MCI progression classification for early diagnosis of Alzheimer's disease using machine learning and deep learning methods



Yu Yan

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2020

Acknowledgement

I would firstly like to thank my supervisor, Professor Vicente Grau, for his constant support during my DPhil studies – for the enthusiasm and encouragement over the past three years.

I would also like to thank my industrial supervisor, Dr Edward Somer, who have helped me along the way. I am grateful to our collaborators at GE healthcare for the provision of data and their valuable advices. Thanks also to my colleagues Hoileong Lee and Hao Xu for their company and support that has made for such an enjoyable working environment.

I must also thank the China Scholarship Council and EPSRC for financial support, and the ADNI investigators for providing the imaging data and demographic information.

Finally, I am forever grateful to my parents and family for their continued support and encouragement throughout my studies.

Abstract

Alzheimer's disease (AD), the most common cause of dementia, affects more than 520,000 people in the UK. It is a progressive disease and its causes still remain unclear. Besides the traditional neuropsychological test, imaging biomarkers are playing an increasingly important role for AD detection and especially for early stage diagnosis including prediction of the transition from mild cognitive impairment (MCI) to AD. The use of structural imaging such as T1 MRI for AD diagnosis has been widely explored, however, the application of PET imaging to AD is relatively novel, with new tracers appearing. In addition, the development of machine learning and deep learning algorithms can help researchers and clinicians build computer-based image analysis tools, which can be used to further enhance the AD diagnosis results by utilising image features that cannot be easily interpreted directly by human eyes. This thesis focuses on the development of such image analysis methods for the prediction of MCI to AD conversion.

We first propose the use of novel features, based on complexity measurement, from volumetric amyloid PET images, which we use as input to a nonlinear Support Vector Machine (SVM) classifier. We tested this method using the novel PET features alone or in combination with standard PET metrics and MRI-derived features, obtaining state of the art accuracy.

Our second main contribution involves the development of a deep learning model using two parallel 3D Convolutional Neural Networks (CNNs) to combine the information from amyloid PET and T1 MRI. Deep learning methods in this area are typically limited by the small size of available PET datasets. We show that prediction accuracy is improved by combining "diagnostically similar" datasets together as a form of augmentation (progressed MCI + pAD

against stable MCI, as well as progressed MCI + pAD against stable MCI + healthy volunteer).

Our third major contribution explores a new method for augmentation using data synthesis through a conditional Generative Adversarial Network (cGAN), which can generate new amyloid PET images from volumetric T1 MRI scans, for which there is much larger availability. The generated amyloid images are used to augment the original progressed MCI dataset to address the problem of data imbalance. The results showed that the accuracy for MCI to AD prediction using generated data augmentation method outperformed traditional augmentation methods substantially.

Publications

Journal articles:

Yu Yan, Edward Somer, and Vicente Grau. "Classification of amyloid PET images using novel features for early diagnosis of Alzheimer's disease and mild cognitive impairment conversion." Nuclear medicine communications 40.3 (2019): 242-248.

Multimodal Deep Residual Network and Amyloid PET Synthesis for Predicting Progression to Alzheimer's disease (in preparation)

Conference Publications:

Yu Yan, Hoileong Lee, Edward Somer, Vicente Grau. Generation of Amyloid PET Images via Conditional Adversarial Training for Predicting Progression to Alzheimer's disease. International Workshop on Predictive Intelligence in Medicine, Springer, Cham, 2018

Patent:

Yu Yan, Vicente Grau, Edward Somer, Jan Wober. Fractal Analysis of Amyloid PET Images with machine learning algorithm (filed in Feb 12, 2018)

Contents

List of Figures 9
List of Tables 12
1 Introduction 14
1.1 Scientific contributions 15
1.2 Thesis outline 17
2 Background 20
2.1 Alzheimer's Disease 21
2.1.1 Non-imaging biomarkers for AD 22
2.1.2 Imaging biomarkers for AD 24
2.1.3 Treatment for AD 27
2.1.4 PET quantification for AD 29
2.2 Machine learning in medical imaging 30
2.2.1 Feature extraction 30
2.2.2 Image classification with machine learning 32
2.3 Deep learning in medical imaging 36
2.3.1 Convolutional neural networks 37
2.3.2 Generative adversarial networks 42
2.4 State of the art in AD classification with imaging biomarkers 45
2.4.1 AD diagnosis using machine learning 46
2.4.2 AD diagnosis using deep learning 49
3 Complexity analysis of 18F PET images and multi modal classification for early diagnosis of AD 56
3.1 Materials 57
3.1.1 Data 57
3.1.2 Pre-processing 59
3.2 Methods 61
3.2.1 Calculation of Fractal Dimension using the Box Counting Method 61
3.2.2 Shannon entropy 63
3.2.3 SVM Classification 64
3.2.4 MRI feature extraction 65
3.2.5 Experiments 66
3.3 Results and discussion 68
3.4 Conclusion 75

4 Multimodal 3D Convolutional Neural Networks for Early Diagnosis of Alzheimer's Disease using MRI and Amyloid Imaging 77
4.1 Materials 78
4.1.1 Data and pre-processing 78
4.2 Methods 79
4.2.1 Feature learning with 3D CNN 79
4.2.2 Multi-modality 3D CNN architecture 80
4.2.3 MCI progression experiment setup 81
4.2.4 Alternative CNN architecture and data dependency experiment 83
4.3 Results and Discussion 83
4.3.1 Model performance monitoring 83
4.3.2 MCI progression classification part 1 83
4.3.3 Alternative CNNs comparison and data dependency exploration 86
4.3.4 MCI progression classification part 2 87
4.3.5 Comparison with previous results and recent literature 89
4.4 Conclusion 92
5 Generation of Amyloid PET Images via Conditional Adversarial Training for Predicting Progression to Alzheimer's disease 95
5.1 Materials 97
5.1.1 Data and pre-processing 97
5.2 Methods 98
5.2.1 Amyloid PET generation with conditional adversarial training 98
5.2.2 MCI progression classification architecture using 3D ResNet 101
5.2.3 Experiments 103
5.3 Results and Discussion 107
5.3.1 Data generation 107
5.3.2 Classification results 108
5.3.3 Tuning Hyperparameters 114
5.3.4 Comparison with previous results 116
5.4 Conclusion 118
6 Conclusions 119
6.1 Summary 119
6.2 Future work 121
6.2.1 Tau imaging 121
6.2.2 Other variants of CNN 123
6.2.3 Other possibilities for future research 125

Reference 128

List of Figures

Figure 2.1.1: Comparison between healthy brain and AD brain with amyloid plaque and tau tangles (from[15]) 22

Figure 2.2.1: Self-similarity (same structure at different scales) 31

Figure 2.2.2: Illustration of linear SVM for two classes, where w is the weight vector and b is the bias. The solid line represents the hyperplane and the margin is the width of separation between support vectors of different classes. 34

Figure 2.2.3: A nonlinear boundary in the input feature space becomes a linear hyperplane in a higher dimensional space with the mapping function ∅ (x) 35

Figure 2.3.1: One neuron perceptron illustration 37

Figure 2.3.2: Plot of ReLU activation function 38

Figure 2.3.3: Convolution operation in CNN 40

Figure 2.3.4: The typical structure of a CNN. Only one layer in each type is shown in the illustration for simplicity. 42

Figure 2.3.5: The architecture of a regular GAN 44

Figure 3.2.1: Amyloid negative and positive axial slices from left to right: T1 MRI images; flutemetamol PET images; hysteresis segmentation of GM and WM region using hysteresis thresholding. The amyloid positive patient shows characteristic tracer uptake in the gray matter region. 62

Figure 3.2.2: Illustration of box counting with three different box sizes on 2D structure. 63

Figure 3.2.3: Fractal dimension measurement with log plot. The slope of the line in the right plot is an estimation of the fractal dimension. 63

Figure 3.2.4: FreeSurfer Analysis recon-all pipeline (adapted from [125]) 66

Figure 3.3.1: Histograms for HV and pAD comparison 69

Figure 3.3.2: Histograms for sMCI and pMCI comparison 69

Figure 3.3.3: 3D FD box plot for three disease states and 3D FD box plot for MCI only (Flutemetamol) 70
Figure 3.3.4: 3D SE box plot for three disease states and 3D SE box plot for MCI only (Flutemetamol) 71
Figure 3.3.5: ROC for 3D FD and SE using 65 flutemetamol images. 71
Figure 3.3.6: 3D FD box plot for three disease states and 3D FD box plot for MCI only (Florbetapir) 71
Figure 3.3.7: 3D SE box plot for three disease states and 3D SE box plot for MCI only (Florbetapir) 72
Figure 3.3.8: ROC for 3D FD and SE using 281 florbetapir images 72
Figure 4.2.1: The network architecture of multi-modality CNN for pMCI and sMCI classification 81
Figure 4.2.2: Examples of Florbetapir PET images for each of four classes 82
Figure 4.3.1: Plot of loss against training epochs for train and test (validation). To prevent overfitting, early stopping is used to specify an arbitrarily large number of training epochs and stop training once the model performance stops improving on the validation dataset. In this work, we stopped training the model when E1 E2 and E3 reached 118,115 and 108 epochs respectively 85
Figure 4.3.2: Example florbetapir images that were correctly classified, as well as others where the algorithm failed 91
Figure 4.4.1: Schematic of the conditional GANs used in this work. The MRI scan is transformed into its corresponding PET by a U-Net working as Generator, while the Discriminator tries to differentiate between the generated and real PETs, in both cases paired to their corresponding MRI 97
Figure 5.2.1: One residual unit architecture. 101
Figure 5.2.2: Complete 152-layer ResNet network architecture (for simplicity, only convolutional layer is shown in each residual block) 105

Figure 5.3.1: Randomly selected generated images in axial plane with corresponding real image 110
Figure 5.3.2: Randomly selected generated images in coronal plane with corresponding real image 111
Figure 5.3.3: Randomly selected generated images in sagittal plane with corresponding real image 112
Figure 5.3.4: Line plot loss for EX2 and EX 3 using early stopping 113
Figure 5.3.5: In EX 3, example PET images (left two columns--correctly classified, right two columns-- incorrectly classified) 114
Figure 5.3.6: The example PET images using different ε with corresponding ground truth (patch size 32) 115
Figure 5.3.7: The example PET images using different patch size with corresponding ground truth (ε=100) 116
Figure 6.2.1: Tau deposition progression based on Braak staging (from[156]) 122
Figure 6.2.2: A 5-layer dense block in DenseNet, each layer takes all preceding feature maps as input (from [158]) 124
Figure 6.2.3: The fractal architecture of the proposed network (from [159]) 125
Figure 6.2.4: Multi-layer Graph Convolutional Network (GCN) with first-order filters (from[163]) 126

List of Tables

Table 2.1.1: Five medications (listed below) to treat the symptoms of Alzheimer's disease. AChEIs (donepezil, rivastigmine and galantamine), NMDA(memantine), and a combination of AChEIs and NMDA. 28

Table 2.3.1: Symbols for the illustration of GAN 43

Table 2.4.1: study size from section 2.4 for both machine learning and deep learning based references 55

Table 3.1.1: Patients demographics and neuropsychological score for GE data 57

Table 3.1.2: Patients demographics and neuropsychological score for ADNI data 59

Table 3.2.1: Different threshold combinations and the corresponding mean FD values across all the subjects 67

Table 3.3.1: SVM classification accuracy for 65 flutemetamol and 281 florbetapir images. The top part of the Table presents results obtained using PET-derived features only, while the bottom part includes MRI-derived features too. 73

Table 3.3.2: Comparison between our proposed methods and previously published studies 74

Table 4.1.1: Demographics for ADNI data used in this study 79

Table 4.3.1: Comparison between different single modalities and multi-modality for pMCI and sMCI classification, 'FDG' and 'Amyloid' represent FDG PET and amyloid PET respectively. 86

Table 4.3.2: Two alternative CNNs trained and tested on E1 dataset using amyloid and MRI images and in comparison with proposed architecture 86

Table 4.3.3: accuracy comparison between each fold when five-fold cross validation is applied to E1 dataset (MRI+Amyloid) 87

Table 4.3.4: Accuracy (%) for each modality and two modal combinations using different training sets. 'FDG' and 'Amyloid' represent FDG PET and amyloid PET respectively. 88

Table 4.3.5: Comparisons between chapter 3 and each of the experiments in this chapter..90

Table 4.4.1: Comparison of pMCI vs sMCI classification accuracy with state-of-the-art methods, 'FDG' and 'Amyloid' represent FDG PET and amyloid PET respectively. ApoE represents Apolipoprotein E. 93

Table 5.2.1: Illustration of data used in EX 1. 103

Table 5.2.2: 152-ResNet architecture for pMCI and sMCI classification..... 104

Table 5.2.3: Illustration of data used in EX 2. 106

Table 5.2.4: Illustration of data used in EX 3. 107

Table 5.3.1: Classification ROC AUC and accuracy with 152-ResNet..... 109

Table 5.3.2: SSIM scores between generated and real PET for different ε (patch size 32). Bold text refers to the chosen regularization term ε 114

Table 5.3.3: SSIM scores between generated and real PET for different patch sizes in discriminator (ε=100). Bold text refers to the chosen patch size in discriminator 115

Table 5.3.4: Comparison between different chapters for MCI prediction performance (tra: traditional augmentation method, syn: synthetic augmentation method)..... 117

1 Introduction

1.1 Scientific contributions 15
1.2 Thesis outline 17

Alzheimer's disease (AD) is the most common cause of dementia. Dementia has a significant cost in the UK (estimated £ 26 billion per year currently) as well as worldwide (over £637 billion a year with the trend of increase)[1]. It is estimated to be responsible for 11.2% of years lived with disability in people over 60 years of age, compared with 9.5% for stroke, 5.0% for cardiovascular disease and 2.4% for cancer [2]. Due to the irreversible nature of AD, early diagnosis of AD is crucial. The progression prediction of the intermediate stage, i.e. mild cognitive impairment (MCI) has become one of the most important and difficult tasks for diagnostic and clinical utility in AD [1].

The uses of imaging in healthcare keep growing: screening, surgery planning, differential diagnosis and the requirements for follow-up [3]. To help the clinicians for the early diagnosis of AD, medical imaging is an essential tool for observing the pathological changes of AD in the brain [4]. For instance, magnetic resonance imaging (MRI) can reveal atrophy in medial temporal lobe structures such as the hippocampus. Positron emission tomography (PET) is considered as another powerful imaging modality for showing the brain function change associated with AD at a molecular level. There are mainly two types of radiotracers that are used in PET imaging: 18F-2-fluoro-2-deoxy-d-glucose (18F-FDG) for measuring the metabolism of glucose in different brain regions, and amyloid radiotracers (such as 18F-florbetapir and 18F-flutemetamol) for measuring the distribution of beta-amyloid deposits. Both types of tracers can be used in PET scans for the prediction of MCI progression to AD.

Machine learning, and deep learning algorithms in particular, have seen an explosion of interest in many research fields and especially in medical image analysis. These algorithms have been used for several challenging tasks, such as breast cancer detection and diagnosis or brain tumour segmentation with MRI [5]. By 2021, deep learning is expected to hold a $300 million medical imaging market [6]. In spite of the large interest in the area, the uses of machine learning and deep learning in medical imaging have evolved more slowly than in other fields, in part because of the lower availability of medical images.

In this thesis, we contribute towards the prediction of MCI progression in AD, through the development of machine learning algorithms. We present three major methodological contributions, in the form of three novel machine learning techniques. We provide more details about the contributions and structure of this thesis in the following sections.

1.1 Scientific contributions

In this thesis, we present a number of research contributions related to the medical image analysis field:

- The application of novel imaging features for early diagnosis of Alzheimer's disease. Two features: Fractal Dimension (FD) and Shannon Entropy (SE) are extracted from datasets containing two types of PET images (18F-florbetapir and 18F-flutemetamol) separately. In addition, the corresponding MRI features are also extracted (lateral ventricles volumes, hippocampal volume and cortical thickness). For comparison, the extracted features from both PET and MRI as well as the traditional feature used for quantification of PET scans (standardized uptake value ratios: SUVr), are used either individually or combined in several different ways to test their respective classification ability. The features are fed to a non-linear SVM classifier for classification of HV (healthy volunteers) vs pAD (probable AD) and pMCI (progressed MCI) vs sMCI (stable MCI). We show that the accuracy of the classification model is competitive with the results from recent published works in the area.

- In order to predict the MCI progression to AD, a multimodal 3D CNN architecture is proposed for pMCI vs sMCI classification. Due to the imbalance of the data (29 pMCI and 134 sMCI subjects), we implemented two extra experiments: (pMCI+pAD) vs sMCI and (pMCI+pAD) vs (sMCI+HV) to observe the effect on the classification performance. We use three imaging modalities here: T1 MRI, FDG-PET and 18F-florbetapir PET. In addition, each imaging type is trained individually and two combinations: MRI+FDG, and MRI+amyloid are trained as well to assess the relative information provided by each. Two parallel volumetric CNNs are implemented and the features are fused to make the final prediction of MCI progression. Our method showed that the modalities can be combined in a 3D CNN to obtain a better result than a single one for predicting MCI progression.
- Due to the limited availability of amyloid PET images compared with MRI images, training deep learning methods for outcome prediction based on PET is challenging. Structural MRI scans, on the other hand, are commonly available. To address this problem, we developed a model to generate amyloid PET from structural MRI using deep generative networks. We compared the generated data with the original data using quantitative metrics. Subsequently, the generated PET scans were used to augment the original datasets during the training process for MCI progression prediction. More specifically, a conditional generative adversarial network (cGAN) was used to generate the amyloid PET volumetric image from its corresponding MRI, and a 3D residual learning framework was implemented for MCI progression classification. We show that the PET scans generated from MRI are visually similar to the real PET. We evaluate this quantitatively by calculating the structural similarity index between real and generated PET, which was also high. In addition, the classification accuracy for sMCI vs pMCI compared favourably with the recent literature and with our own results without augmentation, showing the positive effect of augmentation using synthetised images.

1.2 Thesis outline

The structure of this thesis is as follows:

- Chapter 1 (this Chapter) introduces the main scientific contributions and outlines the thesis structure.
- Chapter 2 first introduces the background relevant to Alzheimer's disease in section 2.1. This is followed by explanations of imaging biomarkers for AD (section 2.1.2), treatment for AD (section 2.1.3) and PET quantification (section 2.1.4). In addition, the state-of-the-art on machine learning applications in medical imaging is reviewed in section 2.2, with a general introduction of different feature extraction techniques (section 2.2.1) and image classification applications (section 2.2.2). In section 2.3, we discuss the most recent deep learning techniques that have been widely applied in medical image analysis including convolutional neural networks (section 2.3.1) and generative adversarial networks (section 2.3.2). Finally, we discuss recent publications in AD diagnosis using machine learning and deep learning methods (section 2.4).
- Chapter 3 introduces the first research contribution chapter, which consists in a non-linear support vector machine classifier using novel features extracted from two types of amyloid PET images. The PhD project was partially funded by GE healthcare and the flutemetamol PET scans with corresponding MRI were also collected as part of a GE study. Although there was no direct involvement of clinicians in the work presented in this Thesis, the blinded image evaluation (BIE) that was used as a baseline for our study was performed by five expert clinicians. The gold standard was assessed at two years follow up for AD conversion by measuring ranges of regional cerebral to cerebellum tracer uptake ratios. Two features, namely fractal dimension (FD) and Shannon entropy (SE) were extracted from two types of PET images. MRI features were also extracted with the use of a brain imaging software package named FreeSurfer. We also calculated the traditional standardized uptake value

ratios (SUVr) as a benchmark feature. Finally, all the features in different combinations (PET features alone, MRI features alone, PET+MRI features) were fed to a non-linear SVM classifier for both HV vs pAD and sMCI vs pMCI classification.

- Chapter 4 introduces the second contribution, which is a multi-modal deep learning framework for MCI Progression prediction. In this chapter, we used three imaging modalities: FDG-PET, amyloid-PET and T1 MRI. Two parallel 3D convolutional neural networks were built and the PET and MRI features were fused before the prediction of MCI progression. Based on [7,8], both pAD and pMCI subjects have similar main symptoms such as cognitive decline and amnestic presentation. Meanwhile, HV subjects can also show mild cognitive issues and some of the sMCI subjects converted back to HV. Incorporating subjects from the HV and pAD classes to the basic pMCI vs sMCI experiments could help alleviate the problem of small datasets. We thus set up three experiments to observe the effect of the mixture of training data on the classifier accuracy. For this purpose we separated the cases into stable (HV and sMCI) and those that eventually developed pAD (pMCI and pAD), i.e we combined the pAD with pMCI subjects and the HV with sMCI subjects
- Chapter 5 is the third contribution chapter, and presents the application of image generation using conditional generative adversarial networks (cGANs). Medical imaging datasets are usually limited in size especially for some types of disease. Data augmentation is a modern way to enrich the dataset but the augmented images may be unrealistic. In this work, we investigate the use of cGANs for producing amyloid PET images from MRI to augment imbalanced datasets. The dataset is then used to train a 3D ResNet to classify images for MCI progression to AD. We compare our augmentation strategy with traditional data augmentation and show higher accuracy for pMCI vs sMCI classification in amyloid PET images.
- Chapter 6 is the final chapter that summaries all the contributions and discusses the future work. We propose a few potential deep learning alternatives and one novel

PET tracer that can be possibly applied to our dataset, as well as other possibilities for future research.

