. “Lazy learning” indicates that the algorithm does not need any training data for a model generation. All training data goes straight to the testing stage, making it very slow. The critical factor of the accurate classification is the number of Ks (nearest neighbours) in the dataset. Their number has to be specified by the observer. A small number of neighbours might be significantly affected by a noise that will impact the classification results. A large number of neighbours makes the computations slow and expensive. It is recommended to perform KNN classification with a smaller number of features for better performance. The KNN algorithm starts by looking for distances between a query and all the data samples in the data space. After that, the algorithm selects the number of the nearest neighbours closest to the query and votes for the most frequent class or label. Euclidean, Hamming, Manhattan and Minkowski distances are often used for distance measures (Hassan et al., 2014). The algorithm's performance can be lower when the algorithm is applied to non-standardized and large dimensional data. For better results, it is recommended to normalize data on the same scale (usually between 0 and 1) and reduce the number of dimensions.

The ***Support Vector Machine*** (SVM) algorithms are supervised machine learning algorithms used for classification and regression analysis. A support vector machine can construct a hyperplane or set of hyperplanes in multi-dimensional space. The optimal separation line for two data classes must be at the midpoint of the channel at a half distance away from the nearest points of both classes. The greatest distance achieves a more substantial separation. The generalization error of this type of classifier will be small. SVM is effective in high-dimensional spaces. Class separation can be performed differently using linear and non-linear mathematical functions known as “kernel”. The popular non-linear SVM functions are polynomial, radial basis function (RBF) and sigmoid (Piccialli and Sciandrone, 2018). Linear SVM classifiers are often compared with logistic regression (LR) classifiers. Both LR and SVM create separating lines between data points, but the classifiers use different optimization criteria. The optimization function of logistic regression is based on minimising the sum of misclassification probabilities for each data point. At the same time, SVM looks for the maximum margin between data points of two classes. The LR and the SVM often demonstrate very close results due to the similarities of their loss functions. The SVM performs better than logistic regression when the classes are well separated in space. In contrast, logistic regression is preferred when classes are overlapped. SVM is very effective in high dimensional space, but it is not suitable for large data sets. Characteristics of classification algorithm are given in Table 11.

Neural network (NN) algorithms included in the table will be discussed later in Sections 5.6, 5.7 and 5.8.

Table 11: Classification algorithms comparison

Characteristics	NB	LD	KNN	SVM	NN
Speed of training	high	high	high	low	low
Tolerance to missing values	high	medium	low	medium	high
Interpretability	easy	easy	hard	easy for linear; hard for other kernels	hard

Model flexibility	low (for Gaussian kernel)	low, creates linear boundaries between classes	medium (for 10 neighbours); coarse (for 100 neighbours)	low (linear kernel); medium (quadratic, cubic, medium); high (fine Gaussian)	high
Tolerance to noise	high	medium	low	medium	high
Tolerance to overfitting	medium/ high	medium/ high	medium/ high	medium	low
Normalization	no	yes	yes	yes	no
Additional characteristics	works well for small data	high performance for normally distributed data	adaptable to the problem; easy to understand	tolerate to overfitting; has linear and non-linear kernels	solves the complex problem; process large data; large number of parameters enhance predictability
Limitations	uses statistically independent attributes	problem to classify categorical variables;	memory intensive; expensive; wrong distance produces an inaccurate result	difficult to determine optimal parameters	computationally expensive; considered as “black box”

5.5 Classification workflow based on classic ML algorithms

This stage aims to verify the robustness of brain asymmetry images and asymmetry features for early diagnosis of cognitive decline and Alzheimer’s Disease. The different classification models were applied to the feature tables created from image sets. Two types of features were used in this line: statistical and Bag-of-Features. The feature engineering process is described in Sections 4.2 and 4.3. Statistical features collected from image asymmetries were enriched with Bag-of-Features to get the most detailed image “signatures” and sent to the classification process. Supervised machine learning algorithms such as Naïve Bayes, Linear Discriminant, Support Vector Machine (linear, quadratic, cubic, Medium Gaussian kernels), and K-Nearest Neighbour (fine, cosine kernels) were tested at this stage of the research. Binary classification tasks were performed without fine-tuning the classification hyperparameters. All models from the list were trained using Matlab Classification Learner App except the NB classifier, which was trained separately. The 10-fold cross-validation procedure was applied to the models before their testing on unseen data. Table 12 presents the model's default hyperparameter values used in the experiments.

Table 12: Hyperparameters of the machine learning algorithms

Model	**Hyperparameters**
NB	Distribution: normal (Gaussian)
LD	Discriminant type: linear
L-SVM	Kernel function: linear Box constraint level:1 Kernel scale mode: auto Standardize data: true
Q-SVM	Kernel function: quadratic Box constraint level:1 Kernel scale mode: auto Standardize data: true
C-SVM	Kernel function: cubic Box constraint level:1 Kernel scale mode: auto Standardize data: true
MG-SVM	Kernel function: medium Gaussian Box constraint level:1 Kernel scale mode: manual Kernel scale: 32 Standardize data: true
Fine-KNN	Number of neighbours: 1 Distance metric: Euclidian Distance weight: equal Standardize data: true
Cos-KNN	Number of neighbours: 10 Distance metric: cosine Distance weight: equal Standardize data: true

5.5.1 Machine learning classification of asymmetry features

In the previous sections we established that brain image asymmetry features provide reliable representations of the asymmetry in the brain during dementia progression. In this section, the robustness of brain asymmetry features for the early diagnosis of dementia will be verified by running a machine learning workflow. To this end, asymmetry features will be used to train different machine learning models to detect cognitive decline and early dementia. At this stage, the focus will be on traditional supervised machine learning methods like SVM, KNN, LD, NB (SVM and KNN tests involved different types of kernels) and a pretrained CNN model was used for comparison. The aim is to investigate whether asymmetry features produce consistent performance across various classification models without any fine-tuning or model optimization performed.

To this end, a set of binary classification problems are solved using the machine learning models mentioned above. The experimental study uses 600 images of brain asymmetries generated from T1-weighted MRIs of 150 subjects of the ADNI database. Obtained images are equally divided into groups of normal cognitively (NC) subjects, subjects with early mild cognitive impairment (EMCI) and subjects with Alzheimer’s Disease (AD). All images are combined into three binary datasets: EMCI vs NC, AD vs NC and AD vs EMCI. Each dataset is composed of two-dimensional images (planes) represented in vertical (frontal) and horizontal (axial) directions.

Statistical features collected from segmented image asymmetries were enriched with Bag-of-Features to get the most detailed image “signatures”. These were used to feed NB, LD, SVM, and KNN classifiers. The performance of trained models was estimated using ten simulation runs of a 10-fold cross-validation procedure. In comparison images of brain asymmetry were inputted directly into a pretrained CNN, the so called AlexNet, and features were generated automatically as a part of the training process of the network. For the deep network, images of segmented asymmetry were resized to 227×227×3 and fed into the model with 80% used for training, 10% for validation, and 10% for testing. The parameters of the CNN were set for 10 epochs, validation data frequency of 50 and mini-batch size of 128. All the machine learning models were tested on unseen data, i.e. unknown images that were not used for training and/or validation.

The performance of the classification models was evaluated in terms of accuracy, sensitivity, specificity, and AUC. Table 13 summarizes the average performance (%) in testing the ADNI database's early mild cognitive impairment, normal cognitively, Alzheimer’s disease datasets. The highest results for each dataset are shown in bold.

Table 13: Average performance (%) of binary classifiers

Datasets	**NB**	**LD**	**L-SVM**	**Q-SVM**	**C-SVM**	**MG-SVM**	**Fine-KNN**	**Cos-KNN**	**CNN**
EMCI vs NC									
Accuracy	77.0	91.0	89.0	**92.5**	**92.5**	88.0	83.0	92.0	75.0
Sensitivity	78.0	91.0	89.0	92.0	95.0	85.0	99.0	96.0	90.0
Specificity	76.0	91.0	89.0	93.0	90.0	91.0	67.0	88.0	60.0
AD vs NC									
Accuracy	78.5	90.0	92.0	92.5	**93.0**	90.0	86.5	89.5	90.0
Sensitivity	78.0	88.0	91.0	90.0	93.0	85.0	98.0	90.0	89.0
Specificity	79.0	92.0	93.0	95.0	93.0	95.0	75.0	89.0	92.0
AD vs EMCI									
Accuracy	78.5	83.0	80.5	**86.5**	**86.5**	80.5	79.0	80.0	81.3
Sensitivity	75.0	85.0	84.0	89.0	88.0	84.0	78.0	83.0	72.5
Specificity	81.0	81.0	77.0	84.0	85.0	77.0	80.0	78.0	90.0

The test results show that features extracted from asymmetries provide consistent performance across different classification models without model-specific fine-tuning of hyperparameters. The SVM variants and the LD method can become methods of choice. They can be easily trained on commodity hardware and demonstrate better accuracy than other alternatives in all cases. SVMs show the best performance amongst the SVM variants with the polynomial cubic and quadratic kernel (C-SVM and Q-SVM). C-SVM accuracy of EMCI vs NC is 92.5%, sensitivity is 95.0%, specificity is 90.0%; accuracy of AD vs NC is 93.0%, sensitivity is 93.0%, specificity is 93.0%; accuracy of AD vs EMCI is 86.5%, sensitivity is 88.0% and specificity is 85.0%. Q-SVM accuracy of EMCI vs NC is 92.5%, sensitivity is 92.0%, specificity is 93.0%; accuracy of AD vs NC is 92.5%, sensitivity is 90.0%, specificity is 95.0%; accuracy of AD vs EMCI is 86.5%, sensitivity is 89.0% and specificity is 84.0%. The prediction results of the CNN are in the same range as those of the other classifiers.

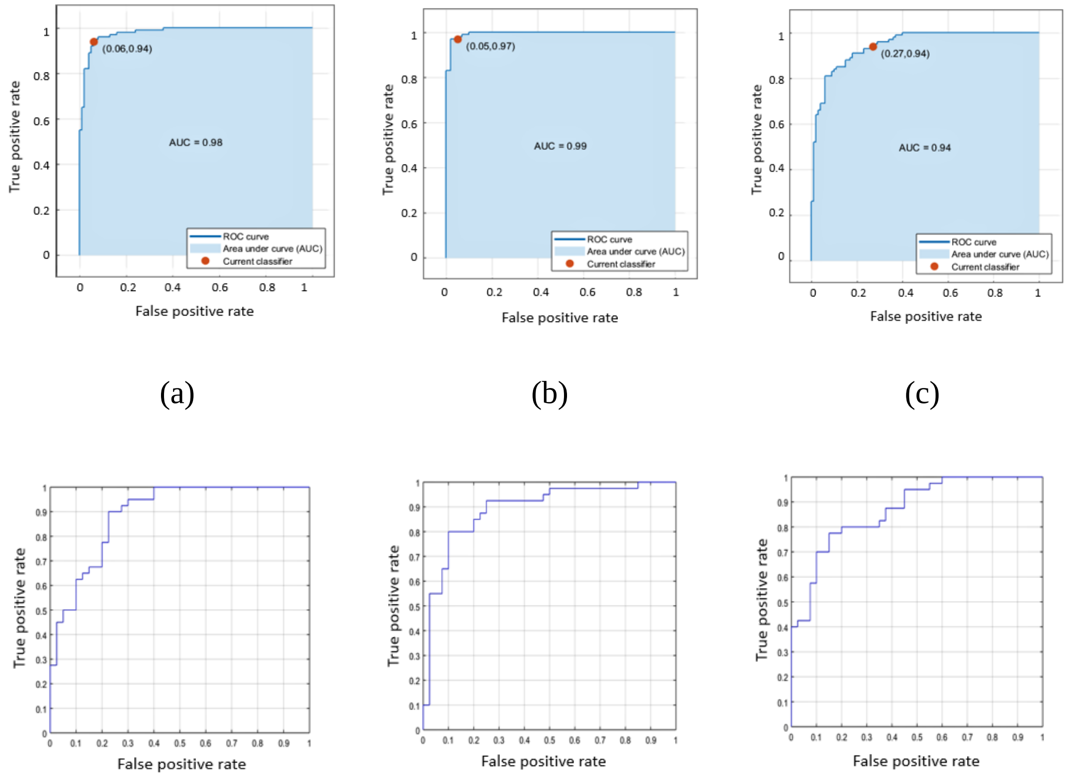
An aggregated measure, the area under the ROC Curve (AUC), shows the relationship between data sensitivity and specificity across different levels of threshold, giving an additional view on the classifier performance. AUC results of the best available models from C-SVM and CNN are presented in Table 14.

Table 14: AUC for cubic-SVM and CNN

Datasets	**C-SVM**	**CNN**

EMCI vs NC	0.98	0.90
AD vs NC	0.99	0.92
AD vs EMCI	0.94	0.88

Figure 29 presents the C-SVM and CNN classifiers' AUC/ROC curves of three binary datasets. All the figures were created on Matlab during testing.



(a) (b) (c)

(d) (e) (f)

Figure 29: AUC/ROC curves of C-SVM (left) and CNN (right) classifiers: (a) and (b) is EMCI vs NC; (c) and (d) is AD vs NC; (e) and (f) is AD vs EMCI

In general, satisfactory performance was obtained in the classification between EMCI and NC, AD and NC, and AD and EMCI. It is worth noting an important difference between the CNN classifier and all other methods in this context The CNN is the only model that operates directly on images of segmented asymmetry, whilst all other models work on images with segmented asymmetry that have been analysed to generate statistical and Bag-of-Features. It is expected that fine-tuning and model optimization could potentially further improve all classification models' performance but operating directly on images adds value to the deep learning approach in order to produce an end-to-end solution. The potential of deep networks will be further investigated in the next section.

5.6 Deep learning approaches used in the thesis

5.6.1 Deep Networks and Learning Algorithms Preliminaries

Deep neural networks allow the processing of a large amount of data. These networks have multiple implications for medical imaging data.

Tensor and tensor operations

Tensor in Deep Learning is a way of data structuring used by machine learning algorithms. Also, a tensor can be considered to be a container for the data and numbers. The tensor might have a different number of dimensions (D). Zero dimensional tensors contain only one number. The array of numbers represents a one-dimensional tensor. For instance, [13, 15, 4, 25, 8] is a 1D tensor of the five-dimensional vector. Two-dimensional tensors are matrices of array vectors with two axes, referred to as rows and columns. Three and higher dimensional tensors are matrices with a number of arrays. Tensor's dimensionality can denote the number of entries along a specific axis or the number of axes. In the first case, it can be a 5D vector. When the number of axes equals 5, it can point to the 5D tensor. It might cause confusion between vectors and tensors. It is correct to speak about the tensor's rank instead of mentioning the tensor dimensions. The tensor rank will point to the number of axes in this situation. There are three main key attributes that identify a tensor. They are the number of axes or rank, shape and data type. The shape is described by the number of dimensions along the tensor's axes. For instance, a scalar does not have a proper shape, and its shape is denoted as an empty parenthesis (). A vector shape contains a single element that can be a number (n). A two-dimensional matrix has a shape with rows and columns (m, n). Three elements (m, n, k) represent a 3D tensor. The different data dimensionalities can also represent the real-world tensors. Vectors with samples and features show a 2D tensor. Sequence or time-series data is a 3D tensor with samples, timesteps and features in the axes. 4D shaped tensors have images with axes of samples, channels, height and width. Videos that are 5D tensors include samples, frames, channels, height and width details in their shapes.

Many operations applied to tensors' numeric data can describe all deep neural network transformations. Mathematically, tensors can be added, subtracted, multiplied and etc. The results of these operations are the dot or tensor products. So, the product of two vectors is a scalar, the product of a vector and matrix is a vector, and the product of two matrices is a matrix. All tensor operations are supported by geometrical interpretation. For example, if there are two vectors a and b with coordinates of point A= [0.5, 1] and point B = [1, 0.25], the resultant vector c will represent the sum of two previous vectors (Figure 30).

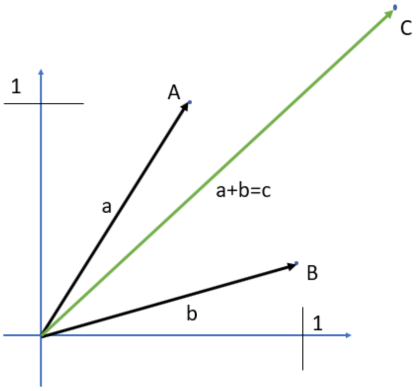


Figure 30: Geometric interpretation of the dot product of two vectors

Elementary geometric operations such as affine transformations, including translation, scaling, reflection, rotation, etc., can be represented by tensor operations. For instance, a rotation of a 2D vector by an angle theta can be calculated through a dot product with a 2 × 2 matrix M = [a, b], where a and b are both vectors of the plane: a = [cos(theta), sin(theta)] and b = [-sin(theta), cos(theta)].

Gradient-based optimization for network training

The different tensor transformations can be applied to the layers of the deep neural networks. Each layer transforms the input data in the following way:

$$output = layer\ name(dot(W, input) + b) \tag{40}$$

where W and b are tensor weights and trainable parameters accordingly.

The weights maintain the data details learned by the neural network. Initial weights can be random values obtained from the observed matrices. Each following stage gradually adjusts the weights according to the feedback signals. This process is called training. During the updating process, the values of the weights' coefficients increase or decrease. The coefficients must be updated to decrease the learning error or network loss.

All network operations are differentiable, and the gradient of loss can be calculated on the base of the network coefficients. Differentiation is a process of computing the function's derivatives (Treiman, 2014). The derivative of a function $y = f(x)$ can be explained geometrically as a slope of the graph of the function f; in other words, the slope of the tangent line in the point p. A function's minimum is a point where the derivative equals 0. Finding the combination of the weight values that give the minimal loss function is a critical point in the training of the neural network. This process is called Stochastic Gradient Descent (SGD).

The stages of the optimization learning process are described below. Stage 1 requires choosing a batch of training samples x and corresponding target values y. Stage 2 runs the network on x values and gets the initial predictions. Stage 3 calculates the loss of the network for the batch that demonstrate the mismatch between the targets y and obtained predictions. Stage 4 calculates the gradient of the loss concerning the parameters of the network. In this stage, a backwards pass approach can be applied. The last Stage 5 reduces the loss on the batch by a slight movement of the parameters in the opposite direction from the calculated gradient. The mathematical expression of this action is described below:

$$W \mathrel{-}= \text{step} * \text{gradient} \tag{41}$$

Choosing a reasonable value for the step factor or learning rate is essential. If the value is too small, the gradient descent will take many iterations down the function's curve. It could force the process to get stuck in the "local minimum" of the curve. If the value of the step is too large, the weight updates may end in the random place of that curve. Graphical representation of the algorithm is provided in Figure 31.

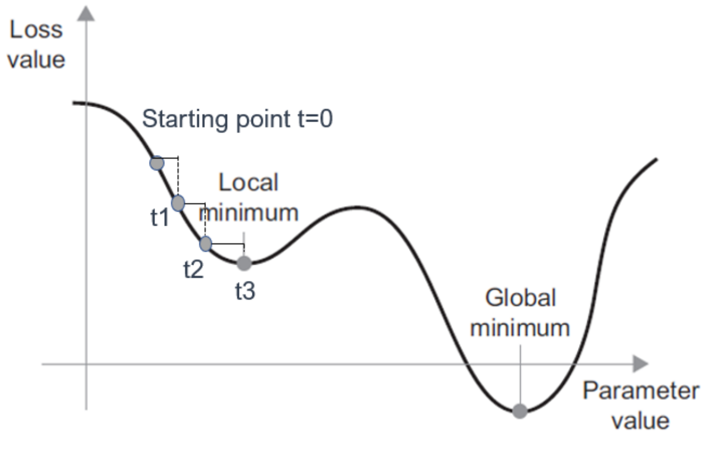


Figure 31: Example of SGD learn down a loss curve of one learnable parameter

Usually, a mini-batch SGD algorithm that processes the data separated into small batches is preferable in the learning process. If the algorithm runs on all available data, it is called “batch SGD”. Each algorithm's update demonstrates a higher accuracy than the mini-batch SGD, but it is more expensive. A reasonable size of mini-batch SGD is an efficient way to balance the accuracy and learning cost.

In real life, the weight coefficient can be presented in n-dimensional space. The number of dimensions can reach tens of thousands and even more. A number of SGD variants differ by computing the previous and the following weight updates rather than looking for the current gradient values. These algorithms are called “optimizers” or “optimization methods.” The representatives of these methods are Momentum, Adagrad, RMSProp, Adam (Ruder, 2016). “Momentum”, for example, addresses two SGD problems: local minima and speed of convergence. From the Figure 31 , it is clear that the minimal loss function should be in the “global minimum.” The movements to the left will increase the loss value, and the change in the direction to the right will decrease it. A low SGD learning rate will end the process in the “local minimum” instead of ending it in the “global minimum”. “Momentum” optimization is one of the ways to reach the global minimum. The development of this technique was inspired by the “moving ball” tracking. Each step of weights updating is based not only on the current acceleration or slope value but also on the previous acceleration or current velocity results.

Backpropagation algorithm

Any neural network joins together many tensor operations. These operations follow specific “chain rules”. One is known as “backpropagation” or “reverse-mode differentiation” (Rojas, 1996). The backpropagation usually starts from the loss value of the output layer. The loss values are obtained by calculating the differences between the expected outputs and forward propagated outputs of the network. The algorithm computes the contribution of each network parameter in the loss value. Then, the algorithm propagates backwards from the output layer towards the hidden layer by assigning error changes and updating weights. Thus, chain operations with known derivatives compute a gradient function for the chain. If there are functions for a single variable $y = f(u)$ and $u = g(x)$, and they are differentiable, the chain rule looks as follows:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} \tag{42}$$

For the function y with multiple variables u_1, $u_2, \ldots, u_m$, differentiable function u_i with variables $x_1, x_2, \ldots, x_n$ and $i = 1, 2, \ldots, n$, the chain rules gives the following equation:

$$\frac{dy}{dx_i} = \frac{dy}{du_1}\frac{du_1}{dx_i} + \frac{dy}{du_2}\frac{du_2}{dx_i} + \cdots + \frac{dy}{du_m}\frac{du_m}{dx_i} \tag{43}$$

Modern network frameworks, such as Tensorflow, operate with symbolic differentiation. They are able to compute the gradient function and map network parameters to gradient values automatically using the backpropagation algorithm.

5.6.2 Deep Learning Algorithms for Image Classification

Deep learning has a structure that is inspired by the neuronal network of the human brain. Each layer is a fundamental network structure. The deep learning network consists of several layers, including input, output, and hidden layers. Input and output layers are tensors consisting of layers' weights. The weights are learned and transformed by optimization algorithms. All layers are chained together. They map the input data to make predictions. The input layer is the network's first layer that accepts the external data and forwards it to the processing stage supported by the hidden layer. Hidden layers can be single or multiple. This type of layer proceeds the main data transformation and analysis. The output layer collects the processed information and makes a prediction.

The deep learning models might have a different topology. The most common instance is composed of a linear stack of layers. But there are many wider varieties of network architectures, such as multi-branch networks, graph networks, and inception blocks models.

Activation functions

The activation function is a crucial part of designing a deep neural network. This function concludes what neurons can be activated by calculating the weighted sum and bias. The activation function uses differentiable operators to transform input layer signals into outputs. There are a number of popular activation functions proposed for deep neural networks, such as Rectified linear unit (ReLU), sigmoid, hyperbolic tangent (tanh), and Softmax.

ReLU demonstrates a straightforward nonlinear transformation and helps archive the neuronal network a good performance (see Figure 32). For a given element *x*, the function can be expressed as the maximum of that element *x* and 0:

$$ReLU(x) = \max\,(x, 0) \tag{44}$$

Thus, the ReLU function reserves only positive values and discards all negative ones. All negative values are converted to 0.

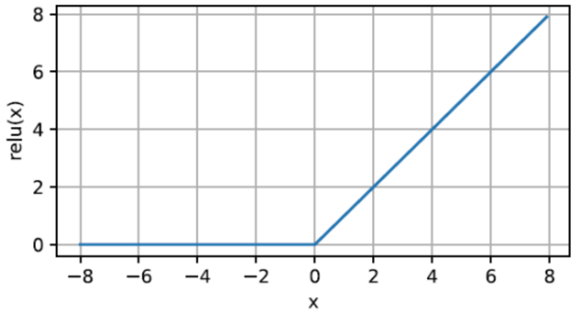


Figure 32: The graphical representation of the ReLU function

There are many advanced variants of the ReLU function. For instance, Leaky ReLU defines the negative values as a significantly small linear component of *x* instead of zero values. Parameterised ReLU introduces a new parameter as a slope of the negative part of the function. Exponential Linear Unit (ELU) modifies the slope of the negative part of the function using a log curve for the negative values instead of a straight line.

The ***sigmoid*** activation function converts input values to output values between 0 and 1. The sigmoid function is known as a “squashing function”, converting any input values of the range minus infinitive to plus infinitive to the value in the field between 0 and 1:

$$sigmoid(x) = \frac{1}{1+\exp\,(-x)} \tag{45}$$

The visualization of the function is provided below in Figure 33:

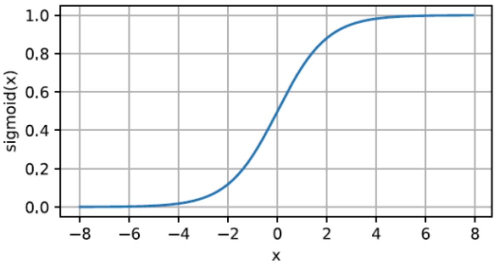


Figure 33: The graphical representation of the sigmoid function

Sigmoid functions are usually used in the outputs when the outputs have to be interpreted as probabilities for binary classification problems. The sigmoid function approaches a linear transformation when the inputs are closed to zero values. In many cases, the ReLU often replaces the sigmoid in hidden layers due to its simple and easily trainable parameters.

A combination of multiple sigmoids can represent the ***softmax*** function:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{46}$$

Where σ is a softmax function, $\vec{z}$ is an input vector, e^{z_i} is a standard exponential function for the input vector, K is a number of classes in the multiclass classifier, e^{z_j} is a standard exponential function for the output vector.

Like a sigmoid function that returns values between 0 and 1and is treated as a probabilistic function, the softmax function also represents probabilities of the targeted classes. The difference is that sigmoid is used for binary classification problems compared to the softmax that builds a network for multiclass problems. The output layer of the softmax has as many neurons as the number of targeted classes. If the outputs from neurons have the values of [1.2, 0.9, 0.75] the softmax function will transform these values to the following probabilities: [0.42, 0.31, 0.27]. The sum of all values always has to be equal to 1.

Model regularization and discovering the generalization patterns

The group of factors that impact the generalizability of a model can be divided into a few groups. The first group is a number of tunable parameters called “degrees of freedom”. If these parameters are large, the model becomes sensible to overfitting. ***Overfitting*** is referred to the model that maintains more parameters than can be justified by the data. The opposite process is called “***underfitting***” when the model cannot adequately collect the underlying data structure. Both overfitting and underfitting processes lead to poor predictive model performance on unseen data. The second factor that influences the generalizability of the model is the range of values taken by parameters. The wide range makes the model more sensitive to overfitting. One more important group affecting the prediction is the number of training examples. A small number is correlated with a high probability to overfit the model. If the dataset has millions of samples, it might require implementing an extremely flexible model.

The over- and underfitting processes can be identified by measuring the validation metrics that include the model accuracy and loss. The validation error increases while the training error progressively decreases if the model is overfitted. Underfitted process is accompanied by a gradual training and validation error decrease. One more error type is a generalization error. The generalization error is the expected error of the model and can only be estimated. It explores the infinitive stream of imaginary data.

The overfitting process can be prevented by the extension of the dataset, data augmentation (archived by the extension of the already existing dataset with slightly modified copies), and ensembling of the predictions from several trained separately models. One more approach to decrease the overfitting is to reduce the model complexity that can be reached by reducing the number of neural network parameters. For instance, the dropout technique is one of the options to do it.

Dropout

Dropout is an efficient way to prevent the neural network from overfitting by applying the regularization technique. The proposed approach approximates the model training and reduces the network's capacity. The Dropout layer randomly sets some inputs to 0 with each update of the training circle. All other inputs are scaled up to 1 such that the sum of all inputs remains the same.

The dropout technique can be used with most types of neural network layers, such as convolutional layers, a long short-term memory layer, recurrent layers, and fully connected layers. The dropout can be applied to the input layer in some situations, but it is never used with the output layer. The most advanced dropout technique specifies the probability at which parameters of the output layer performs the dropout procedure. A standard threshold value for the retaining output is a probability of 0.5 for each hidden node of the layer. The network retains all values above this level. Dropout probability p can be explained with the following formula, where each intermediate activation of the hidden node h is replaced by the random variable h’:

$$H' = \begin{cases} 0 & with\, probability\, p \\ \frac{H}{1-p} & otherwise \end{cases} \tag{47}$$

After the dropout procedure, the network weights will become larger than before. Therefore, the weights are usually scaled by the chosen rate before finalizing the network. Then, the model can be saved and able to make a prediction.

Convolutional Neural Network

Convolutional Neural Network was specially designed for imaging data. A two-dimensional grid of pixels represents each image. Each pixel value and location might be associated with numerical values depending on the black-and-white, greyscale or colour images. While processing imaging data, a neural network architecture must follow the relevant requirements. The first of these requirements is translational invariance. It means that the network layers should respond similarly to the same area regardless of where it appears in the image. The second requirement is based on the principle of "locality", when the earliest layer concentrates mainly on the local regions. The local area representations can aggregate the knowledge about the whole image.

Convolution functions

Image processing with multilayer perceptron requires image conversion to the one-dimensional vector. This conversion impacts the relationship between the pixel and makes the classifier less effective in image processing. In addition, hundreds of thousands of pixels have to process billions of parameters. It takes a lot of computer power and exceeds the training time. The CNN classifier does not have such a problem. It needs fewer parameters than fully-connected networks. Additional advantages of CNN include its possibility to work well with one-dimensional audio, text structures, time-series sequences and graph-structured data.

The main configurational element of CNN is a convolutional layer that operates using a convolutional function. The convolutions between two functions measuring overlap between f and g can be defined as:

$$(f * g)(X) = \int f(z)g(X - z)dz, \tag{48}$$

when one of the functions is flipped and shifted by the distance x.

For discrete objects such as a set of infinite-dimensional vectors, the formula looks as follows:

$$(f * g)(i) = \sum_a f(a)g(i - a) \tag{49}$$

For a 2D tensor, a corresponding sum with indices (a, b) will look as follows:

$$(f * g)(i, j) = \sum_a \sum_b f(a, b) g(i - a, j - b) \tag{50}$$

The image consists of three channels, red, green, and blue in real life. It is possible to say that image is a two-dimensional object and a third-order tensor simultaneously. This type of tensor is characterized by a height, width, and number of channels and has a shape, for instance, of 512×512×3 pixels.

Convolution calculations

The cross-correlation operations are at the base of convolutions. It means that input and output tensors are obtained due to the cross-correlation calculations. An example of a two-dimensional process is provided in Figure 34. The tensor's shape is 3 × 3, and the size of the kernel (convolution filter) is 2 × 2 in the current example.

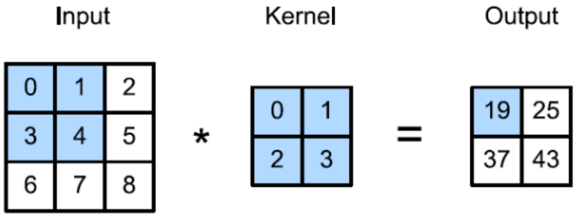


Figure 34: Two-dimensional tensor operations (Source: Zhang et al., 2021)

In the two-dimensional space, the convolution starts at the upper-left corner of the input image (tensor) and slide across it from left to the right and from top to the bottom. With each slide of the convolution filter, the input subimage (subtensor) of the same window size and the filter are multiplied according to the matrix operation rule. The resulting calculations give the output image (tensor) at the corresponding location. In the example, the obtained 2-by-2 output contains four elements received from the two-dimensional cross-correlation operation:

$$0 \times 0 + 1 \times 1 + 3 \times 2 + 4 \times 3 = 19,$$

$$1 \times 0 + 2 \times 1 + 4 \times 2 + 5 \times 3 = 25,$$

$$3 \times 0 + 4 \times 1 + 6 \times 2 + 7 \times 3 = 37,$$

$$4 \times 0 + 5 \times 1 + 7 \times 2 + 8 \times 3 = 43.$$

The output size is calculated as a difference between the input size of $ih \times iw$ and the size of the filter of $fh \times fw$:

$$(ih - fh + 1) \times (iw - fw + 1) \tag{51}$$

To process the whole image, the filter should be moved across it. The number of pixels shifted across the image matrix is called "stride". So, the stride is equal to 1 if the filter moves by one pixel, the stride is equal to 2 if it shifts by two pixels and so on. The final output image of the layer received due to the convolutional operation is called the "image map".

The output size is usually slightly smaller than the input size along each axis. It can be explained by the fact that the kernel's width and height are greater than one, and the cross-correlation operation for the given location can be done only when the filter entirely fits within the image. It leads to some information loss from the image boundaries. An additional technique called "image padding" can be applied to solve this problem. The extra pixel values are added to the image boundaries. Usually, the values of extra pixels are set to zero. In the following Figure 35, the image size increases from 3-by-3 to 5-by-5 pixels after adding the padding. The output image will have a size of 4-by-4 pixels after filtering.

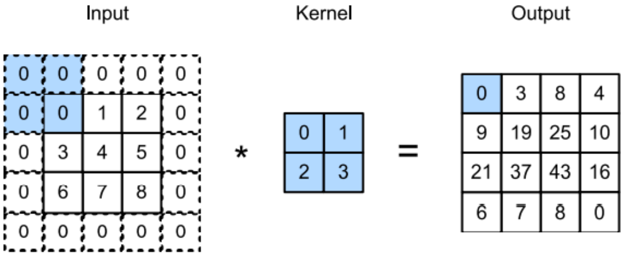


Figure 35: Two-dimensional tensor operations with added padding (Source: Zhang et al., 2021)

When the input image has multiple channels, each colour channel is treated separately with the same cross-correlation rule. The output layer summarizes the results obtained from the channels. Figure 36 demonstrates a two-dimensional cross-correlation operation with two channels. The output image is obtained due to the following computations: (1×1+2×2+4×3+5×4)+(0×0+1×1+3×2+4×3) = 56.

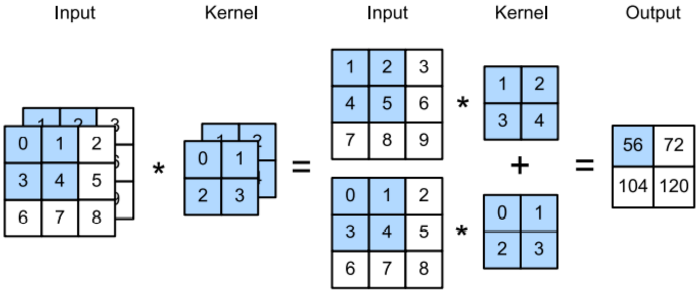


Figure 36: Two-dimensional tensor operations using two channels (Source: Zhang et al., 2021)

The image map remains large from the convolutional operations, and additional feature reduction operations are preferable to be applied. Another technique that gradually reduces the spatial resolution is called "pooling".

Pooling

Pooling layers help to save a global image representation by keeping all advantages of the convolutional layers and other intermediate layers. At the same time, the pooling procedure makes the image size significantly smaller and might reduce the overfitting of the entire neural network.

The pooling operations are performed using a fixed-shaped "pooling window" that moves around the input regions according to the size of the stride. This operation is performed the same as a convolution and begins from the upper-left corner sliding across the input image from left to right, and top to bottom, then stops in the lower right corner. However, unlike the convolutional calculations based on the kernel details, the pooling window does not include any parameters except the size of the window.

The pooling operators are deterministic and usually compute average or maximum values. Respectively, pooling operations are called "averaging pooling" or "max pooling". The pooling operations calculate the average or maximum values at each layer location depending on the pooling function employed. Figure 37 gives the example of the pooling operation.

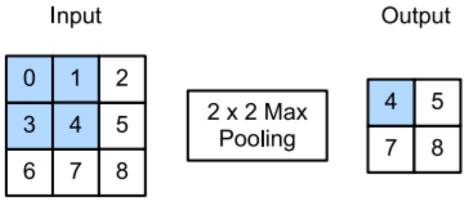


Figure 37: 2D matrix pooling operations

The output is calculated as follows:

max(0, 1, 2, 3)=4,

max(1, 2, 4, 5)=5,

max(3, 4, 6, 7)=7,

max(4, 5, 7, 8)=8.

In this way, the pooling layer significantly reduces the layer size keeping the most significant spatial layer information.

Pretrained convolutional neural network

There are a number of popular CNN architectures that became base models for many research projects. Many of them are winners in the large-scale ImageNet competition, launched in 2010. They demonstrate annual progress in developing news supervised learning models in the area of computer vision. The performance of the pretrained models varies depending on the architecture and choices of the hyperparameters. Pretrained models propose many interesting configurations that can be divided into several groups according to the similarity in their architecture. One of the first successful models is the large-scale AlexNet network. Another group is VGG networks, which were created from the repeating blocks of structural elements. GoogLeNet blocks differ from the previous architectures by concatenation blocks in parallel. Network in networks (NiN) is based on small patch-wised convolutions. Residual networks (ResNet) consist of different numbers of residual blocks and channels. Densely connected networks (DenseNet) have an advanced configuration of the ResNet.

AlexNet

AlexNet includes the first name of his author Alex Krizhevsky from the deep learning research group, which won the ImageNet competition in 2012 (Krizhevsky et al., 2012).

ImageNet competitors trained models on one million images of one thousand object categories. Their idea was based on the complex feature structure in which the hierarchically joint layers can detect learnable parameters. The lowest layers of such architecture are supposed to detect edges, textures and colours resembling the traditional image filters. The hidden layers learn a compact representation of the image whose property can be easily separated into the different data categories.

AlexNet requires the input image size of 224 × 224. The network incorporates the image augmentation approach, such as image clipping, flipping, and usage of colour channels. It makes the model more robust, with many images that successfully reduce overfitting. The AlexNet consists of eight layers, including five convolutional, two fully-connected hidden, and one fully-connected output layer. The network does not require manually-designed features. All the feature detection and extraction procedures are done automatically. The first convolution filter (window) has a size of 11 × 11. It gives the possibility to capture rather big objects. The second and third convolutional layers have reduced filter sizes of 5×5 and 3×3, respectively.

Furthermore, the network has maximum pooling layers inserted after the first, second, and fifth convolutional layers. All max-pooling layers have a window size of 3×3 and slide through the layers with a stride of 2. AlexNet uses a simple and effective ReLU activation function. The network is finalized by two fully-connected layers of 4096 output parameters (8192 parameters were divided between dual GPUs). The model complexity was controlled by adding the dropout function to the fully-connected layer.

VGG

The repeated block structure characterises VGG (Simonyan and Zisserman, 2014). The VGG structure can be divided into two parts integrated convolutional blocks and several fully-connected layers. Each building block includes a sequence of convolutional layers with the kernel of size 3×3 and padding of 1 pixel, a max-pooling layer of size 2×2 and a stride of 2 pixels. The original VGG network (VGG-11) has five blocks of convolutions and 11 layers overall. The first two blocks of the network have one convolutional layer each, and the following three blocks include two convolutional layers each. The number of output channels of the first block is 64. This number doubles with each successive block and reaches 512 in the final one. Similar to AlexNet, VGG uses the ReLU activation function. Output parameters on the fully-connected layer are equal to 4096, 4096 and 100 accordingly. Compared to the AlexNet, the VGG-16 is computationally heavier. The VGG network has several modifications with exceeded number of convolutional layers, VGG-16 and VGG-19.

5.7 Classification workflow based on CNNs

Deep learning models demonstrate exemplary performance in image classification tasks. The diagnostic deep learning workflow includes two classification approaches. The first approach focuses on exploring the potential of transfer learning CNN models such as AlexNet and VGG16. The second diagnostic line examines the proposed based CNN architecture that investigates its performance by changing the type of classification layer. The segmented brain asymmetry images were used in the input for both classification approaches. Features were generated as part of the training process for Convolutional Neural Networks.

In the following Sections, 5.6, 5.7, 5.8 and 5.9, detailed descriptions of DL models used in the research are provided.

5.8 Transfer learning models

Pretrained CNN architectures, like the AlexNet, the VGG, the ResNet variants and architectures based on the Inception models, have been established as a standard pretrained model for deep learning applications in image processing and classification. Pretrained models often perform better than models trained from scratch, and their use significantly reduces training, validation and testing time on new tasks (Tajbakhsh et al., 2016). Transfer learning enables transferring the pretrained model’s knowledge to a newly established task and then creates a new architecture by adding layers to the pretrained/base model. Training the extra layers, retraining the base model, or even training the whole architecture are some of the strategies used to fine-tune the transfer learning architecture using task-related data (Tang, 2013).

Two pretrained CNNs, AlexNet and VGG16, have been chosen as the base models of the system for diagnosing early dementia in binary and multiclass classification setups. They demonstrated a good performance in image classification tasks (see Chapter 2). AlexNet and VGG16 were initially configured and trained for 1000 classes using ImageNet data. The AlexNet architecture consists of 8 layers, has a size of 227MB and includes 61.0 million parameters. The input dimensionality is 224×224×3 (images should be 224 wide, 224 high, and has three colour channels). The VGG16 architecture is much bigger and has a size of 515MB. This network has 16 layers and 138.0 million parameters, and the input dimensionality is 224×224×3 either.

Pretrained models were adapted to the new classification problem. This was done by replacing the last three layers of the pretrained networks with fully connected, Softmax and output layers. Figure 38 demonstrates the architecture of transfer learning models used in the proposed framework.

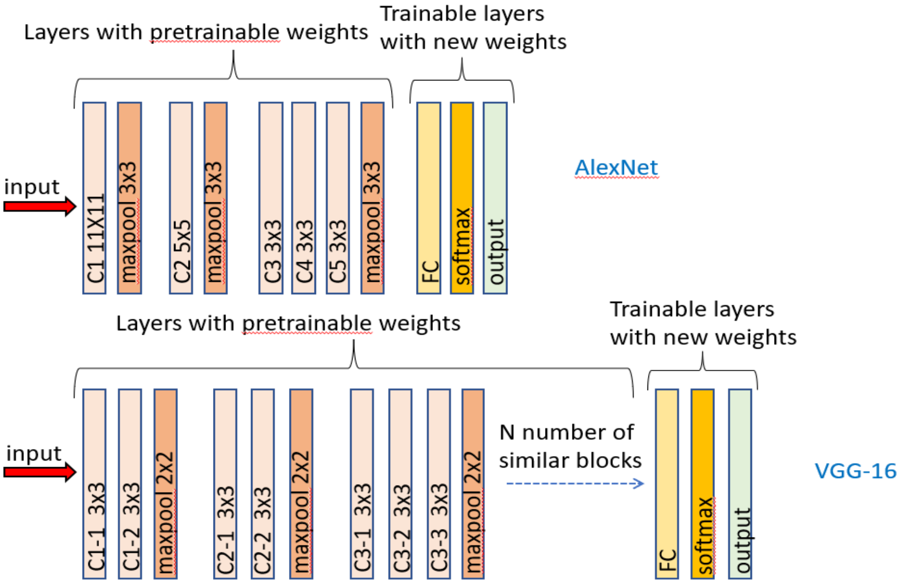


Figure 38: Transfer learning of AlexNet and VGG-16

In addition, transfer learning of AlexNet was used for the building of the two classification models. The first model integrates the Softmax layer to compute the output, while the second one makes a class prediction based on SVM. The fine-tuning models' strategy was based on retraining the whole CNN architecture. The SVM layer was assembled as a separate classification module.

The classification performance obtained due to transfer learning structure is provided in Sections 5.9 and 5.11.

5.9 Experiments with transfer learning architectures

This section exploits the AlexNet CNN classifier by fine-tuning the model's parameters and investigated the potential of transfer learning using an AlexNet core architecture in two configurations: one with Softmax layer and one with an SVM classifier (Herzog and Magoulas, 2022).

A set of five MRI classification problems, four binary and one multiclass, are studied using the two transfer learning AlexNet architectures. Binary classification datasets include 400 images of segmented asymmetries in the ADNI database and 200 images in the OASIS database. The multiclass dataset consists of 600 images from the ADNI database. In all experiments, MRI data were balanced across classes: 600 images of brain asymmetries with a balanced number of AD, EMCI, and NC subjects were combined into the multiclass set and three binary datasets, EMCI vs NC, AD vs NC, and AD vs EMCI. Two hundred images of the OASIS database compose the balanced binary set of non-D vs VM-D subjects. For each task, five independent simulation runs of each architecture were performed, increasing the number of training epochs from 5 to 40 (a total of 300 independent simulation runs). In the rest of the section, the first transfer learning architecture is denoted by M1-Softmax (SoftMax layers), while the second is denoted by M2-SVM (SVM layer). The training parameters were set as: n (n = 5, 10, 15, 20, 30, 40) is the number of epochs each architecture was trained (five independent runs conducted in each case), mini-batch size= 128, validation data frequency= 50, initial learning rate= 0.0001, and the stochastic gradient descent with momentum (SGDM) was used. The two architectures were retrained using segmented asymmetry images, whilst the final SVM layer was fine-tuned after the rest of the architecture was trained. The images of segmented asymmetry of size 256-by-256-by-3 were resized to 224×224×3 and fed into the model with 80% of the images used for training, 10% for validation, and 10% for testing. All the test results presented in this section are for unseen MRI data (unknown patients).

5.9.1 Experiment 1: ADNI repository

Three binary problems (AD vs EMCI, EMCI vs NC, AD vs NC) and one multiclass problem were used in this experiment. The best available classification results on unseen data for the two architectures are presented in Table 15.

Table 15: The best available accuracy (%) in testing for the two transfer learning architectures and the number of epochs each architecture was trained (independent runs) in the classification tasks

Epochs trained	AD vs NC	EMCI vs NC	AD vs EMCI	Multiclass (AD, EMCI, NC)

	M1-Softmax	M2-SVM	M1-Softmax	M2-SVM	M1-Softmax	M2-SVM	M1-Softmax	M2-SVM
5	76.50	91.50	80.25	81.90	79.00	81.25	62.33	73.00
10	94.00	95.25	81.68	82.75	79.00	81.25	72.22	73.75
15	94.25	95.90	81.00	85.25	81.25	81.00	72.22	74.22
20	93.00	93.58	85.50	86.94	81.50	81.25	79.78	79.78
30	92.75	93.33	90.25	90.25	82.75	83.00	82.33	81.75
40	92.70	92.25	88.98	88.98	83.00	83.00	84.75	84.50

Table 15 demonstrates that the SVM layer (M2-SVM) architecture is capable of good performance within a few training epochs. M1-Softmax appears to take advantage of longer training sessions to compensate for its initial shortcomings. The best performance for the AD vs NC task is 95.90%, obtained by an M2-SVM that was trained for 15 epochs; for the EMCI vs NC task, the best performance is 90.25% with an M1-Softmax and M2-SVM trained for 30 epochs; for the AD vs EMCI task, 83% is achieved by an M2-SVM trained for 30 epochs, an M2-SVM trained for 40 epochs and an M1-Softmax trained for 40 epochs. The multiclass task's performance is 84.75%, with an M1-Softmax trained for 40 epochs.

Looking at the average in testing across all tasks, M2-SVM outperforms M1-Softmax by 8.10% (on average), when trained for five epochs, by 1.93% (on average) when trained for 10 epochs, and by 1.31% (on average) trained for 15 epochs. The differences in average test performance per task are analysed in Table 16.

Table 16: The difference in the average classification accuracy (%) between M2-SVM and M1-Softmax depending on the length of the training session (epochs)

Number of epochs trained	AD vs NC	EMCI vs NC	AD vs EMCI	Multiclass
5	10.83	5.81	5.42	10.28
10	4.17	1.25	1.91	0.28
15	1.34	1.41	2.50	~ 0

The two architectures show similar or very close results in testing when they are trained for more than 20 epochs. For example, Figure 39 illustrates the average classification performance (mean classification error rate on unseen patients' MRI data) for the two architectures depending on how long they were trained (epochs) to solve the EMCI vs NC binary classification task.

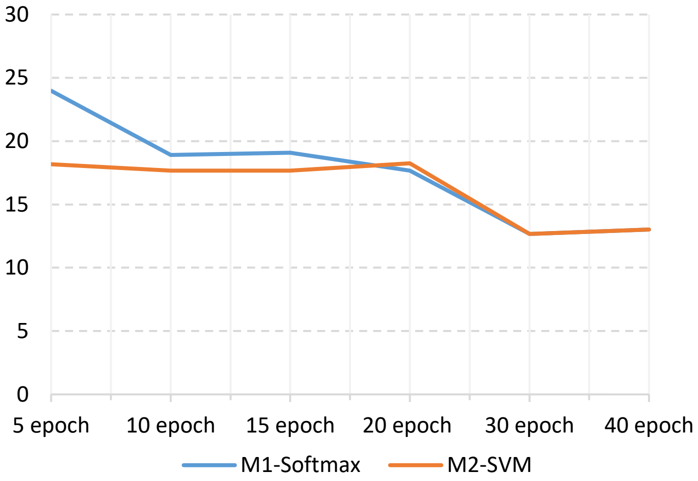


Figure 39: Average test error rate (vertical axis) in the EMCI vs NC binary classification problem with respect to the number of epochs the two architectures were trained for

The Area under the ROC Curve across all runs in the ADNI experiments (binary and multiclass) is in the range of 0.8338 and 0.9763. The best performance received with both transfer learning architectures for diagnosing early dementia (EMCI vs NC) in terms of accuracy is 90.25%, with an AUC of 0.9256 when trained for 30 epochs. Progressive dementia (Alzheimer's Disease) – (AD vs NC) is diagnosed with an accuracy of 95.90% and an AUC of 0.9763 with M2-SVM trained for 15 epochs.

5.9.2 Experiment 2: OASIS database

This database was used as a second source of MRI data to validate the behaviour of two transfer learning architectures and the findings of Experiment 1. The OASIS-2 MRI collection is not organized like the ADNI database, so there are no EMCI and AD images to perform EMCI vs NC or AD vs NC binary classification. The dataset is separated into classes of non-Demented (non-D) patients and patients with Very Mild dementia (VM-D) manually, according to the Clinical Dementia Rating (CDR) score equal to 0.5. The best classification results on unknown patient data are in Table 17.

Table 17: The best available classification accuracy (%) for the two transfer learning architectures in binary classification (non-D vs VM-D)

Epoch number	5	10	15	20	30	40
M1-Softmax	61.50	78.25	71.50	76.00	80.00	79.00
M2-SVM	76.50	80.00	82.50	75.00	80.00	78.50

Table 17 confirms the fast response of the M2-SVM architecture producing better classification accuracy within a few training epochs. M2-SVM outperforms M1-Softmax by 15% when trained for five epochs, 1.75% for ten epochs, and 11% for fifteen epochs. Still, as the training session duration gets longer, the difference diminishes, as happened in Experiment 1. Figure 40 exhibits the average classification performance (mean classification error rate on unseen patients' MRI data) of the two transfer learning architectures with respect to the number of epochs were trained for.

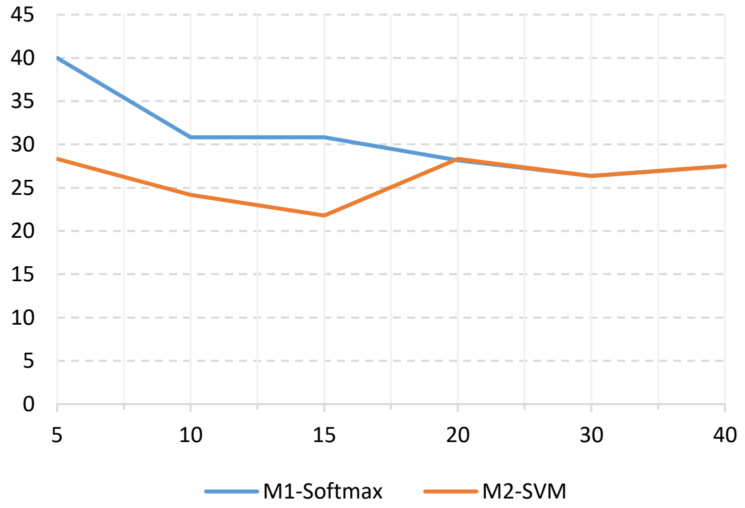


Figure 40 Average test error rate (vertical axis) in the VM-D vs non-D binary classification problem with respect to the number of epochs the two transfer learning architectures were trained for

On average, the best performance is achieved by M2-SVM models trained for 15 epochs: an averaging accuracy of 78.21% and an AUC= 0.8525 for VM-D vs non-D. Time comparison for three datasets is given in Figure 41.

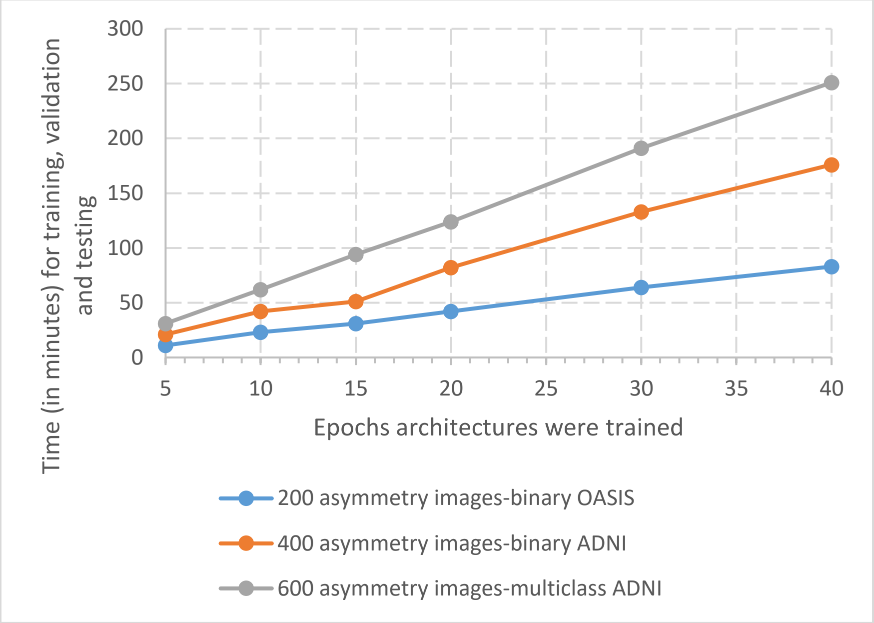


Figure 41: Approximated time (in minutes on average) spent on training, validation and testing across the classification tasks

5.10 Specialized CNN architecture for brain asymmetry

CNNs allow feature generation and selection through learning. In many cases, this helps avoid complicated feature engineering procedures. CNNs have become the method of choice in intelligent medical diagnosis when images are involved (Lundervold A.S. and A. Lundervold, 2016). The CNN architecture has been used widely for processing and classifying MRIs (Yamanakkanavar et al., 2020). Different configurations of CNN were investigated in the literature. Some of them are based on similarities with pretrained networks. Others are included in a hybrid configuration with other types of machine learning or deep learning algorithms.

The current research includes building a CNN architecture, named 5CLNN, for processing and classifying the MRI of brain asymmetries.

At the base of the newly proposed CNN architecture (see Figure 42), the first five convolutional layers play the role of feature detectors, and there is one fully connected layer. Additional layers consist of Local Response Normalization (LRN), Batch Normalization (BN), average pooling, dropout and Softmax layers. Overall, there are 103714 learnable parameters. The LRN uses nonlinear functions, known as Rectified Nonlinearity Units (ReLUs), which speed up training by normalizing the feature maps and removing the negative value. The Batch Normalization layer makes the deep learning process faster and more stable by standardizing the layers' inputs. Recentering and rescaling operations are at the base of this process. Average pooling layers perform dimensionality reduction of the images by implementing a down-sampled procedure on the input layers. The dropout operation clamps to zero the output signals of hidden neurons whose activation probability is 0.5 or lower to reduce overfitting. The first convolutional layer requires an input image of size 256×256×3 (256 pixels wide, 256 pixels high, RGB colour channels used for image augmentation). This layer includes 16 convolutions of size 5×5×3 and uses stride = 1 (the stride indicates the step size with which the filter/convolution moves over the image matrix). The second layer has 32 convolutions of size 3×3×3 with stride = 2, while the third layer is a copy of the second layer but stride =1. The fourth and fifth layers have 64 convolutions of similar sizes of 3×3×3 with stride = 2 and stride = 1. Each convolution layer is followed by Batch Normalization and ReLU layers. Average pooling layers follow the third and the fifth layers' convolutions. A fully connected layer puts together all feature outputs from previous layers and sends them to the activation unit (e.g. Softmax). Figure 59 illustrates the core CNN architecture proposed for the diagnosis of dementia. In this instantiation of the architecture, the Softmax layer normalizes the outputs used as classification probabilities by the output layer for the binary diagnostic task.

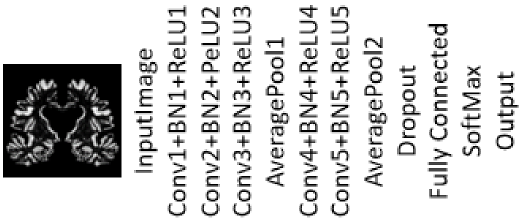


Figure 42: An instantiation of the core CNN architecture equipped with a standard Softmax

The new model shares a CNN core module equipped with separate Softmax, SVM, LD or KNN classification modules (see Figure 43 below).

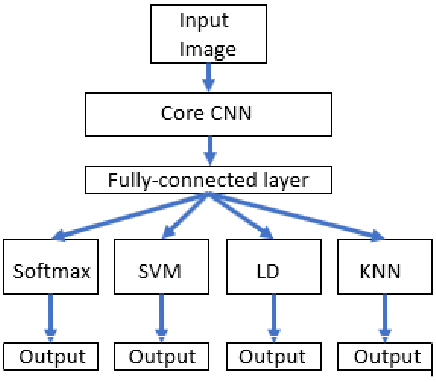


Figure 43*:* CNN core equipped with four classification modules

The diagnostic potential of these modules has been evaluated and compared in the next chapter.

5.11 Deep transfer learning

This section investigates further transfer learning models and how they compare against CNN diagnostic configurations created from scratch (Herzog and Magoulas, 2022). In the first experiment, CNNs from the ImageNet, AlexNet and VGG16 are retrained using transfer learning. Their classification performance is compared with the newly established classifier, 5CLNN, consisting of the CNN core plus Softmax module. The second experiment compares different instances of the CNN core module equipped with Softmax, SVM, LD or KNN modules. This classification module is added to the fully connected layer to produce the output. Similar to the first study, the dataset includes 600 images of segmented asymmetries from 150 patients from the ADNI database. The experiments were run with balanced datasets, i.e., the number of patients and their corresponding MRI images were equally distributed between the three classes. Each binary classification task, namely EMCI vs NC, AD vs NC, and AD vs EMCI, used 400 images, and ten independent simulation runs with each deep learning architecture were performed. The number of training epochs, n, was set to 10, while a mini-batch size of 128 and a validation data frequency equal to 50 were used. Lastly, a small initial learning rate equal to 0.0001 was adopted, whilst training followed the SGDM method. Segmented asymmetry images had a size of 256×256×3, where three indicates the number of colour channels. The data were split into 80% for training, 10% for validation and 10% for unknown MRI images were used for testing.

Experiment 1

In this experiment, the two pretrained neural networks, AlexNet and VGG-16, are fine-tuned by substituting the last three layers of the original architecture with a fully connected layer, SoftMax and output layers which are adjusted to the number of classes in the task. The layer replacement is the first step of fine-tuning, and the training of the whole network follows it. Table 18 provides comparative performance in testing each architecture based on ten independent runs for each classification problem. In this table, the newly established classifier, CNN core plus SoftMax, is denoted by 5CLNN.

Table 18: Average performance in testing based on ten independent runs

Classifiers	AD vs EMCI	EMCI vs NC	AD vs NC
AlexNet			
Acc	0.7867	0.8575	0.8817
F1-score	0.7892	0.8529	0.8800
AUC	0.8750	0.8925	0.9381
VGG16			
Acc	0.7075	0.8225	0.8725
F1-score	0.7060	0.8246	0.8700
AUC	0.8263	0.9019	0.9213
5CLNN			
Acc	0.7981	0.8644	0.9012
F1-score	0.7942	0.8678	0.9020
AUC	0.8738	0.9281	0.9413

Performance measures such as Accuracy, F1-score and AUC demonstrate the overall improvement in the diagnosis of dementia using 5CLNN. The average accuracy across all diagnostic tasks for 5CLNN is higher than the accuracy of AlexNet by 1.27% and VGG-16 by 5.39%. F1-score demonstrates similar differences in performance with 5CLNN better by 1.4% compared to AlexNet, and by 5.45% compared to VGG-16. The 5CLNN achieves the highest AUC in the diagnosis of AD vs NC. Early changes in the brain are diagnosed with an accuracy of 86.44% and an F1 score of 86.78%. The diagnostic performance in AD vs EMCI is 79.81% in terms of accuracy and 79.42 % in the F1-score using 5CLNN.

The performance of AlexNet and 5CLNN was tested further to investigate how consistent the differences were. Thirty additional independent runs were conducted for each subset of 400 images. The obtained results appear in line with the previous ones (see Tables 29-20).

Table 19: Average performance in testing based on 30 independent runs for AlexNet and 5CLNN

Classifiers	AD vs EMCI	EMCI vs NC	AD vs NC
AlexNet			
Acc	0.7710	0.8600	0.8825
F1-score	0.7813	0.8510	0.8942
AUC	0.8800	0.8975	0.9521
5CLNN			
Acc	0.8030	0.8775	0.9105
F1-score	0.7975	0.8700	0.9110
AUC	0.8700	0.9312	0.9475

In terms of accuracy and F1-score, the 5CLNN outperforms the AlexNet by 3.20% and 1.62% for AD vs EMCI, by 1.75% and 1.90% for EMCI vs NC and by 2.80% and 1.68% for AD vs NC sets.

The statistical significance of these changes was evaluated using the Wilcoxon-Mann-Whitney rank-sum test. This pairwise non-parametric test was applied to prove that the accuracy of 5CLNN is better than the accuracy of AlexNet architecture for AD vs EMCI, EMCI vs NC and AD vs NC datasets (see Table 20). The null hypothesis is based on the statement that we cannot easily create a new architecture that will perform at the same level or even outperform the CNN models AlexNet and VGG16. If the null hypothesis is rejected, this opens the opportunity for new configurations with at least the same performance and potential improvement.

Table 20: Statistical significance of testing based on 30 independent runs for AlexNet and 5CLNN

Classifiers	AD vs EMCI	EMCI vs NC	AD vs NC
p	0.01	0.01	0.01
h	1	1	1

The test returns a p-value equal to 0.01 for each imaging subset and shows a logical value of the test decision $h = 1$, which indicates a rejection of the null hypothesis at the 5% level of significance. Thus, the outperformance of 5CLNN is proven by these measures. Also, it is worth reporting that the time spent for training, validation and testing is 29 min for

5CLNN, 36 min for AlexNet and 546 min for VGG-16. The same commodity hardware was used in all cases, as mentioned in previous chapter.

Experiment 2

The second experiment focuses on the comparison of the classification performance of the different instances of the core 5CLNN, which are equipped with SoftMax, Support Vector Machine (SVM), Linear Discriminant (LD), and K-Nearest Neighbour (KNN) classification modules, respectively. Table 21 shows the comparative performance (average from 10 independent simulation runs) for 5CLNN-SoftMax, 5CLNN-SVM, 5CLNN-LD and 5CLNN-KNN in the three classification tasks.

Table 21: Average performance out of 10 runs for the 5CLNN with different classification modules (Softmax, SVM, LD and KNN)

Classifiers	AD vs EMCI	EMCI vs NC	AD vs NC
5CLNN-Softmax			
Acc	0.7981	0.8644	0.9012
F1-score	0.7981	0.8685	0.8983
AUC	0.8963	0.9206	0.9758
5CLNN-SVM			
Acc	0.7887	0.8668	0.8794
F1-score	0.7911	0.8646	0.8832
AUC	0.8738	0.9000	0.9400
5CLNN-LD			
Acc	0.8044	0.8731	0.8856
F1-score	0.8074	0.8685	0.8850
AUC	0.9094	0.9688	0.9200
5CLNN-KNN			
Acc	0.7231	0.8325	0.8742
F1-score	0.7285	0.8286	0.8762
AUC	0.8931	0.9050	0.9025

The core CNN achieves the best average performance over all binary problems with the LD classification module. Its accuracy is higher than the CNN with Softmax by 0.63% for AD vs EMCI and 0.87% for EMCI vs NC. However, 5CLNN-LD architecture shows lower performance than 5CLNNSoftmax in the AD vs NC task. The 5CLNN-KNN architecture demonstrates lower performance than other models in all diagnostic tasks.

A Wilcoxon rank-sum test, as before, was used to check if the performance differences of the CNN-LD model, although small, are significant compared to the second-best model, CNN-Softmax. The null hypothesis (*h*) is that the differences in the performance of CNN-LD and CNN-Softmax are not significant. If the null hypothesis is rejected (h=1), one says that the differences are significant at the 5% level. The outcome of the Wilcoxon test is presented in Table 22.

Table 22: Statistical significance of the results based on ten runs for the 5CLNN with Softmax and LD modules

Classifiers	AD vs EMCI	EMCI vs NC	AD vs NC
p	0.0145	0.0145	0.0145
h	1	1	1
Z	2.4442	2.4442	2.4442

The *p*-value of 0.0145, *h* equal to 1 and *z*-score of 2.4442 confirm the statistical significance of the obtained results.

5.12 Summary and contribution of the chapter

The potential of both feature learning from segmented brain asymmetry and directly from brain asymmetry imaging data was investigated in this chapter. This chapter also described relevant to this work machine learning fundamentals in the processing and classifying of neuroimages. The disease classification stages were included in the diagnostic cycle of the research. In general, machine learning models supported the automatic classification of brain pathologies using both classic machine learning algorithms and deep neural networks. The models utilised the labelled imaging data and were applied to binary and multiclass diagnostic problems.