2 Background

2.1 Alzheimer's Disease 21
2.1.1 Non-imaging biomarkers for AD 22
2.1.2 Imaging biomarkers for AD 24
2.1.3 Treatment for AD 27
2.1.4 PET quantification for AD 29
2.2 Machine learning in medical imaging 30
2.2.1 Feature extraction 30
2.2.2 Image classification with machine learning 32
2.3 Deep learning in medical imaging 36
2.3.1 Convolutional neural networks 37
2.3.2 Generative adversarial networks 42
2.4 State of the art in AD classification with imaging biomarkers 45
2.4.1 AD diagnosis using machine learning 46
2.4.2 AD diagnosis using deep learning 49

This chapter introduces Alzheimer's disease (AD) and its signature imaging biomarkers, the current treatments for slowing down the disease progress and the methods for accurate detection of amyloid plaque for AD diagnosis using amyloid PET imaging. Machine learning as a branch of artificial intelligence has demonstrated its effectiveness in a range of medical image analysis applications, including the diagnosis of several diseases like AD. Feature extraction and image classification using different machine learning algorithms are described in Section 2.2 , with a particular focus on the methods that are used in subsequent chapters of this thesis, i.e. fractal dimension and Shannon Entropy as features and Support Vector Machines for classification. The application of deep learning in medical imaging is subsequently discussed, with a particular focus on Convolutional Neural Networks (CNNs) and Generative adversarial networks (GANs), both of which have been applied in this thesis.

Finally, a review of the literature related to the use of image analysis, machine learning and deep learning for image-based AD classification is included in Section 2.4.

2.1 Alzheimer's Disease

Alzheimer's disease (AD) is one of the most common types of neurodegenerative disease. There are more than 520,000 people in the UK with AD. AD leads to memory lapses, difficulties in thinking, reasoning, perception or communication[9]. The latest statistical estimation predicts that nearly one million new cases will appear each year after 2050, and the total number of patients are going to reach 13.8 million[10].

The total cost of AD in the UK is about £26 billion per year, which is enough to pay the energy bills of every household in the country [11]. At present, definitive diagnosis of AD can only be achieved by autopsy and there is no cure for AD. However, drugs and other treatments may help with both cognitive and behavioural symptoms.

In the early stages of AD, the difficulty is to distinguish normal aging from AD. In the later stages, recent autopsy research shows that AD patients can have multiple types of dementia that will confuse and cause mistakes in the diagnosis[12]. The two most prominent histopathological changes found in high numbers in the Alzheimer's disease brain are the neurofibrillary tangle and amyloid plaque [13]. The soluble building blocks of these structures are tau for tangles and amyloid-β (Aβ) peptides for plaques. Amyloid-β peptides are proteolytic fragments of the transmembrane amyloid precursor protein, whereas tau is a brain-specific, axon-enriched microtubule-associated protein [14]. The Alzheimer's brain is shrunken in the entire cortex and some subcortical regions and there is enlargement of the ventricles and shrinkage of hippocampus. The change of the brain structure with amyloid plaque and tau tangles formation is illustrated in Figure 2.1.1.

Amyloid beta primarily accumulates in the neocortex, the outer layer of the cerebrum responsible for language and sensory processing. It reaches the subcortical structure in later stages. Tau tangles primarily accumulate in the entorhinal cortex (interface between the

hippocampus and neocortex, responsible for memory formulation and consolidation) and the hippocampus. For amyloid PET imaging, the specific gray matter uptake and nonspecific white matter tracer uptake can lead to the complexity difference in gray matter and white matter boundary, in which complexity measurements such as fractal dimension (FD) and Shannon entropy (SE) can be applied to distinguish normal and abnormal subjects. The part of our work that used FD and SE will be discussed in Chapter 3.

Mild cognitive impairment (MCI) is a condition in which a patient has minor problems with cognition. It is estimated that between 5 and 20 per cent of people aged over 65 have MCI. Subjects with MCI, especially MCI involving memory problems, are more likely to develop AD

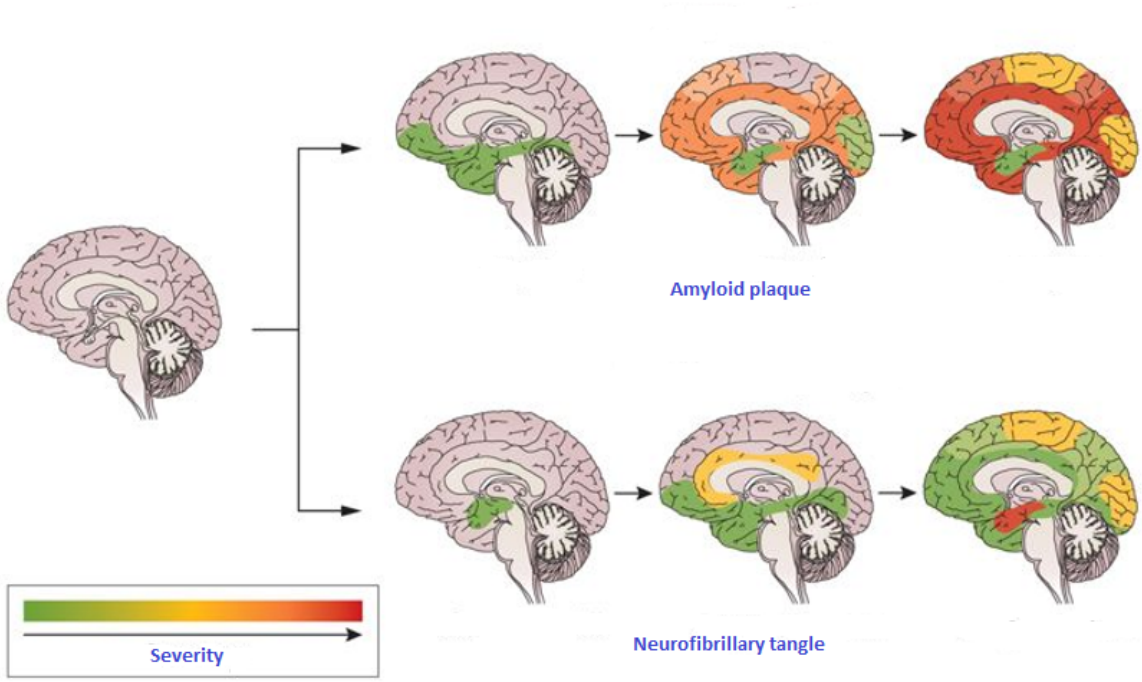


Figure 2.1.1: Comparison between healthy brain and AD brain with amyloid plaque and tau tangles (from[15]*)*

2.1.1 Non-imaging biomarkers for AD

Blood tests:

Blood tests can reliably detects signs of brain damage for people who are on the path to developing AD. Recently, a research team from the US found that blood test can detect

neurofilaments light chain, a structural protein that forms part of the internal skeleton of neurons. When brain neurons are damaged or dying, the protein leaks out into the cerebrospinal fluid and from there, into the bloodstream. They also stated that this test can detect AD even before the signs of confusion and memory loss, and this finding could be used in the clinic in a few years[16].

Neuropsychological assessment:

Neuropsychological assessment is one of the main tools used to detect AD and other dementias, typically through the use of questionnaires. Some commonly used examples include:

Mini mental state examination (MMSE): MMSE is a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment. With a maximum score of 30, a score of 20 to 24 suggests mild dementia, 13 to 20 suggests moderate dementia, and less than 12 indicates severe dementia. On average, the MMSE score of a person with Alzheimer's declines about two to four points each year.

The Alzheimer's Disease Assessment Scale-Cognitive Subscale test (ADAS-Cog): one of the most frequently used tests to measure cognition in research studies and clinical trials for new drugs and other interventions. It's more thorough than the Mini Mental State Exam, and it primarily measures language and memory.

Clinical Dementia Rating (CDR) Scale: The CDR is a 5-point scale used to characterize six domains of cognitive and functional performance applicable to Alzheimer’s disease and related dementias: Memory, Orientation, Judgment & Problem Solving, Community Affairs, Home & Hobbies, and Personal Care.

Genetics:

Researchers have found several genes that increase the risk of Alzheimer's [17]: three genes have been related with autosomal dominant familial AD (APP, PSEN1, and PSEN2)

and one genetic risk factor APOE-e4. Individuals that carry e4 are at increased risk of AD compared to those who carrying e3, while e2 decrease the risk [18].

2.1.2 Imaging biomarkers for AD

Several biomarkers have diagnostic and prognostic value in the early detection of AD. Among different biomarkers, neuroimaging has demonstrated its value in predicting MCI conversion to AD, and in differentiating AD from other types of dementia [19]. Neuroimaging techniques can be mainly classified as structural and functional. The main structural imaging techniques are computed tomography (CT) and magnetic resonance imaging (MRI), whereas popular functional neuroimaging techniques include functional MRI (fMRI) and positron emission tomography (PET) [20] .

STRUCTURAL IMAGING

Computed tomography (CT)

Computed tomography is based on the attenuation of X-rays as they pass through the body. CT scans can be used for the diagnosis of late stages of AD [19]. CT brain imaging can show the atrophy of cerebral cortex and enlargement of ventricles that are late stage features of AD [21]. In clinical practice, the most common use of CT is to look for and rule out other causes of dementia [22].

Structural magnetic resonance imaging

MRI is a non-invasive imaging technique that is widely used for volumetric analysis of the brain. The ability to observe volumetric changes of brain structures is very important for clinical diagnosis of different diseases, including AD [20]. For example, the atrophy of medial temporal lobe is an indicator for early diagnosis of AD that can be observed by a clinical expert from the volumetric MRI image [23]. The requirement of visual assessment is very high: the expert should have a great knowledge of neuroanatomy and skill in depicting the regions of interest. On the contrary, voxel based morphometry (VBM) is an automatic volumetric approach, which allows distinguishing between healthy volunteers and patients

based on the volume of brain and ROIs. SPM, Free surfer and FMRIB Software Library (FSL) are popular software tools for brain tissue image analysis and are commonly used for VBM [24]. Recently, Gomez et al. [25] created a software tool to perform 2D and 3D texture analysis on MRI. The software can extract the gray-level co-occurrence matrix and the gray-level run-length matrix from circular and spherical regions in both right and left hippocampus.

FUNCTIONAL IMAGING

Functional MRI (fMRI)

fMRI is another non-invasive imaging technique that measures brain activity by detecting changes associated with blood flow [26]. fMRI does not require the injection of a contrast agent or exposure to ionising radiation and thus can be repeated many times during a longitudinal study [27]. The primary form of fMRI uses the blood-oxygen-level dependent (BOLD) contrast. This is a type of specialized brain and body scan used to map neural activity in the brain or spinal cord of humans or other animals by imaging the change in blood flow related to energy use by brain cells [28]. Resting and activation state of fMRI studies show that AD patients have lesser coordinated activity in the hippocampus, inferior parietal lobes and cingulate cortex compared with healthy volunteers[29]. fMRI has relatively high spatial and reasonable temporal resolution, and can be acquired in the same session as structural magnetic resonance imaging [30].

Positron emission tomography

PET imaging is a functional imaging technique that can display the brain at the molecular and cellular level. In particular, PET imaging can be classified into two groups: one group utilises metabolic transformation in the brain, such as 18F-FDG. Another group is amyloid PET, which can reversibly bind to the target without being changed structurally.

FDG

18F-2fluoro-2-deoxyD-glucose (FDG) is a metabolic marker in early diagnosis and preclinical detection of dementia [31]. For over 20 years, FDG-PET has been used to measure cerebral metabolic rates of glucose in AD. The rate of brain glucose consumption in AD patients is much lower than in the normal elderly. The extent and regional distribution of hypometabolism may vary across subjects with hemispherical asymmetries. In addition, the glucose change happens prior to the atrophy, which is also important for early detection of disease[32].

Overall, the major strengths of FDG-PET in AD can be summarized as: high sensitivity to distinguish AD from HV and from other neurodegenerative diseases, and to differentiate the patients with higher or lower AD risk, and good quantitative correlation with clinical progression[33].

Amyloid

As a molecular imaging technique, the objective of amyloid PET imaging is to diagnose AD at its earliest stage. More specifically, the tracers used in amyloid PET bind to Aβ. The tracer accumulation in the regions that associate with amyloid plaques, one hallmark of Alzheimer disease, can be detected via imaging. The quantity and pattern of tracer uptake can then be used to indicate different levels of disease status. Several PET tracers for Aβ plaques have been developed in the last few years. The best known tracers are 11C-PiB, 18F- Florbetapir, 18F- Flutemetamol and 18F- Florbetaben[34]. These are discussed in the following subsections.

11C PiB radiotracers

Pittsburgh compound B (PiB) is a radioactive analog of thioflavin T, which can be used in PET scans to image beta-amyloid plaques in neuronal tissue. PiB was the first beta-amyloid tracer developed and used in investigational studies of AD. However, the use of PiB is

limited by the short half-life of carbon-11 (20 minutes), due to which the compound needs to be prepared on-site and used immediately[35].

18F amyloid radiotracers

Florbetapir, Flutemetamol and Florbetaben are radiotracers that bind to beta-amyloid using the fluorine-18 isotope, which has a half-life of 109.75 minutes, in contrast to PiB's radioactive half-life of 20 minutes. The longer half-life of the amyloid tracers allows the tracer to accumulate significantly more in the brains of people with AD, particularly in the regions known to be associated with beta-amyloid deposits, and it can be prepared off-site and transported to the imaging centres [36].

In PET imaging, the PET signal is the sum of specific binding of radioligand, non-specific binding of radioligand and the free ligand in tissue. Specific binding is defined as the tracer binding to the target receptor. Nonspecific binding is the tracer binding to other than the target molecule. For amyloid tracers, the specific grey matter uptake in amyloid positive subjects is higher than in amyloid negative ones. The reason for the high white matter signal in amyloid PET imaging has yet to be satisfactorily explained in the literature [37]. One explanation is that the amyloid PET tracers non-specifically bind to myelin in the white matter with high affinity[38].

2.1.3 Treatment for AD

There is currently no cure for Alzheimer's disease, although medication is available that can temporarily reduce some symptoms or slow down the progression of the condition in some people. There are two types of medication used to treat Alzheimer's disease: acetylcholinesterase inhibitors (AChEIs: donepezil, rivastigmine and galantamine) and N-methyl-D-aspartate (NMDA: memantine) receptor antagonists [39][40]. Five drugs are approved by the FDA (U.S. Food and Drug Administration) for the treatment of AD, which are shown in Table 2.1.1 [41,42].

Table 2.1.1: Five medications (listed below) to treat the symptoms of Alzheimer's disease. AChEIs (donepezil, rivastigmine and galantamine), NMDA(memantine), and a combination of AChEIs and NMDA.

Drug name	Approved For	FDA Approved
donepezil	All stages	1996
rivastigmine	All stages	2000
galantamine	Mild to moderate	2001
memantine	Moderate to severe	2003
donepezil and memantine	Moderate to severe	2014

For three AChEIs inhibitors, one might suit a certain individual better than another, particularly in terms of side effects experienced [42]. The action of memantine is different from that of donepezil, rivastigmine and galantamine. Glutamate is used by nerve cells to send signals to other cells. The generation of glutamate is increased in Alzheimer's disease patients[43][44]. In this case, memantine can protect brain cells from cellular damage due to excess glutamate [45].

The demand for new treatments of AD and the need for better understanding of AD therapies is urgent. Currently, most drugs are disease-modifying therapies (DMTs) [46]. Trials include patients with preclinical AD (cognitively normal with biomarker evidence of AD), prodromal AD (mild cognitive symptoms and biomarker evidence of AD), and AD dementia [41]. Thirteen of 46 phase II DMT trials have amyloid imaging as an entry standard, and 10 of 28 phase III trials use amyloid imaging for diagnosis and entry. This indicates that amyloid imaging is playing an increasingly important role in the development and evaluation of drugs to treat AD.

2.1.4 PET quantification for AD

For PET imaging, standardized uptake value (SUV) is a quantitative method used to make regional comparisons within a subject as well as between subjects and computed as the degree of radiotracer uptake in a target ROI [47]. SUV is a dimensionless measurement and can be calculated as[48]:

$$SUV_{ROI} = \frac{Q_{ROI}}{ID/m} \quad (1)$$

where Q_{ROI} is the radioactivity concentration (in kBq/g) in the regions of interest. ID is the injected dose (in kBq), and m is the body mass (in g).

As can be seen from (1), the SUV normalises for body mass. This can lead to a problem because some tracers such as FDG have very low uptake in fat [49]. To avoid this problem, the lean body mass (mass ignoring fat) can be used in the calculation rather than whole body mass [50]. Another option is standardized uptake value ratios (SUVr). SUVr is calculated as:

$$SUVr_{ROI} = \frac{SUV_{ROI}}{SUV_{REF}} = \frac{Q_{ROI}}{Q_{REF}} \quad (2)$$

where Q_{REF} is the radioactivity concentration in the reference region. The reason for using SUVr is that the body mass is cancelled out in the calculation. In AD, this is typically calculated using part or the entire cerebellum SUV as a reference (SUV_{REF}) because the cerebellum is not affected until the late stages of AD [51]. Some studies show that part or the entire frontal lobe or pons are also good options as reference regions [52]. In addition, the SUVr is calculated when the tracer binding reaches steady state, which can then avoid the effect of blood flow [53].

In summary, SUVr is the most common quantitative method used to make regional comparisons. The combination of SUVr and visual assessment can also decrease inter-reader variability [54]. For the PIB tracer, SUVr values for AD subjects in regions known to

contain amyloid in AD are significantly higher than healthy volunteers. The most significant differences are generally found in lateral temporal lobe and parietal lobe. Meanwhile, there are no significant differences in PIB retention between AD and healthy subjects for regions such as subcortical white matter and pons, which are the regions known as amyloid free in mild to moderate AD subjects [47]. For fluorine-18 tracers such as ^{18}F- Florbetapir, the amyloid deposition is high in cortical target areas such as the frontal cortex, temporal cortex and precuneus, which is reflected in higher SUVr [55].

2.2 Machine learning in medical imaging

Visual assessment is the traditional way to make diagnosis decisions from medical imaging. However, it is time consuming and different experts might arrive at different conclusions. For example, defining the region of interest, reference region, or setting threshold values by researchers and radiologists can cost substantial manual efforts or require extensive clinical trials [56]. In contrast, AI algorithms can find patterns in imaging that are imperceptible to the human eye, particularly when analysing 3D images. Automated tools developed for medical image analysis can be very useful to assist the decision making of radiologists [57].

Machine learning is a branch of artificial intelligence based on the idea that the design of the algorithm can improve their performance automatically through experience. The use of machine learning in image registration, segmentation and classification is becoming increasingly popular in recent years, thanks to the accuracy and robustness of the methods developed. This thesis focuses primarily in its use for image classification, either to detect AD or to predict MCI conversion. In this chapter, we introduce the two most important steps in image classification using machine learning: feature extraction and classification.

2.2.1 Feature extraction

High dimensional data is problematic for classification algorithms due to a number of factors, including its high computational cost and memory usage. For machine learning applications, in a high dimensional space, a very large amount of training data is usually needed. In addition, many of the dimensions are irrelevant or highly correlated. Several dimensionality

reduction techniques have been proposed that aim at extracting relevant features from the data (images in our case) and select those features that are most important for the task at hand. The chosen feature can be as simple as the pixel or voxel value in the image, or more complicated such as hierarchies of image receptive fields [58]. In order to extract discriminative features for AD diagnosis, we compute complexity features on boundary contour between gray matter and white matter, gray matter and CSF, in amyloid PET, which could be potentially used to distinguish different stages of Alzheimer's disease. This is inspired by the complexity difference that exist between HV, MCI and AD brain, due to brain atrophy along AD progression: previous work [59,60] concluded that aging or diseased brains have lower complexity. In this case, this thesis is concerned more specifically about two complexity features named fractal dimension (FD) and Shannon entropy (SE).

Fractal Dimension

The first characterization of FD was proposed in [61] to calculate the length of a coastline: the length is measured by the line segments with various lengths and the total number of line segments that have been used is counted.