Classic machine learning algorithm, trained on features extracted from asymmetry images, demonstrated robust performance with the highest values of 93% obtained by SVMs. Various CNNs, either created from scratch or pretrained, were also used. Experiments provided evidence of the diagnostic potential of these models when images of brain asymmetries are used without applying prior feature extraction and selection. Pretrained CNNs were fine-tuned via a transfer learning strategy. The proposed 5CLNN core architecture was combined with SVM, LD and KNN modules and trained successfully and tested on images of brain asymmetries that were not used before in the state-of-the-art literature. These models performed better than pretrained AlexNet and VGG-16 networks in the tests conducted.

The classification performance achieved when using asymmetry features or images is comparable with the results obtained by other researchers who used MRI images, summarized in Table 2 (Chapter 2) for the ADNI and OASIS databases. In particular, the predictive accuracy of 92.5% for EMCI vs CN and 93% for AD vs CN is high, especially when considering the complexity of the schemes shown in this table (e.g., a larger number of features, a higher number of architectural parameters, and so on). Also, it is worth noting that papers in the literature (cf. with Table 2) report best model performance and do not provide average results produced from several random initialisations of the classifier’s parameters like we do. Also, they do not report how many runs/trials were conducted in order to produce the best model. Since we do not have access to the sample-wise classification results of other classifiers in the literature we cannot perform a McNemar test, or some other test, to compare performances. Moreover, no optimisation was used for our models, so further performance improvement can be potentially achieved with SVM and CNN classifiers by optimization of architectures, as it has been shown in many cases in the literature. Experiments using segmented asymmetry images from the ADNI and OASIS MRI repositories revealed that the M2-SVM architecture outperforms the M1-Softmax architecture in short training cycles (less than 15 epochs). The two architectures show no significant differences in models’ efficiency in longer training cycles. Also, in the experiments, the M2-SVM architecture showed an ability to perform well using relatively small training datasets compared to the M1-Softman architecture.

Averaging the diagnostic performance of 5CLNN shows small but stable improvement, 1.27% in the accuracy and 1.4% in the F1-score, compared to AlexNet and 5.39% in the accuracy and 5.45% in the F1-score compared to VGG-16. Combining the 5CLNN (trained for ten epochs) with different classification modules results in performance differences: in the tests, the 5CLNN with the LD module worked better than the Softmax module, which is typically used in CNN applications.

The next chapter highlight the main contribution of the research, provides a discussion of the proposed methods used for early diagnosis of cognitive decline and dementia compare them with the state-of-the-art methods, and outlines future perspectives of the study.

Chapter 6: Conclusions and Future Work

This Chapter discusses the principal findings of the research and explains their meanings. The thesis investigated the modelling of structural changes in the brain in the early stages of cognitive decline. These findings are compared with signs of dementia due to the development of Alzheimer's Disease. The centre point of the research is the detection and segmentation of the left-rightwards hemispheric asymmetry of the brain from MRI scans of the free publicly available ADNI database. The proposed approach investigated and visualized the differences in shape and patterns of brain asymmetry in two groups of patients with preliminary and severe signs of dementia. The results were compared against the images of cognitively healthy participants. The images of the brain asymmetries were obtained after preprocessing and segmentation stages, which included varieties of thresholding and image transformation techniques.

The second key point of the research is the effectiveness of the digital asymmetry biomarker in a machine learning workflow for the diagnosis of early dementia through the analysis of MRI brain asymmetries. This includes traditional machine learning methods as well deep neural networks. Various Convolutional Neural Networks were trained to solve binary classification problems. Several pretrained CNN models were fine-tuned via transfer learning strategy. Their diagnostic potential was compared with the performance of the proposed CNN model.

Thus, the ***main contributions*** of the work can be summarised as follows:

1. **Digital Biomarker:** The study introduces a digital biomarker for the diagnosis of early dementia based on the analysis of structural brain changes using MRI scans. This biomarker focuses on identifying asymmetries and cross-sectional comparisons between brain hemispheres.
2. **Statistical features of brain asymmetries** are extracted from MRI scans, and then several conventional machine learning methods, such as SVM, KNN, LD and NB, are used to perform binary classification among pairs of the three classes of EMCI, AD and NC.
3. **CNN-based methods** are performed on the same problem as above, with implementation of transfer learning from AlexNet and VGG16, and a simple CNN architecture trained from scratch.

Several ***questions*** were raised at the beginning of the research. The answers can be given now to the restated questions.

Q1. Can changes in structural brain asymmetry be detected by computer vision algorithms?

Based on the functional changes during the development of dementia which are described in the literature (Yang et al., 2017; Liu et.al., 2018), the current research demonstrates the structural differences between the group of people with Early Mild Cognitive Impairment, Alzheimer's Disease and the control group of healthy individuals. The morphological or structural characteristics of the brain can be discovered by computer vision algorithms. In the thesis, the proposed segmentation algorithm detects the structural differences between brain hemispheres among mentioned above groups of patients. The segmentation methods include several image transformation techniques that extract the asymmetrical parts of the brain from initial MRIs. As we can see from the literature (Kim et al., 2012; Wachinger et al., 2016), structural transformation of asymmetry in a single area of the brain does not provide general information about how much the shift between the brain's left and right hemispheres. At the same time, the asymmetry of the entire brain is affected by the regional changes in the cortical and subcortical anatomical structures, which can be detected and visualized with the proposed computer vision technique. The developed algorithm helps to obtain the asymmetry pattern and find the differences between healthy individuals and those with a disease.

Q2. Can statistically engineered features from brain asymmetries distinguish between early and progressive stages of dementia?

The investigation of statistically engineered features of brain asymmetry between three groups of patients confirms that initial and progressive stages of dementia have unequable patterns of asymmetries. The experiments were conducted on several image sets and demonstrated similar pattern of changes. The statistical features collected from images of EMCI patients (early dementia) show more symmetry between brain hemispheres than features collected from AD patients (progressive dementia) and healthy controls. The features collected from MRIs of AD patients demonstrated the highest asymmetry among all examined groups.

Q3. Is it possible to use segmented images of structural brain asymmetry as a diagnostic source to predict early dementia (amnestic MCI) and Alzheimer's Disease?

Analysis of statistical features collected from image asymmetries and machine learning algorithms applied to the datasets confirm the diagnostic power of the MRIs of segmented asymmetries that can be used as an additional biomarker for the robust diagnosis of dementia.

Q4. What machine learning models are the most successful in the diagnosis of early cognitive impairment using images of brain asymmetries?

Among supervised ML algorithms used for the research, SVM demonstrated the best performance. Five-layered CNN overperformed more complex pretrained transfer learning models such as AlexNet and VGG. Moreover, LD and SVM classification modules successfully replaced a popular Softmax function in the network.

A more detailed discussion of the research finding is provided below.

6.1 Features representation of brain asymmetries of three classes of images

The evolution of digital images brings new perspectives to the development of statistical image analysis. Statistical image modelling is recognized as an efficient analytical method that generates the covariance matrices and calculates the likelihood values used to estimate the statistical significance of model testing and evaluation. Even though image processing in computer vision is a well-established area with a collection of techniques, statistical aspects of image evaluation are in the process of development. The statistical data from images can be obtained differently. It can be collected from the histogram, a standard image processing tool. Pixel distribution functions can be used for image analysis.

6.1.1. Statistical features

Some aspects of statistical image analysis were also used for the evaluation of the brain MRI data. The data-analytical approach was applied to MRI scans collected from 150 individuals equally divided between patients with Early Mild Cognitive Impairment, Alzheimer's Disease and a control group of healthy subjects. The feature engineering process includes the extraction of statistical data from image properties. The statistical information is obtained after transforming the pixel values into wavelets. This process was performed by applying the discrete wavelet transformation to the imaging data. DWT is known as one of the image and signal processing methods that give accurate information about image properties (Bengueddoudj and Messali, 2018; Khademi et al., 2011) and can be used to reduce image dimensionality (Fayaz et al., 2021).

The study's initial findings are based on the analysis of the 1st and 2nd order of statistical features obtained from the segmented brain asymmetry images. The statistical approach is particularly valuable for the analysis of imaging microstructures. The 1st order statistics computes the likelihood of observed greyscale values in any random position in the image matrix. The 2nd order statistics determine the likelihood of a pair of grey-level pixels in their random position and orientation in the image (Ruberto and Fodde, 2013).

The analysis of absolute statistical values and their ratio among patients' classes gives an inside into the problem of early dementia. The most valuable information was obtained from MSE, Mean, Std, RMS, Variance, and IDM features. A comparison of statistical features collected from image asymmetries highlights the differences among people with Early Mild Cognitive Impairment, Alzheimer's Disease and a control group of cognitively healthy subjects. The early stages of cognitive decline are characterized by decreased asymmetry between the brain's left and right hemispheres. At the same time, the brain asymmetries of demented people are pronounced compared with healthy subjects and those who have a cognitive impairment. In this vein, analysis of the absolute statistical values gives information about the structural symmetries of MRI scans. The lowest mean value has EMCI class compared to AD and NC data. AD has the highest average value for the same statistical features. The highest ratio has the AD vs EMCI group. The feature values (MSE, Mean, Std, RMS, Variance, and IDM) indicate the substantial differences in the asymmetry between these classes. The smallest but most stable ratio shows the EMCI vs NC group. It confirms the initial degenerative changes in the brain in the early stages of cognitive decline detected by the segmentation algorithm. The statistical differences for AD vs NC classes are also constant for all datasets (Figures 42-48). Similar changes in brain asymmetries were described by Yang et al. (2017). They explored the brain white matter networks and came to a conclusion about the rightward topological asymmetry in patients with AD and the absence of such changes in patients with MCI.

The segmented brain images used in the current study do not give an impression of the type of asymmetry, left- or rightward. They highlight the shape and patterns of the pixel intensity levels in the anatomical brain areas. It is known from other studies that the leftward asymmetry of the healthy brain becomes rightward due to the development of dementia (Yang et al., 2017; Liu et al., 2018). In reality, the statistical features that demonstrate very similar numbers in graphs and bar charts of normal and affected by Alzheimer's Disease brains (see Figures 38-40) are "negatively correlated". In other words, healthy individuals have leftward asymmetry, and individuals with severe dementia have rightward asymmetries between the brain hemispheres (Yang et al., 2017). The structural difference between early signs of dementia and serious disease manifestations can be explained. In the initial stages, the development of dementia leads to a gradual decrease in leftward asymmetry, and the brain becomes more symmetrical. In more severe stages, the asymmetry increases again, leading to rightward hemispheric asymmetry. However, this assumption needs further research. Longitudinal studies of each group of individuals can clarify the obtained results.

Not so many studies describe the general changes in the pattern of asymmetries among individuals with neurogenerative disorders. Most investigators pay attention to the atrophic processes in the local cortical or subcortical areas, such as the hippocampus and amygdala (Wachinger et al., 2015; Kim et al., 2012). Atrophic changes in the regional brain areas might have a different speed of development (Wachinger et al., 2016) that also affect brain functionalities (Liu et al., 2018). Liu et al. discovered the abnormal rightward functional laterality in the brain of AD patients compared to healthy controls. On the other hand, the patients with MCI had a different brain lateralization level than patients with AD and the control group. The researchers assumed that structural features of the brain influenced functional results.

Several other studies describe the structural brain specificities of healthy participants (Segato et al., 2020; Tomasi and Volkow, 2012; Cabeza et al., 2004). Some of them investigate the effect of gender (Koelkebeck et al., 2014), age (Zhou, D. et al., 2013), handedness (Kong et al., 2018), and brain size (Williams et al., 2022) on the hemispheric symmetries.

6.1.2 Gender

The study explores the gender-related changes in asymmetry during cognitive decline and dementia development. MRI scans of 100 male and 50 female individuals were processed separately and compared using the statistical property of asymmetries. Biological gender dimorphism in the cerebral asymmetry of healthy subjects was presented in numerous studies (Koelkebeck et al., 2014; Guadalupe et al., 2015). Their findings confirm the total increase in rightward hemispheric volumetric asymmetry in females compared to males.

More detailed regional differences are discussed in Kong et al. (2018), where substantial distinctions were registered in the medial temporal regions such as the parahippocampal gyrus and the entorhinal cortex. The parahippocampal region demonstrated leftward asymmetry in males, and the entorhinal cortex was rightward oriented in females.

Moreover, these regions are also essential for Alzheimer's Disease, which reflects gender differences and has a prevalence in women (Mazure and Swendsen, 2016; Li and Singh, 2014). For instance, hippocampal atrophy progresses faster in women than in men (Ardekani et al., 2016). Diversity in regional brain atrophy affects the clinical presentation of the disease. Men behave more aggressively and have increased comorbidity and higher mortality rate than women. Women have more tendency to affective disorders and disability, but, at the same time, they survive longer (Sinforiani et al., 2010). It is important to note that the current study confirms the biological dimorphism between the two genders, such as an increase in rightward asymmetry in females compared to males in the control group. The female brain of this group of individuals looks more symmetrical than the male brain (see Figure 44). Nevertheless, the increase in symmetry in the EMCI group and the decrease in symmetry in the AD group were detected in both genders. That's just the degree of asymmetries that looks slightly different.

6.1.3 Age

The age of the patients can impact the level of brain symmetry. Some authors point to a significant correlation between brain asymmetry and cortical thickness in healthy individuals who demonstrate an age-related increase in leftward asymmetry (Kong et al., 2018; Zhou, D. et al., 2013). The age groups of the examined EMCI and NC subjects were narrowed to ages 55 and 65 to minimize the age impact. The age effect on the brain of AD patients is not remarkable compared to the influence of the extensive atrophic process in the brain tissues due to the development of the disease. Thus, the AD dataset consists of the MRIs of 65 years old subjects and 90 years old at the same time.

6.1.4 Handedness

The study did not include an investigation of the patient's handedness and its influence on brain asymmetries. However, according to the study by Kong et al. (2018), handedness is not remarkably associated with cortical asymmetries. Their study is based on MRI data from 17,141 healthy subjects from multiple worldwide resources and provides a detailed description of the typical brain asymmetries.

6.1.5 Brain size

The human brain varies in size and volume. MRI scans give the possibility to calculate brain measures. The average brain volume in men is around 1260 cm^3 and 1130 cm^3 in women. There are substantial individual variations in brain measures (Schoenemann, 2006; Im et al., 2008). Some researchers investigated a correlation between cerebral asymmetry and brain size and came to different conclusions. Thus, Barrick et al. (2005) did not find any evidence supporting this connection. Unlike them, Williams et al. (2022) found that large brains have more significant asymmetries for overall surface area and mean cortical thickness. They suppose that global brain asymmetry impacts local asymmetries and can partially explain them. The current study took into account the differences between the men's and women's brains and investigated the asymmetrical changes in the brain matter separately for each assigned group.

6.1.6 Dataset samples

The statistical information is collected from 5 segmented asymmetry image sets with three classes each. The sets differ in length (number of images), gender and plane of asymmetry (vertical and horizontal). The variability of those parameters did not reveal significant differences in the patterns of changes in brain asymmetries.

6.2 Machine learning workflow for brain asymmetry images

Images of brain asymmetry as a source of imaging data for the classification of dementia were never described in the literature before. Machine learning workflows were added to the imaging source of MRI data to test the segmented asymmetries for their robustness in the detection of cognitive decline. The segmented images of asymmetry were tested for their ability to complement a row of already existing biomarkers used for the diagnosis of dementia. Machine learning models trained on the segmented imaging data became the main diagnostic tools for this medical problem.

6.2.1 Conventional supervised ML algorithms

One of the lines of diagnostic modelling includes the combination of feature engineering processes and classification algorithms. This part of the research focused on testing the various ML algorithms using a fusion of statistical features and Bag-of-Features collected from segmented image asymmetries. The performance of NB, LD, SVM and KNN classifiers was evaluated using a 10-fold cross-validation procedure. The results were compared with the performance of the state-of-the-art methods described in the literature.

It is important to note that the current study and state-of-the-art research papers used the exact source of MRI data obtained from the ADNI database.

The best results among the mentioned above classifiers were obtained with the SVM algorithm using polynomial quadratic and cubic kernels. The average achieved accuracy for the EMCI vs NC classes is 92.5% using two kernels, 86.5% for the AD vs EMCI classes using the same kernels, and 92.5% and 93% for the AD vs NC classes using quadratic and cubic kernels accordingly. The obtained results are very close to those demonstrated in the literature. For instance, Beheshti et al. (2017) diagnosed AD with an accuracy of 94.73% using the SVM model. Moradi et al. (2015) detected the MCI with an accuracy of 82% with the ensemble of the SVM and Random Forest algorithms. Glozman and Le (2014) predicted AD using SVM with a probability of 92%. At the same time, the feature generation processes used by researchers are different. In this situation, the advantage has that method which has less complexity and requires a shorter time for the implementation. Consequently, the approaches proposed by Beheshti et al. (2017) and Moradi et al. (2015) look time-consuming. The first of them had to identify the atrophic brain region with voxel-based morphometry, extract features by applying a 3D mask, rank features according to T-test score, and select them using a specific genetic algorithm. Only after those stages the classification algorithm had been used. The second method consists of the following steps: feature selection using regularized logistic regression algorithm, a semi-supervised low-density feature separation algorithm, a combination of the collected features with cognitive measurements, and classification using a Random Forest algorithm. These two methods are only examples of the complex feature engineering methods described in the literature. The feature extraction procedure used in the current study is more straightforward than many other methods.

Another essential point is that classic algorithms with MRI asymmetry dataset outperformed all CNN models used in the research (their performance will be compared with state-of-the-art methods later in the following subsection). Table 23 shows the highest classification performance archived with classic and CNN models.

Table 23 Classification performance of best classic and CNN models

Classifiers	AD vs EMCI	EMCI vs NC	AD vs NC
C-SVM			
Acc	0.8650	0.9250	0.9300
F1-score	0.8600	0.9200	0.9300

AUC	0.9400	0.9800	0.9900
5CLNN			
Acc	0.8030	0.8775	0.9105
F1-score	0.7975	0.8700	0.9110
AUC	0.8700	0.9312	0.9475

6.2.2 Deep learning: deep transfer learning and specialised CNN architectures

Another line of the classification is the implementation of Artificial Neural Networks and Deep Learning models. These models take some advantages over the other machine learning methods discussed above. They generate features during the training stage, learn them in a supervised or unsupervised mode, and are able to process a large amount of data. Convolutional Neural Network is one of the most adapted for classifying imaging data models. The diagnostic potential of this model and its modifications, including transfer learning, were used for the detection of neurodegenerative disorders.

A set of experiments with deep learning models was conducted. The purpose was to investigate the diagnostic performance of the neural networks using created sets of image asymmetries. The intermediate stages included tests with fine-tuned convolutional models and transfer learning models. Experiments run on commodity hardware to maximize the possibility of integrating these models in routine healthcare checks. For the same reason, the size of the datasets was kept under 1000 samples. It was taken into account that accessible databases containing clinical records and imaging data can be limited.

Significant findings of the study

CNN models demonstrate a very good classification performance on image asymmetry sets. Early changes in the brain (EMCI vs NC) are diagnosed with an accuracy of 87.75 ± 3.10% with the proposed CNN (5CLNN). The model's ability to distinguish between AD and EMCI is 80.30 ± 2.30%. The AD prediction score compared to NC samples is 91.05 ± 3.20%.

The successful competitors of created during the research 5CLNN classifier are pretrained networks. Two of the transfer learning models, Alexnet and VGG16, show the high diagnostic accuracy of early stages of cognitive decline (EMCI vs NC) corresponding to 85.80 ± 1.50% and 82.25 ± 2.00%, respectively. However, the proposed

CNN model performs better than pretrained AlexNet and VGG-16 networks in the tests conducted. The average accuracy across all diagnostic tasks for 5CLNN is higher than the accuracy of AlexNet by 1.27% and VGG-16 by 5.39%. Moreover, the training, validation and testing time of the 5CLNN model is visibly shorter by 7 min (1.24 times) compared to AlexNet and by 517 min (18.83 times) compared to VGG-16.

Another substantial finding of the study is that the classifier's performance can be improved by connecting the various classification modules to the output layer. For instance, the best results for the 5CLNN are obtained when the model works with the LD classification layer. It demonstrates, on average, 87.31% of accuracy for the diagnosis of EMCI. By contrast, in the same model with Softmax (typically used in CNN applications), SVM and KNN classification layers, the accuracy is 86.44%, 86.68%, and 83.25%, accordingly. Another example of the increased performance due to the implementation of the different from the Softmax classification module strategy is a combination of AlexNet with the SVM output layer. It demonstrates 95.90% accuracy in the diagnosis of Alzheimer's Disease, 90.25% in detecting early Mild Cognitive Impairment, and 83% in the separation of early dementia and Alzheimer's Disease.

It is important to note that the efficiency of the deep learning model is correlated to the number of training epochs. Looking at the average testing score across all the tasks, the AlexNet model plus SVM layer outperforms AlexNet plus Softmax layer by 8.10% when trained for five epochs, by 1.93% when trained for ten epochs, and by 1.31% when trained for fifteen epochs.

Lastly, the evaluation of all the classification models highlights the advantage of SVM over the CNN models in diagnosing early dementia based on images of segmented brain asymmetries. On average, the SVM overperforms the CNN by 5% in the prediction of EMCI compared to the control group, by 2% in the discovery of AD and by 6% in the distinguishing between EMCI and AD. These results may indicate that the feature engineering process plays a significant role in more accurate diagnostic predictions. So, it is hard to ignore the fact that the classic machine learning algorithms work pretty well on small data sizes. At the same time, neural networks are able to process a more significant amount of information with fairly high accuracy and offer end-to-end learning. However, the "black box" effect can make it difficult to understand the nature of such types of predictions. It might limit the use of deep learning models in medical practice.

Comparison with the state-of-the-art

The classification results obtained during the research work described in the thesis provide evidence that the segmented asymmetries approach for predicting dementia is valid. Although the diagnostic lines that use images of the segmented parts of the human brain for the detection of structural and functional abnormalities of the human brain require further development, when compared to the methods already described in the literature the proposed approach demonstrates a promising diagnostic potential. For instance, in the study of Chitradevi and Prabha (2020) that evaluated the diagnostic performance of several image segmentation methods in diagnosing Alzheimer's Disease using AlexNet, the highest predictive score of around 90% was achieved with images of the segmented hippocampus. Separated during the study, other parts and tissues of the brain demonstrated lower prediction values than the method based on the image asymmetries.

Images of segmented brain asymmetries can be used independently or as an additional biomarker for the early diagnosis of MCI. In medical practice, the diagnosis of any degenerative brain disease, including Mild Cognitive Impairment and Alzheimer's Disease, is usually supported by several biomarkers that perform the complex analysis of medical data. The combined biomarkers are discussed in many research articles. In this way, Yang et al. (2019) explored the diagnostic potential of three groups of biomarkers such as mental task performance and some sort of digital and imaging biomarkers. Linear discriminant analysis, statistical analysis and CNN were included in the analytical and classification schema. The accuracy obtained with complex biomarkers correlates to the current research results using MRIs of asymmetries only.

Moreover, the highest accuracy of 90.62% in the detection of MCI (Yang et al., 2019) was received using CNN with imaging biomarkers. In their study, the approaches using LDA, and statistical analyses demonstrated 76.67% and 60% accuracy, respectively. It is possible to assume that brain asymmetry images have the potential to increase the prediction of dementia if they are added to any other group of biomarkers.

Another combination of biomarkers was tested with convolutional neural network autoencoders by Martinez-Murcia et al. (2019). They use only non-imaging data, including the age of patients, presence of the Apolipoprotein E4 in chromosome 19, the concentration of Tau protein in cerebrospinal fluid, results of MMSE and Alzheimer's

Disease Assessment Scale cognitive tests, and Clinical Dementia Ratio score. However, the prediction power of non-imaging data was lower than in studies using MRIs. Alzheimer's Disease was diagnosed with an accuracy of 84% only. These findings confirm the advantage of imaging biomarkers in medical diagnosis, including those developed in the current study.

Images of segmented asymmetries have been checked for their robustness in the diagnosis of early and progressive dementia using various machine learning algorithms, including deep neural networks. Pretrained CNNs have gained popularity in the last few years in clinical diagnostics. Ease of use and good performance attract many researchers. A considerable part of the research findings was received from machine learning experiments that involved transfer learning of convolutional neural networks. The number of these networks and their variants are proliferating. For example, Spasov et al. (2019) proposed the combination of the convolutional sub-networks inspired by AlexNet and Xception CNN. The experiments were conducted with the concatenated features collected from the imaging data and those obtained from the clinical and cognitive tests. The combination of MRI and clinical features got the best results of 86% accuracy. The accuracy is comparable with the CNN-based experiments used to classify brain asymmetries. However, it should be noted that current experiments use single imaging biomarkers, which simplify and speed up the machine learning process.

It is worth emphasizing one more significant point, the advantage of computerized diagnostics over human experts in medical practice. This issue is discussed in many publications. A comparative analysis of the diagnostic performance of the Inception-V3 CNN model and human radiology reader was established by Ding et al. (2019). According to their records, DL outperformed the human reader in terms of sensitivity by 43% for the AD and 29% for the MCI classification of PET images. Also, in terms of precision, the algorithm outperformed the specialists by 7% for the MCI diagnosis. At the same time, the medical professional surpassed the DL model in terms of specificity by 9% in the prediction of AD and by 18% in the detection of MCI. Even though the general performance of the DL model was better than the efficiency of human professionals, it tells that more precise machine learning methods are still needed to detect the medical pathology accurately. The current study of asymmetrical brain structures did not aim to prove the advantage of the computer-aided diagnosis but to propose a different approach which helps to understand better the structural changes od the brain tissues during the development of dementia and improve its diagnosis.

The classification performance of the developed method was also compared with the results of 3D neural network models. The 3D VGG16 model, applied by Yagis et al. (2020) to the imaging data of the ADNI and OASIS datasets, demonstrated an average accuracy of 73.4% and 69.9%, respectively, in the AD diagnosis. Another 3D CNN model, used by Oh et al. (2019), established accuracy of 86.60% for detecting AD and 73.95% for predicting a progressive form of MCI. Both three-dimensional models show lower performance than transfer learning CNN architectures used in the current study. Furthermore, the core of the seconds model was a convolutional autoencoder with a GoogLeNet InceptionV2 module specially designed for unsupervised learning, dimensionality reduction, and differentiation between AD and NC. However, the autoencoder was unable to analyse the spatial structure of the imaging data, which is an essential condition for the reliable explanation of classification decisions. In contrast, in the current research, the images of segmented asymmetries give an adjuvant inside to the specific structural details and shapes of the brain hemispheres that help understand the positioning of some useful brain features.

From the following papers describing complicated deep learning architectures, it is possible to conclude that model complexity does not necessarily lead to better classification performance. Several examples of complex neural network models and their results are presented below. Lu et al. (2018) proposed the multiscale neural network-based analysis for early diagnosis of Alzheimer's Disease using metabolic brain activity measures and features collected from multiple MRI regions. The final network was created as a composition of four deep neural networks due to initially pretrained stacked-autoencoders. An accuracy of 82.51 % was obtained in distinguishing between presymptomatic AD and non-progressive MCI. Another example of the complex classification approach is given by Duc et al. (2020). The investigators proposed to join imaging and non-imaging (MMSE) features using 3D CNN and four regression models, including linear least square regression, tree regression, bagging-based ensemble regression and support vector regression models. The CNN model was a modified version of the VGG network with additional batch normalization layers. The mean accuracy for the classification of AD versus cognitively normal subjects was 85.27%. Another example of a more successful configuration implemented for the prediction of MCI is the joint MRI fusion model and two MLP models proposed by Qiu et al. (2018). The MRI fusion model was generated via majority voting from three independently trained VGG11 networks. The attained accuracy of the final multimodal model was

90.9% which is relatively high. Unfortunately, the article does not provide details about the training time for all of these models. From this perspective, supervised machine learning models tested with imaging data from brain asymmetries have a significant advantage over deep learning models.

The classification performance of the pretrained networks and proposed CNN using images of brain asymmetries was compared with several non-convolutional networks described in the state-of-the-art literature (Stamate et al., 2020; Lama et al., 2017; Amoroso et al., 2018). Multi-Layer Perceptron and a Convolutional Bidirectional Long Short-Term Memory model proposed by Stamate et al. (2020) demonstrated the highest accuracy of 86% in the diagnosis of dementia. Another study (Lama et al., 2017) used an unsupervised deep learning method based on a Regularized Extreme Learning Machine algorithm for binary classification of AD vs MCI. The archived accuracy was 80.32%. And lastly, a combined strategy of Random Forest and a feedforward DNN described by Amoroso et al. (2018) demonstrated recall of 87.5% for AD, 52.5 % for HC, and 27.5% for MCI. The results above confirm that the diagnostic performance of non-convolutional neural networks is lower than the convolutional models tested in the current study and those presented in the state-of-the-art literature. This rule is mainly applied to imaging data.

Table 24 summarises the discussion above and shows the validation performance of the proposed CNN architecture and state-of-the-art methods. The 5LCNN with MRI, segmented asymmetries data, outperformed complex models in the detection of EMCI and AD in most cases.

Table 24 The comparison of proposed CNN with state-of-the art methods

Authors	Methods	Results
Herzog and Magoulas (2022)	5LCNN; MRI segmented asymmetry imaging data	Accuracy is 91.05% (AD vs NC), 80.3% (AD vs EMCI), 87.75% (EMCI vs NC)
Lama et al. (2017)	Unsupervised classification learning; MRI data	Accuracy is 80.32% (AD vs MCI)
Stamate et al. (2020)	Deep Learning (complex) models; Multiple biomarkers	Accuracy is 86% (AD vs NC)
Yagis et al. (2020)	3D VGG model; MRI data	Accuracy is 73.4% (AD vs NC)
Qiu et al. (2018)	Combination of deep learning models: MRI + cognitive test scores	Accuracy is 90.9% (MCI vs NC)

Yang et al. (2019)	Complex model Multiple biomarkers	Accuracy is 90.62% with MRI using CNN (MCI vs NC)
Martinez-Murcia et al. (2019)	DL based on convolutional autoencoders; Multiple biomarkers	Accuracy is 84% (AD vs NC)
Duc et al. (2020)	Complex model Multiple biomarkers	Accuracy is 85.27% (AD vs NC)
Li et al. (2019)	Complex model MRI data	Accuracy is 90% (AD vs NC) 76.2% (pMCI vs AD)
Spasov et al. (2019)	AlexNet and Xception CNNs; Multiple biomarkers	Accuracy is 86%, (pMCI vs NC)
Chitradevi and Prabha (2020)	AlexNet CNN; MRI data (Classification of AD, normal)	Accuracy is ~90% for the different segmented brain parts (AD vs NC)
Amoroso et al. (2018)	Complex non-CNN models; Multiple biomarkers	Recall for AD (87.5%), HC (52.5 %), MCI (27.5%) cMCI (57.5%)
Oh et al. (2019)	Complex model; MRI data	Accuracies for the AD is 86.60%, and for pMCI is 73.95%

6.3 Summary of the thesis

The research on the structural MRIs of individuals with early mild cognitive impairment and Alzheimer's Disease proposed a new diagnostic approach based on the analysis of brain asymmetry. It facilitates early detection of initial dementia when clinical symptoms are very mild and challenging.

A growing number of cases of dementia among the elderly population worldwide has become a reason for the active research in this area. Dementia is currently the seventh leading cause of death among all diseases and one of the major causes of disability and dependency among older people. Early and precise diagnosis of the pathology is crucial. Timely and focused therapy help find the appropriate treatment, slowing down the development of pathology and extending the time of independent leaving.

The diagnosis of dementia is based on several clinical and biological criteria, including imaging and non-imaging biomarkers. Computerized diagnostic methods speed up the detection of brain pathology and make it more precise. Categorization of early changes in the brain helps to evaluate the risk of progression of cognitive decline to dementia. Imaging techniques, including MRI scans, allow understanding of the specific degenerative brain processes leading to severe dementia and Alzheimer’s Disease. However, an internal destructive mechanism at some points remains unclear.

6.3.1 Hypothesis restatement and major findings

The current research was based on the hypothesis that changes in the level of asymmetry between the left, and right brain hemispheres might be an additional diagnostic factor (biomarker) for the diagnosis of dementia in the early stage of its development. The study involved the imaging resources (MRI scans) of the well-known ADNI and OASIS medical imaging databases. The experimental study included a cross-sectional classification of early mild cognitive impairment, Alzheimer's disease, and cognitively normal subjects from the ADNI database and demented and non-demented subjects from the OASIS database. The diagnostic pipeline consisted of image processing and classification stages. During the image processing stage, 2D images of brain asymmetries are obtained from MRI scans. The proposed segmentation algorithm highlighted the differences in shape and pattern of asymmetry among patients with cognitive decline, Alzheimer’s Disease and a control group of healthy individuals. Furthermore, the asymmetries of EMCI, AD and NC individuals are analysed using statistical features and classified with machine learning algorithms. The experimental results support the hypothesis that changes in the brain asymmetries during the development of dementia convey essential information, as they were used to generate valuable features for classification. Varieties of applied supervised machine learning algorithms helped evaluate the robustness of the proposed segmentation algorithm in the diagnosis of dementia.

Important to note that in contrast to other methods in the literature (see Chapter 2.3), this study is less complex in terms of image processing. For example, the image processing time on average was 0.1 min per image (3.6 GHz Intel Core i7, 16 GB RAM), compared to the article (Moradi et al., 2015), where it took 8 min (3.4 GHz Intel Core i7, 8 GB RAM). The images of segmented asymmetries require three times less memory space than similar originals. It potentially gives an advantage in terms of computational time spent on training a classifier. However, it is impossible to directly compare with other literature approaches as hardware specifications differ. Nevertheless, it is worth mentioning that the CPU time for an asymmetry features-trained single classifier, including 10-fold cross-validation, was approximately 5 sec on our hardware, which appears considerably low compared to other methods that used more complex feature sets (Breiman, 2001).

If to speak about classification performance, the applied classification models appear comparable with more complex methods described in the state-of-the-art literature and summarized in Table 2 (literature review) for the ADNI database. In particular, the predictive accuracy of 92.5% for EMCI vs CN and 93% for AD vs CN is high, especially when considering the complexity of the schemes shown in this table (e.g., a more significant number of features, a higher number of architectural parameters, and so on). Further performance improvement was achieved with CNN classifiers by optimizing architectures, as demonstrated in other works in the literature (Le, N.Q.K. et al., 2020); (Grassi et al., 2019). Transfer learning models and the proposed convolutional model in combination with SVM, KNN and LD classification modules demonstrated consistent and robust performance in the detection of the brain pathology using images of segmented asymmetries.

6.3.2 Limitation of the study

The study has a limitation in the investigation of the handedness of the patients. The images downloaded from the ADNI database were not separated according to participants' handedness. The data from the OASIS database includes only records of right-handed individuals. Based on the literature, it is known that around 90% of the population is right-handed, and the possible error can be low, especially if taken into account that affected by degenerative processes, areas of the brain might not be connected to handedness at all. However, investigating the impact of the right- and left-handedness on the changes in brain asymmetry during the development of dementia is one of the potential directions for further work.

6.3.3 Future development

It is worth mentioning that the results obtained in current research are interdisciplinary in nature and can contribute to the field of neuroscience, psychology and computer science. The suggested computational framework can be potentially valuable to other researchers working on the diagnosis of brain-related disorders, or on the processing and classification of different medical imaging data, especially in cases with unclear border and texture or low-scale structural changes, which are not visible to the human eye.

On a practical level, the use of image asymmetry and asymmetry features has the potential to contribute to the design of end-user (e.g., physicians and general medical practitioners) applications, which will run on commodity hardware for early diagnosis of cognitive decline and investigation of the nature of the structural changes in the brain. These applications can exploit the generalization properties of machine learning models on unseen MRI data, as demonstrated in this study.

The proposed methodology has a perspective to explore the stages of Alzheimer's disease further. This research can be based on longitudinal studies of patient data. Changes in the asymmetry shape and mapping of these changes to the brain atlas will direct those brain areas affected by the pathological process.

Another point for investigation is the comparison of asymmetries between the grey and white matter of the brain. Some researchers investigated the correlation between the degeneration of the white matter tracts and grey matter atrophy in cases of Alzheimer's disease (Agosta et al., 2011; Frings et al., 2014). They found out that axons of neurons (represent the white matter) can be affected earlier than the neurons themselves (represent the grey matter) and can symbolize the early onset of the disease. The white matter in patients with MCI is affected significantly less. Thus, in the diagnosis of MCI and the transformation of some of its forms to AD, MRI technologies can help computational models in the detection of early degenerative brain changes. This can be explained by the fact that cognitive decline in the case of MCI might have different morphological grounds when the destructive process does not involve white matter only. The nature of MCI is more complex and might have another, for instance, the vascular reason for amnestic and cognitive decline. Only 30% of MCI progress to AD. Additional computer vision segmentation techniques might give a clue to the source of initial tissue deformation, which opens a direction for early prediction and disease prevention.

One more diagnostic direction applied to the classification of diseases is using deep learning models. Transfer learning with a different from Softmax (e.g., SVM, LD, KNN) classification layer may be helpful in situations where large clinical datasets are not readily available. Several studies reported limited access to clinical data (Strongman et al., 2019). Hence, when the imaging datasets are limited in size and the time for testing is restricted, the diagnostic model incorporated, for example, with SVM or LD layers, might get the advantage.

Some neuropsychiatric disorders, such as dyslexia (Altarelli et al., 2014), attention deficit hyperactivity disorder (ADHD) (Shaw et al., 2009), autism (Eyler et al., 2012), psychiatric disorders like obsessive-compulsive disorder (OCD) and Schizophrenia (Crow, 1990; Yücel et al., 2002; Yücel et al., 2003) and mood disorders like Major Depression (Liu, W. et al., 2016; Yucel et al., 2009) may be associated with abnormal cortical asymmetries. These pathologies are still not fully explored (Kong et al., 2018; Kong et al., 2020). For instance, structural and functional MRI of patients with autistic spectrum disorder (ASD) indicates the rightward brain lateralization compared to the healthy controls. These changes in asymmetry are mainly related to the language zone and areas responsible for the brain cortex's cognitive and emotional processes, which are less developed in ASD patients (Lindell and Hudry, 2013). However, some studies have reported similar changes in the white matter tracts of this group of people. Another example of brain asymmetry changes belongs to the mood disorders group is Major Depression. These changes are registered in the frontal and parietal-temporal zones of the brain’s grey matter towards leftward lateralization. Nevertheless, the substrate of the cortical abnormalities is still under investigation. One more psychiatric disorder which demonstrates the atypical structural and functional lateralization is OCD. The changes are found in both cortical and subcortical brain areas, and the pattern of asymmetries might vary from zone to zone. Significantly, most investigators did not find the associations between shifts in brain asymmetry and the length and type of treatment, age in the onset and duration of the disease, or the number of symptoms. Brain asymmetries in ADHD and schizophrenia are still in the initial stages of exploration and analysis.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. and Kudlur, M., 2016. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (pp. 265-283).

Abdi, H. and Williams, L.J., 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), pp.433-459.

Aderghal, K., Afdel, K., Benois-Pineau, J., Catheline, G. and Alzheimer's Disease Neuroimaging Initiative, 2020. Improving Alzheimer's stage categorization with Convolutional Neural Network using transfer learning and different magnetic resonance imaging modalities. *Heliyon*, *6*(12), p.e05652.

Agosta, F., Pievani, M., Sala, S., Geroldi, C., Galluzzi, S., Frisoni, G.B. and Filippi, M., 2011. White matter damage in Alzheimer’s disease and its relationship to grey matter atrophy. *Radiology*, *258*(3), pp.853-863.

Aha, D.W., Kibler, D. and Albert, M.K., 1991. Instance-based learning algorithms. *Machine learning*, *6*(1), pp.37-66.

Alexander, A.L., Lee, J.E., Lazar, M. and Field, A.S., 2007. Diffusion tensor imaging of the brain. *Neurotherapeutics*, *4*(3), pp.316-329.

Al-Kadi, O.S., 2008. Combined statistical and model based texture features for improved image classification.

Altarelli, I., Leroy, F., Monzalvo, K., Fluss, J., Billard, C., Dehaene-Lambertz, G., Galaburda, A.M. and Ramus, F., 2014. Planum temporale asymmetry in developmental dyslexia: Revisiting an old question. *Human brain mapping*, *35*(12), pp.5717-5735.

Amanatiadis, A., Kaburlasos, V.G., Gasteratos, A. and Papadakis, S.E., 2011. Evaluation of shape descriptors for shape-based image retrieval. *IET Image Processing*, *5*(5), pp.493-499.

Amoroso, N., Diacono, D., Fanizzi, A., La Rocca, M., Monaco, A., Lombardi, A., Guaragnella, C., Bellotti, R., Tangaro, S. and Alzheimer's Disease Neuroimaging Initiative, 2018. Deep learning reveals Alzheimer's disease onset in MCI subjects: results from an international challenge. *Journal of neuroscience methods*, *302*, pp.3-9.

Andriluka, M., Roth, S. and Schiele, B., 2008, June. People-tracking-by-detection and people-detection-by-tracking. In *2008 IEEE Conference on computer vision and pattern recognition* (pp. 1-8). IEEE.

Ardekani, B.A., Convit, A. and Bachman, A.H., 2016. Analysis of the MIRIAD data shows sex differences in hippocampal atrophy progression. *Journal of Alzheimer's Disease*, *50*(3), pp.847-857.

Arellano, A.R., Bory-Reyes, J. and Hernandez-Simon, L.M., 2018. Statistical entropy measures in C4. 5 trees. *International Journal of Data Warehousing and Mining (IJDWM)*, *14*(1), pp.1-14.

Armi, L. and Fekri-Ershad, S., 2019. Texture image analysis and texture classification methods-A review. *arXiv preprint arXiv:1904.06554*.

Ashburner, J. and Friston, K.J., 2000. Voxel-based morphometry—the methods. *Neuroimage*, *11*(6), pp.805-821.

Ashburner, J. and Friston, K.J., 2005. Unified segmentation. *Neuroimage*, *26*(3), pp.839-851.

Ashburner, J., Barnes, G., Chen, C.C., Daunizeau, J., Flandin, G., Friston, K., Kiebel, S., Kilner, J., Litvak, V., Moran, R. and Penny, W., 2014. SPM12 manual the FIL methods group (and honorary members).

Avola, D., Cinque, L. and Placidi, G., 2013. Customized first and second-order statistics based operators to support advanced texture analysis of MRI images. *Computational and mathematical methods in medicine*, *2013*.

Bae, E., Yuan, J. and Tai, X.C., 2011. Global minimization for continuous multiphase partitioning problems using a dual approach. International journal of computer vision, 92(1), pp.112-129.

Bair, E., 2013. Semi-supervised clustering methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, *5*(5), pp.349-361.

Baldi, P., 2012, June. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML workshop on unsupervised and transfer learning* (pp. 37-49). JMLR Workshop and Conference Proceedings.

Baldominos, A., Saez, Y. and Isasi, P., 2019. A survey of handwritten character recognition with mnist and emnist. *Applied Sciences*, *9*(15), p.3169.

Baraskar, T.N. and Mankar, V.R., 2019. The DICOM Image Compression and Patient Data Integration using Run Length and Huffman Encoder. In *Coding Theory*. IntechOpen.

Barrick, T.R., Mackay, C.E., Prima, S., Maes, F., Vandermeulen, D., Crow, T.J. and Roberts, N., 2005. Automatic analysis of cerebral asymmetry: an exploratory study of the relationship between brain torque and planum temporale asymmetry. *Neuroimage*, *24*(3), pp.678-691.

Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., Filippi, M. and Alzheimer's Disease Neuroimaging Initiative, 2019. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage: Clinical*, *21*, p.101645.

Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L., 2008. Speeded-up robust features (SURF). *Computer vision and image understanding*, *110*(3), pp.346-359.

Beheshti, I., Demirel, H., Matsuda, H. and Alzheimer's Disease Neuroimaging Initiative, 2017. Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. *Computers in biology and medicine*, *83*, pp.109-119.

Belongie, S., Malik, J. and Puzicha, J., 2000. Shape context: A new descriptor for shape matching and object recognition. *Advances in neural information processing systems*, *13*, pp.831-837.

Bengueddoudj, A. and Messali, Z., 2018, April. Subjective and Objective Evaluation of Noisy Multimodal Medical Image Fusion Using 2D-DTCWT and 2D-SMCWT. In *International Conference on Computer Science and its Applications* (pp. 225-234). Springer, Cham.

Bera, K., Schalper, K.A., Rimm, D.L., Velcheti, V. and Madabhushi, A., 2019. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nature reviews Clinical oncology*, *16*(11), pp.703-715.

Besl, P.J. and Jain, R.C., 1988. Segmentation through variable-order surface fitting. *IEEE Transactions on pattern analysis and machine intelligence*, *10*(2), pp.167-192.

Bharati, M.H., Liu, J.J. and MacGregor, J.F., 2004. Image texture analysis: methods and comparisons. *Chemometrics and intelligent laboratory systems*, *72*(1), pp.57-71.

Bianco, S., Ciocca, G., Guarnera, G.C., Scaggiante, A. and Schettini, R., 2014, March. Scoring recognizability of faces for security applications. In *Image Processing: Machine Vision Applications VII* (Vol. 9024, pp. 159-168). SPIE.

Bieniek, A. and Moga, A., 2000. An efficient watershed algorithm based on connected components. *Pattern recognition*, *33*(6), pp.907-916.

Biratu, E.S.S., Schwenker, F., Debelee, T.G.G., Kebede, S.R.R., Negera, W.G.G. and Molla, H.T.T., 2021. Enhanced Region Growing for Brain Tumor MR Image Segmentation. *Journal of Imaging*, *7*(2), p.22.

Bitar, M. and Barry, G., 2018. Multiple innovations in genetic and epigenetic mechanisms cooperate to underpin human brain evolution. *Molecular biology and evolution*, *35*(2), pp.263-268.

Böhning, D., 1992. Multinomial logistic regression algorithm. *Annals of the institute of Statistical Mathematics*, *44*(1), pp.197-200.

Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T. and Vogler, C., 2019, October. Sign language recognition, generation, and translation: An interdisciplinary perspective. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 16-31).

Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp.5-32.

Brentnall, A.R. and Cuzick, J., 2018. Use of the concordance index for predictors of censored survival data. *Statistical methods in medical research*, *27*(8), pp.2359-2373.

Brinker, T.J., Hekler, A., Enk, A.H., Klode, J., Hauschild, A., Berking, C., Schilling, B., Haferkamp, S., Schadendorf, D., Holland-Letz, T. and Utikal, J.S., 2019. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer*, *113*, pp.47-54.

Bühlmann, P. and Hothorn, T., 2007. Boosting algorithms: Regularization, prediction and model fitting. *Statistical science*, *22*(4), pp.477-505.

Bunce, S.C., Izzetoglu, M., Izzetoglu, K., Onaral, B. and Pourrezaei, K., 2006. Functional near-infrared spectroscopy. *IEEE engineering in medicine and biology magazine*, *25*(4), pp.54-62.

Cabeza, R., 2002. Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychology and aging*, *17*(1), p.85.

Cabeza, R., Daselaar, S.M., Dolcos, F., Prince, S.E., Budde, M. and Nyberg, L., 2004. Task-independent and task-specific age effects on brain activity during working memory, visual attention and episodic retrieval. *Cerebral cortex*, *14*(4), pp.364-375.

Cabezas, M., Oliver, A., Lladó, X., Freixenet, J. and Cuadra, M.B., 2011. A review of atlas-based segmentation for magnetic resonance brain images. *Computer methods and programs in biomedicine*, *104*(3), pp.e158-e177.

Carrazza, S., Forte, S., Kassabov, Z., Latorre, J.I. and Rojo, J., 2015. An unbiased hessian representation for Monte Carlo PDFs. *The European Physical Journal C*, *75*(8), pp.1-20.

Catal, C., 2012. Performance evaluation metrics for software fault prediction studies. *Acta Polytechnica Hungarica*, *9*(4), pp.193-206.

Chen, P.H.C., Gadepalli, K., MacDonald, R., Liu, Y., Kadowaki, S., Nagpal, K., Kohlberger, T., Dean, J., Corrado, G.S., Hipp, J.D. and Mermel, C.H., 2019. An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. *Nature medicine*, *25*(9), pp.1453-1457.

Chitradevi, D. and Prabha, S., 2020. Analysis of brain sub regions using optimization techniques and deep learning method in Alzheimer disease. *Applied Soft Computing*, *86*, p.105857.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

Colantonio, S., Coppini, G., Giorgi, D., Morales, M.A. and Pascali, M.A., 2018. Computer vision for ambient assisted living: Monitoring systems for personalized healthcare and wellness that are robust in the real world and accepted by users, carers, and society. In *Computer Vision for Assistive Healthcare* (pp. 147-182). Academic Press.

Coley, D.A., 1999. *An introduction to genetic algorithms for scientists and engineers*. World Scientific Publishing Company.

Cox, D.R., 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), pp.187-202.

Crow, T.J., 1990. Temporal lobe asymmetries as the key to the etiology of schizophrenia. *Schizophrenia bulletin*, *16*(3), pp.433-443.

Dalitz, C., Brandt, C., Goebbels, S. and Kolanus, D., 2013. Fourier descriptors for broken shapes. *EURASIP Journal on Advances in Signal Processing*, *2013*(1), pp.1-11.

De Boer, R., Vrooman, H.A., Van Der Lijn, F., Vernooij, M.W., Ikram, M.A., Van Der Lugt, A., Breteler, M.M. and Niessen, W.J., 2009. White matter lesion extension to automatic brain tissue segmentation on MRI. *Neuroimage*, *45*(4), pp.1151-1161.

Despotović, I., Goossens, B. and Philips, W., 2015. MRI segmentation of the human brain: challenges, methods, and applications. *Computational and mathematical methods in medicine*, *2015*.

Ding, S., Xu, X. and Nie, R., 2014. Extreme learning machine and its applications. *Neural Computing and Applications*, *25*(3), pp.549-556.

Ding, Y., Sohn, J.H., Kawczynski, M.G., Trivedi, H., Harnish, R., Jenkins, N.W., Lituiev, D., Copeland, T.P., Aboian, M.S., Mari Aparici, C. and Behr, S.C., 2019. A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain. *Radiology*, *290*(2), pp.456-464.

Di Ieva, A., 2011. Brain anatomy-from a clinical and neurosurgical perspective: a clinically oriented manual of neuroanatomy.

Dolz, J., Ayed, I.B., Yuan, J. and Desrosiers, C., 2018, April. Isointense infant brain segmentation with a hyper-dense connected convolutional neural network. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 616-620). IEEE.

Dorado, J., del Toro, X., Santofimia, M.J., Parreño, A., Cantarero, R., Rubio, A. and Lopez, J.C., 2019. A computer-vision-based system for at-home rheumatoid arthritis rehabilitation. *International Journal of Distributed Sensor Networks*, *15*(9), p.1550147719875649.

DSM-V (Diagnostic and Statistical Manual of Mental Disorders) 2013 updated https://www.psychiatry.org/psychiatrists/practice/dsm. Accessed 10 June 2020.

Du, W. and Tibshirani, R., 2018. A pliable lasso for the Cox model. *arXiv preprint arXiv:1807.06770*.

Duchenne, O., Bach, F., Kweon, I.S. and Ponce, J., 2011. A tensor-based algorithm for high-order graph matching. *IEEE transactions on pattern analysis and machine intelligence*, *33*(12), pp.2383-2395.

Ehsan, S., Clark, A.F., Rehman, N.U. and McDonald-Maier, K.D., 2015. Integral images: efficient algorithms for their computation and storage in resource-constrained embedded vision systems. *Sensors*, *15*(7), pp.16804-16830.

El Hajjar, A. and Rey, J.F., 2020. Artificial intelligence in gastrointestinal endoscopy: general overview. *Chinese medical journal*, *133*(3), p.326.

Esteva, A. and Topol, E., 2019. Can skin cancer diagnosis be transformed by AI?. *The Lancet*, *394*(10211), p.1795.

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J. and Socher, R., 2021. Deep learning-enabled medical computer vision. *npj Digital Medicine*, *4*(1), pp.1-9.

Evans, A.J., Bauer, T.W., Bui, M.M., Cornish, T.C., Duncan, H., Glassy, E.F., Hipp, J., McGee, R.S., Murphy, D., Myers, C. and O'Neill, D.G., 2018. US Food and Drug Administration approval of whole slide imaging for primary diagnosis: a key milestone is reached and new questions are raised. *Archives of pathology & laboratory medicine*, *142*(11), pp.1383-1387.

Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J. and Zisserman, A., 2015. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, *111*(1), pp.98-136.

Evgeniou, T. and Pontil, M., 1999, July. Support vector machines: Theory and applications. In *Advanced Course on Artificial Intelligence* (pp. 249-257). Springer, Berlin, Heidelberg.

Eyler, L.T., Pierce, K. and Courchesne, E., 2012. A failure of left temporal cortex to specialize for language is an early emerging and fundamental property of autism. *Brain*, *135*(3), pp.949-960.

Farrokhi, F., Buchlak, Q.D., Sikora, M., Esmaili, N., Marsans, M., McLeod, P., Mark, J., Cox, E., Bennett, C. and Carlson, J., 2020. Investigating risk factors and predicting complications in deep brain stimulation surgery with machine learning algorithms. *World neurosurgery*, *134*, pp.e325-e338.

Fayaz, M., Torokeldiev, N., Turdumamatov, S., Qureshi, M.S., Qureshi, M.B. and Gwak, J., 2021. An Efficient Methodology for Brain MRI Classification Based on DWT and Convolutional Neural Network. *Sensors*, *21*(22), p.7480.

Feichtenhofer, C., Pinz, A. and Zisserman, A., 2017. Detect to track and track to detect. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3038-3046).

Felzenszwalb, P.F., Girshick, R.B., McAllester, D. and Ramanan, D., 2009. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, *32*(9), pp.1627-1645.

Frejlichowski, D., 2012. Application of the curvature scale space descriptor to the problem of general shape analysis. *Przeglad Elektrotechniczny*, *88*, pp.209-212.

Freund, Y. and Schapire, R.E., 1996, July. Experiments with a new boosting algorithm. In *icml* (Vol. 96, pp. 148-156).

Frings, L., Yew, B., Flanagan, E., Lam, B.Y., Hüll, M., Huppertz, H.J., Hodges, J.R. and Hornberger, M., 2014. Longitudinal grey and white matter changes in frontotemporal dementia and Alzheimer’s disease. *PloS one*, *9*(3), p.e90814.

Fukushima, K. and Miyake, S., 1982. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets* (pp. 267-285). Springer, Berlin, Heidelberg.

Gabryel, M., 2018. The Bag-of-Words Method with Different Types of Image Features and Dictionary Analysis. *J. Univers. Comput. Sci.*, *24*(4), pp.357-371.

Gainotti, G., 2019. Emotions and the right hemisphere: can new data clarify old models? *The Neuroscientist*, *25*(3), pp.258-270.

Gainotti, G., 2019. The role of the right hemisphere in emotional and behavioral disorders of patients with frontotemporal lobar degeneration: an updated review. *Frontiers in aging neuroscience*, *11*, p.55.

Galatsanos, N.P., Segall, C.A. and Katsaggelos, A.K., 2003. Digital image enhancement. *Encyclopedia of optical engineering*, pp.388-402.

García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J.M. and Herrera, F., 2016. Big data preprocessing: methods and prospects. *Big Data Analytics*, *1*(1), pp.1-22.

Gargiulo, P., Árnadóttir, Í., Gíslason, M., Edmunds, K. and Ólafsson, I., 2017. New directions in 3D medical modeling: 3D-printing anatomy and functions in neurosurgical planning. *Journal of healthcare engineering*, *2017*.

Ghorbani, A., Ouyang, D., Abid, A., He, B., Chen, J.H., Harrington, R.A., Liang, D.H., Ashley, E.A. and Zou, J.Y., 2020. Deep learning interpretation of echocardiograms. *NPJ digital medicine*, *3*(1), pp.1-10.

Glozman, T. and Le, R.K. Classification of Alzheimer’s Disease Based on White Matter Architecture 2014, pdfs.semanticscholar.org

González, P., Álvarez, E., Díez, J., López-Urrutia, Á. and del Coz, J.J., 2017. Validation methods for plankton image classification systems. *Limnology and Oceanography: Methods*, *15*(3), pp.221-237.

Grassi, M., Rouleaux, N., Caldirola, D., Loewenstein, D., Schruers, K., Perna, G., Dumontier, M. and Alzheimer's Disease Neuroimaging Initiative, 2019. A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer's disease using socio-demographic characteristics, clinical information, and neuropsychological measures. *Frontiers in neurology*, p.756.

Gu, Y., Pandit, S., Saraee, E., Nordahl, T., Ellis, T. and Betke, M., 2019. Home-based physical therapy with an interactive computer vision system. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (pp. 0-0).

Guadalupe, T., Zwiers, M.P., Wittfeld, K., Teumer, A., Vasquez, A.A., Hoogman, M., Hagoort, P., Fernandez, G., Buitelaar, J., van Bokhoven, H. and Hegenscheid, K., 2015. Asymmetry within and around the human planum temporale is sexually dimorphic and influenced by genes involved in steroid hormone receptor activity. *Cortex*, *62*, pp.41-55.

Guarneri, F., Vaccaro, M. and Guarneri, C., 2008. Digital image compression in dermatology: format comparison. *Telemedicine and e-Health*, *14*(7), pp.666-670.

Guerra, B.M.V., Ramat, S., Beltrami, G. and Schmid, M., 2020. Automatic pose recognition for monitoring dangerous situations in Ambient-Assisted Living. *Frontiers in Bioengineering and Biotechnology*, *8*, p.415.

Gui, L., Lisowski, R., Faundez, T., Hüppi, P.S., Lazeyras, F. and Kocher, M., 2012. Morphology-driven automatic segmentation of MR images of the neonatal brain. *Medical image analysis*, *16*(8), pp.1565-1579.

Gunčar, G., Kukar, M., Notar, M., Brvar, M., Černelč, P., Notar, M. and Notar, M., 2018. An application of machine learning to haematological diagnosis. *Scientific reports*, *8*(1), pp.1-12.

Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), pp.1157-1182.

Hadar, P.N., Kini, L.G., Coto, C., Piskin, V., Callans, L.E., Chen, S.H., Stein, J.M., Das, S.R., Yushkevich, P.A. and Davis, K.A., 2018. Clinical validation of automated hippocampal segmentation in temporal lobe epilepsy. *NeuroImage: Clinical*, *20*, pp.1139-1147.

Haenssle, H.A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Hassen, A.B.H., Thomas, L., Enk, A. and Uhlmann, L., 2018. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of oncology*, *29*(8), pp.1836-1842.

Haffner, P., Bottou, L., Howard, P.G. and LeCun, Y., 1999, September. DjVu: Analyzing and compressing scanned documents for internet distribution. In *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318)* (pp. 625-628). IEEE.

Han, X., Laga, H. and Bennamoun, M., 2019. Image-based 3D object reconstruction: State-of-the-art and trends in the deep learning era. *IEEE transactions on pattern analysis and machine intelligence*.

Harini, R. and Chandrasekar, C., 2012, March. Image segmentation using nearest neighbour classifiers based on kernel formation for medical images. In *International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)* (pp. 261-265). IEEE.

Harris, C.G. and Stephens, M., 1988, August. A combined corner and edge detector. In *Alvey vision conference* (Vol. 15, No. 50, pp. 10-5244).

Hassan, D., Aickelin, U. and Wagner, C., 2014, July. Comparison of distance metrics for hierarchical data in medical databases. In *2014 International Joint Conference on Neural Networks (IJCNN)* (pp. 3636-3643). IEEE.

He, K., Zhang, X., Ren, S. and Sun, J., 2016, October. Identity mappings in deep residual networks. In *European conference on computer vision* (pp. 630-645). Springer, Cham.

Herzog, N.J. and Magoulas, G.D., 2021. Brain asymmetry detection and machine learning classification for diagnosis of early Dementia. *Sensors*, *21*(3), p.778.

Herzog, N.J. and Magoulas, G.D., 2021, June. Deep Learning of Brain Asymmetry Images and Transfer Learning for Early Diagnosis of Dementia. In *International Conference on Engineering Applications of Neural Networks* (pp. 57-70). Springer, Cham.

Herzog, N.J. and Magoulas, G.D., 2022. Machine Learning-Supported MRI Analysis of Brain Asymmetry for Early Diagnosis of Dementia. In *Medical Informatics and Bioimaging Using Artificial Intelligence* (pp. 29-52). Springer, Cham.

Herzog, N.J. and Magoulas, G.D., 2022. Convolutional Neural Networks-Based Framework for Early Identification of Dementia Using MRI of Brain Asymmetry. *International Journal of Neural Systems*, pp.2250053-2250053.

Hoops, S., Nazem, S., Siderowf, A.D., Duda, J.E., Xie, S.X., Stern, M.B. and Weintraub, D., 2009. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson's disease. *Neurology, 73*(21), pp.1738-1745.

Hope, T., Resheff, Y.S. and Lieder, I., 2017. *Learning tensorflow: A guide to building deep learning systems*. " O'Reilly Media, Inc.".

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

Hubel, D. and Wiesel, T., 2012. David Hubel and Torsten Wiesel. *Neuron*, *75*(2), pp.182-184.

ICD-11 (International Classification of Diseases) for Alzheimer and Dementia 2018 released and endorsed in May 2019, https://icd.who.int/en. Accessed 27 January 2021.

Iizuka, T., Fukasawa, M. and Kameyama, M., 2019. Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies. *Scientific reports*, *9*(1), pp.1-9.

Im, K., Lee, J.M., Lyttelton, O., Kim, S.H., Evans, A.C. and Kim, S.I., 2008. Brain size and cortical structure in the adult human brain. *Cerebral cortex*, *18*(9), pp.2181-2191.

Isensee, F., Kickingereder, P., Wick, W., Bendszus, M. and Maier-Hein, K.H., 2017, September. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In *International MICCAI Brainlesion Workshop* (pp. 287-297). Springer, Cham.

Işık, Ş., 2014. A comparative evaluation of well-known feature detectors and descriptors. *International Journal of Applied Mathematics Electronics and Computers*, *3*(1), pp.1-6.

Isles, A.R., 2018. Epigenetics, chromatin and brain development and function. *Brain and neuroscience advances*, *2*, p.2398212818812011.

James, S., Konidaris, G. and Rosman, B., 2017, February. An analysis of monte carlo tree search. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1).

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. NeuroImage 62, 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015.

Jin, Y., Dou, Q., Chen, H., Yu, L., Qin, J., Fu, C.W. and Heng, P.A., 2017. SV-RCNet: workflow recognition from surgical videos using recurrent convolutional network. *IEEE transactions on medical imaging*, *37*(5), pp.1114-1126.

Johansen, A.M., Evers, L. and Whiteley, N., 2010. Monte carlo methods. *Lecture notes*, *200*.

Joshi, A.S., Kulkarni, O., Kakandikar, G.M. and Nandedkar, V.M., 2017. Cuckoo search optimization-a review. *Materials Today: Proceedings*, *4*(8), pp.7262-7269.

Jović, A., Brkić, K. and Bogunović, N., 2015, May. A review of feature selection methods with applications. In *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 1200-1205). Ieee.

Kaestner, R., Maye, J., Pilat, Y. and Siegwart, R., 2012, May. Generative object detection and tracking in 3d range data. In *2012 IEEE International Conference on Robotics and Automation* (pp. 3075-3081). IEEE.

Kaganami, H.G. and Beiji, Z., 2009, September. Region-based segmentation versus edge detection. In *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (pp. 1217-1221). IEEE.

Karami, E., Shehata, M. and Smith, A., 2017. Image identification using SIFT algorithm: performance analysis against different image deformations. *arXiv preprint arXiv:1710.02728*.

Kass, M., Witkin, A. and Terzopoulos, D., 1988. Snakes: Active contour models. *International journal of computer vision*, *1*(4), pp.321-331.

Kaur, D. and Kaur, Y., 2014. Various image segmentation techniques: a review. *International Journal of Computer Science and Mobile Computing*, *3*(5), pp.809-814.

Khademi, A., Krishnan, S. and Venetsanopoulos, A., 2011. Shift-invariant DWT for medical image classification. *DISCRETE WAVELET TRANSFORMS Ț THEORY AND APPLICATIONS*, p.179.

Khairandish, M.O., Gupta, R. and Sharma, M., 2020. A hybrid model of faster R-CNN and SVM for tumor detection and classification of MRI brain images. *Int. J. Mech. Prod. Eng. Res. Dev*, *10*(3), pp.6863-6876. Khan, T., 2016. *Biomarkers in Alzheimer's disease*. Academic Press.

Khan, T., 2016. *Biomarkers in Alzheimer's disease*. Academic Press.

Kim, J.H., Lee, J.W., Kim, G.H., Roh, J.H., Kim, M.J., Seo, S.W., Kim, S.T., Jeon, S., Lee, J.M., Heilman, K.M. and Na, D.L., 2012. Cortical asymmetries in normal, mild cognitive impairment, and Alzheimer's disease. *Neurobiology of aging*, 33(9), pp.1959-1966.

Koelkebeck, K., Miyata, J., Kubota, M., Kohl, W., Son, S., Fukuyama, H., Sawamoto, N., Takahashi, H. and Murai, T., 2014. The contribution of cortical thickness and surface area to grey matter asymmetries in the healthy human brain. *Human Brain Mapping*, *35*(12), pp.6011-6022.

Kong, X.Z., Mathias, S.R., Guadalupe, T., ENIGMA Laterality Working Group, Glahn, D.C., Franke, B., Crivello, F., Tzourio-Mazoyer, N., Fisher, S.E., Thompson, P.M. and Francks, C., 2018. Mapping cortical brain asymmetry in 17,141 healthy individuals worldwide via the ENIGMA Consortium. *Proceedings of the National Academy of Sciences*, *115*(22), pp.E5154-E5163.