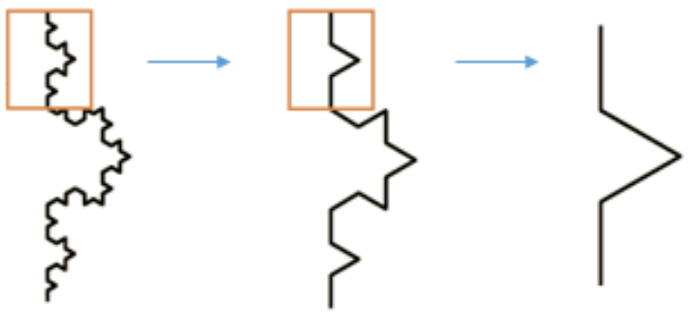


Figure 2.2.1: Self-similarity (same structure at different scales)

Fractal dimension analysis has been applied to structural imaging such as MRI. For example, Bullmore et al. [62] used fractal dimension to measure brain MRI boundary for a study of schizophrenic and manic-depressive patients. Significant differences in mean FD between diagnostic groups were demonstrated by analysis of covariance, with age and volumetric measures of brain size as covariates. In addition, King et al.[63] calculated FD of

three cortical models using a cube counting algorithm, and the results indicate that the HV and AD subjects can be accurately classified using FD of cortical ribbon.

Shannon Entropy

Shannon entropy (SE) is an alternative complexity measure to FD that quantifies the information content of the data. The concept of SE was first introduced by Claude Shannon in his paper “A Mathematical Theory of Communication” [64]. There are also some other varieties of entropy such as Renyi, Aczel and Varma [65–67].

Recently, SE was investigated as a diagnostic tool in patients with mild traumatic brain injury (TBI) using diffusion tensor imaging, to differentiate between those with and without posttraumatic migraines [68]. The results show that white-matter damage from a mild traumatic injury may cause some decrease of image complexity that results in lower SE.

In addition to SE, statistical complexity (SC) and excess entropy (EE) are also complexity measures. A higher SC represents an increase in locally correlated patterns. Meanwhile, EE measures convergence rate of entropy. A higher EE represents increased long range correlations across regions. For example, Young and Schuff proposed [69] a complexity estimation method based on Shannon entropy, EE and SC. This method is implemented on T1-weighted brain MRI for automatic detection and classification of changes in cortical thickness.

As discussed in the previous paragraphs, both FD and SE show promising classification capacity on MRI. However, the application of FD and SE to PET and especially amyloid PET has not yet been widely explored before the conduction of the experiment in Chapter 3. For feature extraction in PET, we explore both FD and SE approaches and the details are stated in Chapter 3.

2.2.2 Image classification with machine learning

After feature extraction from the training data, a classifier can be used to predict which category the image belongs to. Among the wide variety of classifiers, SVM is one of the most

widely used machine learning algorithms that are applied to solve binary classification problems [70][71]. In addition, SVM has shown promising results for early diagnosis of AD compare with other learning based models, when using T1 MRI or FDG PET for training the classifiers. For instance, Magnin et al. [72] used SVM for classification between pAD and HV. They reached mean accuracy 94.5% (mean specificity, 96.6%; mean sensitivity, 91.5%), which is trained on 38 3D T1 MRI images (16 pAD and 22 HV). Another work [73] confirmed that the support vector machine showed the best performance in prediction of future development of AD from MCI. They applied SVM, general linear model (GLM), subprofile modeling (SSM) to FDG PET (94 AD and 111 HV), which were collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI). More importantly, traditional machine learning methods usually work better than deep neural networks when we do not have a large amount of data to train the model [74]. For these reasons, we used SVM as the binary classifier for our dataset.

A two-class support vector machine classifier separates points in a multidimensional space into two classes by building a hyperplane that maximises the margin, which is the distance between the closest points on either side of the boundary. These points closest to the boundary are named support vectors [75]. The original SVM is suited to linearly separable training data, as illustrated in Figure 2.2.2. For data that are not linearly separable, the soft-margin method as well as the kernel trick have been proposed to create nonlinear classifiers.

Linear SVM

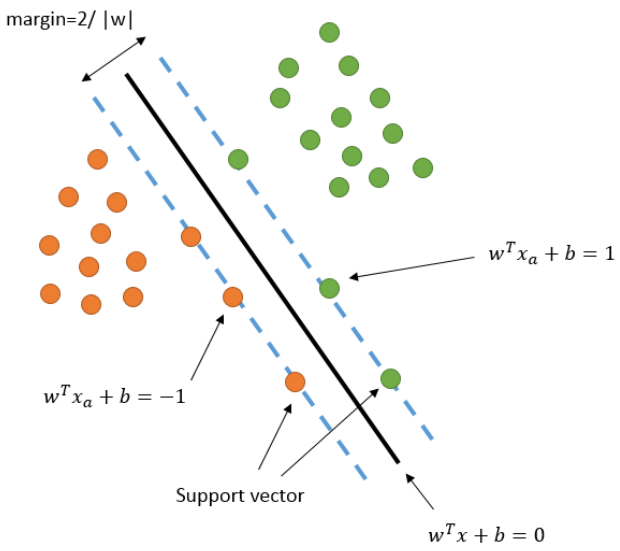


Figure 2.2.2: Illustration of linear SVM for two classes, where w is the weight vector and b is the bias. The solid line represents the hyperplane and the margin is the width of separation between support vectors of different classes.

The decision surface of a linear SVM classifier is described by y(x) = $w^T x + b$=0. The feature weight vector w (normal to the separating hyperplane) and bias b are then chosen such that the margin is maximised. As shown in Figure 2.2.2, the values are calculated to make the support vectors lie on two parallel hyperplanes described as y(x) = 1 and y(x) = -1, so that the distance between them is 2/||w||. The maximisation of the margin can then be expressed as the constrained optimisation:

$$\min_{w} \ \frac{1}{2} w^T w \qquad (3)$$

$$\text{subject to } t_i(w^T x_i - b) \geq 1,$$

where x_i and t_i are feature vector and corresponding label.

Soft-margin SVM

The hard margin SVM explained in the previous section can only work without any errors when data are completely linearly separable. Soft margin SVM was proposed [75] to solve this problem by introducing slack variables. Slack variables ξ_i allow misclassification of

difficult or noisy training examples by measuring the degree of misclassification of feature vectors. In addition, slack variables also allow training examples to exist in the space between the hyperplanes formed by the support vectors. The optimisation then becomes a trade-off between maximising the margin and minimising the degree of misclassification. This trade-off is controlled by the regularisation parameter C, such that the constrained optimisation may be expressed as:

$$\min_{w,\xi} \left\{ \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \right\} \tag{4}$$

subject to $t_i(w^T x_i - b) \geq 1 - \xi_i$ and $\xi_i \geq 0$.

Non-linear SVM

For data that are not linearly separable in the input feature space, the original feature space can be mapped to some higher-dimensional space where the training set is linear separable. As shown in Figure 2.2.3, in this example data are linearly separable by the hyperplane after the mapping function $\emptyset\ (x)$ was applied.

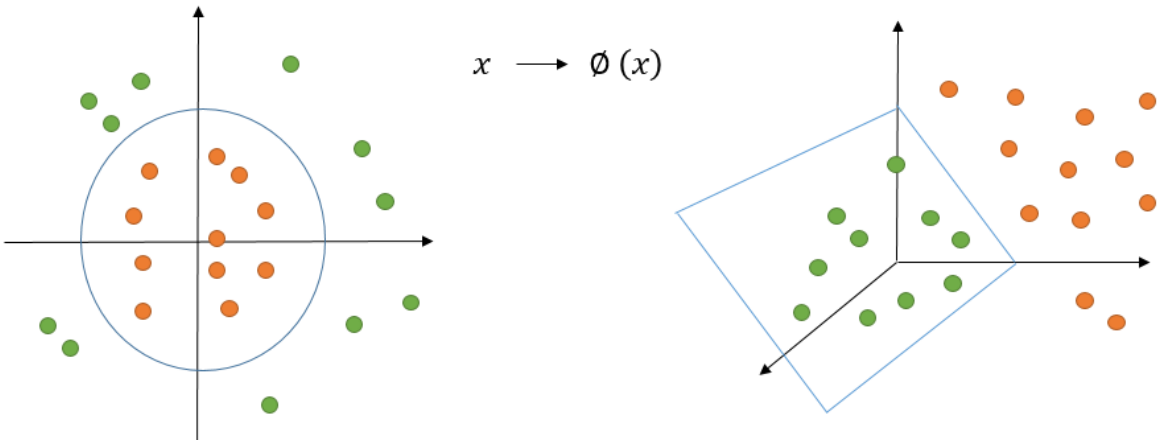


Figure 2.2.3: A nonlinear boundary in the input feature space becomes a linear hyperplane in a higher dimensional space with the mapping function Ø (x)

In the higher dimensional space, the optimisation problem is thus related to the inner products of the transformed feature vectors. More specifically, $\mathrm{k}\left(x_i, x_j\right) = \emptyset\left(x_i\right)^T \emptyset\left(x_j\right)$ is the kernel function that can represent the inner products with a nonlinear mapping function $\emptyset\left(x\right)$. It is not necessary to explicitly perform the mapping. The most commonly used kernel is the Gaussian radial basis kernel, given by $\mathrm{k}\left(x_i, x_j\right) = \exp\left(-\gamma||x_i - x_j||^2\right)$, where $\gamma > 0$ describes the width.

Due to the fact that SVMs try to classify the data in higher dimensional space, rather than minimise the error rate directly, they are relatively insensitive to the relative numbers of each class. For example, adding extra data that are far behind the hyperplanes does not change the support vectors or classification boundary [76]. This is advantageous for classification cases in which the data are imbalanced. For this reason, in Chapter 3 we use a nonlinear SVM to separate healthy subjects from AD patients as well as stable MCI from progressive MCI.

2.3 Deep learning in medical imaging

Deep Learning is a subfield of machine learning that uses a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. Learning can be supervised, semi-supervised or unsupervised. The term deep learning was introduced to the machine learning community by Rina Dechter in 1986 [77], and to artificial neural networks by Igor Aizenberg and colleagues in 2013, in the context of Boolean threshold neurons [78]. The deep learning revolution in image and object recognition starts from 2011 to 2012 following a series of wins in the international contests such as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Since then, the use of deep learning in image analysis has grown to include image classification, object detection, segmentation, registration, and many other tasks.[79][80]

Deep learning algorithms, in particular convolutional neural networks (CNNs), have rapidly become a methodology of choice for analysing medical images. The architecture of CNNs is introduced in chapter 2.3.1

2.3.1 Convolutional neural networks

Convolutional neural networks (CNNs), as the most successful type of models for image analysis to date [81]. In computer vision, deep convolutional networks have become the default choice.

In order to understand how CNNs work, we first introduce the perceptron, the most basic neural network. As can be seen from Figure 2.3.1, feature vectors $x_1 \cdots x_n$ are inputs in the network. Each input in the feature vector is assigned its own relative weight ($w_1 \cdots w_n$), which determines the impact that the particular input needs in the summation function. A bias (w_0) is also added to the summation function.

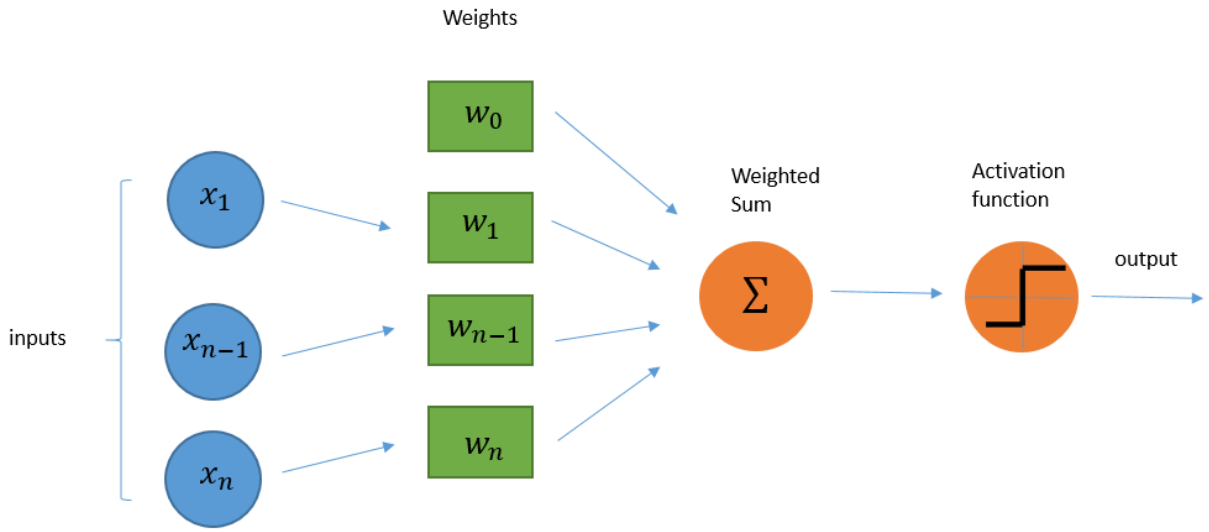


Figure 2.3.1: One neuron perceptron illustration

$$\text{weighted sum} = w_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n \quad (5)$$

The weighted sum from the summation function is transformed to the desired output by a nonlinear function called activation function, which can bring non-linearity to the neural networks so that they can approximate any functions, linear or non-linear. Commonly used activation functions include Rectified Linear Unit (ReLU), sigmoid and tanh, however, ReLU units have been found to offer some advantage since they significantly decrease training times without making a significant difference to the accuracy [82]. The ReLU layer simply applies the function $f(x) = \max(0, x)$ to all the values in the input, which changes all negative values to 0. The plot of the ReLU function is shown in Figure 2.3.2:

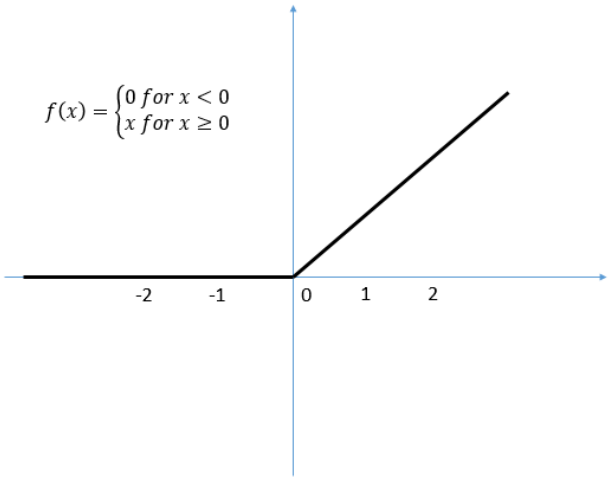


Figure 2.3.2: Plot of ReLU activation function

The loss function is an important part in learning networks, which is used to measure the inconsistency between predicted value and actual label. It is a non-negative value, where the accuracy of the model increases along with the decrease of the value of loss function. Different loss functions will give different errors for the same prediction, and thus have a considerable effect on the performance of the model. One of the most widely used loss functions is mean squared error, which calculates the square of the difference between actual value and predicted value. Different loss functions are used to deal with different type of tasks, e.g. regression and classification.

The error that needs to be minimized is related to internal parameters (weights w and bias b). During training of the network, both parameters are updated by a method called back propagation. More specifically, the current error is propagated backwards to the previous layer to adjust the weights and bias in the previous layer. The weights are modified using an optimization function.

Among the different optimization functions, stochastic gradient descent (SGD) is fast but recent studies have suggested that other algorithms such as Adam (Adaptive moment estimation) can lead to more accurate results [83]. Adam is an adaptive learning rate optimization algorithm that has been designed specifically for training deep neural networks. This algorithm leverages the power of adaptive learning rate methods to find individual learning rates for each parameter. However, Adam is more computationally expensive and hence the trade-off between more computational power and more accurate results has to be considered.

Convolutional neural networks are based on the main concepts of the perceptron. In mathematics, the convolution of two functions f and g is defined as the integral of the product of the two functions after one is reversed and shifted. In this case, it is a special kind of integral transform. The convolution operator is typically denoted with an asterisk:

$$(f * g)(t) = \int_0^t f(t-\tau)g(\tau)d\tau \tag{6}$$

More specifically, the convolution operation can be visualised as placing the convolution kernel (filter) on top of the input image, positioned so their upper left corners overlap, then multiplying the values of the input matrix with the corresponding values in the convolution filter. All the multiplied values are then added together to a single scalar, which is placed in the first position of the resulting tensor (feature map). The kernel is then moved certain amount of steps to the right, where the amount of steps is denoted as stride. The process of

multiplication and summation is then repeated, filling the next value in the feature map. After the entire row is filled in the feature map, we shift down the columns with the same stride length until all entries in the input image has been covered. This process is shown in the figure below:

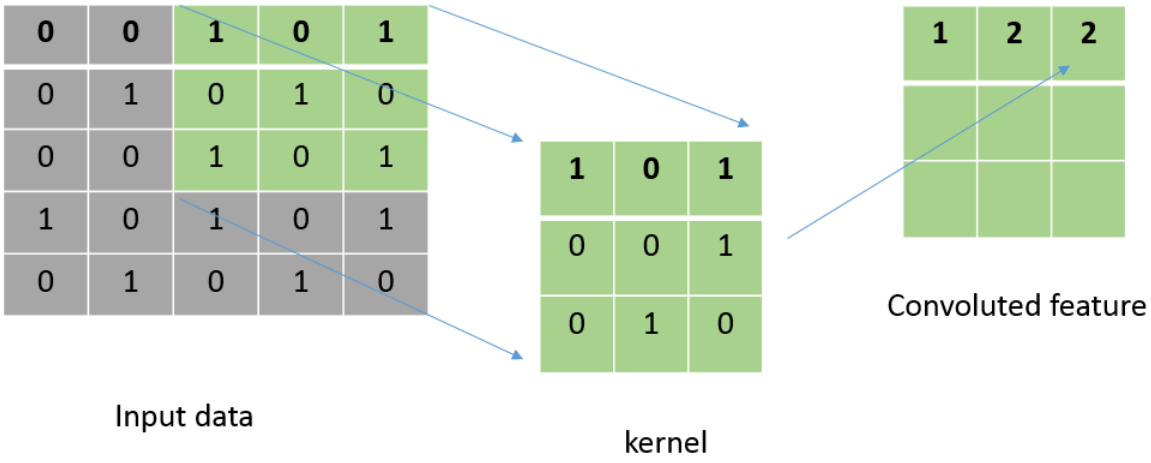


Figure 2.3.3: Convolution operation in CNN

In some applications, it may be preferable to keep the same input and output size in the convolution. This is usually solved by padding. The formula to calculate the output size is:

$$o = \frac{w - k + 2p}{s} + 1 \qquad (7)$$

Where o is the output height, w is the input height, k is the filter size, p is the padding and s is the stride. If o is not an integer, then the strides are set incorrectly.

After the convolution layer and activation function, a pooling layer is often applied, which is sometimes referred to as a downsampling layer. Pooling layers are used to reduce the size of feature maps from the previous convolutional layer, while providing positional invariant

feature extraction. The stride is usually the same length as filter length, resulting in non-overlapping operations. Two types of pooling operations are widely used: max pooling and average pooling, which calculate the maximum value and average value in the pooling region respectively.

After multiple layers of convolution, activation function and pooling, in networks designed for many applications including classification we reach the fully connected layers. These work similarly to the multi-layer perceptron used in traditional neural networks, where the input to a fully connected layer includes all the outputs of the previous layer. Fully connected layers are used to detect specific global features after the input has been reduced to a compact representation of features. This is where most of the weights are located in the network, which means that the fully connected layers are the most computationally expensive part.

For applications such as image classification, the fully connected layers are usually ended with a softmax function. The softmax layer outputs a probability distribution, i.e. the values of the output sum to 1. The softmax equation is defined by the formula:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{8}$$

For $i = 1,2 \ldots, K$ and $z = (z_1, \ldots, z_K) \in R^K$

In summary, a CNN normally consists of a combination of five kinds of layers: input layer, convolutional layer, pooling layer, fully connected layer and output layer. The structure of a typical CNN in detail is shown in Figure 2.3.4.