Kong, X.Z., Postema, M.C., Guadalupe, T., de Kovel, C., Boedhoe, P.S., Hoogman, M., Mathias, S.R., Van Rooij, D., Schijven, D., Glahn, D.C. and Medland, S.E., 2022. Mapping brain asymmetry in health and disease through the ENIGMA consortium. *Human brain mapping*, *43*(1), pp.167-181.

Kornilov, A.S. and Safonov, I.V., 2018. An overview of watershed algorithm implementations in open source libraries. *Journal of Imaging*, *4*(10), p.123.

Kotoulas, L. and Andreadis, I., 2005, October. Image analysis using moments. In *5th Int. Conf. on Technology and Automation, Thessaloniki, Greece* (Vol. 360364).

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, pp.1097-1105.

Lama, R.K., Gwak, J., Park, J.S. and Lee, S.W., 2017. Diagnosis of Alzheimer’s disease based on structural MRI images using a regularized extreme learning machine and PCA features. *Journal of healthcare engineering*, *2017*.

Larobina, M. and Murino, L., 2014. Medical image file formats. *Journal of digital imaging*, *27*(2), pp.200-206.

Larobina, M. and Murino, L., 2014. Medical image file formats. *Journal of digital imaging*, *27*(2), pp.200-206.

Lazli, L., Boukadoum, M. and Mohamed, O.A., 2020. A survey on computer-aided diagnosis of brain disorders through MRI based on machine learning and data mining methodologies with an emphasis on Alzheimer disease diagnosis and the contribution of the multimodal fusion. *Applied Sciences*, *10*(5), p.1894.

Le, N.Q.K., Do, D.T., Hung, T.N.K., Lam, L.H.T., Huynh, T.T. and Nguyen, N.T.K., 2020. A computational framework based on ensemble deep neural networks for essential genes identification. *International journal of molecular sciences*, *21*(23), p.9070.

Le, X. and Gonzalez, R., 2009, September. Pattern-based corner detection algorithm. In *2009 Proceedings of 6th International Symposium on Image and Signal Processing and Analysis* (pp. 238-243). IEEE.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, *521*(7553), pp.436-444.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), pp.2278-2324.

LeCun, Y., Touresky, D., Hinton, G. and Sejnowski, T., 1988, June. A theoretical framework for back-propagation. In *Proceedings of the 1988 connectionist models summer school* (Vol. 1, pp. 21-28).

Lee, V.E., Liu, L. and Jin, R., 2014. Decision trees: Theory and algorithms. In *Data Classification* (pp. 115-148). Chapman and Hall/CRC.

Levantesi, S. and Pizzorusso, V., 2019. Application of machine learning to mortality modeling and forecasting. *Risks*, *7*(1), p.26.

Li, H., Habes, M., Wolk, D.A., Fan, Y. and Alzheimer's Disease Neuroimaging Initiative, 2019. A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's & Dementia*, *15*(8), pp.1059-1070.

Li, N., Zhou, S., Wu, Z., Zhang, B. and Zhao, G., 2020. Statistical modeling and knowledge-based segmentation of cerebral artery based on TOF-MRA and MR-T1. *Computer methods and programs in biomedicine*, *186*, p.105110.

Li, R. and Singh, M., 2014. Sex differences in cognitive impairment and Alzheimer’s disease. *Frontiers in neuroendocrinology*, *35*(3), pp.385-403.

Liebelt, J. and Schmid, C., 2010, June. Multi-view object class detection with a 3d geometric model. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1688-1695). IEEE.

Lin, M., Chen, Q. and Yan, S., 2013. Network in network. *arXiv preprint arXiv:1312.4400*.

Lindell, A.K. and Hudry, K., 2013. Atypicalities in cortical structure, handedness, and functional lateralization for language in autism spectrum disorders. *Neuropsychology Review*, *23*(3), pp.257-270.

Liu, B., Lv, Y., Gu, Y. and Lv, W., 2020. Implementation of a Lightweight Semantic Segmentation Algorithm in Road Obstacle Detection. *Sensors*, *20*(24), p.7089.

Liu, D., Xiong, Y., Pulli, K. and Shapiro, L., 2011, August. Estimating image segmentation difficulty. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition* (pp. 484-495). Springer, Berlin, Heidelberg.

Liu, W., Mao, Y., Wei, D., Yang, J., Du, X., Xie, P. and Qiu, J., 2016. Structural asymmetry of dorsolateral prefrontal cortex correlates with depressive symptoms: evidence from healthy individuals and patients with major depressive disorder. *Neuroscience bulletin*, *32*(3), pp.217-226.

Liu, W., Sun, X. and Li, D., 2019. Robust object tracking via online discriminative appearance modeling. *EURASIP Journal on Advances in Signal Processing*, *2019*(1), pp.1-9.

Liu, Y., Collins, R.T. and Rothfus, W.E., 2001. Robust midsagittal plane extraction from normal and pathological 3-D neuroradiology images. *IEEE transactions on medical imaging*, *20*(3), pp.175-192.

Liu, Y., Hel-Or, H., Kaplan, C.S. and Van Gool, L., 2010. Computational symmetry in computer vision and computer graphics. *Foundations and Trends® in Computer Graphics and Vision*, *5*(1–2), pp.1-195.

Liu, Y., Jain, A., Eng, C., Way, D.H., Lee, K., Bui, P., Kanada, K., de Oliveira Marinho, G., Gallegos, J., Gabriele, S. and Gupta, V., 2020. A deep learning system for differential diagnosis of skin diseases. *Nature Medicine*, *26*(6), pp.900-908.

Liu, Y., Teverovskiy, L., Carmichael, O., Kikinis, R., Shenton, M., Carter, C.S., Stenger, V.A., Davis, S., Aizenstein, H., Becker, J.T. and Lopez, O.L., 2004, September. Discriminative MR image feature analysis for automatic schizophrenia and Alzheimer's disease classification. In *International conference on medical image computing and computer-assisted intervention* (pp. 393-401). Springer, Berlin, Heidelberg.

Liu, Y., Teverovskiy, L.A., Lopez, O.L., Aizenstein, H., Meltzer, C.C. and Becker, J.T., 2007, April. Dicovery of" Biomarkers" for Alzheimer's Disease prediction from structural MR images. In *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 1344-1347). IEEE.

Liu, Z., Wang, S., Di Dong, J.W., Fang, C., Zhou, X., Sun, K., Li, L., Li, B., Wang, M. and Tian, J., 2019. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics*, *9*(5), p.1303.

Liu, H., Zhang, L., Xi, Q., Zhao, X., Wang, F., Wang, X., Men, W. and Lin, Q., 2018. Changes in brain lateralization in patients with mild cognitive impairment and Alzheimer’s disease: a resting-state functional magnetic resonance study from Alzheimer’s disease neuroimaging initiative. *Frontiers in neurology*, 9, p.3.

Lowe, D.G., 1999, September. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision* (Vol. 2, pp. 1150-1157). Ieee.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), pp.91-110.

Lu, D., Popuri, K., Ding, G.W., Balachandar, R., Beg, M.F. and Alzheimer's Disease Neuroimaging Initiative, 2018. Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. Medical image analysis, 46, pp.26-34.

Lu, Y., Jiang, T. and Zang, Y., 2003. Region growing method for the analysis of functional MRI data. *NeuroImage*, *20*(1), pp.455-465.

Lundervold, A.S. and Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*, *29*(2), pp.102-127.

Makropoulos, A., Counsell, S.J. and Rueckert, D., 2018. A review on automatic fetal and neonatal brain MRI segmentation. *NeuroImage*, *170*, pp.231-248.

Manjunath, B.S., Ohm, J.R., Vasudevan, V.V. and Yamada, A., 2001. Colour and texture descriptors. *IEEE Transactions on circuits and systems for video technology*, *11*(6), pp.703-715.

Marr, D. and Poggio, T., 1979. A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *204*(1156), pp.301-328.

Marr, D., 1982. Vision: A computational investigation into the human representation and processing of visual information.

Martinez-Murcia, F.J., Górriz, J.M., Ramírez, J. and Ortiz, A., 2018. Convolutional neural networks for neuroimaging in parkinson's disease: is preprocessing needed?. *International journal of neural systems*, *28*(10), p.1850035.

Martinez-Murcia, F.J., Ortiz, A., Gorriz, J.M., Ramirez, J. and Castillo-Barnes, D., 2019. Studying the manifold structure of Alzheimer's disease: A deep learning approach using convolutional autoencoders. *IEEE Journal of Biomedical and Health Informatics*, *24*(1), pp.17-26.

Materka, A. and Strzelecki, M., 1998. Texture analysis methods–a review. *Technical university of lodz, institute of electronics, COST B11 report, Brussels*, *10*(1.97), p.4968.

Mazure, C.M. and Swendsen, J., 2016. Sex differences in Alzheimer's disease and other dementias. *The Lancet. Neurology*, *15*(5), p.451.

McAndrew, A., 2004. An introduction to digital image processing with matlab notes for scm2511 image processing. *School of Computer Science and Mathematics, Victoria University of Technology*, *264*(1), pp.1-264.

McDougall, F., Edgar, C., Mertes, M., Delmar, P., Fontoura, P., Abi-Saab, D., Lansdall, C.J., Boada, M. and Doody, R., 2021. Psychometric Properties of the Clinical Dementia Rating—Sum of Boxes and other Cognitive and Functional Outcomes in a Prodromal Alzheimer's Disease Population. *The Journal of Prevention of Alzheimer's Disease*, *8*(2), pp.151-160.

McInerney, T. and Terzopoulos, D., 1996. Deformable models in medical image analysis: a survey. *Medical image analysis*, *1*(2), pp.91-108.

McManus, C., 2019. Half a century of handedness research: Myths, truths; fictions, facts; backwards, but mostly forwards. *Brain and neuroscience advances*, *3*, p.2398212818820513.

Mehmood, A., Yang, S., Feng, Z., Wang, M., Ahmad, A.S., Khan, R., Maqsood, M. and Yaqub, M., 2021. A transfer learning approach for early diagnosis of Alzheimer’s disease on MRI images. *Neuroscience*, *460*, pp.43-52.

Meir, R. and Rätsch, G., 2003. An introduction to boosting and leveraging. In *Advanced lectures on machine learning* (pp. 118-183). Springer, Berlin, Heidelberg.

Michalak, H. and Okarma, K., 2019. Improvement of image binarization methods using image preprocessing with local entropy filtering for alphanumerical character recognition purposes. *entropy*, *21*(6), p.562.

Mikołajczyk, A. and Grochowski, M., 2018, May. Data augmentation for improving deep learning in image classification problem. In *2018 international interdisciplinary PhD workshop (IIPhDW)* (pp. 117-122). IEEE.

Milletari, F., Navab, N. and Ahmadi, S.A., 2016, October. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)* (pp. 565-571). IEEE.

Mirjalili, S., Mirjalili, S.M. and Lewis, A., 2014. Grey wolf optimizer. *Advances in engineering software*, *69*, pp.46-61.

Mitani, A., Huang, A., Venugopalan, S., Corrado, G.S., Peng, L., Webster, D.R., Hammel, N., Liu, Y. and Varadarajan, A.V., 2020. Detection of anaemia from retinal fundus images via deep learning. *Nature biomedical engineering*, *4*(1), pp.18-27.

Mittal, H. and Saraswat, M., 2020. A new fuzzy cluster validity index for hyper-ellipsoid or hyper-spherical shape close clusters with distant centroids. *IEEE Transactions on Fuzzy Systems*.

Mittal, H., Pandey, A.C., Saraswat, M., Kumar, S., Pal, R. and Modwel, G., 2021. A comprehensive survey of image segmentation: clustering methods, performance parameters, and benchmark datasets. *Multimedia Tools and Applications*, pp.1-26.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J. and Alzheimer's Disease Neuroimaging Initiative, 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage*, *104*, pp.398-412.

Moreno, J.C., Prasath, V.S., Proenca, H. and Palaniappan, K., 2014. Fast and globally convex multiphase active contours for brain MRI segmentation. *Computer Vision and Image Understanding*, *125*, pp.237-250.

Murero, M., Vita, S., Mennitto, A. and D’Ancona, G., Artificial Intelligence for Severe Speech Impairment: Innovative approaches to AAC and Communication.

Muthukrishnan, R. and Radha, M., 2011. Edge detection techniques for image segmentation. *International Journal of Computer Science & Information Technology*, *3*(6), p.259.

Nakawala, H., Ferrigno, G. and De Momi, E., 2017. Toward a knowledge-driven context-aware system for surgical assistance. *Journal of Medical Robotics Research*, *2*(03), p.1740007.

Nanni, L., Interlenghi, M., Brahnam, S., Salvatore, C., Papa, S., Nemni, R., Castiglioni, I. and Alzheimer's Disease Neuroimaging Initiative, 2020. Comparison of transfer learning and conventional machine learning applied to structural brain MRI for the early diagnosis and prognosis of Alzheimer's disease. *Frontiers in neurology*, p.1345.

Nanni, L., Paci, M., Santos, F.L.C.D., Brahnam, S., Hyttinen, J. and Alexander, S., 2016. Review on texture descriptors for image classification. *Computer Vision and Simulation: Methods, Applications and Technology*.

Nie, D., Lu, J., Zhang, H., Adeli, E., Wang, J., Yu, Z., Liu, L., Wang, Q., Wu, J. and Shen, D., 2019. Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. *Scientific reports*, *9*(1), pp.1-14.

Nitsch, J., Klein, J., Moltz, J.H., Miller, D., Sure, U., Kikinis, R. and Meine, H., 2019, March. Neural-network-based automatic segmentation of cerebral ultrasound images for improving image-guided neurosurgery. In *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling* (Vol. 10951, p. 109511N). International Society for Optics and Photonics.

Nogueira, J., Freitas, S., Duro, D., Tábuas-Pereira, M., Guerreiro, M., Almeida, J. and Santana, I., 2018. Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): Normative Data for the Portuguese Population. *Acta medica portuguesa*, *31*(2), pp.94-100.

Nyma, A., Kang, M., Kwon, Y.K., Kim, C.H. and Kim, J.M., 2012. A hybrid technique for medical image segmentation. *Journal of Biomedicine and Biotechnology*, *2012*.

O’Donnell, L.J. and Westin, C.F., 2011. An introduction to diffusion tensor image analysis. *Neurosurgery Clinics*, *22*(2), pp.185-196.

Oishi, K., Chang, L. and Huang, H., 2019. Baby brain atlases. *Neuroimage*, *185*, pp.865-880.

Ortiz, A., Gorriz, J.M., Ramirez, J. and Salas-Gonzalez, D., 2014. Improving MR brain image segmentation using self-organising maps and entropy-gradient clustering. *Information Sciences*, *262*, pp.117-136.

Park, S.C., Cha, J.H., Lee, S., Jang, W., Lee, C.S. and Lee, J.K., 2019. Deep learning-based deep brain stimulation targeting and clinical applications. *Frontiers in neuroscience*, *13*, p.1128.

Patel, N. and Singh, D., 2015. An algorithm to construct decision tree for machine learning based on similarity factor. *International Journal of Computer Applications*, *111*(10).

Perona, P. and Malik, J., 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, *12*(7), pp.629-639.

Piccialli, V. and Sciandrone, M., 2018. Nonlinear optimization and support vector machines. *4OR*, *16*(2), pp.111-149.

Pierre, F., Amendola, M., Bigeard, C., Ruel, T. and Villard, P.F., 2021. Segmentation with Active Contours. *Image Processing On Line*, *11*, pp.120-141.

Plataniotis, K.N. and Venetsanopoulos, A.N., 2013. *Colour image processing and applications*. Springer Science & Business Media.

Pohl, K.M., Fisher, J., Grimson, W.E.L., Kikinis, R. and Wells, W.M., 2006. A Bayesian model for joint segmentation and registration. *NeuroImage*, *31*(1), pp.228-239.

Poplin, R., Varadarajan, A.V., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G.S., Peng, L. and Webster, D.R., 2018. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, *2*(3), pp.158-164.

Porwik, P. and Lisowska, A., 2004. The Haar-wavelet transform in digital image processing: its status and achievements. *Machine graphics and vision*, *13*(1/2), pp.79-98.

Qiu, P., 2009. Image restoration. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(1), pp.110-113.

Qiu, S., Chang, G.H., Panagia, M., Gopal, D.M., Au, R. and Kolachalama, V.B., 2018. Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*, pp.737-749.

Quinlan, J.R., 1986. Induction of decision trees. *Machine learning*, *1*(1), pp.81-106.

Raileanu, L.E. and Stoffel, K., 2004. Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence*, *41*(1), pp.77-93.

Ranjitkar, H.S. and Karki, S., 2016. Comparison of A*, Euclidean and Manhattan distance using Influence map in MS. Pac-Man.

Remedios, S., Roy, S., Blaber, J., Bermudez, C., Nath, V., Patel, M.B., Butman, J.A., Landman, B.A. and Pham, D.L., 2019, March. Distributed deep learning for robust multi-site segmentation of CT imaging after traumatic brain injury. In *Medical Imaging 2019: Image Processing* (Vol. 10949, p. 109490A). International Society for Optics and Photonics.

Rezaei, M., Nafchi, H.Z. and Morales, S., 2013, October. Global Haar-like features: A new extension of classic Haar features for efficient face detection in noisy images. In *Pacific-Rim Symposium on Image and Video Technology* (pp. 302-313). Springer, Berlin, Heidelberg.

Ricard, J., Coeurjolly, D. and Baskurt, A., 2004, October. Generalization of angular radial transform. In *2004 International Conference on Image Processing, 2004. ICIP'04.* (Vol. 4, pp. 2211-2214). IEEE.

Ries, M.L., Carlsson, C.M., Rowley, H.A., Sager, M.A., Gleason, C.E., Asthana, S. and Johnson, S.C., 2008. Magnetic resonance imaging characterization of brain structure and function in mild cognitive impairment: a review. *Journal of the American Geriatrics Society*, *56*(5), pp.920-934.

Roberts, L., 1965. ``Machine perception of 3D solids”, Chapter 9 in JT Tippett, et al.(eds), Optical and Electro-Optical Information Processing.

Robnik-Šikonja, M. and Kononenko, I., 2003. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine learning*, *53*(1), pp.23-69.

Rohithkumar, A., Rajath, T. and Bipin Nair, B.J., 2020. Document Image Binarization. *Journal of Seybold Report ISSN NO*, *1533*, p.9211.

Rojas, R., 1996. The backpropagation algorithm. In *Neural networks* (pp. 149-182). Springer, Berlin, Heidelberg.

Rolls, E.T., Huang, C.C., Lin, C.P., Feng, J. and Joliot, M., 2020. Automated anatomical labelling atlas 3. *Neuroimage*, *206*, p.116189.

Romo, C., Conde, F.A., John, N.W. and Torres, J.C., 2017, September. Modeling Deformable Objects for Medical Training with Haptic Devices. In *2017 International Conference on Cyberworlds (CW)* (pp. 202-205). IEEE.

Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

Rosentzweig, A., 2019. Montreal Cognitive Assessment (MoCA) test for dementia.

Ruberto, C.D. and Fodde, G., 2013, September. Evaluation of statistical features for medical image retrieval. In *International Conference on Image Analysis and Processing* (pp. 552-561). Springer, Berlin, Heidelberg.

Ruder, S., 2016. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Rueda, A., Arevalo, J., Cruz, A., Romero, E. and González, F.A., 2012, September. Bag of features for automatic classification of Alzheimer's disease in magnetic resonance images. In *Iberoamerican Congress on Pattern Recognition* (pp. 559-566). Springer, Berlin, Heidelberg.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1986. Learning representations by back-propagating errors. *nature*, *323*(6088), pp.533-536.

Ruppert, G.C., Teverovskiy, L., Yu, C.P., Falcao, A.X. and Liu, Y., 2011, March. A new symmetry-based method for mid-sagittal plane extraction in neuroimages. In *2011 IEEE international symposium on biomedical imaging: from nano to macro* (pp. 285-288). IEEE.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. and Berg, A.C., 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*(3), pp.211-252.

Sabanayagam, C., Xu, D., Ting, D.S., Nusinovici, S., Banu, R., Hamzah, H., Lim, C., Tham, Y.C., Cheung, C.Y., Tai, E.S. and Wang, Y.X., 2020. A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *The Lancet Digital Health*, *2*(6), pp.e295-e302.

Salahat, E. and Qasaimeh, M., 2017, March. Recent advances in features extraction and description algorithms: A comprehensive survey. In *2017 IEEE international conference on industrial technology (ICIT)* (pp. 1059-1063). IEEE.

Salscheider, N.O., 2021. Object Tracking by Detection with Visual and Motion Cues. *arXiv preprint arXiv:2101.07549*.

Sánchez-Marono, N., Alonso-Betanzos, A. and Tombilla-Sanromán, M., 2007, December. Filter methods for feature selection–a comparative study. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 178-187). Springer, Berlin, Heidelberg.

Schmid, C., Mohr, R. and Bauckhage, C., 2000. Evaluation of interest point detectors. *International Journal of computer vision*, *37*(2), pp.151-172.

Schneider, K.M., 2005, February. Techniques for improving the performance of naive bayes for text classification. In *International conference on intelligent text processing and computational linguistics* (pp. 682-693). Springer, Berlin, Heidelberg.

Schober, P., Boer, C. and Schwarte, L.A., 2018. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), pp.1763-1768.

Schoenemann, P.T., 2006. Evolution of the size and functional areas of the human brain. *Annu. Rev. Anthropol.*, *35*, pp.379-406.

Segato, A., Marzullo, A., Calimeri, F. and De Momi, E., 2020. Artificial intelligence for brain diseases: A systematic review. *APL bioengineering*, *4*(4), p.041503.

Sengupta, S., Basak, S. and Peters, R.A., 2019. Particle Swarm Optimization: A survey of historical and recent developments with hybridization perspectives. *Machine Learning and Knowledge Extraction*, *1*(1), pp.157-191.

Sezgin, M. and Sankur, B., 2004. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, *13*(1), pp.146-165.

Shalev-Shwartz, S. and Ben-David, S., 2014. *Understanding machine learning: From theory to algorithms*. Cambridge university press.

Shannon, C.E., 1948. A mathematical theory of communication. *The Bell system technical journal, 27*(3), pp.379-423.

Shardlow, M., 2016. An analysis of feature selection techniques. *The University of Manchester*, *1*, pp.1-7.

Shaw, P., Lalonde, F., Lepage, C., Rabin, C., Eckstrand, K., Sharp, W., Greenstein, D., Evans, A., Giedd, J.N. and Rapoport, J., 2009. Development of cortical asymmetry in typically developing children and its disruption in attention-deficit/hyperactivity disorder. *Archives of general psychiatry*, *66*(8), pp.888-896.

Shrivakshan, G.T. and Chandrasekar, C., 2012. A comparison of various edge detection techniques used in image processing. *International Journal of Computer Science Issues (IJCSI)*, *9*(5), p.269.

Shusterman, E. and Feder, M., 1995. Image compression via Quadtree decomposition of Wavelets. In *Conference Record of The Twenty-Ninth Asilomar Conference on Signals, Systems and Computers* (Vol. 2, pp. 1424-1428). IEEE.

Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sinforiani, E., Citterio, A., Zucchella, C., Bono, G., Corbetta, S., Merlo, P. and Mauri, M., 2010. Impact of gender differences on the outcome of Alzheimer's disease. *Dementia and geriatric cognitive disorders*, *30*(2), pp.147-154.

Singh, S.A. and Majumder, S., 2020. Short and noisy electrocardiogram classification based on deep learning. In *Deep Learning for Data Analytics* (pp. 1-19). Academic Press.

Snider, B., Phillips, P., MacLean, A., McBean, E.A., Gadsden, A. and Yawney, J., 2020. Artificial intelligence to predict the risk of mortality from COVID-19: Insights from a Canadian Application. *medRxiv*.

Song, Z., Sun, J. and Yu, J., 2017. Object Tracking by a Combination of Discriminative Global and Generative Multi-Scale Local Models. *Information, 8*(2), p.43.

Spasov, S., Passamonti, L., Duggento, A., Lio, P., Toschi, N. and Alzheimer's Disease Neuroimaging Initiative, 2019. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *Neuroimage*, *189*, pp.276-287.

Sprawls, P., 2000. *Magnetic resonance imaging: principles, methods, and techniques*. Madison, Wisconsin: Medical Physics Publishing.

Srinidhi, C.L., Ciga, O. and Martel, A.L., 2020. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, p.101813.

Stamate, D., Smith, R., Tsygancov, R., Vorobev, R., Langham, J., Stahl, D. and Reeves, D., 2020, June. Applying Deep Learning to Predicting Dementia and Mild Cognitive Impairment. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 308-319). Springer, Cham.

Štencel, M.I.K.U.L.Á.Š., Janáček, J.I.Ř.Í., Lechnerová, R., Saxl, I. and Beneš, V., 2006. On calculation of chamfer distance and Lipschitz covers in digital images. *Proceedings S4G. Prague: Union of Czech Mathematicians and Physicists*, pp.517-522.

Strongman, H., Williams, R., Meeraus, W., Murray-Thomas, T., Campbell, J., Carty, L., Dedman, D., Gallagher, A.M., Oyinlola, J., Kousoulis, A. and Valentine, J., 2019. Limitations for health research with restricted data collection from UK primary care. *Pharmacoepidemiology and drug safety*, *28*(6), pp.777-787.

Štruc, V. and Pavešić, N., 2017. Image normalization techniques for robust face recognition. In *Proceedings of the 8th WSEAS International Conference on Signal Processing Robotics and Automation* (pp. 155-160).

Stulp, F. and Sigaud, O., 2015. Many regression algorithms, one unified model: A review. *Neural Networks*, *69*, pp.60-79.

Sundaresan, V., Zamboni, G., Le Heron, C., Rothwell, P.M., Husain, M., Battaglini, M., De Stefano, N., Jenkinson, M. and Griffanti, L., 2019. Automated lesion segmentation with BIANCA: impact of population-level features, classification algorithm and locally adaptive thresholding. *NeuroImage*, *202*, p.116056.

Szaflarski, J.P., Rajagopal, A., Altaye, M., Byars, A.W., Jacola, L., Schmithorst, V.J., Schapiro, M.B., Plante, E. and Holland, S.K., 2012. Left-handedness and language lateralization in children. *Brain research*, *1433*, pp.85-97.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

Szeliski, R., 2010. *Computer vision: algorithms and applications*. Springer Science & Business Media.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B. and Liang, J., 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning?. *IEEE transactions on medical imaging*, *35*(5), pp.1299-1312.

Tang, Y., 2013. Deep learning using linear support vector machines. *arXiv preprint arXiv:1306.0239*.

Teverovskiy, L.A., Becker, J.T., Lopez, O.L. and Liu, Y., 2008, May. Quantified brain asymmetry for age estimation of normal and AD/MCI subjects. In *2008 5th IEEE international symposium on biomedical imaging: From nano to macro* (pp. 1509-1512). IEEE.

Teverovskiy, L. and Li, Y., 2006, April. Truly 3D midsagittal plane extraction for robust neuroimage registration. In *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.* (pp. 860-863). IEEE.

Tharwat, A., Gaber, T., Ibrahim, A. and Hassanien, A.E., 2017. Linear discriminant analysis: A detailed tutorial. *AI communications*, *30*(2), pp.169-190.

Thung, K.H. and Raveendran, P., 2009, December. A survey of image quality measures. In *2009 international conference for technical postgraduates (TECHPOS)* (pp. 1-4). IEEE.

Toga, A.W. and Thompson, P.M., 2003. Mapping brain asymmetry. *Nature Reviews Neuroscience*, *4*(1), pp.37-48.

Tomasi, D. and Volkow, N.D., 2012. Laterality patterns of brain functional connectivity: gender effects. *Cerebral Cortex*, *22*(6), pp.1455-1462.

Tonutti, M., Gras, G. and Yang, G.Z., 2017. A machine learning approach for real-time modelling of tissue deformation in image-guided neurosurgery. *Artificial intelligence in medicine*, *80*, pp.39-47.

Treiman, J.S., 2014. Rules for Finding Derivatives. In *Calculus with Vectors* (pp. 87-120). Springer, Cham.

Triggs, B., 2004, May. Detecting keypoints with stable position, orientation, and scale under illumination changes. In *European conference on computer vision* (pp. 100-113). Springer, Berlin, Heidelberg.

Tripepi, G., Jager, K.J., Dekker, F.W. and Zoccali, C., 2008. Linear and logistic regression analysis. *Kidney international*, *73*(7), pp.806-810.

Tuytelaars, T. and Mikolajczyk, K., 2008. *Local invariant feature detectors: a survey*. Now Publishers Inc.

Tuytelaars, T., Turina, A. and Van Gool, L., 2003. Noncombinatorial detection of regular repetitions under perspective skew. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(4), pp.418-432.

Udayakumar, E., Santhi, S., Gowrishankar, R., Ramesh, C. and Gowthaman, T., 2016. Region growing image segmentation for newborn brain MRI. *Biotechnology: An Indian Journal Trade Sci Inc*, *12*, pp.1-8.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A. and Bottou, L., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, *11*(12).

Viola, P. and Jones, M., 2001, December. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001* (Vol. 1, pp. I-I). IEEE.

Wang, G., Wu, M., Wei, X. and Song, H., 2020. Water identification from high-resolution remote sensing images based on multidimensional densely connected convolutional neural networks. *Remote Sensing*, *12*(5), p.795.

Wang, Y.T., Hung, D.Y. and Cheng, S.H., 2011. Monocular SLAM for a small-size humanoid robot. *Journal of Applied Science and Engineering*, *14*(2), pp.123-129.

Watanabe, T. and Wolf, D.F., 2019, June. Instance Segmentation as Image Segmentation Annotation. In *2019 IEEE Intelligent Vehicles Symposium (IV)* (pp. 432-437). IEEE.

Weiskopf, N., Lutti, A., Helms, G., Novak, M., Ashburner, J. and Hutton, C., 2011. Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). *Neuroimage*, *54*(3), pp.2116-2124.

Weissler, E.H., Naumann, T., Andersson, T., Ranganath, R., Elemento, O., Luo, Y., Freitag, D.F., Benoit, J., Hughes, M.C., Khan, F. and Slater, P., 2021. The role of machine learning in clinical research: transforming the future of evidence generation. *Trials*, *22*(1), pp.1-15.

Welling, M., 2005. Fisher Linear Discriminant Analysis‖ University of Toronto. *Technical Note*.

Willemink, M.J., Koszek, W.A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L.R., Summers, R.M., Rubin, D.L. and Lungren, M.P., 2020. Preparing medical imaging data for machine learning. *Radiology*, *295*(1), pp.4-15.

Williams, C.M., Peyre, H., Toro, R. and Ramus, F., 2022. Comparing brain asymmetries independently of brain size. *NeuroImage*, *254*, p.119118.

Witter, M.P., Doan, T.P., Jacobsen, B., Nilssen, E.S. and Ohara, S., 2017. Architecture of the entorhinal cortex a review of entorhinal anatomy in rodents with some comparative notes. *Frontiers in systems neuroscience*, *11*, p.46.

Wolpert, D.H. and Macready, W.G., 1997. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, *1*(1), pp.67-82.

Wong, D.F., Maini, A., Rousset, O.G. and Brašić, J.R., 2003. Positron emission tomography: a tool for identifying the effects of alcohol dependence on the brain. *Alcohol Research & Health*, *27*(2), p.161.

Xu, C., Pham, D.L. and Prince, J.L., 2000. Image segmentation using deformable models. *Handbook of medical imaging*, *2*(20), p.0.

Yagis, E., Citi, L., Diciotti, S., Marzi, C., Atnafu, S.W. and De Herrera, A.G.S., 2020, July. 3D Convolutional Neural Networks for Diagnosis of Alzheimer's Disease via Structural MRI. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 65-70). IEEE.

Yamanakkanavar, N., Choi, J.Y. and Lee, B., 2020. MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer’s disease: a survey. *Sensors*, *20*(11), p.3243.

Yamashita, R., Nishio, M., Do, R.K.G. and Togashi, K., 2018. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, *9*(4), pp.611-629.

Yanase, J. and Triantaphyllou, E., 2019. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Systems with Applications*, *138*, p.112821.

Yang, D., Hong, K.S., Yoo, S.H. and Kim, C.S., 2019. Evaluation of neural degeneration biomarkers in the prefrontal cortex for early identification of patients with mild cognitive impairment: an fNIRS study. *Frontiers in human neuroscience*, *13*, p.317.

Yazdani, S., Yusof, R., Karimian, A., Mitsukira, Y. and Hematian, A., 2016. Automatic region-based brain classification of MRI-T1 data. *PloS one*, *11*(4), p.e0151326.

Yodogawa, E., 1982. Symmetropy, an entropy-like measure of visual symmetry. *Perception & Psychophysics*, *32*(3), pp.230-240.

Yousefi, J., 2011. Image binarization using otsu thresholding algorithm. *Ontario, Canada: University of Guelph*.

Yu, C., Liu, J., Nemati, S. and Yin, G., 2021. Reinforcement learning in healthcare: A survey. *ACM Computing Surveys (CSUR)*, *55*(1), pp.1-36.

Yucel, K., McKinnon, M., Chahal, R., Taylor, V., Macdonald, K., Joffe, R. and MacQueen, G., 2009. Increased subgenual prefrontal cortex size in remitted patients with major depressive disorder. *Psychiatry Research: Neuroimaging*, *173*(1), pp.71-76.

Yücel, M., Stuart, G.W., Maruff, P., Wood, S.J., Savage, G.R., Smith, D.J., Crowe, S.F., Copolov, D.L., Velakoulis, D. and Pantelis, C., 2002. Paracingulate morphologic differences in males with established schizophrenia: a magnetic resonance imaging morphometric study. *Biological psychiatry*, *52*(1), pp.15-23.

Yücel, M., Wood, S.J., Phillips, L.J., Stuart, G.W., Smith, D.J., Yung, A., Velakoulis, D., McGorry, P.D. and Pantelis, C., 2003. Morphology of the anterior cingulate cortex in young men at ultra-high risk of developing a psychotic illness. *The British Journal of Psychiatry*, *182*(6), pp.518-524.

Zaudig, M., 1992. A new systematic method of measurement and diagnosis of "mild cognitive impairment" and dementia according to ICD-10 and DSM-III-R criteria. *International Psychogeriatrics*, *4*(4), pp.203-219.

Zhang, A., Lipton, Z.C., Li, M. and Smola, A.J., 2021. Dive into deep learning. *arXiv preprint arXiv:2106.11342*.

Zhang, D., Shen, D. and Alzheimer's Disease Neuroimaging Initiative, 2012. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage*, *59*(2), pp.895-907.

Zhang, J., Gajjala, S., Agrawal, P., Tison, G.H., Hallock, L.A., Beussink-Nelson, L., Lassen, M.H., Fan, E., Aras, M.A., Jordan, C. and Fleischmann, K.E., 2018. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation*, *138*(16), pp.1623-1635.

Zhang, J., Zhong, Y. and Gu, C., 2017. Deformable models for surgical simulation: a survey. *IEEE reviews in biomedical engineering*, *11*, pp.143-164.

Zhang, Y., 1993. Deformable model for image segmentation. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY AND REMOTE SENSING*, *29*, pp.720-720.

Zheng, A. and Casari, A., 2018. *Feature engineering for machine learning: principles and techniques for data scientists*. " O'Reilly Media, Inc.".

Zhou, D., Lebel, C., Evans, A. and Beaulieu, C., 2013. Cortical thickness asymmetry from childhood to older adulthood. *Neuroimage*, *83*, pp.66-74.

Zhou, K., He, W., Xu, Y., Xiong, G. and Cai, J., 2018. Feature selection and transfer learning for Alzheimer's disease clinical diagnosis. *Applied Sciences*, *8*(8), p.1372.

Zhu, X., Suk, H.I., Wang, L., Lee, S.W., Shen, D. and Alzheimer's Disease Neuroimaging Initiative, 2017. A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. *Medical image analysis*, *38*, pp.205-214.

Zou, K., Abdullah, M. and Michikawa, M., 2020. Current biomarkers for Alzheimer's disease: From CSF to blood. *Journal of Personalized Medicine*, *10*(3), p.85.

Appendix A: Matlab code for image processing and classification

MRI brain preprocessing and segmentation

```
close all
clc
clear
%Load and read image from a file
[filename,pathname] =
uigetfile({'*.*';'*.jpg';'*.bmp';'*.tif';'*.gif';'*.png'},'Pick an
Image File');
[im, map] = imread([pathname,filename]);

%Resize, modify and adjust an image
I = imresize(im,[256,256]);
R = I(:,:,1);
G = I(:,:,2);
B = I(:,:,3);
newI = zeros(size(I,1),size(I,2),'uint8');
for x=1:size(I,1)
for y=1:size(I,2)
newI(x,y)=(R(x,y)*.3)+(G(x,y)*.6)+(B(x,y)*.1);
end
end
returnedI = newI;
figure, imshow(returnedI);
img = imadjust(returnedI);%image normalize [0,1]
figure, imshow(img);

%% Brain partition
img(195:end, :) = 0;
figure,imshow(imadjust(img));

%% Threshold
lb = 80;
ub = 270;
mriAdjust = img;
mriAdjust(mriAdjust <= lb) = 0;
mriAdjust(mriAdjust >=ub) = 0;
mriAdjust(195:end, :) = 0;
bw = logical(mriAdjust);
figure,imshow(bw);

%% Find centroid
L = bwlabeln(bw);
stats = regionprops(L, 'Area', 'Centroid');
LL = L + 1;
cmap = hsv(length(stats)); cmap = [0 0 0;cmap];
sizeI = size(img);
LL = cmap(LL, :);LL = reshape(LL, [sizeI, 3]);
figure, imshow(LL);

%% Find the largest blob
A = [stats.Area];
biggest = find(A ==max(A));
mriAdjust(L ~= biggest) = 0;
imA = imadjust(mriAdjust);
figure, imshow(imA);
```

Segmentation of brain asymmetry and statistical features collection

```
%%Process segmented brain images (find centroid of the brain,
%%translate the brain to the center of the image, rotate the brain,
find asymmetry between the right and the left parts of the brain) and
calculate the statistics from the image asymmetry
close all
clc
clear
%Load and read image from a file
[filename,pathname] =
uigetfile({'*.*';'*.jpg';'*.bmp';'*.tif';'*.gif';'*.png'},'Pick an
Image File');
[I, map] = imread([pathname,filename]);
%%
% Get the dimensions of the image.
% numberOfColourChannels should be = 1 for a grey scale image, and 3
for an RGB colour image
% Convert the image from 3 to 2 dimensional
[rows, columns, numberOfColourChannels] = size(I);
R = I(:,:,1);
G = I(:,:,2);
B = I(:,:,3);
newI = zeros(size(I,1),size(I,2),'uint8');
for x=1:size(I,1)
for y=1:size(I,2)
newI(x,y)=(R(x,y)*.3)+(G(x,y)*.6)+(B(x,y)*.1);
end
end
bin = newI;

%% Binary image with centroid and axes of symmetry
binaryImage = logical(bin);

% Label the image
labeledImage = logical(binaryImage);

% Make the measurements
stats = regionprops('table',labeledImage,'Centroid',...
    'MajorAxisLength','MinorAxisLength','Orientation');
xCentroid = stats.Centroid(1);
yCentroid = stats.Centroid(2);

% Find the halfway point of the image
middlex = columns/2;
middley = rows/2;

%% Translate the image
deltax = middlex - xCentroid;
deltay = middley - yCentroid;
translatedImage = imtranslate(I,[deltax, deltay],'OutputView','full');

%% Rotate the image
angle = -stats.Orientation;
rotatedImage = imrotate(translatedImage, -0.14, 'crop');

%% Find asymmetry in the image
F = fliplr(rotatedImage);
D = imabsdiff(rotatedImage,F);
```

```
% Display the image asymmetry
figure,imshow(D);

%% MSE
MSE = immse(rotatedImage,F);
%% Discrete Wavelets transform from image differences:
signal1 = D(:,:,:);
[cA1,cH1,cV1,cD1] = dwt2(signal1,'db4');
[cA2,cH2,cV2,cD2] = dwt2(cA1,'db4');
[cA3,cH3,cV3,cD3] = dwt2(cA2,'db4');
DWT_feat = [cA3,cH3,cV3,cD3];%3? level of DWT
%% 1-st and 2-nd level of statistics:
Mean = mean2(DWT_feat);
Standard_Deviation = std2(DWT_feat);
Entropy = entropy(DWT_feat);
RMS = mean2(rms(DWT_feat));%Root-mean-square level
Variance = mean2(var(double(DWT_feat)));
a = sum(double(DWT_feat(:)));
Smoothness = 1-(1/(1+a));
Kurtosis = kurtosis(double(DWT_feat(:)));
Skewness = skewness(double(DWT_feat(:)));
% Inverse Difference Movement:
m = size(DWT_feat,1);
n = size(DWT_feat,2);
in_diff = 0;
for i = 1:m
    for j = 1:n
        temp = DWT_feat(i,j)./(1+(i-j).^2);
        in_diff = in_diff+temp;
    end
end
IDM = double(in_diff);

%% Vector of statistics values calculated from an image asymmetry
V =
[MSE,Mean,Standard_Deviation,Entropy,RMS,Variance,Smoothness,Kurtosis,
Skewness,IDM];
%% Plot data
%load StatTbl.mat;
%gplotmatrix(V,[],label);
```

Extraction and collection the Bag-of-Features

```
%Load image sets
outputFolder = fullfile('\\winfs\home\nitsa\My Documents\MCI\DBO
Dementia Test');
rootFolder = fullfile('\\winfs\home\nitsa\My Documents\MCI\DBO
Dementia Test');
imgSets = [ imageSet(fullfile(rootFolder, 'ADaxial_sym')), ...
            imageSet(fullfile(rootFolder, 'EMCIaxial_sym')), ...
            imageSet(fullfile(rootFolder, 'NCaxial_sym')) ];
        minSetCount = min([imgSets.Count]); %count the sets nmber
        imgSets = partition(imgSets, minSetCount, 'randomize');

%partition sets according sets data
        [trainingSets, validationSets] = partition(imgSets, 0.3,
'randomize');%partition sets for the training and validation data
        bag = bagOfFeatures(trainingSets);%creat the visual vocabulary
for each image in the datasets (SURF extraction and K-means
clustering)
```

```
%img = imread('\\winfs\home\nitsa\My Documents\MCI\DBO Dementia
Test\AD\ad1.png');
%featureVector = encode(bag, img);

% Plot the histogram of visual word occurrences
%figure
%bar(featureVector)
%title('Visual word occurrences')
%xlabel('Visual word index')
%ylabel('Frequency of occurrence')
```

Training Bayesian classifier and creating the confusion matrix

```
%%Train Bayesian Classifier
load Bigdata.mat %Load features table
X = featureData;    % Predictors
Y = label; % Response
Mdl = fitcnb(X,Y,...
    'ClassNames',{'AD','EMCI','NC'});
%Result prediction, displated for 10 observations
lbl = resubPredict(Mdl);
rng(1); % Control random number generation (for reproducibility)

%Confusion matrix
ConfusionMat = confusionmat(Y,lbl);
%Plot confusion
labels = {'AD', 'EMCI', 'NC'};
numlabels = size(ConfusionMat,1);
imagesc(ConfusionMat);
title(sprintf('Confusion Matrix. Accuracy: %.2f%%',
100*trace(ConfusionMat)/sum(ConfusionMat(:))));
ylabel('Target Class'); xlabel(' Output Class');
%colourmap(flipud(grey);
colourmap(flipud(parula));
textStrings = num2str([ConfusionMat(:),0.01*ConfusionMat(:)],
'%.1f%%\n%.2f\n');
textStrings = strtrim(cellstr(textStrings));
[x,y] = meshgrid(1:numlabels);
hStrings = text(x(:),y(:),textStrings(:), ...
    'HorizontalAlignment','center');
midValue = mean(get(gca,'CLim'));
textColours = repmat(ConfusionMat(:) > midValue,1,3);
set(hStrings,{'Colour'},num2cell(textColours,2));
set(gca,'XTick',1:numlabels,...
    'XTickLabel',labels,...
    'YTick',1:numlabels,...
    'YTickLabel',labels,...
    'TickLength',[0 0]);
```

Creating compact Bayesian classifier

```
load Bigdata.mat
X = featureData;
Y = label;
rng(1);
n = size(X,1);
%Partition the data set into two sets: one is the training set,
%and the other is new unobserved data. Reserve 10 observations for the
new data set.
```

```
newInds = randsample(n,10);
inds = ~ismember(1:n,newInds);
XNew = X(newInds,:);
YNew = Y(newInds);
%Creating standard and compact classifier.
Mdl = fitcnb(X(inds,:),Y(inds),...
    'ClassNames',{'AD','EMCI','NC'});
CMdl = compact(Mdl);

%whos('Mdl','CMdl')
%CMdl.ClassNames
%Estimation posterior probability and misclassification cost
[lbls,PostProbs,MisClassCost] = predict(CMdl,XNew);
table(YNew,lbls,PostProbs,'VariableNames',...
    {'TrueLabels','PredictedLabels',...
    'PosteriorProbabilities'});
%MisClassCost
%Test any image from outside the image sets
% load FV.mat;
% lblTest = predict(CMdl,FV);
%FV is an image feature vector, created by Bag-Of-Features, and
converted to double.
%From the beginning the image should be processed:
%imread, resize, conver to greyscale, adjust and only after that it
should be sent to the Bad-of-Features.
```

Cross-validation procedure for Bayesian classifier

```
load Bags_Of_Ax_NormPlusSym.mat
X = T1;
Y =label;
rng(1);% For reproducibility
%Train and cross-validate a naive Bayes classifier using the default
options and k-fold cross-validation.
%CVMdl1 is a ClassificationPartitionedModel.
CVMdl1 = fitcnb(X,Y,...
    'ClassNames',{'AD','EMCI','NC'},...
    'CrossVal','on');
%Create a default naive Bayes binary classifier template, and train an
error-correcting, output codes multiclass model.
t = templateNaiveBayes();
CVMdl2 = fitcecoc(X,Y,'CrossVal','on','Learners',t);%CVMdl2 is a
ClassificationPartitionedECOC model.
%Compare the out-of-sample k-fold classification error.
classErr1 = kfoldLoss(CVMdl1,'LossFun','ClassifErr');
classErr2 = kfoldLoss(CVMdl2,'LossFun','ClassifErr');
```

New data labelling

```
%Load and read image from a file
[filename,pathname] =
uigetfile({'*.*';'*.jpg';'*.bmp';'*.tif';'*.gif';'*.png'},'Pick an
Image File');
[im, map] = imread([pathname,filename]);

%Resize, modify and adjust an image
I = imresize(im,[256,256]);
R = I(:,:,1);
G = I(:,:,2);
B = I(:,:,3);
```

```
newI = zeros(size(I,1),size(I,2),'uint8');
for x=1:size(I,1)
for y=1:size(I,2)
newI(x,y)=(R(x,y)*.3)+(G(x,y)*.6)+(B(x,y)*.1);
end
end
returnedI = newI;
img = imadjust(returnedI);%image normalize [0,1]
load Bag.mat % load Bag-of-Features

featureVector = encode(bag, img);%create feature vector for image
vector = double(featureVector);
%First you need to train classifier Mdl and create compact classifier
CMdl
load Bigdata.mat
X = featureData;
Y = label;
rng(1);
n = size(X,1);
newInds = randsample(n,10);
inds = ~ismember(1:n,newInds);
XNew = X(newInds,:);
YNew = Y(newInds);
Mdl = fitcnb(X(inds,:),Y(inds),...
    'ClassNames',{'AD','EMCI','NC'});
CMdl = compact(Mdl);

%Label unknown image
lblTest = predict(CMdl,vector);
```

CNN tests with AlexNet and VGG16 using transfer learning

```
datapath='Images';
imds=imageDatastore(datapath, ...
'IncludeSubfolders',true, ...
'LabelSource','foldernames');
total_split=countEachLabel(imds);
train_percent=0.80;
[imdsTrain,imdsTest]=splitEachLabel(imds,train_percent,'randomize');
valid_percent=0.1;
[imdsValid,imdsTrain]=splitEachLabel(imdsTrain,valid_percent,'randomiz
e');
train_split=countEachLabel(imdsTrain);
%net=vgg16();
net=alexnet;
%analyzeNetwork(net);
layersTransfer=net.Layers(1:end-3);
clear net;
numClasses=numel(categories(imdsTrain.Labels));
layers=[
    layersTransfer
fullyConnectedLayer(numClasses,'Name','fc','WeightLearnRateFactor',20,
'BiasLearnRateFactor',20)
    softmaxLayer('Name', 'softmax')
    classificationLayer('Name', 'classOutput')];
lgraph = layerGraph(layers);
plot(lgraph);
imdsTrain.ReadFcn=@(filename)preprocess_images(filename,[layers(1).Inp
utSize(1), layers(1).InputSize(2)]);
```

```
imdsValid.ReadFcn=@(filename)preprocess_images(filename,[layers(1).Inp
utSize(1), layers(1).InputSize(2)]);
options=trainingOptions('sgdm', ...
    'MiniBatchSize',128, ...
    'MaxEpochs',11, ...
    'Shuffle','every-epoch', ...
    'InitialLearnRate',1e-4, ...
    'ValidationData',imdsValid, ...
    'ValidationFrequency',50,'ValidationPatience',4, ...
    'Verbose',false, ...
    'Plots','training-progress');
netTransfer=trainNetwork(imdsTrain,layers,options);
imdsTest.ReadFcn=@(filename)preprocess_images(filename,[layers(1).Inpu
tSize(1), layers(1).InputSize(2)]);
[predicted_labels,posterior]=classify(netTransfer,imdsTest);
actual_labels=imdsTest.Labels;
ConfusionMat = confusionmat(actual_labels,predicted_labels);
ConfusionMat =
bsxfun(@rdivide,ConfusionMat,sum(ConfusionMat,2));%Convert into % form
mean(diag(ConfusionMat));%Display the mean accuracy
test_labels=double(nominal(imdsTest.Labels));
[fp_rate,tp_rate,T,AUC]=perfcurve(test_labels,posterior(:,1),1);
figure;
plot(fp_rate,tp_rate,'b-');hold on;
grid on;
xlabel('False Positive Rate');
ylabel('Detection Rate');

%train CVM
%layer = 'fc';
%featuresTrain =
activations(netTransfer,imdsTrain,layer,'OutputAs','rows');
%featuresTest =
activations(netTransfer,imdsTest,layer,'OutputAs','rows');
%YTrain = imdsTrain.Labels;
%YTest = imdsTest.Labels;
%mdl = fitcsvm(featuresTrain,YTrain);
%YPred = predict(mdl,featuresTest);
%accuracy = mean(YPred == YTest);

%mdl =
fitcsvm(featuresTrain,YTrain,'Standardize',true,'KernelFunction','RBF'
,...
  %  'KernelScale','auto'); % for binary output
%mdl = fitcecoc(featuresTrain,YTrain);% for multiclass output
```

Image preprocessing for CNN tests

```
function Iout=preprocess_images(filename,desired_size)
% This function preprocesses images using colour constancy
% technique and later reshapes them to an image of desired size
% Author: Barath Narayanan

% Read the Image
I=imread(filename);
% Some images might be greyscale, replicate the image 3 times to
% create an RGB image.
if ismatrix(I)
  I=cat(3,I,I,I);
end
```

```
% Conversion to Double for calculation purposes
I=double(I);
% Mean Calculation
Ir=I(:,:,1);mu_red=mean(Ir(:));
Ig=I(:,:,2);mu_green=mean(Ig(:));
Ib=I(:,:,3);mu_blue=mean(Ib(:));
mean_value=(mu_red+mu_green+mu_blue)/3;
% Scaling the Image for Colour constancy
Iout(:,:,1)=I(:,:,1)*mean_value/mu_red;
Iout(:,:,2)=I(:,:,2)*mean_value/mu_green;
Iout(:,:,3)=I(:,:,3)*mean_value/mu_blue;
% Converting it back to uint8
Iout=uint8(Iout);

% Resize the image
Iout=imresize(Iout,[desired_size(1) desired_size(2)]);
end
```

Proposed 5-layers CNN

```
datapath='Images';
imds=imageDatastore(datapath, ...
'IncludeSubfolders',true, ...
'LabelSource','foldernames');
total_split=countEachLabel(imds);
train_percent=0.80;
[imdsTrain,imdsTest]=splitEachLabel(imds,train_percent,'randomize');
valid_percent=0.1;
[imdsValid,imdsTrain]=splitEachLabel(imdsTrain,valid_percent,'randomiz
e');
train_split=countEachLabel(imdsTrain);
numClasses=numel(categories(imdsTrain.Labels));
layers = [
    imageInputLayer([256 256 3],'Name','input')
    convolution2dLayer(5,16,'Padding','same','Name','conv_1')
    batchNormalizationLayer('Name','BN_1')
    reluLayer('Name','relu_1')

convolution2dLayer(3,32,'Padding','same','Stride',2,'Name','conv_2')
    batchNormalizationLayer('Name','BN_2')
    reluLayer('Name','relu_2')
    convolution2dLayer(3,32,'Padding','same','Name','conv_3')
    batchNormalizationLayer('Name','BN_3')
    reluLayer('Name','relu_3')
    averagePooling2dLayer(2,'Stride',2,'Name','avpool_1')

convolution2dLayer(3,64,'Padding','same','Stride',2,'Name','conv_4')
    batchNormalizationLayer('Name','BN_4')
    reluLayer('Name','relu_4')
    convolution2dLayer(3,64,'Padding','same','Name','conv_5')
    batchNormalizationLayer('Name','BN_5')
    reluLayer('Name','relu_5')
    averagePooling2dLayer(2,'Stride',2,'Name','avpool_2')
    dropoutLayer('Name','drop')
    fullyConnectedLayer(2,'Name','fc')
    softmaxLayer('Name','softmax')
    classificationLayer('Name','classOutput')];
lgraph = layerGraph(layers);
figure
plot(lgraph);
```

```
imdsTrain.ReadFcn=@(filename)preprocess_images(filename,[layers(1).Inp
utSize(1), layers(1).InputSize(2)]);
imdsValid.ReadFcn=@(filename)preprocess_images(filename,[layers(1).Inp
utSize(1), layers(1).InputSize(2)]);
options=trainingOptions('sgdm', ...
    'MiniBatchSize',128, ...
    'MaxEpochs',11, ...
    'Shuffle','every-epoch', ...
    'InitialLearnRate',1e-4, ...
    'ValidationData',imdsValid, ...
    'ValidationFrequency',50,'ValidationPatience',4, ...
    'Verbose',false, ...
    'Plots','training-progress');
net=trainNetwork(imdsTrain,layers,options);
imdsTest.ReadFcn=@(filename)preprocess_images(filename,[layers(1).Inpu
tSize(1), layers(1).InputSize(2)]);
[predicted_labels,posterior]=classify(net,imdsTest);
actual_labels=imdsTest.Labels;
ConfusionMat = confusionmat(actual_labels,predicted_labels);
ConfusionMat = bsxfun(@rdivide,ConfusionMat,sum(ConfusionMat,2));
mean(diag(ConfusionMat));
test_labels=double(nominal(imdsTest.Labels));
[fp_rate,tp_rate,T,AUC]=perfcurve(test_labels,posterior(:,1),1);
figure;
plot(fp_rate,tp_rate,'b-');hold on;
grid on;
xlabel('False Positive Rate');
ylabel('Detection Rate');

%SVM layer
%layer = 'fc';
%featuresTrain = activations(net,imdsTrain,layer,'OutputAs','rows');
%featuresTest = activations(net,imdsTest,layer,'OutputAs','rows');
%YTrain = imdsTrain.Labels;
%YTest = imdsTest.Labels;

%mdl = fitcsvm(featuresTrain,YTrain);
%YPred = predict(mdl,featuresTest);
%accuracy = mean(YPred == YTest);

%mdlLinear = fitcdiscr(featuresTrain,YTrain);
%YPredL = predict(mdlLinear,featuresTest);
%accuracyL = mean(YPredL == YTest);

%mdlKNN = fitcknn(featuresTrain,YTrain);
%YPred = predict(mdlKNN,featuresTest);
%accuracyK = mean(YPred == YTest);
```

Appendix B: List of abbreviations

AAL - Ambient Assisted Living

AAL - Automated Anatomical Labelling

Ab42 - Amyloid-beta 42

AD - Alzheimer’s Disease

ADAS Alzheimer’s Disease Assessment Scale

ADAS-Cog - AD Assessment Scale-Cognitive Subscale

ADHD - attention deficit hyperactivity disorder

ADNI - Alzheimer’s Disease Neuroimaging Initiative

AE - autoencoder

AI - Adobe Illustrator

AIBL - Australian Imaging Biomarkers and Lifestyle Study of Aging

aMCI - amnestic Mild Cognitive Impairment

ANN - artificial neural network

AR - autoregressive

ART - Angular Radial Transform

ASD - autistic spectrum disorder

ASM - angular second moment

AUC - area under the curve

AUROC - Area Under the Receiving Operating Curve

BN - batch normalization

BOF - Bag-of-Features

BraTS - Brain Tumor Segmentation

CART - Classification and Regression Tree

CC - Corpus Callosum

CDR - Clinical Dementia Ratio

CIS - cingulate island sign

CLD - colour layout descriptor

CNN - Convolutional Neural Network

ConvBLSTM - Convolutional Bidirectional Long Short-Term Memory

CS - Cuckoo Search

CSF - cerebrospinal fluid

CSS - Curvature Scale Space

CT - computer tomography

CV – Computer Vision

DenseNet - Densely connected networks

Dicom - Digital Imaging and Communications in Medicine

DLB - dementia with Lewy bodies

DNN deep neural network

DPM - Deformable Part Model

DT - Decision Tree

DTI - Diffusion Tensor Imaging

DTR - dynamic treatment regimes

EEG – Electroencephalography

EMCI - early mild cognitive impairment

EPS - Encapsulated PostScript

FC - fully connected

FD - Fourier descriptors

FDA - Food and Drug Administration

FDG-PET - fluorodeoxyglucose positron emission tomography

fMRI - functional Magnetic Resonance Imaging

FN – False Negative

fNIRS - functional near-infrared spectroscopy

FR – False Positive

FSIM - Feature Similarity Index Metrics

GA - Genetic Algorithm

GF - Gabor filters

gICA - group independent component analysis

GIF - Graphics Interchange Format

GLCM - grey level co-occurrence matrix

GM - grey matter

GMS - Geriatric Mental State Examination

GradCAM - Gradient-weighted class activation mapping

GRU - Gated Recurrent Units

GWO - Grey Wolf Optimization

HC – healthy controls

HC - Hippocampus

HIS - hue-saturation-intensity

HMMD - hue-max-min-difference

HOG - histogram of oriented gradients

HOGM - High-Order Graph Matching

HSV - hue-saturation-value

ICBM - Internet Brain Segmentation Repository

IDM - Inverse difference moment

ILSVRC - ImageNet Large Scale Visual Recognition Competition

IM - Image Moments

ITML - information theory metric learning

IVM - import vector machine

JPEG - Joint Photographic Expert Group

KNN - K-nearest neighbour

KNN - K-nearest neighbour

LASSO - Least Absolute Shrinkage and Selection Operator

LCSM - laser confocal scanning microscope

LD - Linear Discriminant

LDA - linear discriminant analysis

LDA - Linear Discriminant Analysis

LDS - low-density separation

LDS - low-density separation

Linear-MSVM - linear metric-based support vector machine

LR – Linear Regression

LRN - Local Response Normalization

M3T - Multi-Modal Multi-Task

MDNN - multiscale deep neural network

Minc - Medical Imaging NetCDF

ML – Machine Learning

MLP - Multi-Layer Perceptron

MMSF - Mini-Mental State Examination

MoCA - Montreal Cognitive Assessment

MRF - Markov random fields

MRI - magnetic resonance imaging

MSE - Mean Squared Error

MVPA - multivariate pattern analysis

NACC - National Alzheimer Coordinating Center

NB - Naïve Bayes

NC - normal controls

NetCDF - Network Common Data Format

Nifti - Neuroimaging Informatics Technology Initiative

NiN – Network

NLP - natural language processing

non-D - non-demented

OASIS - Open Affective Standardized Image Set

OCD - obsessive-compulsive disorder

OCT - optical coherence tomography

PCA - Principal Component Analysis

PDF - Portable Document Format

PET - positron emission tomography

PNG - Portable Network Graphics

PSO - Particle Swarm Optimization algorithm

RAG - region adjacency graph

RELM - Regularized Extreme Learning Machine

ReLUs - rectified linear units

ResNet - Residual networks

RF - random forest

RGB - red-green-blue

RLG - regularized logistic regression

RLM - run-length matrix

RNN - recurrent Neural Network

ROC - receiver operating characteristic curve

ROI - regions of interest

SAE - stacked-autoencoder

SBM - surface-based morphometry

SGD - Stochastic Gradient Descent

SIFT - Scale Invariant Feature Transform

SIFT - Scale-invariant feature transform

sMRI - structural Magnetic Resonance Imaging

SPECT - single-photon emission computer tomography

SSIM - Structure Similarity Index Metrics

STD - standard deviation

SURF - speeded up robust features

SVM - support vector machine

SVM-RFE - support vector machine recursive feature eliminator

SVR - Support Vector Regression

TIF - Tagged Image File

TN – True Negative

TP - True Positives

TrAdaboost - transfer learning Adaboost

T-tau - total tau

UID - Unique Item Identification

UNICORT - Unified segmentation-based correction of R1 brain maps for RF transmit field inhomogeneities

VBM - voxel-based morphometry

VGG - Visual Geometry Group

VM-D - very mildly demented

VOC - Visual Object Classes

WM - white matter

WT - wavelet transform

Appendix C: Brain anatomy and diagnosis of MCI

Brain micro-anatomy

The anatomical and functional unit of the brain is a neuron (Di Ieva, 2011). The human brain has between eighty to a hundred billion neuronal cells, subdivided into several categories depending on the level and area of functionality. Histologically they are organized into layers which belong to the cortical, paleo-cortical or subcortical zone. The bodies of numerous neurons form the grey matter of the brain. The neurons connect via axons and dendrites and create a branchy network called white matter. The axons and dendrites propagate the electrical signal and provide communication between neurons. Another type of interaction between neurons is established via chemical signals. Neurotransmitters transfer the chemical signals between gaps and synapses of neurons. The neurotransmitters can be non-specific or be responsible for certain functions. Some axons are surrounded by the fibre, composed of glial cells responsible for the signal propagation. This type of cell also supports the trophic of neurons and creates a barrier between neurons and blood. One type of glial cell called the ependymal cell produces cerebrospinal fluid. Scientists disagree on the number of glial cells in the brain. Some claim the ratio between neuron and glial cells is equal to 10:1; others confirm the 1:1 ratio.

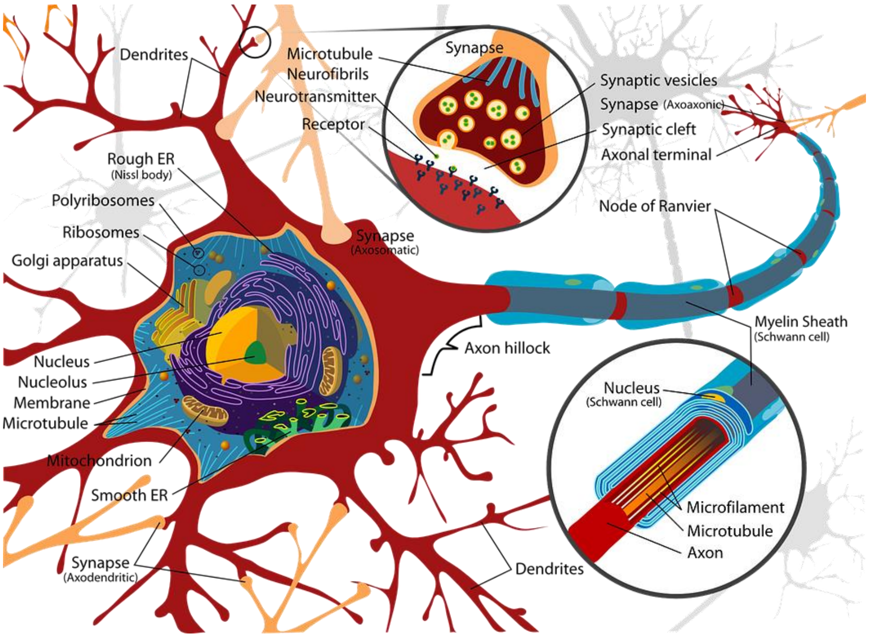


Figure A. Neuron, synapses and myelin (glial cells) (Image source: http://en.wikipedia.org/wiki/Image:Complete_neuron_cell_diagram_en.svg).

Brain macro-anatomy

The paleocortex and neocortex represent the brain cortex. Paleocortex composes of the limbic system and hippocampus. Histologically, the paleocortex has three or four levels of neuronal cells. The paleocortex's limbic system regulates emotions, and the extrapyramidal nuclei of the hippocampus are responsible for muscular tone and movements. Hippocampus is one of the main brain structures which supports long-term memory processing and distribution to specific cortex areas. Several theories explain how memory occurs. One of them says that memory is a product of an interaction between protein and RNA structures which facilitate DNA methylation. According to another theory, the memory process is supported by constant electric signals between the groups of neurons that neurotransmitters might enhance. All parts of the paleocortex participate in complex interconnected circuits of the brain. The Corpus callosum, which connects two cerebral hemispheres, represents the brain's neocortex. It is a part of white matter organized into six layers of neurons. The brain's surface has convolutions and furrows called gyri and sulci.