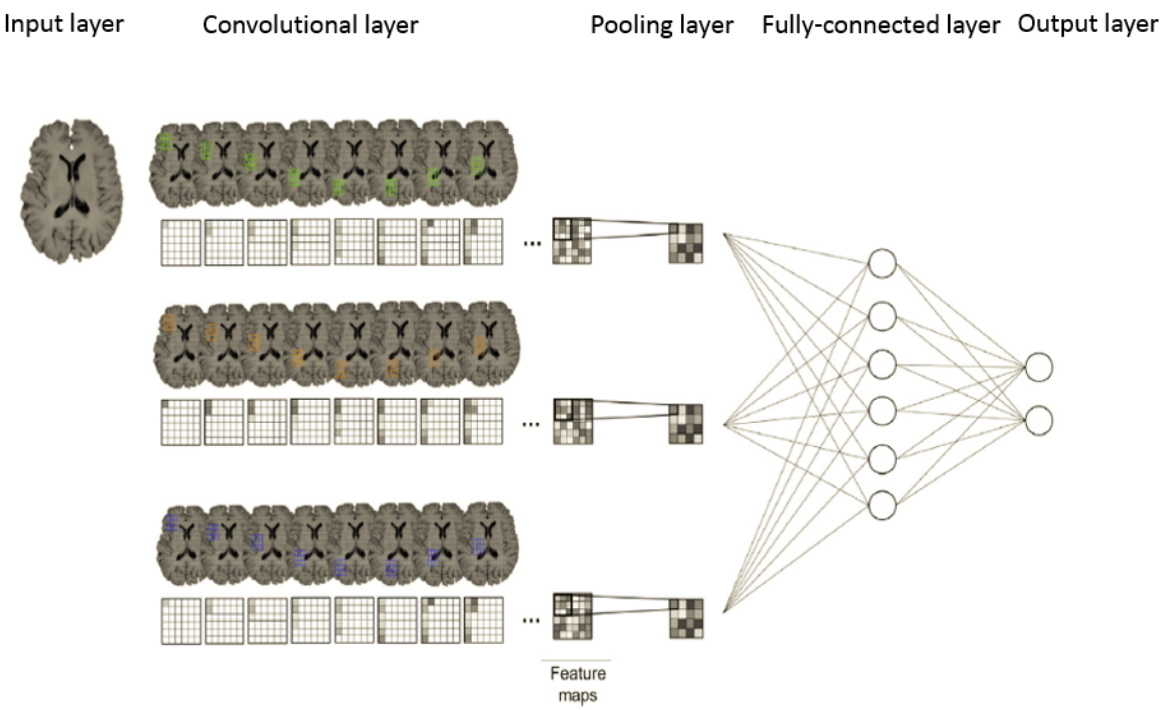


Figure 2.3.4: The typical structure of a CNN. Only one layer in each type is shown in the illustration for simplicity.

2.3.2 Generative adversarial networks

Generative modelling is the use of artificial intelligence, statistics and probability in applications to produce a representation or target variables that can be calculated from observations, and can then be used to generate new samples corresponding to the learned distribution. Generative adversarial networks (GANs) are powerful generative models with a wide range of applications including image synthesis, semantic image editing, style transfer, image super-resolution and classification [84]. GANs are inspired by the two-player zero-sum game, a mathematical representation of a situation in which one player's gain or loss of utility is exactly balanced by the losses or gains of the utility of another player, and the total gains of the two players add up to zero.

GANs normally comprise two parts: generator and discriminator. Both are trained simultaneously, and in competition with each other [85]. The generator is used to extract the distribution from the real data and generate new data, and the discriminator tries to

distinguish real data from the generated data. We name generator and discriminator as G and D respectively, and give the symbol and meaning in Table 2.3.1:

Table 2.3.1: Symbols for the illustration of GAN

symbol	meaning
p_z	Data distribution over noise input z
p_g	The generator distribution over data x
p_r	Data distribution over real sample x

On one hand, the algorithm tries to make sure the discriminator D's decisions over real data are accurate by maximizing $E_{x\sim p_r(x)}[logD(x)]$. Meanwhile, given a fake sample $G(z), z\sim p_z(z)$, the discriminator is expected to output a probability, $D\big(G(z)\big)$, close to zero by maximizing $E_{z\sim p_z(z)}[\log\Big(1-D\big(G(z)\big)\Big)]$. On the other hand, the generator is trained to increase the chances of D producing a high probability for a fake example, thus to minimize $E_{z\sim p_z(z)}[\log\Big(1-D\big(G(z)\big)\Big)]$.

When combining both aspects together, G and D are playing a minimax game in which we should optimize the following loss function:

$$\begin{aligned}\min_G \max_D L(D,G) &= E_{x\sim p_r(x)}[logD(x)] + E_{z\sim p_z(z)}\Big[\log\Big(1-D\big(G(z)\big)\Big)\Big] \\ &= E_{x\sim p_r(x)}[logD(x)] + E_{x\sim p_g(x)}\big[\log\big(1-D(x)\big)\big]\end{aligned} \tag{9}$$

Where E is the expectation operation and z is random Gaussian noise.

Overall, the discriminator tries to maximize the loss function while the generator tries to minimize it, for which gradient ascent and gradient descent are applied respectively. By alternating between gradient ascent and descent on the objective function, the network can

be trained. During training, there are two stages and they are performed sequentially. In the first stage, the discriminator is trained and the generator is frozen (i.e. the network does only forward pass and no backpropagation is applied). In the second stage, the generator is trained and the discriminator is frozen.

More specifically, first the discriminator is trained on real data for a certain number of epochs, and then the discriminator is trained on generated data from the generator; finally, the generator is trained on the output of the discriminator. These three steps repeat until there is no visually perceived improvement in the generated images. The GAN model converges when the discriminator and the generator reach a Nash equilibrium. The regular architecture of a GAN is shown in Figure 2.3.5.

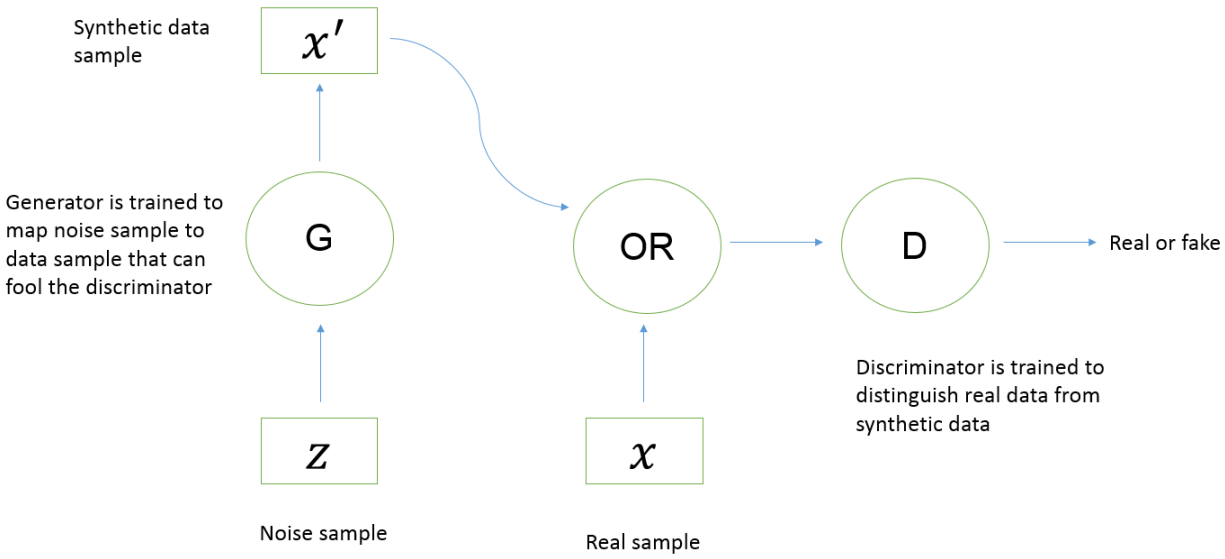


Figure 2.3.5: The architecture of a regular GAN

The very first version of GAN used fully connected layers for both generator and discriminator [86]. However, the application of this model was limited to relatively simple datasets such as MINIST, CIFAR-10 and the Toronto Face Dataset. Afterwards, the deep convolution GAN (DCGAN) was developed which allows training a pair of deep convolutional

generator and discriminator networks [85]. It mainly composes of convolution layers without max pooling or fully connected layers. It uses convolutional stride and transposed convolution for the down-sampling and the up-sampling respectively. In 2016, a new GAN model named conditional generative adversarial networks (cGANs) was proposed for image to image translation [87]. This network can learn both the mapping between input image and output image and a loss function to train the mapping, and it has achieved excellent results for synthesizing photos from label maps, reconstructing objects from edge maps, and colorizing images. Its application for medical imaging synthesizing will be discussed in Chapter 5.

2.4 State of the art in AD classification with imaging biomarkers

Over the past decade, classification frameworks using machine learning and deep learning have been successfully applied to neuroimaging data, including AD detection and MCI conversion prediction. The essential steps for a classification model are the feature extraction and classification algorithms, while other steps such as feature selection and dimensionality reduction can be added if needed. In the feature extraction step, neuroimaging data from multiple modalities such as structural MRI, functional MRI, diffusion tensor imaging (DTI), amyloid-PET, and FDG-PET can be used to extract AD-related features. The extracted features from one modality can be used alone or in combination with features from other modalities to make use of the complementary information provided by several modalities. A classification algorithm is then trained on the extracted features to provide diagnostic support in differentiating healthy and diseased subjects. In terms of the two essential steps for classification, the choice of extracted features and imaging modality are considered more important than the choice of classification algorithms for achieving a higher classification accuracy [88]. In this chapter, recent works for both single modal and multimodal AD/MCI classification framework using machine learning and deep learning are discussed.

2.4.1 AD diagnosis using machine learning

This section explores the application of machine learning methods for classifying different AD stages to improve accuracy. Three imaging modalities (structural MRI, FDG PET and amyloid PET) as well as multi modalities are discussed in the following section.

Structural MRI based studies

Among the various neuroimaging modalities, structural MRI has been the most frequently used. In this section, the related studies are divided into two categories: cortical surface based works and pre-defined regions-based works.

For the cortical surface, changes in temporal and parietal regions are generally observed for AD patients [89,90]. In [89], the authors scanned two independent samples of cognitive normal patients. In sample 1, 8 subjects that converted from healthy status to AD with an average conversion year 11.1 are compared with 25 subjects who remained cognitive normal. In sample 2, 7 CN-AD converters with an average conversion year 7.1 are compared with 25 CN-stable. Finally, they concluded that cortical atrophy measurements such as cortical thickness thinning can detect dementia nearly a decade before the standard diagnosis, which indicated the importance of preclinical diagnosis. Similarly, Dickerson et al. [90] also proposed that the thinning in cortical regions can be used to detect the earliest stages of clinical symptoms, especially for amyloid-positive asymptomatic individuals. The data for this study involved 380 participants (336 using structural MRI and 44 using amyloid-PIB PET and structural MRI).

Although these changes are hard to measure in the early stage of AD, some classification frameworks are able to detect them. Cortical surface measures are extracted from the points (vertices) of a surface, which include two main categories: vertices as feature-based methods and reduced vertices as feature-based methods.

Vertices as feature-based methods simply refer to features that are calculated from all the vertices of the cortical surface. For example, volumetric (cortical thickness, surface area, and

grey matter volume) and geometric (metric distortion and mean curvature) measurements can be calculated at each vertex on the pial surface, which can subsequently be used for MCI classification [91]. The participants included 24 MCI and 26 heathy volunteers. They used SVM as classifier and the regions identified by their method could potentially be useful for clinical diagnosis. However, due to the large dimensionality of the extracted features, reduced vertices as features-based methods were developed by using a pre-defined atlas or feature reduction methods.

Accurate selection of pre-defined regions allows the introduction of prior knowledge, which leads to a more effective and robust classification framework. For example, in [92], the authors studied 19 probable AD patients and 16 healthy volunteers, and mapped gray matter density by means of voxel-base morphometry on T1-weighted MRI images. They reported significant gray matter loss which affecting mainly the medial temporal structures while the frontal lobe is slightly affected.

Among the different structures in the brain, the hippocampus, a brain area located in medial temporal lobe, is critical for learning and memory [93]. The hippocampus is especially vulnerable to damage at early stages of Alzheimer's disease [94]. The shape and volume of hippocampus has been shown to be an effective biomarker for AD and MCI classification. For instance, Sørensena et al. [95] extracted different MRI features from ADNI and AIBL MRI data for multi class classification (HV, pAD and MCI). They collected 544 studies from ADNI, 145 studies from AIBL, and 384 studies from CADDementia challenge data. They found that the hippocampal texture was the most important feature in the algorithm followed by hippocampal volume, ventricular volume, and parietal lobe thickness.

For PET based studies, there are two major classes of tracer for brain PET imaging: metabolism PET tracers and amyloid PET tracers. The relevant studies that applied machine learning approaches to PET using different tracers are discussed below.

FDG PET based studies

One example for the first category is 18F-FDG PET tracer. The uptake of 18F-FDG by tissues is a marker for the tissue uptake of glucose, which in turn is closely correlated with certain types of tissue metabolism [96].

FDG PET is extensively and increasingly used to support the clinical diagnosis in the examination of patients with suspected neurodegenerative disorders, especially AD. For instance, Cabral et al. [97] developed an ensemble method named favourite class ensemble for classification of AD, MCI and HV. They randomly selected a class balanced subset of 177 FDG PET volumes from ADNI database. In this method, each base classifier in the ensemble is assigned with a subset of features and optimized for one of the three classes. They tested the approach on FDG PET brain imaging and used SVM and random forests as base classifier. The results indicate that the ensembles outperformed the corresponding single classifier and the SVM ensemble reached the highest accuracy compared with the random forest ensemble. In addition, Katako et al. [73] compared three classification methods (general linear model, scaled subprofile modelling and SVM) for AD prediction using FDG PET brain imaging from the ADNI database. In total, 763 subjects from the ADNI database who had available PET and MRI scans were included in this study. They applied the same algorithm to four different datasets (HV, AD, Stable MCI, Prodromal AD) and found that SVM showed the best performance for MCI to AD prediction. They also conducted a retrospective imaging study to demonstrate the applicability of proposed classification methods in a real clinical setting.

Amyloid PET based studies

A recent review [98] highlights that the use of novel features from amyloid PET to classify MCI to AD progression has not been widely explored. Tracing back to 2013, Vandenberghe et al. [99] proposed a framework that used the intensity values of all the voxels of 18F-flutemetamol PET scans as features for SVM-based classification of AD vs HV, and pMCI vs sMCI subjects. They used 18F-flutemetamol scans and volumetric MRI scans from 72 cases from the 18F-flutemetamol phase 2 study (27 clinically probable AD, 20 MCI, 25 HV). For

distinguishing AD from NC, they showed that the 18F-flutemetamol based classifier was able to duplicate the same accuracy as expert readings. They also compared an 18F-flutemetamol based classifier with a structural MRI (gray matter volume maps as features) based classifier. In the MCI group, the performance by using 18F-flutemetamol based classifier outperformed the MRI based classifier.

Multimodal based studies

Multimodal classification frameworks benefit from the complementary information gained by different imaging modalities. For MCI conversion prediction, Cheng et.al [100] proposed a multi-modal manifold-regularized transfer learning for MCI conversion prediction. Only ADNI subjects with all corresponding MRI, CSF and FDG-PET baseline data are included. The data including 202 subjects (51 AD, 43 pMCI, 56 sMCI, and 52 HV). They found that the multimodality method shows considerably better performance, compared to the case of using biomarkers from a single modality. Similarly, Zhang et al. [101] proposed a new multimodal data fusion and classification method based on kernel combination for AD and MCI. They used biomarkers from three modalities, i.e., MRI, FDG-PET, and CSF biomarkers, to discriminate between AD / MCI and healthy volunteers. They used the same dataset as [100] and achieved a classification accuracy of 93.2% when combining all three modalities of biomarkers, and only 86.5% when using even the best individual modality of biomarkers.

2.4.2 AD diagnosis using deep learning

In the following sections, we explore the application of deep learning methods for classifying different AD stages to improve accuracy. Recently, related studies have used a range of neuroimaging modalities including structural MRI, PET and a combination of different modalities.

Structural MRI based studies

Hosseini-Asl et al. [102] proposed a deeply supervised adaptable 3D convolutional network to detect AD. The 3D CNN is built upon a 3D convolutional autoencoder, which is pre-trained to extract the related features from structural MRI. They randomly selected 210 subjects (70 NC, 70 MCI and 70 AD) with corresponding T1 MRI images from ADNI. Their framework achieved 94.2% accuracy for MCI and HV classification. In addition, Payan et al. [103] built a deep learning model using sparse autoencoders and 3D CNNs based on MRI scans of the brain. They collected 755 patients in each one of the three classes (AD, MCI, HV) from ADNI. Their results demonstrated that 3D CNNs outperform several other classifiers such as logistic regression. Moreover, Shi et al. [104] developed a novel nonlinear metric learning method with a denoising sparse autoencoder to improve the classification performance for AD and MCI on structural MRI dataset. They collected 338 subjects including 94 patients with AD, 121 with MCI and 123 HV from ADNI. They achieved 89% and 81.7% accuracy for HV against AD and HV against MCI respectively.

FDG PET based studies

Singh et al. [105] proposed a deep learning based classification method for AD diagnosis using FDG PET brain imaging data. The data consists of baseline data of subjects including 186 healthy volunteers 336 MCI subjects with 158 Late MCI and 178 Early MCI, and 146 AD patients from the ADNI database. They used probabilistic principal component analysis (PCA) on max-pooled and average-pooled data for dimensionality reduction, and a multilayer perceptron for binary classification. Ding et al. [106] applied a variant of CNN named Inception V3 to FDG PET from ADNI dataset. For training, they collected 1000+ patients with 236 AD, 406 MCI and 360 healthy volunteers. For testing, they collected 40 patients in total including 7 AD, 7 MCI and 26 healthy volunteers. The results indicate that the area under the ROC curve outperformed reader performance for predicting the final clinical diagnosis of AD. With further large scale external validation on the data and model

modification, the algorithm may be integrated into clinical workflow and serve as an important decision support tool to aid radiology readers and clinicians with early prediction of AD from FDG PET imaging studies. In addition, Liu et al. [107] developed a classification framework based on the combination of 2D CNNs and RNNs for AD vs HV and MCI vs HV classification. They acquired FDG PET from 339 ADNI subjects including 93 AD, 146 MCI and 100 HV. They achieved area under ROC curve of 95.3% for AD vs HV classification and 83.9% for MCI vs HV classification.

Amyloid PET based studies

Compared to FDG PET and MRI, amyloid PET has not yet been widely explored with state of the art deep learning methods. However, from the limited research papers that we can find, it shows great potential as a strong predictor for early diagnosis of AD. For instance, Choi et al. [108] proposed an automatic image interpretation system based on a deep CNN using both FDG and florbetapir PET images as input volumes. They acquired 139 patients with AD, 171 patients with MCI and 182 HV subjects from ADNI database. The proposed deep CNN approach reached accuracy of prediction (84.2%) for conversion to AD in MCI patients. The authors expect the model can be validated in clinical trials with large scale datasets.

To our knowledge, at the time of writing this thesis, chapter 4 and chapter 5 present the first investigation of using amyloid PET images alone as well as the combinations with MRI for the MCI progression prediction using 3D CNNs.

Multimodal based studies

Similar to multimodal based machine learning methods, in deep learning, combining multiple different modalities that providing complementary information could help improve early diagnosis of AD. For example, Lu et al. [109] developed a multimodal and multiscale deep convolutional neural network for early AD diagnosis. They collected 1242 subjects while each subject with both a T1-weighted MRI image and an FDG PET image. The data are then

split into five categories: stable HV, progressive HV, stable MCI, progressive MCI and AD. The proposed method reached 82.4% accuracy in identifying pMCI subjects 3 years before conversion. They show that deep neural network classifiers may be useful as a potential tool in support of the clinical AD diagnosis. Suk et al. [110] developed a deep sparse multi-task learning for feature selection in AD diagnosis. They acquired baseline MRI, FDG PET, and CSF data with 51 AD, 99 MCI, and 52 HV subjects from ADNI. The classification accuracy for HV against AD using the proposed multi-task learning method is 95.1%. Liu et al. [111] proposed a novel data fusion method to extract complementary information from both MRI and FDG PET. They then chose a stacked autoencoder as the classifier for HV/AD and HV/MCI classification. In total, 331 subjects are collected including 85 AD, 67 pMCI, 102 sMCI and 77 HV from ADNI. The classification accuracy is 91.4% (HV/AD) and 82.1% (HV/MCI). With further improvement of the proposed model, they can potentially build a multiclass classification framework that can be applied to clinical use. Overall, all these studies agreed on the importance and advantage of using multimodal data compare to single modality data.

Conclusion

In this chapter, we have presented information about AD pathology and biomarkers, as well as the treatment and PET quantification (SUVr) for AD. In addition, we have introduced the concept of fractal dimension and Shannon entropy, and discussed SVM and two deep leaning models (CNN and GAN) used in this thesis. Finally, we discussed the state of the art in AD classification with imaging biomarkers using both machine learning and deep learning methods.

Based on the literature review in section 2.4, we summarize the study size using Table 2.4.1. Among traditional machine learning based methods, most of the literature used MRI or FDG PET or a combination of these two. For MRI, Li et al. [91] developed a SVM based AD classifier using MRI images from 24 MCI and 26 HV subjects (50 subjects in total), which is smaller than our data size (24 HV, 23 pAD and 18 MCI - 65 subjects for both PET and MRI)

as they are using single modality only. In addition, Magnin et al. [72] presented an automated method based on SVM for AD classification. They used the whole brain sMRI data (16 pAD and 22 HV, 38 subjects in total) which is smaller than our data size. For amyloid PET, the authors [99] proposed a SVM based AD classifier using 27 pAD, 20 MCI and 25 HV (72 subjects in total) flutemetamol PET. This work used the same PET radiotracer and similar method as our study in chapter 3. Even though the data size is relatively small, we believe that it is sufficient for conducting the experiment because the data is in three dimensions and our study size is similar to other published studies that have used similar methods.

For deep learning based methods, we can see that the data size for training deep learning models is higher compare to the traditional machine leaning models. Similarly, most literatures are focusing on the MRI and FDG PET. In our case, the data collected from ADNI including 36 pAD, 82 HV, 163 MCI (281 for each modality, 843 for all three modalities) florbetapir PET with corresponding MRI and FDG, which is comparable to deep learning based multimodal studies such as [110,111] in terms of data size (202 and 331 subjects for each modality, 606 and 662 for all modalities, respectively). Additionally, Korolev et al. [112] proposed a 3D CNN AD classification framework using sMRI alone and they obtained 50 pAD, 120 MCI and 61 HV (231 subjects in total) from ADNI which is smaller than our study size. In [113], the authors also proposed a 3D CNN for AD diagnosis based on sMRI and DTI modalities. They obtained 48 pAD, 108 MCI and 58 HV (214 subjects in total for each modality) from ADNI, which is also smaller than our study size. In this case, we believe that the available volumetric data in three modalities for training a deep neural network is sufficient while performance could be improved with increasing training data size.

Generally, the ultimate goal for computer aided diagnosis is to translate the work into routine clinical practice. In this chapter, some works have shown the feasibility of implementation of their research findings in clinical practice [73,91,106,108,109,111]. However, machine learning methods focus on the improvement of the proposed model while deep learning

methods focus on both the enlargement of data validation and model modification before they can be translated into clinical practice.

The applications using amyloid PET brain imaging for AD classification and MCI conversion prediction in particular, are very limited. For this reason, in the following chapter, we propose a multimodal classification framework using two novel complexity measurements for amyloid status classification. In chapter 4, we propose another multimodal classification framework using 3D CNNs. Both works show competitive results especially for MCI conversion prediction compared with recent published literature.

In addition, we can also observe that the availability of structural MRI and FDG PET is much higher than that of amyloid PET, which is also the case in our study. Therefore, we addressed this problem by using generative neural network (eg. GANs) to generate synthetic amyloid PET from the corresponding structural MRI and the details are included in chapter 5.

Table 2.4.1: study size from section 2.4 for both machine learning and deep learning based references

Machine learning based references			
Reference	Source	Modality	Study size (# of subjects)
[89–92,95]	Mainly hospitals and AD centres	sMRI	33-544
[73,97]	ADNI	FDG	177, 763
[99]	Hospital	Amyloid (Flutemetamol)	72
[100]	ADNI	Multimodal (MRI, CSF, FDG-PET)	202
Deep learning based references			
Reference	Source	Modality	Study size (# of subjects)
[102–104]	ADNI	sMRI	210-338
[105–107]		FDG	339-1042
[108]		Amyloid (Florbetapir)	492
[109–111]		Multimodal (MRI, FDG, CSF)	202-1242

3 Complexity analysis of 18F PET images and multi modal classification for early diagnosis of AD

3.1 Materials 57
3.1.1 Data 57
3.1.2 Pre-processing 59
3.2 Methods 61
3.2.1 Calculation of Fractal Dimension using the Box Counting Method 61
3.2.2 Shannon entropy 63
3.2.3 SVM Classification 64
3.2.4 MRI feature extraction 65
3.2.5 Experiments 66
3.3 Results and discussion 68
3.4 Conclusion 75

New positron emission tomography (PET) tracers could have a substantial impact on early diagnosis of Alzheimer's disease (AD), particularly if they are accompanied by optimised image analysis and machine learning methods. Fractal dimension (FD) analysis, a measure of shape complexity, has been proven useful in MR imaging [114] [115], but its application to fluorine-18 amyloid PET has not yet been demonstrated. Shannon entropy (SE) has also been proposed as a measure of image complexity in DTI imaging, but it is not yet widely used in radiology.

In this study, one volumetric FD method and one volumetric SE method were developed and applied to 18F-flutemetamol and 18F-florbetapir 3D amyloid images from 65 and 281 subjects respectively, including Healthy Volunteers (HV), and patients with probable Alzheimer's Disease (pAD) or Mild Cognitive Impairment (MCI).

The group average fractal dimension of white matter surface and Shannon entropy of white matter volume for HV subjects were higher than for pAD subjects. Both FD and SE are effective in the identification of MCI patients who progress to pAD during the two-year follow-up (ground truth).

Finally, we developed a support vector machine (SVM) multi-modal classification framework using both PET and MRI features, which showed higher accuracy compared to traditional SUVr or using PET alone. The classification accuracy for flutemetamol and florbetapir is 88.9% and 83.3% respectively for MCI progression, which is competitive with existing literature.

The results presented in this study demonstrate the potential of FD and SE methods for the analysis of brain PET scans in early AD diagnosis and in the prediction of MCI-AD conversion. The work presented in this chapter was partly published as a journal article in 2019. (Yu Yan, Edward Somer, and Vicente Grau. "Classification of amyloid PET images using novel features for early diagnosis of Alzheimer's disease and mild cognitive impairment conversion." Nuclear medicine communications 40.3 (2019): 242-248.)

3.1 Materials

3.1.1 Data

Two sets of data were used in this work. The first was obtained from GE Healthcare, and included 18F-flutemetamol PET and corresponding MRI scans for 65 subjects (24 HV, 23 pAD and 18 MCI, divided into 9 sMCI and 9 pMCI). All patients had to be older than 55 years. Patient demographics and neuropsychological score are shown in Table 3.1.1.

Table 3.1.1: Patients demographics and neuropsychological score for GE data

	AD	sMCI	pMCI	HV
Gender(M/F)	12/11	4/5	5/4	12/12
Age	69.1±7	72.3±7.1	72.3±7.6	68.5±7.2

Years of education	14.1±4.1	14.7±3	14.7±3	14.4±3.5
MMSE	23.2±2.2	28.1±0.9	27.2±1.1	29.2±0.9

T1-weighted MR images were acquired using a standard three-dimensional Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) protocol at 3T (scanner: Philips Achieva MRI). A screening MRI scan was performed for each subject prior to the PET scan. The maximum interval between the PET and the MRI was 30 days. Subjects then underwent a dynamic PET scan.

On the day of PET imaging, each subject received an intravenous dose of 18F-flutemetamol as a bolus injection. The target activity of a single administration of 18F-flutemetamol injection was 185 MBq. Five subjects with probable AD received two administrations of 18F-flutemetamol injection and received the second 18F-flutemetamol injection 7 days after the first one. In this study, the PET scanners used were Siemens Biograph 16 PET/CT.

To diagnose pAD and assess MCI progression, there was an initial assessment at study enrolment (baseline) with a subsequent assessment 2 years later. The clinical assessment included physical symptoms, blood tests and a variety of mental tests. The flutemetamol image dimensions are 128×128×81 with a voxel size of 2.13mm×2.13mm×2mm. The PET images were acquired 30 minutes after the injection of radiotracer, and were analysed by five expert clinicians for blinded image evaluation (BIE). The consensus of this BIE is available for our study.

More specifically, five experts were trained on 10 flutemetamol scans that were collected in a previous phase 1 study. The experts were asked to take into account both increased cortical and subcortical gray matter signals when assigning flutemetamol to a control or progressive category. Following the training, these independent experts who were blinded to the clinical diagnosis assigned the 65 flutemetamol scans from the current study to either progressive or normal uptake categories based on visual assessment. For each expert, 8

randomly selected scans from 65 scans were shown twice to assess intra-rater agreement. If readers disagreed, the assignment was achieved by majority vote, resulting in a final decision for the visual assessment of each scan.

The gold standard was assessed at two years follow up for AD conversion by measuring ranges of regional cerebral to cerebellum tracer uptake ratios.

For 18F-florbetapir data, 36 pAD, 82 HV, 29 pMCI and 134 sMCI 18F-florbetapir images with corresponding MRI were obtained from the ADNI database [116]. The MRI dataset included standard T1-weighted images obtained with a 3T Philips scanner using a three-dimensional MPRAGE sequence. The PET scanners used were Philips Gemini PET/CT scanners. The PET and MRI volumes for each subject selected for this study were acquired no more than 12 months apart. All the subjects have at least two years follow up, which we use as gold standard for AD conversion (sMCI if the status of MCI remains throughout 2 years follow up and pMCI if MCI label changes to AD in the follow up). Patient demographics and neuropsychological score are shown in Table 3.1.2.

Table 3.1.2: Patients demographics and neuropsychological score for ADNI data

	HV	pAD	sMCI	pMCI
Number of subjects	82	36	134	29
Age	74.3±5.1	76±8.3	73.3±7.5	73±8.1
Gender(M/F)	41/41	21/15	83/51	18/11
Education (years)	16.5±2.7	15.2±2.6	15.7±2.8	16.3±2.6
MMSE	29±1.2	22.9±2.1	28.1±1.6	27±1.7

3.1.2 Pre-processing

All the florbetapir images had been already pre-processed by ADNI: MRI and PET scans from each subject were co-registered, and the volumetric image from their baseline PET scan was then reoriented into a standard 160×160×96 voxel image grid, comprising 1.5 mm

cubic voxels. This image grid was oriented such that the anterior-posterior axis of the subject is parallel to the AC-PC line. The corresponding MRI images have dimensions 256×256×196 with a voxel size of 1mm×1mm×1.2mm.

Flutemetamol images were obtained in their original state from the GE study and thus needed to be pre-processed as follows. As a first step, the segmentation of the brain in the PET images was performed using a hysteresis thresholding method[117]. This segmentation method performs a dual thresholding operation on a grayscale image using two threshold values. We followed this by an erosion operation. The brain masks calculated in this way were used in two ways: a) to binarise the original images for the calculation of FD; and b) to remove all image values outside of the mask (by making them zero), while keeping the original image intensities inside the mask: the images obtained in this way were used for the calculation of SE. Examples of original T1 MRI, flutemetamol PET images, with their corresponding segmentation are shown in Figure 3.2.1.

Secondly, the PET volumes were rigidly registered to their corresponding MR volumes using Statistical Parametric Mapping, version 12 (SPM12) [118]. The MR volumes were then affinely registered to Montreal Neurological Institute (MNI) space using FSL's FLIRT software [119]. Finally, the resulting transformations were applied to the corresponding PET volumes to create affinely registered PET volumes.

To evaluate the registration quality, an overlap agreement measure was used to check how well the segmentation of the brain overlaps. More specifically, the segmentation was propagated using the resulting transformation, and volume overlap was measured using the Dice metric. The Dice metric quantifies the similarity between the fixed and moving volumes[120]:

$$Dice = 2 * \frac{\sum |X \cap Y|}{\sum (|X| + |Y|)} \quad (10)$$

where X refers to a transformed volume from the moving image, and Y refers to the corresponding volume from the fixed image. During the pre-processing of the data, the MR

images were affinely registered to the MR template in MNI space. The resulting inverse deformation fields were used to transform the template space to native space. In this case, the transformations applied to MR volumes were also applied to the corresponding template. Subsequently, Dice coefficients were calculated between the MR template and each of the registered images. The average Dice overlap was 0.786 ± 0.125 when computed across all subjects, which indicates reasonably good registration quality in the pre-processing steps.

3.2 Methods

3.2.1 Calculation of Fractal Dimension using the Box Counting Method

Due to the non-specific white matter uptake in the control group, hysteresis thresholding results in delineation of the white/gray matter boundary. On amyloid positive subjects, on the other hand, specific gray matter uptake will result in the delineation of the gray matter/CSF contour. These two boundaries exhibit substantially different complexity, which can be measured by fractal dimension.

In order to calculate features based on fractals, here we applied a box counting method to compute the fractal dimension from binary images.

The original box counting method consists of partitioning the structure space with a three dimensional fixed grid of cubic boxes with edges of equal size r. The box sizes correspond to powers of 2, with the exponent ranging in our case from 0 to 4 [121]. In fractal structures, the number N_r of nonempty boxes of size r needed to completely cover the fractal structure depends on r according to the following equation:

$$N_r \sim r^{-FD} \tag{11}$$

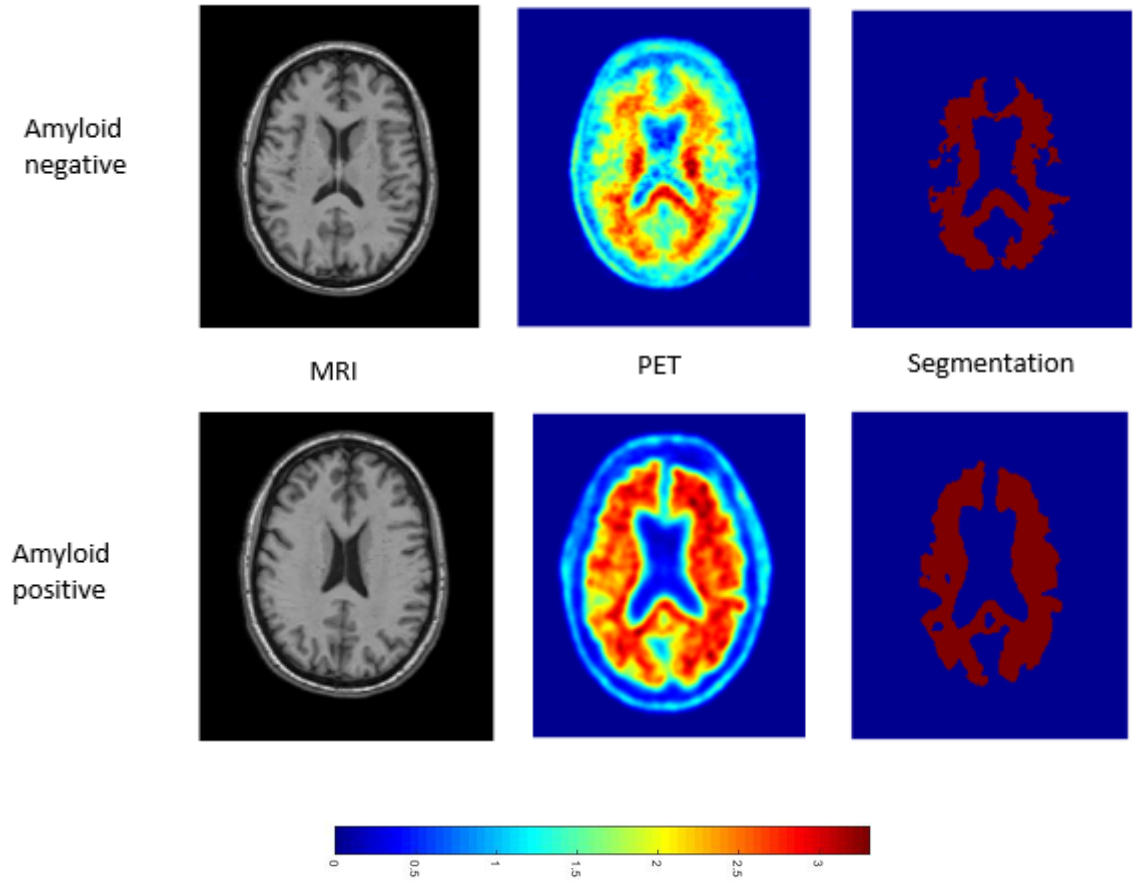


Figure 3.2.1: Amyloid negative and positive axial slices from left to right: T1 MRI images; flutemetamol PET images; hysteresis segmentation of GM and WM region using hysteresis thresholding. The amyloid positive patient shows characteristic tracer uptake in the gray matter region.

The box counting algorithm hence counts the number N_r for different values of r and plots the log of the value N_r versus the log of the box size r. The value of FD is then estimated from the slope of the least squares best fitting curve:

$$\mathrm{FD} = \lim_{r \to 0} \frac{\log (N_r)}{\log (1/r)} \qquad (12)$$

In practical applications, when studying objects that are not pure fractals, the $\lim_{r \to 0}$ is not attainable. Instead, N_r is measured for a limited range of r values corresponding to intermediate resolutions of the object under study. In our case, each voxel in the brain surface was replaced by a cube with size r. Subsequently, fractal dimension is then estimated by approximating (12). As the size of the measuring box is scaled progressively smaller, the complexity of the object contour can be measured by the box in a more precise

way (shown in Figure 3.2.2). To illustrate how fractal dimension is measured from a 3D structure, we also plot the number of counted, filled cubes at each cube size (left), and the plot in log-log space (right), which are shown in Figure 3.2.3. The fractal dimension is the slope of the line in log-log space.



Figure 3.2.2: Illustration of box counting with three different box sizes on 2D structure.

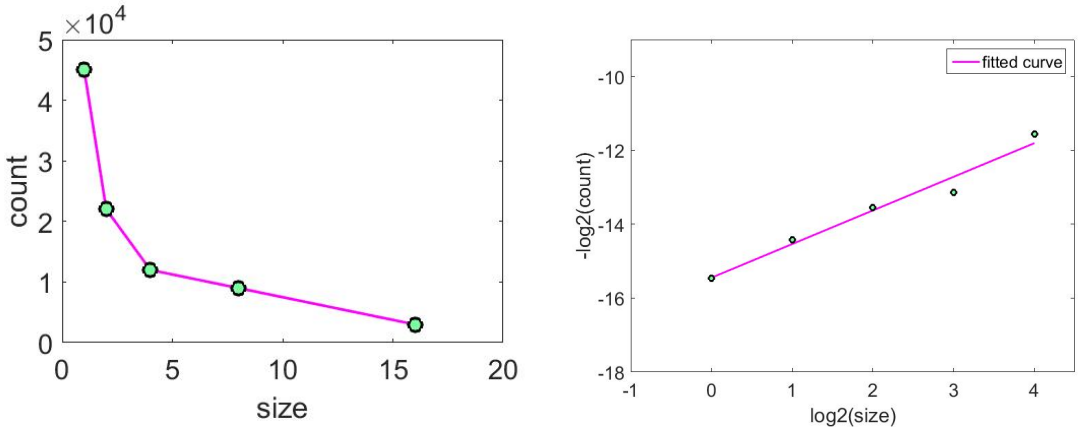


Figure 3.2.3: Fractal dimension measurement with log plot. The slope of the line in the right plot is an estimation of the fractal dimension.

3.2.2 Shannon entropy

Shannon introduced another complexity measure [64], which weighs the information per outcome by the probability of that outcome occurring. Given events $e_1, \ldots, e_m$ occurring with probabilities $p_1, \ldots, p_m$, the Shannon entropy H is defined as

$$H = \sum_i p_i log_2 \frac{1}{p_i} = -\sum_i p_i log_2 p_i \quad (13)$$

In the medical image processing field, entropy has been used as a statistical metric to characterize texture distribution [122]. To calculate the Shannon entropy of the brain volume, the probability of occurrence for each intensity value in the brain volume is needed. We estimated the probability distribution values p_i by normalizing the histogram of voxel intensities in the brain volume.

3.2.3 SVM Classification

In this study, a two class SVM classifier is used to classify different groups of subjects depending on their disease state. In its simplest form, the SVM method constructs a hyperplane that maximizes the margin, defined as the distance between the closest points on either side of the calculated boundary.

Here we apply a SVM with a non-linear kernel using Radial Basis Function (RBF), using the implementation by the SVM package named LIBSVM, running under Matlab version 8.6 [123]. More specifically, given a training set $(x_i, y_i), i = 1 \cdots l$ where $x_i \in R^n$ and $y \in \{1, -1\}^l$, the SVM requires the solution of the optimization problem in equation (). SVM finds a linear separating hyperplane with the maximum margin in the higher dimensional space. Meanwhile, the RBF kernel has the form:

$$K(x_i, x_j) = e^{-\gamma | x_i - x_j|^2} \quad (14)$$

The C parameter in equation () trades off classification accuracy against maximization of decision function's margin. For larger values of C, a smaller margin will be accepted if the decision function classifies the training samples correctly. A lower C will encourage a larger margin which will lead to a simpler decision function but at the cost of training accuracy. In other words, C behaves as a regularization parameter in SVM.

Intuitively, the gamma parameter defines how far the influence of a single training example reaches, with low values meaning high influence and high values meaning low influence. The gamma parameters can be seen as the inverse of the radius of support vectors influence.