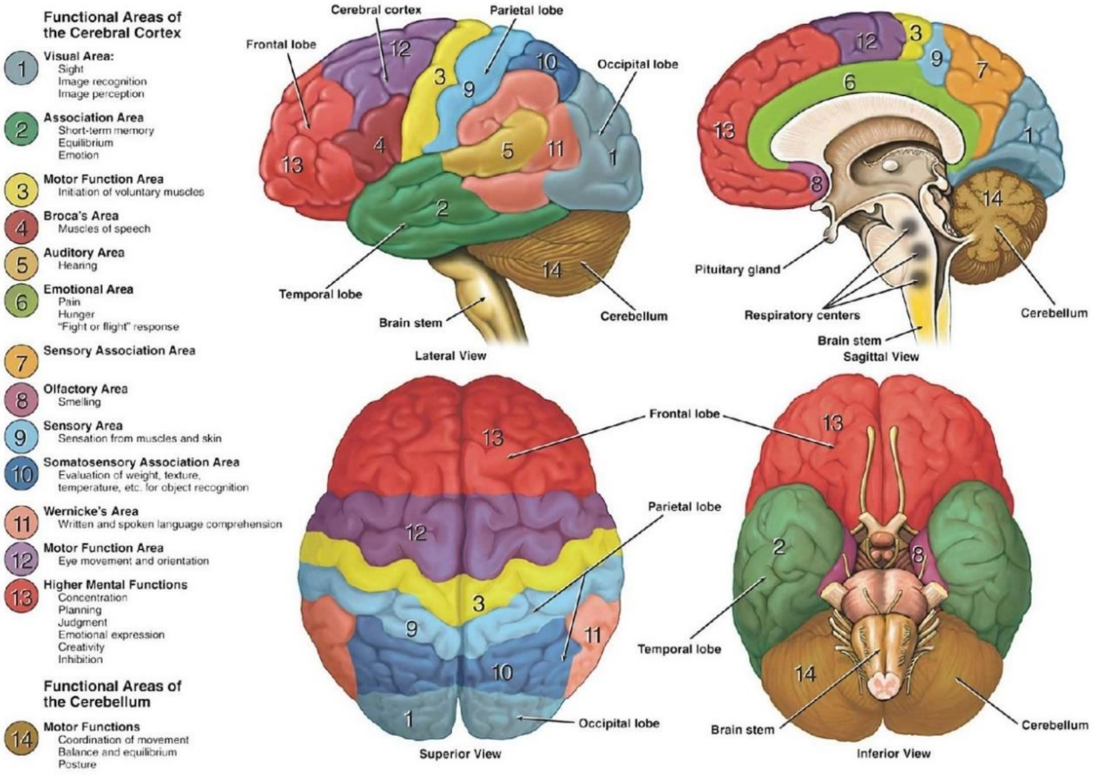


Figure B. Schematic anatomy of the brain and its functionality (Image source: https://dana.org/article/neuroanatomy-the-basics/).

Cortex is subdivided into several anatomical macro-regions: frontal, parietal, temporal, and occipital lobes. Each region includes a few zones with different functionalities. The frontal part is responsible for intellectual processing, the parietal zone synthesises the tasks, the occipital area forms the visual perception, and the temporal lobe is associated with auditory, memory and emotional processes.

Brain pathology leading to cognitive decline and dementia

The anatomical and functional plasticity of the brain plays a protective role in developing brain disorders. However, there is a number of conditions which can cause brain disease accompanied by dementia. The pathological process might take a few years to develop.

The general cause of functional impairments is structural damage to corresponding brain areas. The direct damage to the brain tissues might come through trauma, haemorrhagic insult, infection, or brain tumour. Medical conditions accompanied by the pathological processes in the immunological, humoral, metabolic or endogenous systems can also trigger the disease. The indirect impact can be obtained from a damaged blood supply system, as in ischemic lesions. Systemic medical disorders may affect all body tissues, including the brain if the blood-brain barrier becomes damaged.

Brain disorders are categorized according to the presence of several specific criteria. World Health Organization (WHO) established the global standards for the diagnosis of health conditions which is called the “International Classification of Diseases” (ICD-11), where 11 is a current version. The standards for mental disorders are named “The Diagnostic and Statistical Manual of Mental Disorders” or DSM-5 and were created by American Psychological Association in 2013. They characterise the evidence of cognitive decline based on one or more cognitive domains and evaluate the level of cognitive deficits, including independence in everyday activities.

The main known courses of cognitive decline and dementia are listed as follows:

- Alzheimer’s disease
- Pseudobulbar affect
- Parkinson’s disease
- Frontotemporal lobar degeneration

- Lewy body disease
- Vascular disease
- Traumatic brain injury
- Substance/medication use
- HIV infection
- Prion disease
- Huntington’s disease
- Another medical condition
- Multiple aetiologies
- Unspecified

Alzheimer’s disease is one of the main factors of dementia associated with ageing. While the main reason for the condition is not entirely apparent, the leading cause is the death or apoptosis of neurons and progressive atrophy of the cortical and some subcortical areas. Due to some internal genetic factors leading to the production of the pathological misfolding proteins, and external immune mechanisms, the number of neurons and synapses decreases. It impacts cognitive functions and memory.

Pseudobulbar affect (PA) is an affective disorder characterized by uncontrolled outbursts of emotions. PA is developed due to damage to the neuron of the prefrontal cortex. Multiple factors can cause the damage.

Parkinson's disease is a neurological condition which might be accompanied by dementia in 20% of cases. Damage to Substantia Nigra located in basal Ganglia leads to dopamine deficiency and the manifestation of neurological symptoms. The disease's diagnostic criteria are the presence of alpha-synuclein protein, called Lewi bodies, in the neuronal cells.

Lewy body disease is a group of neurodegenerative conditions accompanied by cognitive impairment and neuropsychiatric symptoms. It is caused by mentioned above

Lewy bodies that form the deposits of protein in neurons. Another disease from this group is Multiple System Atrophy (MSA), known as Shy-Drager Syndrome. MSA can lead to dementia in the frontotemporal zone.

Frontotemporal lobar degeneration (FTLD) is a group of neurodegenerative conditions caused by the development of atypical proteins leading to frontotemporal cortical atrophy. There are tau-positive known as Pick disease, FUS-positive and ubiquitin-positive variants, which differ by the type of pathological protein. They can produce or clear dementia or dementia with progressive aphasia (language/speech difficulties). Other pathological conditions caused by misfolding of specific PrP protein and leading to multiple dysfunctions of the central nervous system and dementia are Creutzfeldt-Jakob disease (CJD), Gerstmann-Sträussler-Scheinker syndrome, kuru, Bovine Spongiform Encephalopathy (BSE) and Fatal Familial Insomnia (FFI).

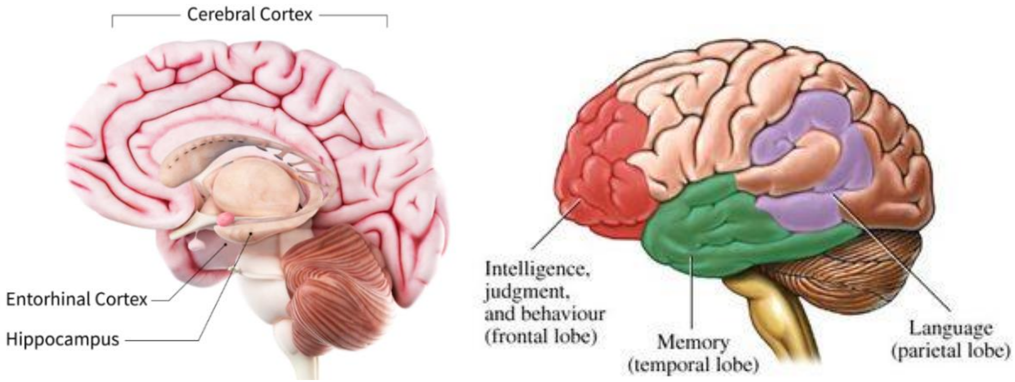
Huntington's disease is a neurodegenerative genetic disorder accompanied by neuropsychiatric symptoms, including memory loss and dementia.

Alcohol abuse or Alcohol Use Disorder (AUD) may cause encephalopathy, damaging the frontal lobe. It is caused by direct neurotoxicity of alcohol and liver dysfunction, which leads to vitamin deficiency in group B. This deficiency is associated with dementia. The Wernicke-Korsakoff syndrome often develops due to alcohol abuse and secondary B1(thiamine) deficiency. B3 (niacin) and B6 (pyridoxine) deficits are also associated with other types of dementia. Another alimentary depletion of vitamin D (D2 and D3) may exacerbate or cause dementia.

Other reasons for dementia can have different aetiology as mentioned above—for instance, traumatic, vascular, and infectious (HIV) reasons, substance or medication abuse. A side effect of some medications also may cause dementia. For example, these were reported cases of cortical dementia developed due to prolonged treatment by neuroleptic used for psychotic conditions.

Brain areas affected in dementia

Cortical and subcortical brain areas can be affected in MCI and during the development of dementia. Frontal, parietal and temporal cortical zones are the most affected areas. Subcortical brain areas that can be involved in the degenerative process during disease development are the amygdala and hippocampus. The destruction of neurons characterizes the initial stages of Alzheimer’s disease, and their connections in the areas involved the memory. These areas are located in the entorhinal cortex and hippocampus. The entorhinal cortex (EC) is placed in the medial temporal lobe between the hippocampus and neocortex, creating the hub for memory, navigation, and time perception. The amygdala is the main target of EC (Witter et al., 2017). Volumes of parahippocampal gyri and amygdala decrease in both cases of MCI and AD. Later, during AD progression, the brain grey matter responsible for reasoning, language, and social behaviour is affected (Ries et al., 2008).



(a) (b)

Figure C. The brain areas affected by dementia: (a) subcortical, (b) cortical (Image source: https://www.nia.nih.gov/health/what-alzheimers-disease; https://myhealth.alberta.ca/Health/pages/conditions.aspx?hwid=tp12408).

The signs of AD are visible on MRI and PET. On the histological level, the presence of the amyloid-beta and tau-protein confirms the diagnosis. The trigger factors of the development the AD are still under investigation by researchers. Among the principal reasons are proteostasis and damage to lymphatic circulation, which interrupt the protein clearance through the meningeal and cisterns systems. There is a comprehensive discussion about the diagnostic value of beta-amyloid in the distribution and progression of AD.

MCI and cognitive tests

The first step in the diagnosis of MCI is cognitive tests. Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) are the most popular. The MMSE and the MoCA are routine screening tests rated on a 30-point scale. They are short. The MMSE takes approximately seven to eight minutes; the MoCA takes about 10 to 12 minutes to administer. MCI MMSE scores ranged from 23-27 according to a DSM and 23-28 for ICD (Zaudig, 1992). For MoCA, a score of 26 and higher is considered normal. People with MCI have scored an average of 22.1 compared to normal controls with an average score of 27.4 (Rosentzweig, 2019). The diagnostic performance of MoCA was found to be a bit higher than MMSE for the detection of MCI and dementia in Parkinson's disease (Hoops et al., 2009). Another cognitive measure often used in clinical practice is the Clinical Dementia Rating (CDR) score. The CDR is scaled from zero to three: CDR = 0 denies dementia, CDR = 0.5 points to questionable dementia, CDR = 1 detects MCI, CDR = 2 characterizes moderate cognitive impairment, and CDR = 3 indicates severe cognitive impairment (Khan, 2016).

Clinical diagnosis of MCI

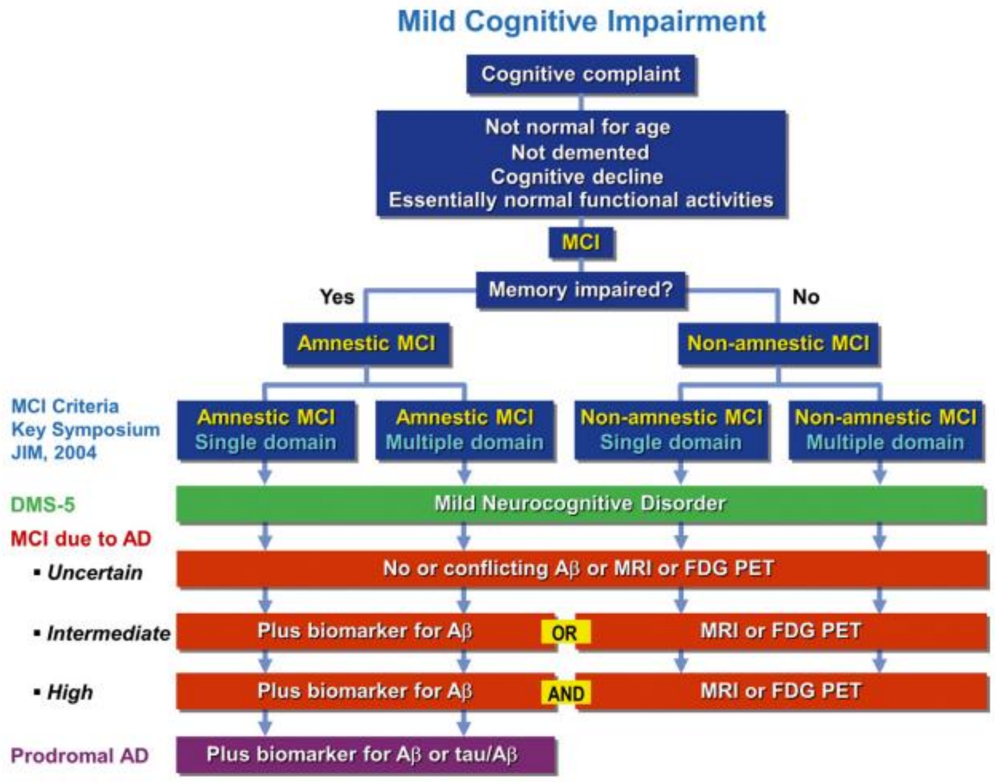


Figure D: Image source: 2014 The Association for the Publication of the Journal of Internal Medicine Journal of Internal Medicine, 2014, 275; 214–228.

The schema demonstrates the composition of imaging and non-imaging criteria for the diagnosis of MCI. MCI criteria proposed at the Key Symposium in 2004 (highlighted in blue) include amnestic and non-amnestic forms of memory impairment based on the single or multiple domains. Other criteria (highlighted in orange), such as amyloid-beta protein, and data from MRI or FDG-PET scans, were taken from a guide created by American Psychiatric Association for healthcare professionals, DSM-5. The classification of MCI based on DSM-5 includes three stages of the development of cognitive impairment: uncertain, intermediate, and high, which can be diagnosed using a particular combination of criteria. In the case of the detection of tau-globulin, the MCI is categorized as a prodrome of Alzheimer's Disease.

Appendix D: Medical image formats

Analyze

The commercial software Analyze 7.5 (AnalyzeAVW is a new version) was developed by the Biomedical Imaging Resource in the Mayo Clinic, USA, in the late 1980s. The software was designed for multidimensional data. The files can be stored in 3D or 4D data with temporary information represented by the fourth dimension. An imaging data in Analyze format have two binary files. The first one has an extension ".img" and contains the raw voxel data. The second file is a header file with an extension ".hdr". It contains the metadata with the data type, number of pixels in three directions and voxel size. The header of Analyse 7.5 includes a header key and image dimension, has a size of 348 bytes. An optional header feature is a data history. Analyze header can be adapted with new data types defined by the user. The structure of Analyse format is written in the C programming language. This file format has some limitations. It does not specify the pixel orientation in space and time and does not support some basic data types, including the unsigned 16 bits pixel depth format. This format is considered an old one, but it is still used and supported by many image processing software packages such as FreeSurfer (https://surfer.nmr.mgh.harvard.edu/fswiki/), SPM (https://www.fil.ion.ucl.ac.uk/spm/), Mango (https://ric.uthscsa.edu/mango/) and MRIcro (https://people.cas.sc.edu/rorden/mricro/mricro.html). The files in the format of Analyze 7.5 can be obtained from Positron Emission Tomography.

NIFTI

NIFTI-1 was created as neuroimaging format to solve the weaknesses of the Analyze format. It was developed at the beginning of the 2000s by the National Institutes of Health, USA. It supports the data types that are not considered by the Analyze format, output stores volumetric information about image orientation that is important in the neuroimaging study. NIFTI-1 format allows two ways to store the header and pixel imaging data. The first way of storage keeps the header and pixel data in separate ".hdr" and ".img" files and has a size of 348 bytes. The second way merges and saves all the image and header data in a single ".nii" file of a size of 352 bytes with additional four bytes at the end. Header parameters of the NIFTI-1 format are very similar to Analyze 7.5. The first three dimensions store spatial data, and the fourth dimension is preserved for a specific time. NIFTI-1 has an advantage over Analyze files in storing two files per 3D scan instead of keeping multiple ones. In addition, the NIFTI-1 file can store key acquisition and experimental design parameters.

In neuroimaging research, the NIFTI format has quickly replaced the Analyze and become the default and widely-used format in some public domain software packages like SPM (https://www.fil.ion.ucl.ac.uk/spm/), FSL(https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) and AFNI (https://afni.nimh.nih.gov/). Many modern image analysis software such as ImageJ (https://imagej.nih.gov/ij/), 3D Slicer (https://www.slicer.org/), OsiriX (https://www.osirix-viewer.com/), Nibabel (https://nipy.org/nibabel/) and R (https://www.r-project.org/) support the NIFTI format.

An updated version of the NIFTI-1, the NIFTI-2 format, was released in 2011. It was developed to operate with large data sets. This new version encodes an image matrix with a 64-bit integer for a single dimension instead of a 16-bit integer per dimension in the previous version.

MINC

MINC was designed by Montreal Neurological Institute (MNI), Canada, in 1992 (http://www.bic.mni.mcgill.ca/ServicesSoftware/MINC). The first version, MINC 1.0, provides a flexible format to store medical data. This version is based on the standard Network Common Data Format (NetCDF) and includes patient data, imaging and acquisition information, etc. A new version of MINC, MINC 2.0, was constructed to support large hierarchical data of HDF5 format with the possibility of internal data compression, 64-bit file size processing and some other features. Header or metadata of

MINC 2.0 can contain any number of entries and construct a hierarchy of datasets. There are three data subgroups available in MINC 2.0: image, info and dimension. The image subgroup holds actual image data such as minimum and maximum values of image slices named image-min and image-max, respectively. The info subgroup includes patient identification details, information about experimental settings, type of scans and modalities. The properties of the MINC file like space, time, and vector dimensions are kept in the dimension subgroup.

DICOM

DICOM file format and a network communication protocol were developed by the American College of Radiology and the National Electric Manufacturers Association in 1993. DICOM standards include image retrieval, storage, visualization and analysis, image compression and exchange protocols. A secure exchange protocol helps to transfer medical images between multiple entities. DICOM network communication protocols integrate image acquisition devices, printers, servers, and workstations used by different manufacturers. The standards of this format are accepted by numerous imaging modalities, including CT, MRI, and ultrasound.

As a generic file format, DICOM consists of header and image data merged into a single file represented as '.dcm'. DICOM header contains acquisition parameters and their modalities, image dimensions and matrix size, operator identification and patient information. Patient personal details include name, age, gender, weight, and height. The file header has a size of 128 bytes with additional four bytes of DICOM annexe. In the DICOM format, the pixel values are stored as integers. This format supports various types of data and metadata. However, it cannot store the pixel data in floating points. The files in DICOM format can be easily converted into digital image formats such as '.jpg', '.bmp', '.tif'. DICOM supports lossless image compression like JPEG (JPEG-LS, JPEG-2000, JPEG-XR), MPEG (MPEG2, MPEG4), run-length encoding (RLE), Huffman coding (Baraskar and Mankar, 2019).

Appendix E: Image segmentation methods

This section concentrates on contemporary brain segmentation methods (Despotović et al., 2015). MRI segmentation can be performed on 2D images separated from an image sequence or on 3D series. If 2D images are segmented slice-by-slice, they can be connected to a 3D volume or a continuous surface. In brain MRI, there are three main tissue types: grey matter, white matter, and cerebrospinal fluid. But typically, the scanned image also incorporates the skull and a nonbrain area. Therefore, the standard approach of MRI processing involves the extraction of nonbrain tissues before the brain segmentation methods can be used.

There is no single method that can be suitable for all images because of the image diversity, presence of noise and artefacts. However, most of the segmentation methods developed for one image class can be easily adapted to another class.

Segmentation methods with application to brain MRI can be divided into the following groups:

- manual segmentation;
- intensity-based methods (thresholding, region-growing or region-based, edge-based, and clustering);
- morphology-based methods (watershed segmentation);
- atlas-based methods;
- deformable model (active contour-based or snake-based methods);
- classification methods;
- hybrid segmentation methods.

Manual segmentation

Manual segmentation refers to an image segmentation process accomplished by a human expert (medical practitioner), which segments and labels the image by hand. Segmented images look more accurate compared to the automatic methods, but the process is very intensive and time-consuming. The results of the manual methods are difficult to reproduce. However, manual segmentation is still actively used for defining a correct delineation of an object and the quantitative assessment of automated segmentation methods. Furthermore, manual segmentation of different brain structures is a fundamental pace in the development of brain atlases and atlas-based segmentation approaches.

Intensity-based methods

Intensity-based methods have a very wide variety. Three main MRI brain components, WM, GM and CSF can be easily distinguished due to different pixel intensity levels (see Figure E). However, the presence of noise, artefacts, overlapped objects, inhomogeneity of the tissues is an objective factor that can require the incorporation of additional tools and implementation of advanced techniques.

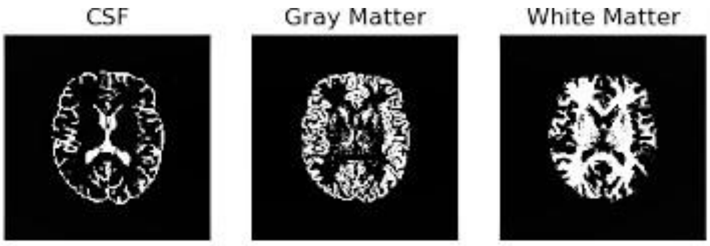


Figure E: MRIs with segmented CSF, GM and WM of the brain

There are several commonly used intensity-based techniques. Generally, they build a pixel (voxel) intensity histogram, which is based on pixel distribution functions. As a result, the probability of pixel intensities can be assigned to a certain type of tissue. The incorporation of the additional information received from the neighbouring pixels helps to allocate homogeneous regions in the resulting segmentation. Some popular intensity-based methods are described below.

Thresholding

Thresholding is a well-known image segmentation method that separates an object from its background (Sezgin and Sankur, 2004). The method divides pixels of the image according to their intensity levels, and the output of the thresholding operation is a binary image. The thresholding has many variations that roughly can be divided into global (single), locally adaptive, and multiple thresholding (Kaur D. and Kaur Y., 2014). In the case of *global thresholding*, a single value of the threshold level T is chosen for the whole image. The value of T is constant and separates pixels into two classes. Segmentation of an original image $I(i, j)$ is defined as:

$$I'(i,j) = \begin{cases} 1, \ if \ I(i,j) > T, \\ 0, \ if \ I(i,j) \leq T, \end{cases} \tag{1}$$

where $I'(i, j)$ is an output image, pixels labelled with the number 1 correspond to the segmented object, and pixels labelled with the number 0 correspond to the background.

L*ocally adaptive thresholding* computes the threshold value for subregions of the image so that each subregion might have a different threshold value. Adaptive thresholding depends on some local statistics like range, variance, or surface-fitting parameters of the neighbouring pixels.

Multiple thresholding is based on several threshold values like T0 and T1. Segmentation of the image can be defined as:

$$I'(i,j) = \begin{cases} m, \; if \; I(i,j) > T1, \\ n, \; if \; I(i,j) \leq T1, \\ o, \; if \; I(i,j) \leq T0, \end{cases} \tag{2}$$

Thresholding is actively used in image ***binarization*** (Rohithkumar et al., 2020) when a binary image is obtained by quantization of the grey level values into black and white. The binarization technique is usually applied when the object needs to be extracted from the background. The main problem in image binarization is to determine the optimal grey threshold value that separates the foreground from the background correctly. Due to the process of binarization, the pixels above the determined threshold obtain the maximum intensity values equal to “one”. The other group of pixels are assigned to a “zero” value. A number of binarization algorithms are proposed for image segmentation. Among them is *Otsu’s* thresholding technique (Yousefi, 2011). The algorithm looks for the optimal threshold value that can minimize the intraclass variance of the fore- and background. The intraclass variance is determined as a weighted sum of these two classes. The Otsu method exploits the statistical information obtained from the image, such as variances and probabilities. The best threshold value t is calculated by finding the minimum intraclass variance. The mathematical expression for a weighted sum of intraclass variances:

$$\sigma_w^2 = w_b(t) \cdot \sigma_b^2(t) + w_f(t) \cdot \sigma_f^2(t) \tag{3}$$

where w states for weight, σ - for standard deviation, b - for the background and f - for the foreground.

Mathematical formulas for weights (w), mean values (μ) and standard deviation (σ) for each class and probability (P) are presented in formula (8). $P(i)$ is a probability of pixel *i* that should be in the range from *i* to *t*, where *t* is a threshold value. The pixel will be excluded from the visualisation if it is not in this range:

$$w_b(t) = \sum_{i=1}^{t} P(i) \tag{4}$$

$$w_f(t) = \sum_{i=t+1}^{l} P(i) \tag{5}$$

$$\mu_b(t) = \frac{\sum_{i=1}^{t} i \cdot P(i)}{w_b(t)} \tag{6}$$

$$\mu_f(t) = \frac{\sum_{i=t+1}^{l} i \cdot P(i)}{w_f(t)} \tag{7}$$

$$\sigma_b^2(t) = \frac{\sum_{i=1}^{t} (i - \mu_b(t))^2}{w_b(t)} \cdot P(i) \tag{8}$$

$$\sigma_f^2(t) = \frac{\sum_{i=t+1}^{l} (i - \mu_f(t))^2}{w_f(t)} \cdot P(i) \tag{9}$$

The application of the Otsu binarization technique used for object segmentation is shown in Figure F.

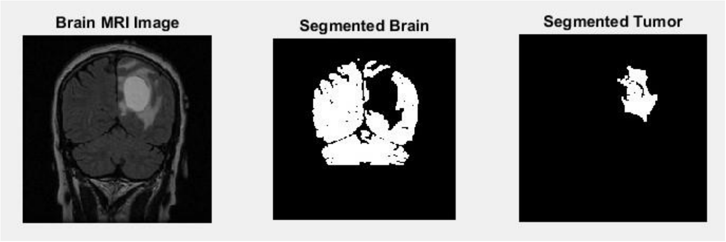


Figure F: Binary thresholding for segmentation of brain tumor

Thresholding methods are fast and computationally efficient, but they are sensitive to noise and image textural inhomogeneity. Low-contrast images after thresholding are keen to produce disconnected regions and require the application of additional postprocessing algorithms that can improve or fully restore connectivity. If the size of the segmented object is not proportionate with a scene, it also makes the segmentation process more complicated.

Region-based methods

Region-based methods highlight areas of the image according to predefined uniformity or homogeneity criteria of pixel intensity values (Lu et al., 2003). The region-based methods can combine region growing, region merging and region splitting procedures. The region merging starts from a “seed point” (a single pixel or group of pixels) that belongs to the object of interest. As soon as the “seed point” is manually or automatically initialized, additional neighbouring pixels satisfying the similarity criteria of the tissue can be added to the growing area. The region-growing procedure is repeated until no more matching pixels are left.

In practice, there is several questions that must be solved before starting the procedure. These questions are: How to choose the “seed” point? What similarity criteria needs to be applied? The choice of a “seed” point depends on the grey-level value of pixels and their positioning in the image. It is recommended to compute the image histogram and the strongest intensity pics to consider as “seed” points. The similarity criteria can be based on any characteristics of the image region, such as average intensity, variance intensity, texture, colour, shape and size.

One of the popular ***region-growing*** approaches is based on simple surface fitting. The algorithm is described by Besl and Jain (1988). The idea is to fit an appropriate low-order surface that can be planar or biquadratic over the image data of a region. It was suggested to use a function that calculates the error in approximating the pixel data. The error should be less than some threshold levels. In the following formulas *x*, *y*, *a* and *m* state for a region coordinates; *i* and *j* are pixel values.

$$f(x, y, a, m) = \sum_{i+j \le m} a_{ij} x^i y^j \tag{10}$$

If the errors are small, it might indicate that the pixel values belong to the same region of interest.

$$E(R, a, m) = \sum_{(x,y) \in R} [g(x, y) - f(x, y, a, m)]^2 \tag{11}$$

The detailed stages of the region-based methods are described below.

(a) Partitioning the image into the “seed” regions $R_i^{(0)}$ of the size of *n* × *n* where *n* is between 5 and 9.

(b) Fitting a planar model to each “seed” region. If $E(R_i^{(0)}, a, m)$ is small enough, accept $R_i^{(0)}$ and its model; otherwise, reject $R_i^{(0)}$.

(c) Finding neighbouring points compatible with each “seed” region

$$C_i^{(k)} = [(x, y): \left(g(x, y) - f(x, y, a, m)\right)^2 < \varepsilon \ and \ (x, y) is \ a \ 4 - neighbor \ of \ R_i^{(k)} \tag{12}$$

(d) In case when no compatible points are found, the *m* should be extended to *m+1*. If *m* > *M*, the region $R_i^{(k)}$ can't be grown further, otherwise, the previous stage must be repeated.

(e) Forming a new region $R_i^{(k+1)} = R_i^{(k)} \cup C_i^{(k)}$, refitting the model to a new region $R_i^{(k+1)}$, and computing error $E(R_i^{(k+1)}, a, m)$

(f) Computing the difference in errors

$$P^{(k)} = E\left(R_i^{(k+1)}, a, m\right) - E(R_i^{(k)}, a, m) \tag{13}$$

(g) If the error difference is less than the level of the threshold $P^{(k)} < T_1$, the region can be increased *m* = *m+1*; if the *m* > *M*, the region cannot grow further.

(h) Refitting the region at the new model *f (x, y, a, m).* If the error of fit decreases, a new model can be accepted, and the region can be grown until all the neighbouring pixels are joined to the region according to the initial condition.

This method is successfully used in medical imaging for the segmentation of tissues, anatomical structures and organs, lesions in the human body. For instance, advanced region-growing approaches are implemented for brain tumor segmentation (Biratu et al., 2021), analysis and segmentation of new-born brain MRI (Udayakumar et al., 2016), automatic classification of MRI data (Yazdani et al., 2016).

The region-growing method has some disadvantages that must be counted when the method is applied. The first of them is a sensitivity to the initialization of the "seed point". By displacement of the initial point, segmentation results can be significantly different. Suppose the "seed point" is chosen incorrectly, or the homogeneity criteria are not properly defined. In that case, the growing region can incorporate the areas that do not belong to the object of interest. In some cases, the sensitivity of the method to noise makes parts of the segmented area partially disconnected, full of holes or, conversely, segmented parts can become connected to the regions that do not belong to the area of interest.

After initial intensity-based region segmentation, it very often requires additional segmentation techniques. Some of those include ***region splitting and region merging***. Region splitting and merging operations are a combination of region-based techniques. The splitting technique separates an image into regions that have similar characteristics. The merging approach contributes to combining similar regions into a smaller number of larger regions (Kaganami and Beiji, 2009). Similarity, for instance, can be structural, topological, or semantic. The choice of the algorithm will depend on the image properties. There are several approaches to performing region merging. Most of them are based on grey-pixel values of regions or the weakness of boundaries between the regions. Two main approaches to merging are based on the comparison of mean intensity values of a region or probability distribution of the intensity values. The mean intensity approach joins the regions with similar mean intensities that do not exceed the predetermined level. The probability-based approach considers the statistical characteristics of the adjacent regions.

The ***region merging*** approach is composed of the detection of the initial regions, preparation of the region adjacency graph (RAG), checking the RAGs for similarity and merging those with close criteria stages. After each merging round, the graph needs to be modified. The process is repeated until no more regions can be merged. Figure G represents example of RAG.

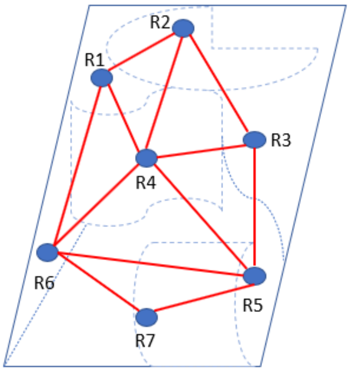


Figure G: Merging technique using the region adjacency graph

The regions are merged based on their similarities in pixel intensities. If the intensity values are received from a probability distribution, hypothesis testing is applied to judge the similarity of adjacent regions. If image regions have constant, statistically independent grey-scale values and zero-mean Gaussian noise, they are distributed normally. The null hypothesis (Ho) assumes that both regions belong to the same object. On this occasion, the pixel intensities are received from a single Gaussian distribution with parameters $\mu_{0,}\sigma_0^2$. The null hypothesis is rejected if the pixel intensities of the regions are obtained from two separate Gaussian distributions with parameters $\mu_{1,}\sigma_1^2$ and $\mu_{2,}\sigma_2^2$. Pixel probability distribution in the region is estimated according to the following equation, where n pixels in the region have grey levels g_i with $i = 1,2, \ldots n$:

$$p(g_{i)=}\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(g_i-\mu)^2}{2\sigma^2}} \tag{14}$$

The Maximum Likelihood estimation is given by

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}g_i \tag{15}$$

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(g_i-\hat{\mu})^2 \tag{16}$$

The joint probability density under the Ho hypothesis for all pixels received from single distribution $N(\mu_0,\sigma_0^2)$ is described by the following equation:

$$p\left(g_1,g_2,\dots,g_{m_1+m_2}\middle|H_0\right)=\prod_{i=1}^{m_1+m_2}p(g_i\,|H_0)=\prod_{i=1}^{m_1+m_2}\frac{1}{\sqrt{2\pi\sigma_0}}e^{-\frac{(g_i-\mu_0)^2}{2\sigma_0{}^2}}=$$

$$\frac{1}{(\sqrt{2\pi\sigma_0})^{m_1+m_2}}e^{-\frac{\sum_{i=1}^{m_1+m_2}(q_i-\mu_0)^2}{2\sigma_0{}^2}}=\frac{1}{(\sqrt{2\pi\sigma_0})^{m_1+m_2}}e^{-\frac{(m_1+m_2)}{2}} \tag{17}$$

Under the H1 hypothesis, the joint density function for the pixel *m1* of region 1 with distribution $N=(\mu_1,\sigma_1^2)$ and pixel *m2* of region 2 with distribution $N=(\mu_2,\sigma_2^2)$ is written according to the following equation:

$$p\left(q_{1,}q_{2,\dots,}q_{m_{1,}}q_{m_1+1,\dots,}q_{m_1+m_2}\middle|H_1\right)=\frac{1}{(\sqrt{2\pi\sigma_1})^{m_1}}e^{\frac{-m_1}{2}}\;\frac{1}{(\sqrt{2\pi\sigma_2})^{m_2}}e^{\frac{-m_2}{2}} \tag{18}$$

From the ratio of probability densities of the two hypotheses is calculated the likelihood ratio L:

$$L=\frac{p(g_1,g_{2,\dots}|H_1)}{p(g_1,g_{2,\dots}|H_0)}=\frac{\sigma_0^{(m_1+m_2)}}{\sigma_1^{m_1}\cdot\sigma_2^{m_2}} \tag{19}$$

If the likelihood ratio L is below the threshold value, it provides evidence that regions 1 and 2 can be merged.