The behaviour of the model is very sensitive to the gamma parameter. If gamma is too large, the influence of the support vectors only affect the support vector itself and the regularization term C will fail to prevent overfitting. When gamma is very small, the model is too constrained and cannot capture the complexity of the data. In this work, C and gamma are set using a ten-fold cross validation resulting in C=1 and gamma=0.1.

3.2.4 MRI feature extraction

FreeSurfer is a set of software tools for processing and analyzing human brain MRI images [124]. The standard pipeline used here is named *recon-all*, which performs all the cortical surface reconstruction process.

More specifically, recon-all first removes the skull from the volume, the volume is then segmented into cortex, white matter and sub-cortical regions. Initially, recon-all attempts to find the boundary between the tissues, and then makes a second refinement to find the white matter boundary (called white surface). Once the boundary is established, recon-all expands outwards to the outer boundary of gray matter (called pial surface). The distance between the white surface and pial surface is the gray matter thickness. These surfaces are then inflated into a sphere, and then registered to a template sphere. The cortical parcellations of the template are then mapped back on to the individual subject and adjusted for small variations in the individual's anatomy. These parcellations are taken from two atlases that come with FreeSurfer. Statistics can then be calculated from the atlas parcellations. The illustration of analysis pipeline recon-all is shown in Figure 3.2.4.

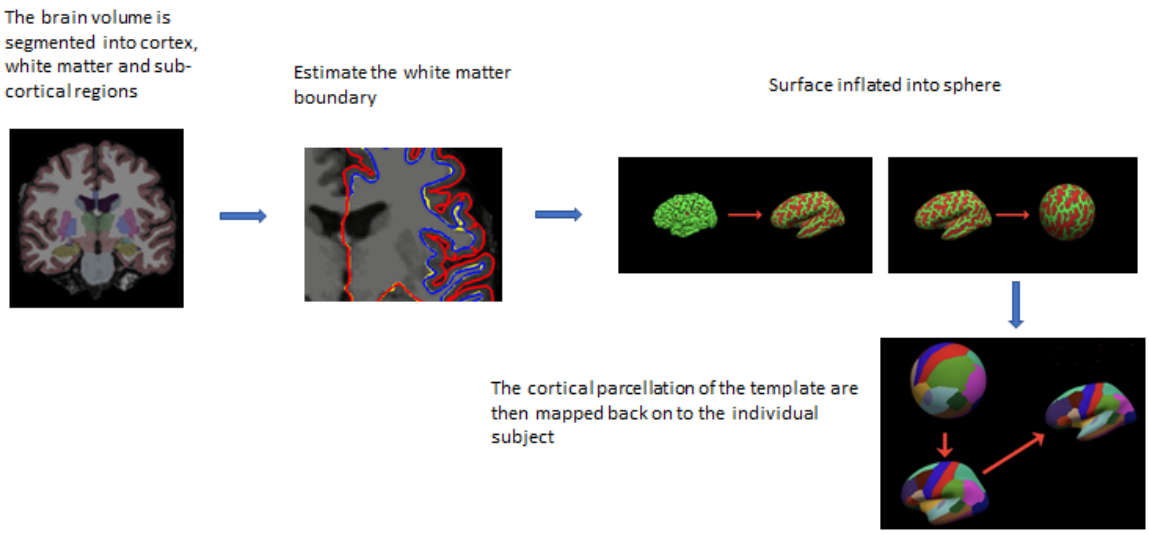


Figure 3.2.4: FreeSurfer Analysis recon-all pipeline (adapted from [125] *)*

3.2.5 Experiments

65 flutemetamol images (23 pAD, 24 HV, 9 pMCI and 9 sMCI) and 281 florbetapir images (36 pAD, 82 HV, 134 sMCI and 29 pMCI) were used for this study.

Firstly, we applied the hysteresis segmentation to the pre-processed PET images. The hysteresis function performs a dual thresholding operation on the 3D grayscale image using two threshold values. The lower and upper threshold values $t_1 and\ t_2$ are between 0 and 1. Three-dimensional 6-connected neighborhoods were searched from each object voxel. In order to estimate how sensitive the fractal dimension is to changes in segmentation parameters, we change the thresholds and calculate the fractal dimension each time.

In this study, we tested 36 combinations (t1<t2, minimum t1 is 0.1 and maximum t2 is 0.9) when using two thresholds and compute the average FD across all the subjects. The combinations can be found in Table 3.2.1. The statistical analysis (R^2=0.0067 and p=0.6345) proved that there was no correlation between the segmentation thresholds and the fractal dimension. In this case, we empirically choose the 17th combination (t1=0.3, t2=0.5) as the lower and upper threshold value respectively.

Table 3.2.1: Different threshold combinations and the corresponding mean FD values across all the subjects

T1 / T2	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0.2	2.17							
0.3	2.25	2.25						
0.4	2.25	2.15	2.29					
0.5	2.13	2.23	**2.16**	2.20				
0.6	2.19	2.23	2.21	2.23	2.12			
0.7	2.18	2.13	2.14	2.27	2.12	2.15		
0.8	2.22	2.12	2.25	2.29	2.15	2.26	2.28	
0.9	2.24	2.19	2.15	2.20	2.26	2.148	2.16	2.13

Binary erosion was applied to the segmented binary image to extract the surface, using a cube structuring element with size 3.

After this step, box counting and Shannon entropy were applied to the brain surface (binarized) and brain volume in original intensity respectively.

Subsequently, SUVr values were calculated from the ratio of florbetapir and flutemetamol retention in each of four lobes (frontal lobe, parietal lobe, occipital lobe, and temporal lobe) to the cerebellum (reference region) from the registered amyloid PET images. In this work, the target and reference regions were based on the MNI atlas regions. The resulting SUVr values from all lobes were fed to a SVM classifier to make a decision for pAD and HV, pMCI and sMCI.

Finally, FD, SE and MRI features (hippocampus volume, lateral ventricle volume and cortical thickness generated from the standard FreeSurfer pipeline) were then combined with SUVr [126].

Due to the limited number of flutemetamol images, a leave one out cross validation was performed for these data sets: one subject was taken as a test set and the remaining as

training set, and this process was repeated 65 times. For each of the leave-one-out experiments, a 3-fold cross validation was used on the 64 subjects in the training set to find the optimal parameters C and γ for the nonlinear support vector machine with a Gaussian radial basis function kernel. The resulting classifier was then used on the single testing set.

The florbetapir set was larger (n=281) and thus we chose to perform 10-fold cross validation on it. 253 subjects were taken as training set and the remaining 28 subjects as test set. 3-fold cross validation was applied within the training set to find the optimal parameters as before.

In this study, PET and MRI features were computed from all 65 flutemetamol images and 281 florbetapir images. Both PET and MRI features were combined with SUVr to predict HV and pAD, pMCI and sMCI. The classification results were compared with BIE and most recent results from literature. In all cases, the gold standard for pAD classification was clinical diagnosis, and for MCI progression was two-year follow-up.

3.3 Results and discussion

For SE, the histograms for four chosen volumetric images in different categories are shown in Figure 3.3.1 and Figure 3.3.2. As can be seen from the figures, the peaks are approximately in the middle of the intensity range for HV and sMCI, while the peaks are located on the right side of the histogram for pAD and pMCI group. To explain this, we suggest that, in pAD and pMCI group, the high specific uptake in gray matter will result in high intensity values in the image and move the peak of the histogram to the right, which also produced lower SE values, compare to HV and sMCI group.

The flutemetamol classification results between pAD and HV were compared with the BIE majority read (88% sensitivity and 96% specificity, gold standard two years followup). In addition, the MCI group at baseline were analysed using the proposed methods and compared with progression status in the follow-up evaluation. In this case, BIE achieved 78% sensitivity and 78% specificity.

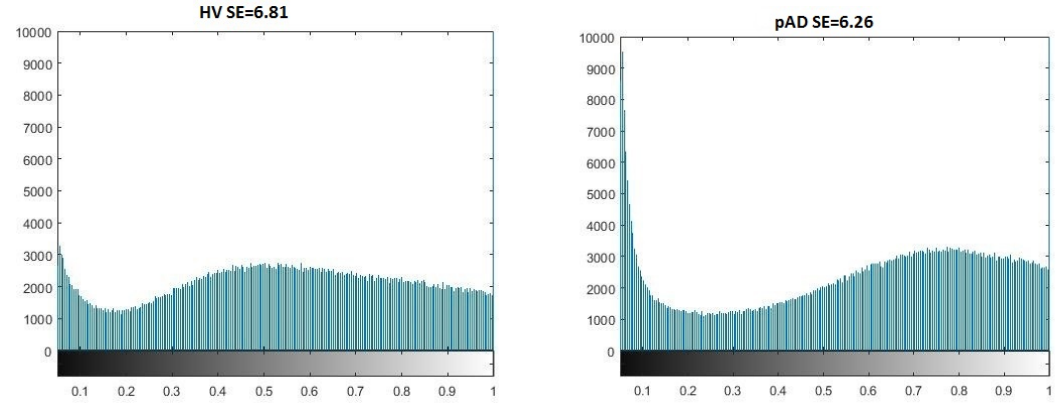


Figure 3.3.1: Histograms for HV and pAD comparison

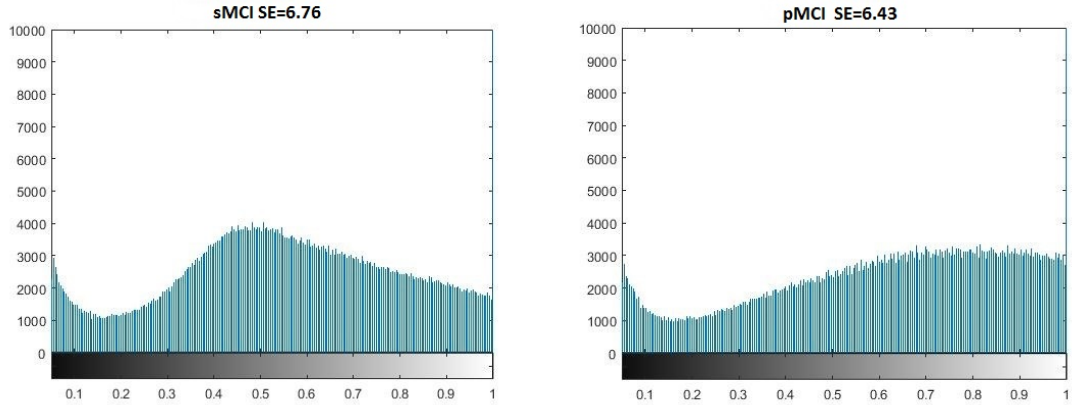


Figure 3.3.2: Histograms for sMCI and pMCI comparison

Both FD and SE performed well for the classification of flutemetamol images as pAD or healthy volunteers. The sensitivity and specificity of FD were 85.7% and 89% respectively, while the sensitivity and specificity of the SE method were 87% and 92.6%. The performance of the classification of MCI subjects as progressed or stable was also evaluated. For FD, sensitivity and specificity were 78% and 89% respectively. SE reached sensitivity 89% and specificity 78%, which compared favourably with the BIE results. Figure 3.3.3, Figure 3.3.4 and Figure 3.3.5 show the variation in FD and SE between the three groups, as well as the ROCs for MCI conversion classification using 65 Flutemetamol images. Figure 3.3.6, Figure 3.3.7 and Figure 3.3.8 show the variation in FD and SE

between the three groups, as well as the ROCs for MCI conversion classification using 281 Florbetapir images.

For PET features comparison, the SVM classification accuracy for both flutemetamol and florbetapir are shown in Table 3.3.1. The top part of the Table presents results achieved with PET images only. As can be seen, both FD and SE alone outperformed the traditional SUVr method in pAD/HV and pMCI/sMCI classifications. Several feature combinations produced the same highest accuracy (91.5%) for pAD vs HV classification. The highest accuracy for pMCI vs sMCI classification was obtained by using flutemetamol PET features, which reached 83.3% with the combination of SUVr, FD and SE. The accuracy decreased slightly to 80.4% when using 281 florbetapir images with the same feature combination.

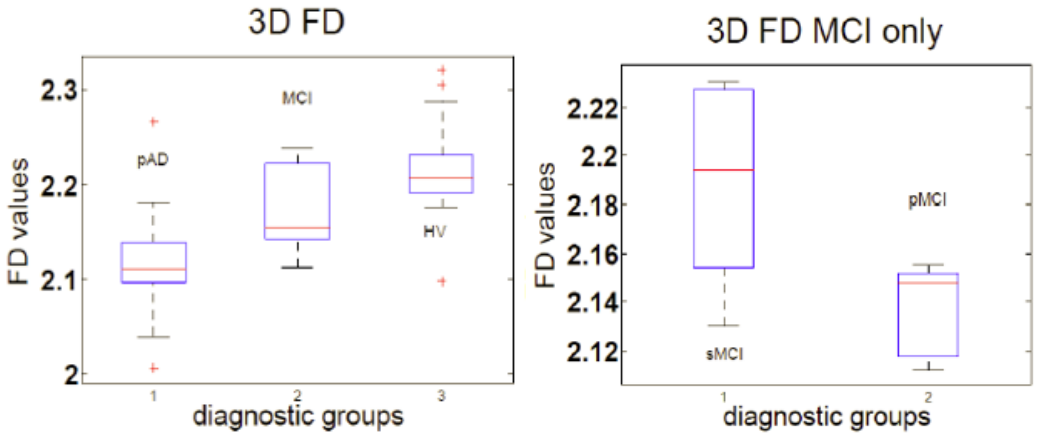


Figure 3.3.3: 3D FD box plot for three disease states and 3D FD box plot for MCI only (Flutemetamol)

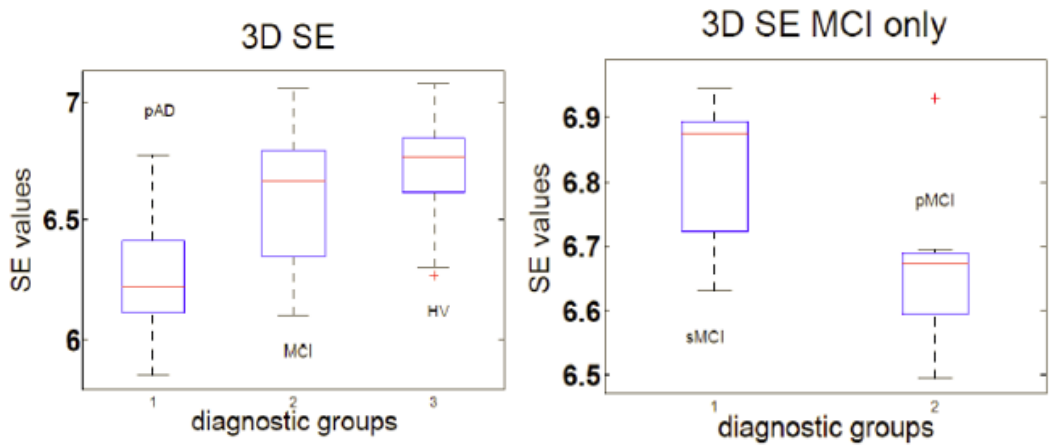


Figure 3.3.4: 3D SE box plot for three disease states and 3D SE box plot for MCI only (Flutemetamol)

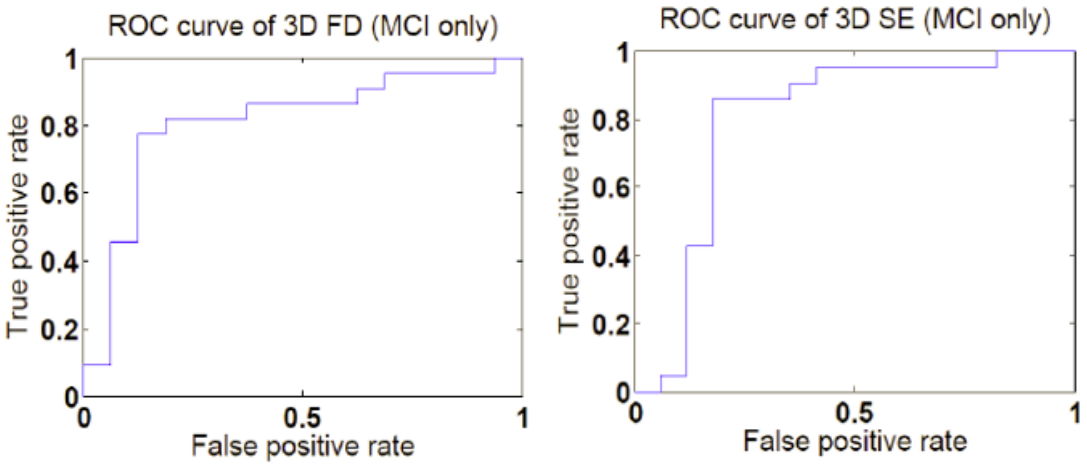


Figure 3.3.5: ROC for 3D FD and SE using 65 flutemetamol images.

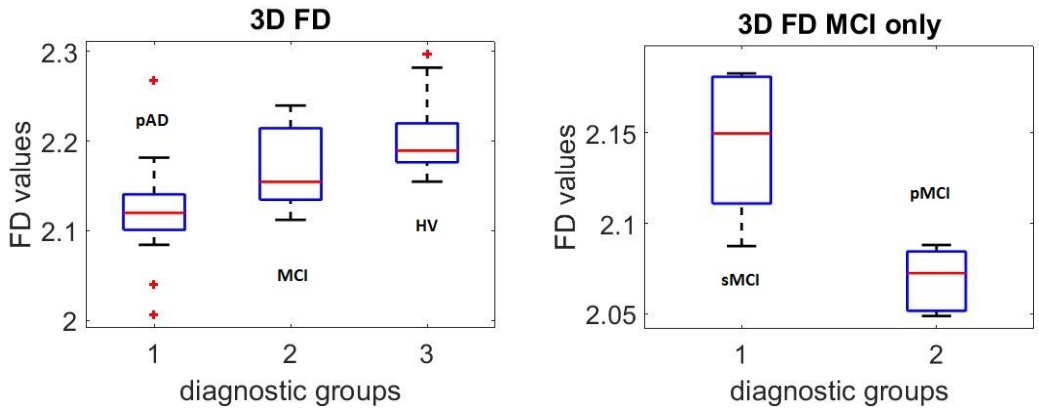


Figure 3.3.6: 3D FD box plot for three disease states and 3D FD box plot for MCI only (Florbetapir)

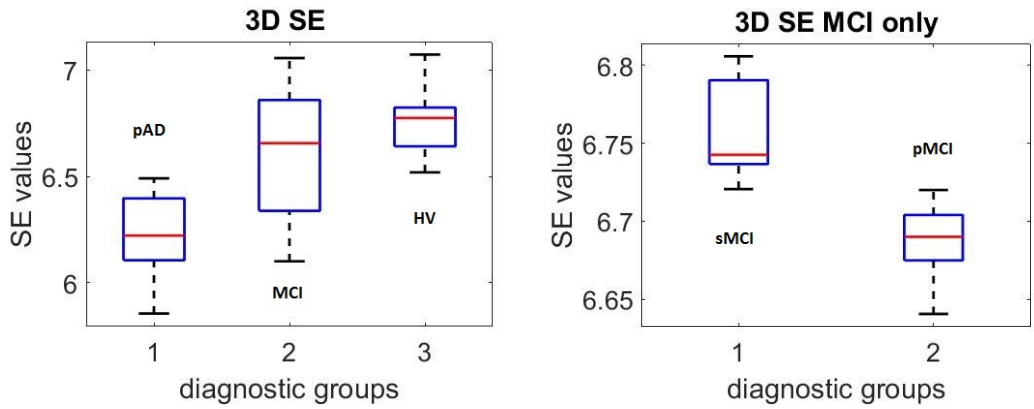


Figure 3.3.7: 3D SE box plot for three disease states and 3D SE box plot for MCI only (Florbetapir)

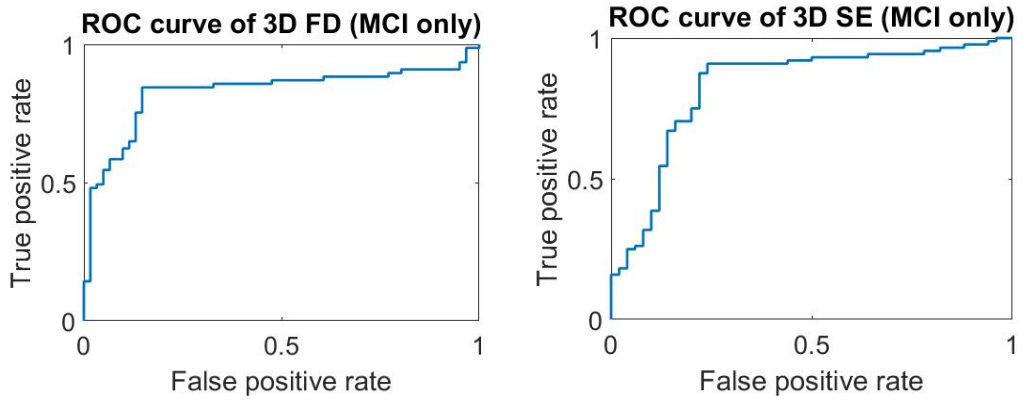


Figure 3.3.8: ROC for 3D FD and SE using 281 florbetapir images

In addition, the results of combining PET features and corresponding MRI features from FreeSurfer are shown in the bottom part of Table 3.3.1. As expected, the highest accuracy was achieved when combining all PET and MRI features. The highest classification accuracy for pAD vs HV was 93.6% and 93.2% for the flutemetamol and florbetapir groups, respectively. The highest classification accuracy between pMCI and sMCI was 88.9% and 82.8% for the flutemetamol and florbetapir groups, respectively. We compared these results with recent state of the art in image-based (i.e. excluding non-imaging tests) classification [127][128][129][100]. As shown in Table 3.3.2, our method outperformed recent results from previously published studies in the prediction of MCI conversion, and obtained competitive

accuracy in the classification of pAD vs HV, when comparing versions using PET information only as well as those combining PET and MRI features.

Table 3.3.1: SVM classification accuracy for 65 flutemetamol and 281 florbetapir images. The top part of the Table presents results obtained using PET-derived features only, while the bottom part includes MRI-derived features too.

	Number of features per subject	pAD and HV accuracy		pMCI and sMCI accuracy	
		Flutemetamol	Florbetapir	Flutemetamol	Florbetapir
FD	1	89.4	86.4	77.8	77.9
SE	1	89.4	87.3	72.2	77.3
FD+SE	2	**91.5**	90.7	77.8	79.1
SUVr	4	85.1	84.7	66.7	71.8
SUVr+SE	5	89.4	89	77.8	78.5
SUVr+FD	5	**91.5**	**91.5**	**83.3**	79.8
SUVr+FD+SE	6	**91.5**	**91.5**	**83.3**	**80.4**
MRI	6	78.7	76.3	66.7	71.8
MRI+SUVr	10	87.2	86.4	72.2	77.3
MRI+FD+SE	8	91.5	90.7	83.3	**82.8**
MRI+FD+SE+ SUVr	12	**93.6**	**93.2**	**88.9**	**82.8**

Table 3.3.2: Comparison between our proposed methods and previously published studies

Reference	Modalities	Classifier	Number of subjects	Accuracy pAD vs HV	Accuracy pMCI vs sMCI
[127]	FDG-PET	T-sum score	108	N/A	68.5
[128]	Flutemetamol PET	SVM	72	Sensitivity 85.2 Specificity 92	N/A
Proposed method	Flutemetamol PET	SVM	65	**91.5**	**83.3**
	Florbetapir PET		281	**91.5**	80.4
[129]	MRI FDG-PET	SVM	202	**95.6**	69.8
[100]	MRI FDG-PET CSF	SVM	202	N/A	76.5
Proposed method	MRI Flutemetamol PET	SVM	65	93.6	**88.9**
	MRI Florbetapir PET		281	93.2	82.8

To our best knowledge, we suppose the reason for the different results between flutemetamol and florbetapir is due to the fact that flutemetamol is similar to PIB and hence has structural similarities to Thioflavin-S, which may improve its classification capabilities

when compared with florbetapir [130]. Unfortunately, since flutemetamol is a more recently developed tracer the number of samples available is still small, which makes it difficult to use with methods requiring larger datasets such as deep learning. For this reason, in the remainder of this thesis we focus on florbetapir scans.

3.4 Conclusion

In this chapter two complexity measurements were proposed for amyloid status classification, namely fractal dimension and Shannon entropy methods. Both FD and SE outperformed the traditional SUVr, both for the identification of pAD vs HV subjects and for identification of subjects likely to convert from MCI to pAD. Individual application of FD and SE improved on the individual application of SUVr. When FD and SE were combined, an additional modest improvement was observed, suggesting some complementarity of the information from each. The best results for PET-only classification were obtained when SUVr, FD and SE were combined together. As an additional advantage, in our framework FD and SE are calculated directly on the images on a straightforward manner, while SUVr was parcellated into four lobes with the help of an external atlas.

The SVM classification results indicated the ability of using 18F PET features alone, and also pointed out the importance of introducing MR features, which resulted in a marked increase of the accuracy when compared to PET-only calculation in both the flutemetamol and florbetapir groups. These clearly introduced some information complementary from that in the PET as shown in the fact that the combination between MRI, FD, SE and SUVr (calculated on a lobar basis) resulted in the best classification overall. Results were competitive with existing literature, particularly on the prediction of MCI conversion.

Recently, deep learning methods have become the trend when solving classification problems. They are the current state of the art when we deal with complex problems such as image classification, natural language processing and speech recognition. In addition, another important advantage of deep learning algorithms is that they try to learn high level

features from data in an incremental manner. This avoids the need of domain expertise and hard core feature extraction.

In the next chapter, we continued using the combination of MRI and PET, however, by means of deep learning rather than traditional machine learning approach. More specifically, two parallel 3D CNNs were built and the features learnt by the two networks are concatenated to make a final decision.

4 Multimodal 3D Convolutional Neural Networks for Early Diagnosis of Alzheimer's Disease using MRI and Amyloid Imaging

4.1 Materials 78
4.1.1 Data and pre-processing 78
4.2 Methods 79
4.2.1 Feature learning with 3D CNN 79
4.2.2 Multi-modality 3D CNN architecture 80
4.2.3 MCI progression experiment setup 81
4.2.4 Alternative CNN architecture and data dependency experiment 83
4.3 Results and Discussion 83
4.3.1 Model performance monitoring 83
4.3.2 MCI progression classification part 1 83
4.3.3 Alternative CNNs comparison and data dependency exploration 86
4.3.4 MCI progression classification part 2 87
4.3.5 Comparison with previous results and recent literature 89
4.4 Conclusion 92

In the work presented in this Chapter, we developed a multimodal 3D convolutional neural network (CNN) for MCI progression classification using structural MRI, FDG PET and amyloid PET. Three experiments are conducted: pMCI vs sMCI, (pMCI+pAD) vs sMCI, and (pMCI+pAD) vs (sMCI + HV), which are also referred as E1, E2 and E3. In E1, the experiment results showed that the accuracy is 78.9% when MRI and FDG are combined, and the highest accuracy (81.5%) is achieved when MRI and amyloid PET are combined. We also performed three more experiment to compare with the proposed architecture and explore the data dependency of the model. In E2 and E3, the highest accuracy was 83.3%

and 83.6% respectively when MRI and amyloid are combined. All three experiments showed competitive results for MCI progression classification compared with results from chapter 3 and recent published literature.

4.1 Materials

4.1.1 Data and pre-processing

All image data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) [116] [131].

In this study, we used the same dataset as chapter 3 but also including corresponding FDG PET images. All the PET images were pre-processed in the same way as in chapter 3. Similar to the florbetapir images, all the FDG images had been already pre-processed by ADNI: MRI and PET scans from each subject were co-registered, and the image from their baseline PET scan was then reoriented into a standard 160×160×96 voxel image grid, comprising 1.5 mm cubic voxels. In this study, we used 82 HV, 36 pAD, 134 sMCI and 29 pMCI florbetapir images with corresponding FDG PET and T1 MRI which were obtained from the ADNI database. Among the 134 sMCI subjects, 63 subjects returned to cognitively normal state after follow up. The PET scanners used were Philips Gemini PET/CT scanners.

The MRI dataset included standard T1-weighted images obtained with 3T Philips scanner using a three-dimensional MPRAGE sequence. The PET and MRI volumes for each subject selected for this study were acquired no more than 12 months apart. All the subjects have at least two years follow up, which we use as gold standard for AD conversion (sMCI if the status of MCI remains throughout 2 years follow up and pMCI if MCI label changes to AD in the follow up). More specifically, MRI images were pre-processed by ADNI as follows: first, the images were corrected for gradient non-linearity using a system-specific correction [131]. Subsequently, the image intensity non-uniformity was corrected using B1 calibration scans. Finally, a histogram peak sharpening algorithm was applied to the MRI images which reduce intensity non-uniformity [131]. The MRI images have dimensions 256×256×196 with a voxel

size of 1mm×1mm×1.2mm. Patient demographics and neuropsychological score are shown in Table 4.1.1.

Table 4.1.1: Demographics for ADNI data used in this study

	HV	pAD	sMCI	pMCI
Number of subjects	82	36	134	29
Age	74.3±5.1	76±8.3	73.3±7.5	73±8.1
Gender(M/F)	41/41	21/15	83/51	18/11
Education	16.5±2.7	15.2±2.6	15.7±2.8	16.3±2.6
MMSE	29±1.2	22.9±2.1	28.1±1.6	27±1.7

4.2 Methods

4.2.1 Feature learning with 3D CNN

Convolutional Neural Networks (CNN) are a category of neural networks that have proven very effective in areas such as image recognition and classification [132]. There are four basic layers that combine to form a CNN: input layer, convolutional layer, pooling layer and fully connected layer. The primary purpose of convolution is to extract features from the input images. Convolution preserves the spatial relationship between pixels by learning image features using small squares of input data. The size of the generated feature map from the convolution operation is controlled by three parameters: number of filters, stride and zero-padding. Pooling reduces the dimensionality of each feature map while retaining the most important information. Meanwhile, ReLU is used after every convolution in order to introduce non-linearity in the network. The fully connected layer is a traditional multilayer perceptron that uses a softmax activation function in the output layer [133].

In this study, each CNN is built with 6 convolutional layers (with batch normalization and dropout), 5 max pooling layers, and 2 fully connected layers. The filter size is set to 3x3x3 for each convolutional layer and 2x2x2 for each pooling layer. The number of filters for the first

two convolutional layers is 32 and for the last four convolutional layers is 64. Finally, a softmax classification layer is used to predict the class probabilistic score.

4.2.2 Multi-modality 3D CNN architecture

In the past few years, multi-modal imaging data has gained great popularity for the classification problem. [20][110][111] The availability of different imaging modalities for the study of Alzheimer's has been greatly increased thanks to ADNI, and the complementary information from these modalities can be combined in order to further improve the accuracy of the results. In addition, convolutional neural networks have achieved superior results for medical image classification [134][135]. Thus, we propose to combine volumetric PET images and volumetric MRI images to build an ensemble 3D CNN classification model for predicting progression from MCI to AD (i.e. to classify between sMCI and pMCI subjects). In addition, we have also performed two extra experiments: 1) make use of all the diseased data to make the training more balanced; and 2) utilise all the dataset to observe the effect of introducing HV and pAD to the MCI progression classification. In this work, we combine T1 MRI with two different PET imaging datasets and compare the classification results with three single modality models. The CNN configuration is the same for each modality and the extracted features are fused in the fully connected layer. The loss function used here is binary cross entropy and the optimiser used is Adam. In order to monitor the training process and avoid overfitting, we used early stopping of training, which will stop training once triggered by the performance measure specification. The network architecture is shown in Figure 4.2.1.

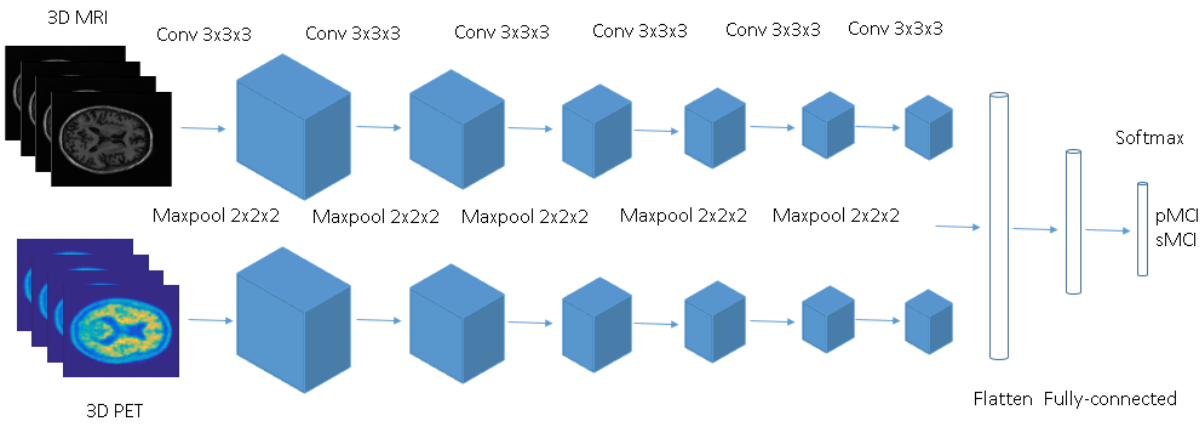


Figure 4.2.1: The network architecture of multi-modality CNN for pMCI and sMCI classification

4.2.3 MCI progression experiment setup

Clinically, the progressive MCI subjects indicate increasing memory impairment near the conversion to AD [136]. There are many similarities in diagnosis and recognition of progressive MCI and mild AD [137]. pMCI subjects have some neuroimaging time points with a clinical diagnosis of probable AD [138]. In our study, all pMCI subjects have a clinical diagnosis of probable AD after two years follow up. Taking these similarities into account, and in order to increase the size and variability of the class, we combined them with pAD group for the assessment of classifier accuracy. In addition, a subset of stable MCI subjects in our study returned to a cognitively normal state after follow-up; again, following a similar reasoning, we combined them with the HV group for accuracy assessment [139].

In the first experiment (E1), the classification of pMCI vs sMCI was attempted. Additionally, we performed two more experiments, aiming at the classification of: 1) sMCI vs (pMCI+pAD) 2) (pMCI+pAD) vs (HV+sMCI), which are named E2 and E3 respectively. In order to ensure that each fold has the same proportion of observations with a given categorical value, a stratified version of 10-fold cross-validation was applied to the whole dataset. For training we used a batch size of 1 with a learning rate of 0.0001 for 150 epochs.

We used the Keras framework to implement the proposed CNN architecture. The experiments were performed on a computer cluster within IBME, University of Oxford. The

cluster has one headnode and three nodes. Each node has four Nvidia GeForce GTX1080 Ti GPUs, two Intel Xeon E5-2667 v4 3.2GHz CPUs and a storage server (1TB). Node 1 has 512 GB memory while Node 2 and Node 3 have 256 GB memory. All the nodes run on CentOS 7.3. This work was conducted with 3 GPUs using Node 1. For visualization purposes, here we also showed examples of images for four classes in Figure 4.2.2.

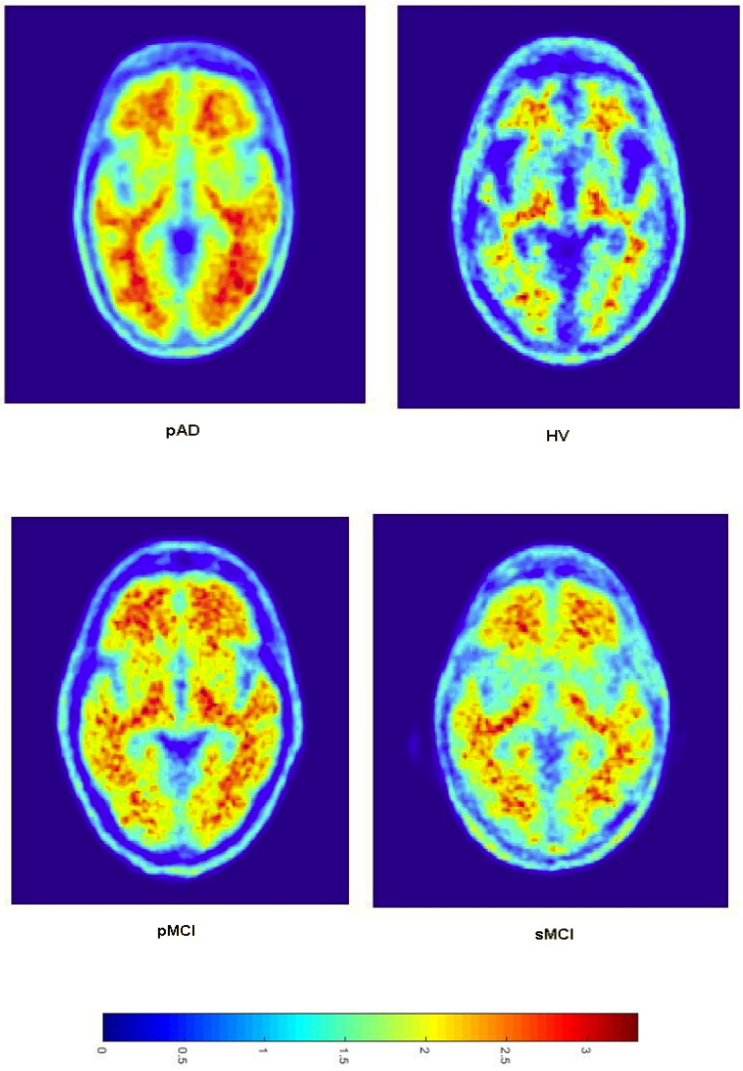


Figure 4.2.2: Examples of Florbetapir PET images for each of four classes

4.2.4 Alternative CNN architecture and data dependency experiment

In addition to the proposed architecture, we also built two CNNs with different architectures for comparison. One alternative CNN is built with increased convolutional layers from 6 to 9. We also changed the filter size from 3x3x3 to 5x5x5 for each convolutional layer to build another alternative CNN and observe the difference in performance. The remaining parameters of these two alternative architectures are the same as the proposed CNN. We applied both alternative CNNs to the E1 dataset (MRI and Amyloid) with stratified 10-fold cross validation and the results are reported in Section 4.3.3.

To explore the data dependency of our proposed model, we perform train/test split five times (five-fold cross validation) on the E1 dataset (MRI and Amyloid) and report the accuracy of each fold independently. The performance comparison is also shown in Section 4.3.3.

4.3 Results and Discussion

4.3.1 Model performance monitoring

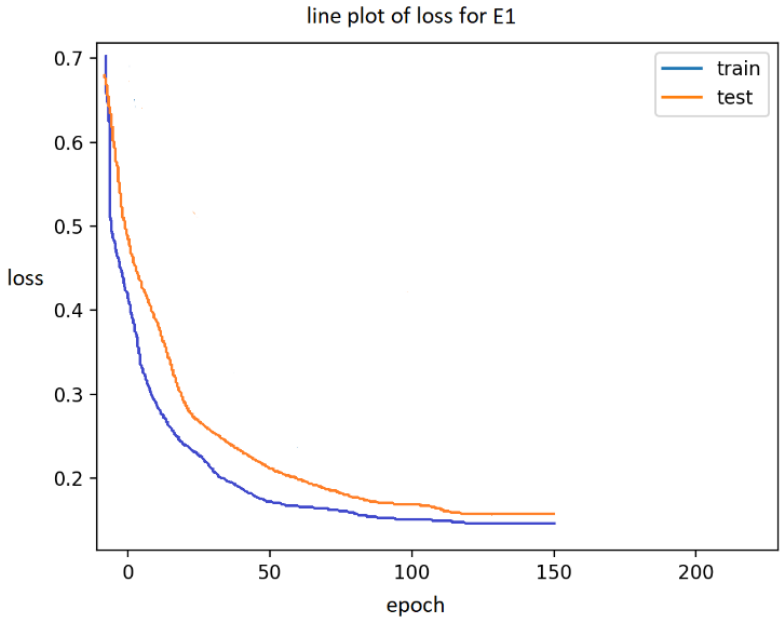
A problem with training neural networks is in the choice of the number of training epochs to use. Too many epochs can lead to overfitting of the training dataset, whereas too few may result in an underfit model. Early stopping is a regularization method to specify an arbitrarily large number of training epochs and stop training once the model performance stops improving on the validation dataset. The plot of train loss and test loss (validation loss) against number of training epochs for three experiments are shown in Figure 4.3.1. In this work, we stopped and saved the trained model when E1, E2 and E3 reached 118,115 and 108 epochs respectively.

4.3.2 MCI progression classification part 1

As can be seen from , three baseline experiments (MRI, FDG and amyloid PET as a single input to the neural network) have shown the ability to predict MCI progression using the proposed method. The accuracy for MRI, FDG and amyloid PET is 72.4%, 75.5% and 80.8%

respectively, when used in a single-modality framework. The highest accuracy is achieved using amyloid PET with an increase of 7% and 11.6% compared to FDG PET and T1 MRI. In addition, both the amyloid based and FDG based method achieved higher sensitivity with respect to MRI. However, FDG has lower specificity compared with amyloid and MRI. For AUC, three imaging modalities achieved similar results while the values for amyloid-based and FDG-based classification were higher than MRI.