The "***region splitting"*** procedure usually starts with large regions. For each image region, the variance of grey values is computed. If the variance is above the threshold level, the region is split along the appropriate boundary. To determine the best boundary to divide the region is to calculate the edge strength within the region. The easiest technique which is used for image splitting is image decomposition, when the image is divided into a fixed number of equal-sized subregions. An example of a regular decomposition is a quart tree approach. Wavelet decomposition lies at the base of this approach (Shusterman and

Feder, 1995). Splitting and merging techniques can be used together for the segmentation of complex scenes.

Edge-based method

Edge-based image segmentation is based on rapid changes in pixel intensity values to determine the object boundaries. This method, unlike the previous one (region-based method), is focused on finding the differences, instead of similarities, between pixels. The method is computationally fast and strongly sensitive to the notable variation of the pixel grey level values.

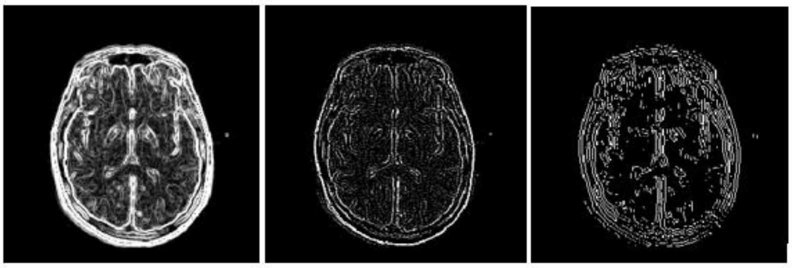
There are plenty of segmentation techniques developed for edge detection (Shrivakshan and Chandrasekar, 2012). The approach based on a comparison of statistical probability pixel distributions in two regions (see region merge approaches) can also be used for edge detection. The likelihood ratio indicates the boundaries between two regions or detects the presence of edges. A comparison of some popular edge detection techniques is provided in Table A below.

Table A: Edge detection technique comparison

Technique	Advantages	Disadvantages
Classical (Sobel, Prewitt, Roberts)	Provides a simple approximation to the gradient magnitude Detects edges and their orientations	Sensitivity to noise leads to a decrease in the magnitude of edges which lowers the performance
Zero-Crossing (Laplacian)	Detects edges and their orientations Has fixed characteristics in all directions	Sensitivity to the noise Edge detection gets diffracted by some other edges in the presence of noise
Gaussian (Gabor Filter)	Performs well in specific spatial locations for different types of texture representation Can be designed for multiple image dilations and rotations	Can be time-consuming (a filter bank of Gabor filters is created to overcome the problem)
Marr-Hildreth (Laplacian of Gaussian (LoG))	Can be tested and processed in the wide area around the pixels	Detection of edge orientation is reduced in places of corner and curves malfunctioning due to the variability of the grey level intensity function

Gaussian (Canny)	Uses "smoothing" concept (finds errors by using the probability distribution) Improves the signal-to-noise ratio Well detects edges in a noisy image using thresholding methods	Complex computation of the angle gradient in the image Time-consuming

Figure H shows examples of the most popular image segmentation algorithms.



(a) (b) (c)

Figure H: Examples of the edge segmentation algorithm: (a) Sobel operator, (b) Laplacian operator, (c) Canny operator

The edge-based method has some limitations. One of them is that the resulting segmented edges are not connected at some points. It might happen in the presence of noise or with a smooth or low-contrasted image. Usually, such sorts of problems are resolved by performing the additional postprocessing steps focused on linking corresponding to single boundary edges into the chains. These techniques improve the representation of edges in the image.

Clustering

Clustering is an image segmentation method that partitions images into clusters of pixels or voxels with similar characteristics. There are two clustering approaches, ***hierarchical*** and ***partitional***. The hierarchical approach uses the concept of tree-like clustering, which split the data using a divisive or agglomerative way. The partitional approach groups the data into clusters according to the objective similarity functions of data elements. Thus, the data elements of one cluster are more similar to the data elements of another cluster. These types of clustering might use ***hard*** and ***soft*** approaches to the division of data elements inside the image. In the case of hard clustering, the image is partitioned into a set of clusters such that each pixel can only belong to one cluster. In soft clustering, the division of the pixels is not so strict. The pixels inside the clusters have partial membership when one pixel can belong to many clusters based on the degree of membership values. The hard partitional clustering methods can be divided into ***K-means based*** methods, ***histogram-based*** thresholding and ***metaheuristic-based*** clustering. The diagram of various clustering methods is provided below in Figure I.

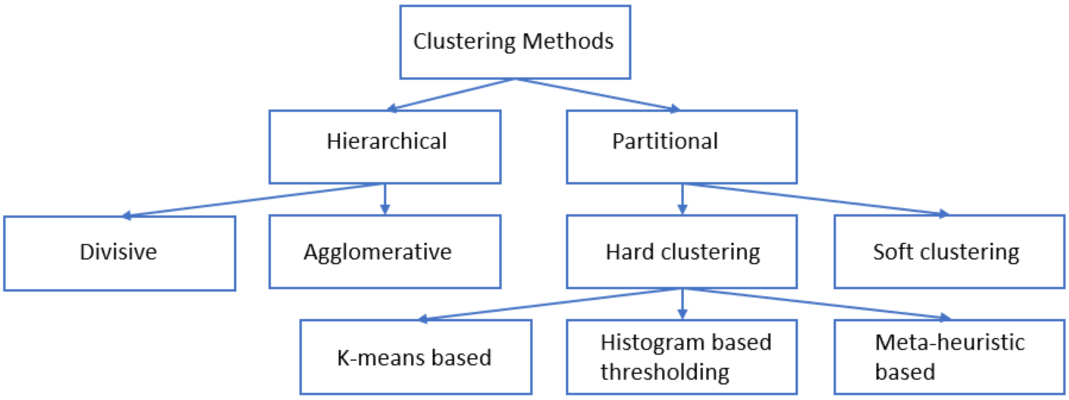


Figure I: Clustering methods

Hierarchical clustering is applied in two directions: top-down for divisive and bottom-up for agglomerative. For divisive clustering, all the data points belong initially to a single cluster which further splits into smaller sub-clusters. This process continues until each data point forms its own cluster or termination criteria will be satisfied. Agglomerative clustering merges data hierarchically. Initialisation usually starts from single data points, which are aggregated into bigger clusters until a single cluster is formed or termination criteria are met. Divisive clustering is efficient and usually more accurate than agglomerative one. At the same time, the divisive way takes more time to process the data and often needs a time-optimization algorithm to be applied. Both types of clustering require the initialization of the number of clusters. They are time-consuming and computationally expensive for large high-dimensional datasets. The hierarchical methods employ a greedy approach and do not reconsider the data if it is already clustered. It can increase the level of misclassification errors. Both of these methods have lack robustness and poorly perform in the presence of noise, outliers, and on occasions when clusters are overlapping.

Partitional clustering methods are efficient methods of data clustering. They overcome some problems of hierarchical clustering and are widely used for large datasets. Objective functions such as similarity measurements are applied for data clustering. Computing the Euclidian distances between the observed data points is one of the popular approaches used for data clustering. Like in hierarchical clustering, the number of clusters needs to be defined before the start of the clustering process.

During ***soft clustering***, each data point is assigned to one or a few clusters according to the degree of membership. The degree of membership is a continuous or discrete value in the interval [0, 1]. One of the popular soft clustering methods is fuzzy c-means (FCM) clustering (Mittal et al., 2020). The algorithm starts from the initialization of the number of clusters. The cluster centroids are initialized randomly. The algorithm returns a set of clusters by minimising the objective function. This function is defined in the following equation:

$$\sum_{i=1}^{N} \sum_{k=1}^{K} \mu_{ik}^{m} \left|\left|x_i - v_k\right|\right|^2, m \geq 1 \tag{20}$$

Where μ_{ik} ϵ [0 1] and demonstrates the membership degree of any *i* pixel with any *k* cluster, and *x* states for the positioning of the pixel.

If the clustering conditions are not satisfactory or the centroids do not change optimization technique is used.

Equation 20 is optimized by updating μ_{ik} that is used to compute the fuzzy partition matrix:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{K} \left(\frac{\left|\left|x_i - v_k\right|\right|}{\left|\left|x_i - v_j\right|\right|}\right)^{\frac{2}{m-1}}} \tag{21}$$

Cluster centroid v_k is updated with each iteration:

$$v_k = \frac{\sum_{i=1}^{N} \mu_{ik}^{m} x_i}{\sum_{i=1}^{N} \mu_{ik}^{m}} \tag{22}$$

FCM clustering is fast and accurate in handling incomplete or heterogeneous data and generating approximate solutions. However, it has low scalability, is very sensitive to noise and outliers, and requires initial knowledge about the number of clusters which can be challenging.

Hard clustering methods can be roughly divided into K-means based, histogram-based and metaheuristic-based (Mittal et al., 2021).

K-means is a low complexity clustering method when a randomly chosen centroid is updated each time with new entries. Iterations continue until no data points are left outside the cluster or some clustering conditions are met. The quality of this method depends on the initialization parameters that include the number of chosen clusters and the positioning of the centroid. The presence of noise and outliers reduce the clustering performance. Usually, data partitioning is performed using similarity criteria which is a sum of squared error. Criterion function J is minimised with each iteration:

$$J = \sum_{i=1}^{k} sum_{x_{j \in X}} \left|\left|x_j - m_i\right|\right|^2 \tag{23}$$

where K is the number of clusters for a set of data points $x = \{x_i, \ldots, x_n\}$. m_i is a cluster centroid for data points $M = \{m_i, \ldots, m_k\}$ which is updated with data points

$C = \{c_i, \ldots, c_k\}$ to form a new centroid. This is defined by the following equations:

$$x_i \in c_l, if\ 1 = argmin_{l=1}^{k} \left|\left|x_j - m_i\right|\right|^2 \tag{51}$$

$$m_i = \frac{\sum_{x_i \in c_l} x_i}{|c_l|} \tag{52}$$

for $1 \leq i \leq N$ and $1 \leq l \leq k$

The time complexity of the algorithm depends on the number of data points, number of clusters and maximum iteration needed to form these clusters.

Histogram-based methods belong to the group of thresholding methods that is generally described above in this section. The clustering is performed by constructing the histogram with respect to the frequency of pixel intensity values and their derivatives, average pixel intensity and pixel gradient. The clustering is achieved by grouping the pixels with similar intensity values. Usually, all histogram's peaks, valleys, and curvatures are analyzed to perform the segmentation. The mathematical expression of histogram-based clustering starts from the calculation of the probability distribution of pixel intensity values in an image:

$$p_i = \frac{n_i}{N}, 0 \leq i \leq L - 1 \tag{24}$$

where N corresponds to the number of pixels with intensities {*0, 1, 2, …, L-1*}, n_i corresponds to the number of pixels for i^{th} intensity level.

From the formula above, the mean intensity value is defined as:

$$\mu = \sum_{i=1}^{L} ip_i \tag{25}$$

After that, the pixels of the image are grouped into clusters n {*C1, C2, …, Cn*} using *n-1* threshold level $\{t_1, t_2, \dots, t_{n-1}\}$:

$$v(x, y) = \begin{cases} 0, & v(x, y) \le t_1 \\ \frac{t_1+t_2}{2}, & t_1 < v(x, y) \le t_2 \\ & . \\ & . \\ & . \\ \frac{t_{n-2}+t_{n-1}}{2}, & t_{n-2} < v(x, y) \le t_{n-1} \\ L-1, & v(x, y) > t_{n-1} \end{cases} \tag{26}$$

Where *v* corresponds to pixel intensity at *x* and *y* location in the image M x N. The *Cj* cluster (*1<j<n*)is partitioned to pixels with intensity more than t_{j-1} and less than or equal t_j.

The frequency of the cluster *Cj* is defined as:

$$w_j = \sum_{i=t_{j-1}+1}^{t_j} p_i \tag{27}$$

The mean for *Cj* is computed as:

$$\mu_j = \sum_{i=t_{j-1}+1}^{t_j} ip_i / w_j \tag{28}$$

The inter-class variance for *Cj* is calculated as:

$$\sigma^2 = \sum_{j=1}^{n} w_j \, (\mu_j - \mu)^2 \tag{29}$$

The maximization of inter-class variance helps to form the clusters. The objective function which maximizes the fitness function is defined as follows:

$$\emptyset = max_{1<t_l<t_{n-1}<L}\{\sigma^2(t)\} \tag{30}$$

Histogram-based clustering methods are efficient. But, as in previous methods, they require prior knowledge about the number of clusters. They are highly sensitive to noise, especially if the overlapping regions of the histogram are presented, or peaks and valleys are hardly identified.

Metaheuristic clustering is a hybrid method used for the optimization of real-world problems. The algorithm might start from random centroid initialization. Optimal clusters are obtained by updating the data according to the mathematical formulation of the objective function and optimality criteria. The searching behaviour of the algorithm is based on probabilistic theories and leads to the approximate solution of complex real-world problems. The objective function of the optimization algorithm ($f(x)$) performs either maximization or minimization procedures with a given set of constraints. The mathematical expression of the maximization optimization problem is given below:

$$Maximize_{\{x \in \mathbb{R}^d\}} f(x) \tag{31}$$

$$\text{such that: } a_i(x) \geq 0,$$

$$b_j(x) = 0,$$

$$c_k(x) \leq 0$$

where $x = (x_1, x_2, x_3, \dots, x_d)^T$ is a set of decision variables, $\mathbb{R}^d$ is a search space of the problem in *d* dimensions, $a_i(x), b_j(x), and c_k(x)$ is a set of constraints that belong to the optimization problem.

There is no single algorithm able to solve every heuristic problem (Wolpert and Macready, 1997). These methods are complex and often present a combination of genetic and K-means algorithms. More than sixty solutions have been proposed for physical, biological, and evolutionary tasks in the last decades. These approaches help to explore the research space and intensify the local solutions to achieve the global goal.

Morphology-based methods (watershed segmentation)

Watershed segmentation in medical imaging is an image partitioning method based on morphological differences in the biological tissues. The greyscale image is represented as a topological water-filled relief, where watersheds distinguish the areas of the different basins (Kornilov and Safonov, 2018). The main task of this method is to determine the position of all morphological basins and watershed lines. Each water basin responds to a separate segment of the image. There are two principal watershed segmentation approaches: watershed by flooding and watershed by rain falling (Bieniek and Moga, 2000).

"Watershed by the flooding" is based on the principle of the gradual feeling of the water basin, which starts from the minimum level of water. The process ends when the water reaches the maximum peak of the relief, and as a result, every catchment basin gets covered by the watershed lines.

Watershed by flooding

The watershed by flooding algorithm operates in the following order. At the beginning of the process, the local minima in the image are found. Then, the segmentation process goes according to the priority queues, which are based on image element (pixel/voxel) values. The sequence of queues is scanned from smaller values to larger ones. The selection process starts from the first non-empty element of the queue. The algorithm terminates when all queues from the priority queue are empty. After that, the selected element is removed from the queue, and the algorithm looks for the next neighbouring element to be marked and processed. All the marked neighbours are placed in the priority queues again. The process is repeated until all elements of the image are selected. The described above algorithm segments the image into basins. In the case of watershed segmentation with lines, each image element needs to be marked additionally. So, each element has a unique marker and an additional marker.

Watershed by rain falling

Watershed by rain falling approach creates a model that simulates the rain falling process, which is based on the connectivity components. Each drop of water that is falling on the surface goes down along the slope to the appropriate valley. Usually, the water drop chooses the steepest path that connects the initial point of fall and the valley. All connected to the same valley elements establish one catchment basin. The levels of catchment basins can be different. It involves the additional steps to process the segmentation. This type of algorithm uses a simpler structure compared to the watershed by flooding model. The runtime of the watershed by rain falling model is short because it does not depend on the range of the pixel/voxel values of the image.

The main problems of the watershed methods are under- and over-segmentation. It is proposed to apply additional preprocessing and postprocessing techniques for overcoming these limitations. The preprocessing technique is usually based on the morphological image transformation before the segmentation process. The postprocessing technique helps to connect each unmarked image region with the nearest marked one using distance measure algorithms.

The watershed segmentation might be applied to any kind of imaging data including MRIs. The example of it is shown in Figure J.

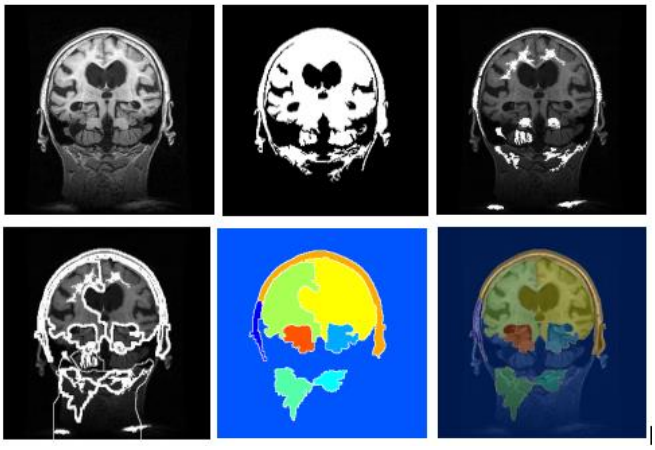


Figure J: Example of the watershed segmentation algorithm applied to the MRI of the brain

Atlas-based methods

Atlas-based segmentation methods (Cabezas et al., 2011) are built using prior knowledge about brain anatomy. The dominance of these methods is the chance to segment any part of the brain known by atlas. There are two types of atlas-based segmentation methods: topological (other names “single-subject”, “deterministic”) and probabilistic (also cited as “statistical”, “population-based”). Nowadays, most deterministic atlases are created from single-subject image acquisition and used as imaging templates for different types of pathology. Probabilistic atlases were constructed to represent the anatomical variability in the human population. Population-based atlases can be divided into sub-groups according to gender, age, handedness, etc. Additional attention is received by the disease-based atlases. The construction of them has increased in the last decade (for example, Alzheimer’s disease template is provided by the ICBM (Internet Brain Segmentation Repository), http://www.cma.mgh.harvard.edu/ibsr/). Disease-related atlases allow qualitative and quantitative examination of the evolutionary changes due to disease progression. They also allow the evaluation of the quality of clinical treatment.

At the beginning of a segmentation procedure, each image needs to be aligned or registered with the existing averaging anatomical atlas. So, the quality of image segmentation strongly depends on the registration method. An example of an atlas-based segmentation method applied to the human brain is given in Figure 31. The well-known aligning image method is affine registration. However, affine alignment can be insufficient if the brain anatomy significantly differs from the standard anatomical atlas. To avoid these problems, the researchers proposed to use segmentation, bias correction, and nonrigid registration of the probabilistic atlas simultaneously (Ashburner and Friston, 2005). Some approaches suggest selecting only those subset of pixel/voxel samples that have a high probability per class (De Boer et al., 2009). The unified segmentation approach that was based on the correction of the brain maps even without the need for mapping was presented by Weiskopf et al. (2011) in their work “Unified segmentation-based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT)”.

For probabilistic atlas-based segmentation, the probabilities of pixel/voxel values are integrated as a part of a statistical framework. For example, the Bayesian framework can be defined as:

$$S(x) = argmax(I(x)|c \cdot p(c), \tag{32}$$

where p(I(x) | c) represents the conditional probability of the pixel intensity values, and p(c) are the priors of the class.

Probabilistic atlases can be used with different types of parametric and non-parametric frameworks that can be expressed by the following equation:

$$S(x) = argmax(E_d + \lambda \cdot E_s), \tag{33}$$

where Ed is the energy of data, Es is the energy of smoothness, and lambda is a user-defined measure.

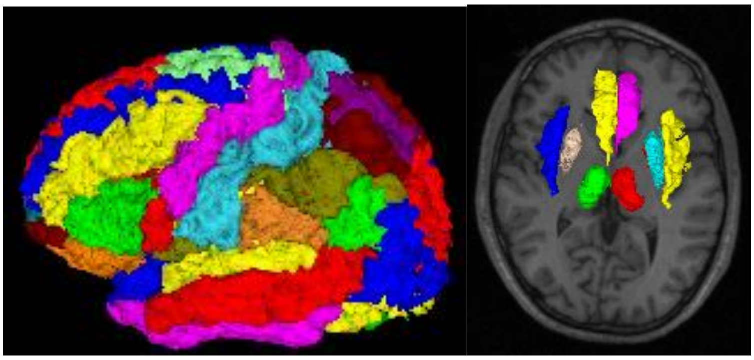


Figure K: An example of atlas-based segmentation (left) and mapping of some atlas-based elements on the original image (right). The mapping is done on Matlab using SPM software package (IBASPM (Individual Brain Atlases using Statistical Parametric Mapping) toolbox) developed by Wellcome Department of Cognitive Neurology, London, UK, https://www.fil.ion.ucl.ac.uk/spm/ext/

Anatomical variability requires the further development of new strategies and algorithms. For instance, segmentation of deep anatomical structures of the brain and tissue segmentation of fetus and newborn is a subject of main research interest (Oishi et al., 2019; Dolz et al., 2018). The segmentation of the neonatal brain from the MRI and other imaging methods is more complicated than in adults. This problem occurs due to high anatomical variability between infants and their fast and active growth. The imaging quality is also frequently insufficient. Consequently, the development of a dynamic, probabilistic atlas is vital for separate stages of the evolution of the neonatal brain.

Deformable model

Deformable models belong to surface-based methods, which includes a variety of segmentation algorithms such as active contour or snake-based model (Pierre et al., 2021). These models were introduced in 1988 (Kass et al., 1988). Deformable models outline the object boundaries by closed parametric curves (for 2D objects) or surfaces (for 3D objects). The surface regularity is controlled by external and internal image forces. These forces are represented by the fusion of geometry and physics that cooperate according to the rules of approximation theory. Geometry describes the shape of objects. Physics defines the boundaries of changes in shape which may vary over space and time. Internal forces that hold the curve together are called “elasticity forces”. Other internal forces that keep the curve from extremal bending are called “bending forces”. Internal image forces, in traditional models, rely on the edge-located intensities information and are designed to retain the model smooth during the deformation. However, it makes the model sensitive to noise and dependent on the primary estimate. More advanced algorithms use the global region information integrated into the deformable model (Zhang, 1993). External forces are designed to attract the curve towards the boundary of the object of interest within the image. There are several types of external forces: multiscale Gaussian potential force, pressure force, distance potential force, dynamic distance force, interactive force.

Multiscale Gaussian potential force. While operating, the Gaussian potential force uses different ranges of standard deviation. The principal idea is to start with a large value of sigma and to create a potential energy function that helps to allocate the boundaries around the object broadly. The large-scale force attracts the deformable contour towards the desired boundaries from a large distance. Then, the value of sigma is gradually reduced to allow tracking of the boundary with higher precision.

Pressure force. “Balloons” is another name for deformable models that use pressure forces. They are often used together with Gaussian potential forces. When they operate together, the pressure force should be slightly smaller than the Gaussian potential force applied to the same edges. At the same time, the pressure forces have to be large enough to pass through those edges, which can be labelled as weak or false. During the model deformation, it can be inflated or deflated by the pressure force until the last one is stopped by the Gaussian potential force.

Distance potential force. Distance potential force is used as an additional approach for extending the distance range. It can be done by defining the potential energy function. For this reason, a pixel distance map needs to be created. The pixel values in the map are obtained by calculating the distances between each pixel and the closest point of the boundary. Examples of calculated distances are Euclidean distance, Manhattan distance (Ranjitkar and Karki., 2016), Chamfer distance (Stencel and Janacek, 2006).

The potential energy function constructed on the distance map helps to compute a potential force field with a large range of distances. Distance potential forces have the gradient vector flow (GVF) field that works horizontally in opposite directions. This prevents the converging of the contour into the boundary concavity.

Dynamic distance force. Dynamic distance force is similar to distance potential force, but it operates in a dynamic environment when distances are recalculated each time following the model deformation. Desired points in the boundary can be found using several criteria. One of them is to use those edge points that have the highest gradient magnitude of pixel intensity values. A threshold needs to be applied to specify the maximum search distance. This approach eliminates the outliers and minimizes the computation time of the algorithm.

Interactive force. In some cases, the deformable model operates better if an operator is able to interact with it. This applies to clinical situations when the interaction is vital for high-quality segmentation or automated forces fail to adjust the model to the desired result.

At the beginning of segmentation, the closed surface/curve S is positioned near the desired object boundary in an image. During the iterative relaxation process, the external and internal forces deform this curve/surface. The mathematical expression of the energy forces is provided below:

$$F(\mathcal{S}) = F_{int} + F_{ext} \tag{34}$$

where F_{int} states for internal forces and F_{ext} for external.

Figure L shows an example of the visual representation of the deformable model in image segmentation.

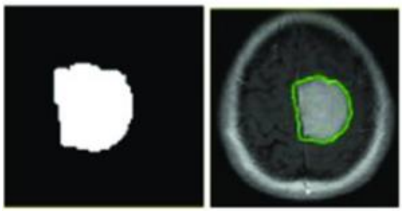


Figure L: Brain tumor segmentation using active countour deformable model

Parametric deformable and geometric deformable models

Conditionally deformable models are divided into parametric deformable models and geometric deformable models (Xu et al., 2000). The parametric deformable model represents curves and surfaces using changeable parametric forms during deformation. It allows fast interaction with the model in a real-time environment. However, this model has limitations in splitting and merging curves. The geometrical deformable model represents surfaces and curves as a set of higher-dimensional scalar functions. The model parameters are calculated after completed deformation. Because of this ability, these types of models are topological highly adaptable.

Multiphase Active Contour model

The advanced level of the development of deformable models is a Multiphase Active Contour model. This type of models was developed for imaging data with non-overlapping different intensities or other characteristics regions. Those algorithms include the multiple level set functions. They are robust to variation of images, have topological flexibility. They are adaptive to functional energies and able to detect the object boundaries with high precision. Many recent works were focused on the development of new multiphase active contours for more challenging problems (Bae et al., 2011). Some investigations were dedicated to the problems of computational complexity reduction of the multiphase active contour models (Moreno et al., 2014).

The interactive deformable models are actively used in medical diagnosis, surgical treatment and medical training with haptic devices (McInerney and Terzopoulos, 1996; Zhang et al., 2017; Romo et al., 2017).

Classification methods

Classification methods are used as automated segmentation methods that operate in the image feature spaces. They can use supervised or unsupervised learning. Supervised machine learning segmentation used manually segmented images as references. After that, new images are segmented automatically with ML algorithms. Supervised segmentation methods have some limitations. They do not consider the physiological and anatomical variability between different groups of subjects. Also, the manual segmentation step can be time-consuming.

K-nearest neighbour

One of the simplest classifiers that take part in the segmentation process is the K-nearest neighbour (KNN) (Harini and Chandrasekar, 2012). This is a non-parametric classifier which does not consider the statistical structure of the data. According to the KNN classification rules, the pixel/voxel classification is based on similarity measures and the majority vote of the closest neighbouring data. In medical imaging, the KNN segmentation procedure requires a large number of training samples collected from each tissue class. This algorithm works well with a large amount of data and variabilities in the anatomical data structure. However, it shows some sort of errors when dealing with intensity variation naturally presented within each class of tissues.

Bayesian classifier

Another popular classifier is the Bayesian classifier (Pohl et al., 2006). Unlike the previous one, this is a parametric classifier. For the image segmentation process, the Bayesian algorithm calculates the probabilistic relationships between class variables and the set of attributes. Then, the classifier estimates the class probability of any unknown variable. All pixel/voxel intensities are considered to be strongly independent. All intensities represent a mixture of Gaussian probability distributions. The Bayesian framework consists of four main components: the prior independent probability *P*(y), the evidence of independent probability *P*(x), the likelihood of conditional probability *P*(x | y) and the posterior conditional probability *P*(y | x):

$$P(y|x) = \frac{s(x|y)P(y)}{P(x)} \tag{35}$$

After calculating the posterior probability, it is possible to select the hypothesis with the highest probability. The highest posterior probability is called the maximum a posterior (MAP):

$$Y_{\{map\}} = argmax\{Y\}(P(X|Y)) \tag{36}$$

The P(X) is a constant, and it can be used for data normalization. It is possible to drop it in a formula for the MAP calculation. The formula gets the following view:

$$MAP(Y) = \max\,(P(X|Y)P(Y)) \tag{37}$$

Often, the probability distribution has an exponential function. The probability can be calculated by using a logarithmic transform:

$$Y_{\{map\}} = argmax(logP(X|Y) + logP(Y)) \tag{38}$$

Training the Bayesian classifier is a fast process because only the probability of each class and the probability of different input (x) values in the class need to be calculated. The class probabilities are calculated as the frequency of samples that belong to each class divided by the total number of samples. For the simple binary classification, each class will have the probability of 0.5 or 50% if there is the same number of samples in each class.

ANN-based segmentation

ANN-based segmentation became the most popular segmentation method because of its ability to learn from historical data and generate new rules automatically. The pixel intensities and texture features from images are collected automatically in the input layer, and then they are processed through a series of hidden layers. A set of mathematical operations are performed within each layer. The output layer provides a final result of segmentation. The best neural network used for image segmentation is the convolutional neural network. This network was designed to work especially with imaging data. CNN can reduce the number of features without significant loss of the quality of segmentation (Dolz et al., 2018).

Hybrid segmentation methods

The choice of the appropriate segmentation technique for a given task can be difficult. In this case, the combination of methods can bring a solution. Hybrid segmentation methods are continuously explored. Multiple hybrid combinations are introduced for complex segmentation problems (Nyma et al., 2012; Ortiz et al., 2014). These methods received popularity in neonatal medicine. The methods combining thresholding, Fuzzy C-means clustering, active contours, and morphology-driven automatic methods were applied for the segmentation of the various anatomical regions of the neonatal brain (Makropoulos et al., 2018; Gui et al., 2012). The main disadvantage of the hybrid method in comparison to each single segmentation approach integrated into it is the increased complexity. Lower processing time and a higher number of computation parameters make the hybrid method limited in use for a vast range of applications. On the other hand, a wisely and carefully designed method can give a valuable solution to a complex problem or problem that require a high segmentation precision.

Table B below illustrates the comparison of the segmentation method described in this chapter.

Table B: Comparison of segmentation methods

Segmentation method	Advantages	Disadvantages
Manual	Accurate	Time-consuming

Thresholding	Fast; Efficient	Depends on pick values; Sensitive to noise and texture inhomogeneity
Region-based	Low sensitivity to noise; Works well when similarity criteria are defined	Time-consuming; Required a lot of computer memory; Sensitive to initialization “seed” point
Edge-based	Fast; Segment well if the objects have significant contrast between the each other	Sensitive to the significant variation in pixel/voxel intensity values; Does not suitable for objects with weak, false or multiple edges
Clustering	More useful for real problems when having partial membership (Fuzzy clustering)	It depends on initialization parameters: the number of clusters and positioning of the “centroid”; Membership determination is not always accurate
Morphology-based	Detected boundaries are continuous and stable	Complex calculations of the intensity gradients; Over- or under-segmentation
Atlas-based	Easy to apply if the part of the body is known by atlas initially	Not accurate for deformable objects
Deformable model (active contour-based)	High quality of segmentation; Able to segment multiple objects at the same time	Needs initial contour parameters and balanced external and internal forces; Takes long runtime
Classification-based	Does not require complex programming; Operates with a large amount of data and variabilities in anatomy	Model training is time-consuming
Hybrid	High precision	High computational complexity; Time-consuming

All the described above segmentation methods can be adapted and used for MRI segmentation of the brain tissue.