As expected, the classification performance was further improved when MRI and PET images are combined. The accuracy is raised to 78.9% when MRI and FDG are combined, and the highest accuracy (81.5%) is achieved when MRI and amyloid PET are combined. In addition, we observed that the specificity increased significantly for both multimodal methods (81.3% and 82.2%) compared with the three single modal methods (74.3%, 72.4% and 75.4%). In terms of sensitivity, the highest among all the cases is achieved by using amyloid PET alone. For AUC, both multimodal methods reached above 80%, which are higher than the single modal methods as well.



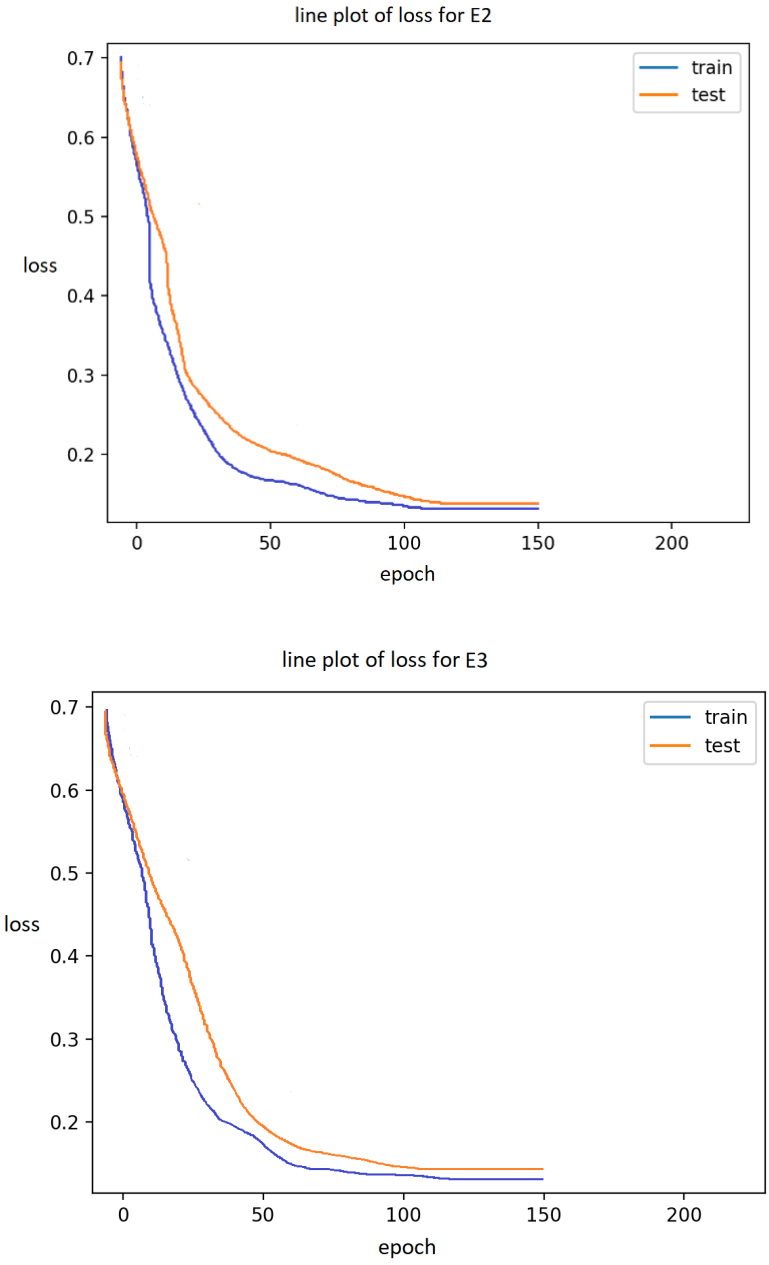


Figure 4.3.1: Plot of loss against training epochs for train and test (validation). To prevent overfitting, early stopping is used to specify an arbitrarily large number of training epochs and stop training once the model performance stops improving on the validation dataset. In this work, we stopped training the model when E1 E2 and E3 reached 118,115 and 108 epochs respectively.

Table 4.3.1: Comparison between different single modalities and multi-modality for pMCI and sMCI classification, 'FDG' and 'Amyloid' represent FDG PET and amyloid PET respectively.

pMCI vs sMCI (E1)				
Modality	AUC (%)	SEN (%)	SPE (%)	ACC (%)
MRI	76.6	70.8	74.3	72.4
FDG	78.1	78.2	72.4	75.5
Amyloid	78.2	82.6	75.4	80.8
MRI+ FDG	80.8	76.4	81.3	78.9
MRI+ Amyloid	80.4	79.8	82.2	**81.5**

4.3.3 Alternative CNNs comparison and data dependency exploration

To compare with our proposed CNN architecture, we also built two alternative architectures: 1) enlarging the kernel size from 3x3x3 to 5x5x5 and 2) increasing the number of layers from 6 to 9. In this case, both alternative CNNs are trained and tested on the E1 data (MRI and amyloid) using stratified 10-fold cross validation and the comparison is shown below.

Table 4.3.2: Two alternative CNNs trained and tested on E1 dataset using amyloid and MRI images and in comparison with proposed architecture

	Number of layers	Kernel size	ACC (%)	AUC
Proposed	6	3x3x3	81.5	0.804
Alternative 1	6	5x5x5	80.8	0.798
Alternative 2	9	3x3x3	81.9	0.811

As can be seen from the table, we conclude that enlarging the size of the kernel or increasing the number of the convolutional layers in our experiments does not affect the performance significantly. In consideration of the high computation power required in this work, we applied the proposed architecture to E2 and E3 datasets and the results are shown in the following section.

Meanwhile, in Table 4.3.3, we show the results using five-fold cross validation with five train/test splits that applied to E1 dataset (MRI and Amyloid, 163 images for each modality), the accuracy and area under the ROC curve is reported. As can be seen from the table, we achieved similar accuracy between each fold (accuracy range from 79.7%-80.9%), and hence demonstrated the low data dependency of our proposed model.

Table 4.3.3: accuracy comparison between each fold when five-fold cross validation is applied to E1 dataset (MRI+Amyloid)

Five train/test splits	Test ACC (%)	AUC
Fold 1	80.4	0.802
Fold 2	79.7	0.783
Fold 3	80.9	0.818
Fold 4	80.8	0.797
Fold 5	79.9	0.805

4.3.4 MCI progression classification part 2

In E2, the training dataset is more balanced due to the additional pAD data. For the single modality experiment, the amyloid based method reached the highest accuracy (81.8%) compared to MRI based (73.7%) and FDG based (77.3%). The amyloid based method also presented the highest sensitivity and specificity among the three imaging modalities. Interestingly, both E1 and E2 showed that the FDG based method has higher sensitivity but lower specificity compared with MRI. For AUC, both PET based methods showed high area under the ROC curve (close to 80%) while MRI reached 77.5%.

Table 4.3.4: Accuracy (%) for each modality and two modal combinations using different training sets. 'FDG' and 'Amyloid' represent FDG PET and amyloid PET respectively.

(pMCI+pAD) vs sMCI (E2)				
Modality	AUC (%)	SEN (%)	SPE (%)	ACC (%)
MRI	77.5	71.4	74.1	73.7
FDG	79.4	78.5	71.5	77.3
Amyloid	79.3	82.4	76.3	81.8
MRI+ FDG	81.4	75.9	80.9	80.4
MRI+ Amyloid	82.6	78.5	84.5	**83.3**
(pMCI+pAD)vs(HV+sMCI) (E3)				
Modality	AUC (%)	SEN (%)	SPE (%)	ACC (%)
MRI	76.8	75.2	72.1	74.2
FDG	79.1	79.3	73.4	77.1
Amyloid	80.1	81.8	78.3	81.2
MRI+ FDG	81.7	76.1	80.9	79.4
MRI+ Amyloid	83.2	80.2	85.2	**83.6**

For the two multimodality cases, MRI combined FDG reached 80.4% accuracy while MRI combined amyloid reached 83.3% accuracy, which is consistent with the findings of E1 results (MRI+amyloid>amyloid>MRI+FDG>FDG>MRI). The highest specificity is achieved by using MRI combined amyloid (84.5%) and the highest sensitivity is achieved by using amyloid alone (82.4%). For AUC, the MRI combined amyloid again outperformed all other cases in E2 (82.6%).

In E3, we further added HV subjects to a subset of sMCI subjects (63 out of 134 reverted back to cognitively normal stage). In this case, we found that amyloid also reached the highest accuracy when using single imaging modality. Similar to E1 and E2, the amyloid

based method also achieved the highest sensitivity and specificity (81.8% and 78.3% respectively) with respect to MRI (75.2% and 72.1%) and FDG (79.3% and 73.4%).

Meanwhile, MRI combined with amyloid again outperformed MRI combined FDG and all other single modality cases in E3. More specifically, the accuracy for MRI combined with amyloid is 83.6% and the second highest is achieved by using amyloid alone (81.2%). In addition, MRI combined with amyloid presented the highest specificity (85.2%) while the highest sensitivity is achieved by amyloid alone (81.8%). The highest AUC is achieved again by the MRI combined amyloid method which is 83.2%.

To conclude, most of the highest performance metrics were achieved by the combination of MRI and amyloid PET. In addition, using amyloid PET alone also achieved the highest sensitivity among three experiments. Finally, the results also indicate the possibility that using pAD/HV data to augment the shortage of pMCI/sMCI data (or vice versa) and balance the training dataset is feasible and produces favourable results.

In Figure 4.3.2, some florbetapir PET examples that were classified correctly and examples where the algorithm failed are shown. As can be seen from the figure, the images for which the method failed to provide the correct classification showed significantly different patterns compare to more typical examples from the same class. For example, most of the failed cases in the pMCI group have shown low amyloid deposition in the gray matter while most of the failed cases in the sMCI group showed high amyloid deposition in the gray matter.

4.3.5 Comparison with previous results and recent literature

In Chapter 3, for MCI prediction classification, we showed that using Florbetapir PET alone achieved a competitive classification accuracy (80.4%) while using both Florbetapir PET and MRI achieved 82.8% accuracy. In comparison, the deep learning model developed in this chapter (81.8% accuracy using amyloid alone, and 83.6% accuracy using MRI+ amyloid) outperformed the machine learning model in previous chapter. More comparisons can be seen in Table 4.3.5.

Table 4.3.5: Comparisons between chapter 3 and each of the experiments in this chapter

Accuracy (%) (No of subjects)	Chapter 3 (163)	E1 (163)	E2 (199)	E3 (210)
MRI	71.8	72.4	73.7	**74.2**
amyloid	80.4	80.8	**81.8**	81.2
MRI+amyloid	82.8	81.5	83.3	**83.6**

According to Table 4.3.5, we can see that the accuracy of SVM based method in Chapter 3 outperformed CNN based E1 when combined both MRI and amyloid. However, after the enrichment of training data by additional pAD and HV data (E2 and E3), the proposed method in this chapter outperformed Chapter 3 method when using MRI, amyloid alone or MRI amyloid combination. Performance improvement of E2 and E3 in this chapter compare to Chapter 3 indicates the great potential for deep learning classifier when the data set is arbitrarily large.

For traditional machine learning studies, in 2013, Young et al. [140] proposed a multi-kernel SVM (MKSVM) based classification model for prediction of conversion to AD. So far, studies that use MKSVM have focused on MRI and FDG PET. In 2017, Zhu et al. [141] developed a novel relational regularization feature selection method to reduce the data size, and then built two support vector regression models to predict clinical scores and a support vector classification model for MCI conversion prediction. For deep learning studies, in 2014, Suk et al. [142] proposed a novel latent and shared representation of brain MRI and FDG using deep Boltzmann machine (DBM) for sMCI vs pMCI classification. DBM is a deep network with a restricted Boltzmann machine (RBM) as a building block. In addition, in 2016, Korolev et al. [143] proposed a probabilistic, kernel-based pattern classification approach for AD prediction. Beside the imaging biomarker (MRI), they also incorporated risk factors, cognitive measurements and plasma data. Our multi model 3D CNN achieved 83.6% (MRI+ amyloid) and 80.4% (MRI+ FDG) accuracy for MCI to AD conversion prediction, hence outperformed

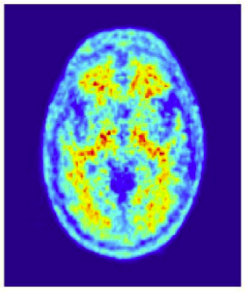
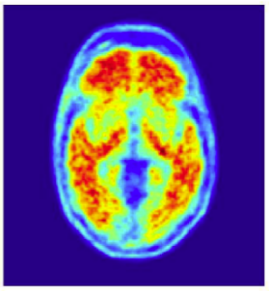
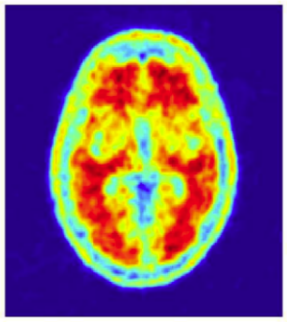
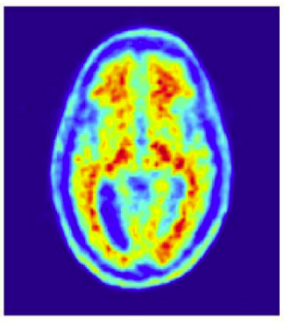
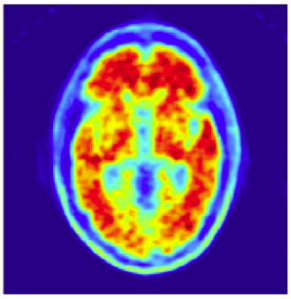
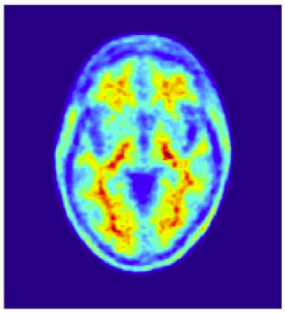
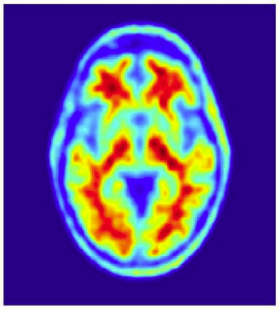
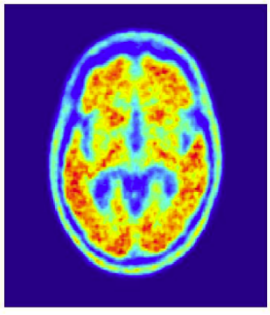
	Correctly classified	Classification failed
HV		
pAD		
pMCI		
sMCI		

Figure 4.3.2: Example florbetapir images that were correctly classified, as well as others where the algorithm failed

[140] and [143], which used genetics information and cognitive measurement respectively. In Table 4.4.1, we can see that our model outperformed the recent works that used MRI and FDG alone for classification, while the results from our multimodal work also close to the results in [108]. However, our model only used imaging biomarkers, which indicated the strong prediction ability of our model.

In summary, we conclude the advantages of the handcrafted feature based method that proposed in Chapter 3: 1) the requirement of data amount is lower compared to deep learning methods. 2) the requirement for hardware is lower, i.e. no need for GPUs. 3) training time is usually shorter. 4) we can interpret the learning process more easily.

Meanwhile, the superiority of deep learning model used in this chapter compare to Chapter 3: 1) no requirement for time consuming features extraction and expertise 2) enlarging the data can improve the performance by a great amount.

4.4 Conclusion

In this Chapter we presented a novel multi-modality deep learning model for MCI progression prediction using structural MRI, FDG PET and amyloid PET. Three experiments were performed: pMCI vs sMCI, (pMCI+pAD) vs sMCI, and (pMCI+pAD) vs (sMCI + HV), which are referred to E1, E2 and E3 in this study. The best performance was reached using MRI with amyloid with added pAD to the pMCI group and HV to the sMCI group. All three experiments showed that the use of multimodal images outperformed the single modality results in terms of accuracy. Meanwhile, both E2 and E3 achieved higher classification performance compare to E1 with E3 showing the highest accuracy (83.6%) for disease prediction. In the future, we could extend the application of this method to novel data such as tau imaging.

Table 4.4.1: Comparison of pMCI vs sMCI classification accuracy with state-of-the-art methods, 'FDG' and 'Amyloid' represent FDG PET and amyloid PET respectively. ApoE represents Apolipoprotein E.

Methods	Modality	methods	Subjects	Accuracy (%)
Young et al. [140]	MRI	MK-SVM	143	64.3
Zhu et al. [141]	MRI	SVM	99	71.8
Proposed	MRI	3D CNN	210	**74.2**
Young et al. [140]	FDG	MK-SVM	143	65
Zhu et al. [141]	FDG	SVM	99	71.2
Proposed	FDG	3D CNN	199	77.3
Proposed	Amyloid	3D CNN	199	**81.8**
Young et al. [140]	MRI+ FDG+ ApoE	MK-SVM	143	69.9
Zhu et al. [141]	MRI+ FDG	SVM	99	72.4
Suk et al. [142]	MRI+ FDG	DBM	204	75.9
Korolev et al. [143]	MRI+cognitive measure	Probabilistic multi-kernel learning	259	80
Choi et al. [108]	Amyloid+FDG	3D CNN	171	**84.2**
Proposed	MRI+ FDG	3D CNN	199	80.4
Proposed	MRI+ Amyloid	3D CNN	210	**83.6**

In this chapter, the presented results show the necessity and demonstrate the feasibility of adding similar categories to the training data to improve the performance. In the next chapter, we will hence address the problem of dataset shortage and imbalance, by developing a novel data augmentation deep learning modal using conditional generative adversarial networks (cGANs) for the improvement of MCI conversion classification. Due to

the limited availability of amyloid PET in the pMCI class, we use MRI in the pMCI class to generate corresponding amyloid PET in the same class. The results show that the synthetic amyloid PET images that were generated from structural MRI using cGANs can improve the classification performance significantly.

5 Generation of Amyloid PET Images via Conditional Adversarial Training for Predicting Progression to Alzheimer's disease

5.1 Materials 97
5.1.1 Data and pre-processing 97
5.2 Methods 98
5.2.1 Amyloid PET generation with conditional adversarial training 98
5.2.2 MCI progression classification architecture using 3D ResNet 101
5.2.3 Experiments 103
5.2.3.1 Data generation setup 103
5.2.3.2 MCI Classification setup 106
5.3 Results and Discussion 107
5.3.1 Data generation 107
5.3.2 Classification results 108
5.3.3 Tuning Hyperparameters 114
5.3.4 Comparison with previous results 116
5.4 Conclusion 118

To realize the full potential of deep learning for PET imaging, large datasets are required for training. However, dataset sizes are restricted due to a number of reasons, including patient confidentiality and high cost, particularly in the case of PET scans. Meanwhile, most of the AD classification studies have been based on structural MRI rather than PET. A possibility that has been successful in other deep learning applications is to generate new samples to add to the original ones for augmentation [144][145][146]. In our case, given the much higher availability of MR than PET scans, we hypothesise that generating new PET images from unseen MRIs could provide this much needed augmentation of the training sets.

However, to the best of our knowledge, direct generation of 3D amyloid PET imaging from structural MR has not yet been attempted. In this study, we focus on the application of conditional Generative Adversarial Networks (cGANs) to generate high quality volumetric florbetapir PET images from corresponding MRI images. In this way, we expect the natural variability in MRI scans and the image characteristics in PET to be combined. We use these augmented datasets to train a deep learning classification framework similar to the one introduced in Chapter 4, with the main difference being that here we build a more sophisticated network: a 152-layer ResNet classification model, to distinguish between pMCI and sMCI subjects. Our experiments quantify the difference in performance caused by the use of a more advanced network as well as the addition of these synthetic datasets in training, showing that generated PET images can be used for synthetic data augmentation, and improve the performance of 3D CNN for predicting progression to AD. The summary of the data generation model in this work is shown in Figure 4.4.1.

Our method is applied to a dataset of 163 18F-florbetapir PET images (134 sMCI, 29 pMCI), obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. This is the same dataset that was used in Chapters 3 and 4, allowing us to directly compare results. We use the trained cGANs to generate synthetic PET images from an additional 105 MRI datasets from pMCI patients (not used in previous chapters), which increases the total number of pMCI PET scans to 134 (the same number as sMCI), including real and generated scans. In this way we also address the problem of class imbalance between sMCI and pMCI. We evaluate the quality of generated PET images by comparison to the real PET images. Finally we build a 152-layer ResNet to compare the MCI classification performance using both a traditional data augmentation method and our proposed synthetic data augmentation method. The mean Structural Similarity (SSIM) index was 0.94 when the generated PET scans were compared with their corresponding real PET scans. For MCI progression classification, the traditional data augmentation method showed 82.7% accuracy while the synthetic data augmentation proposed here improved this to 84.5%.

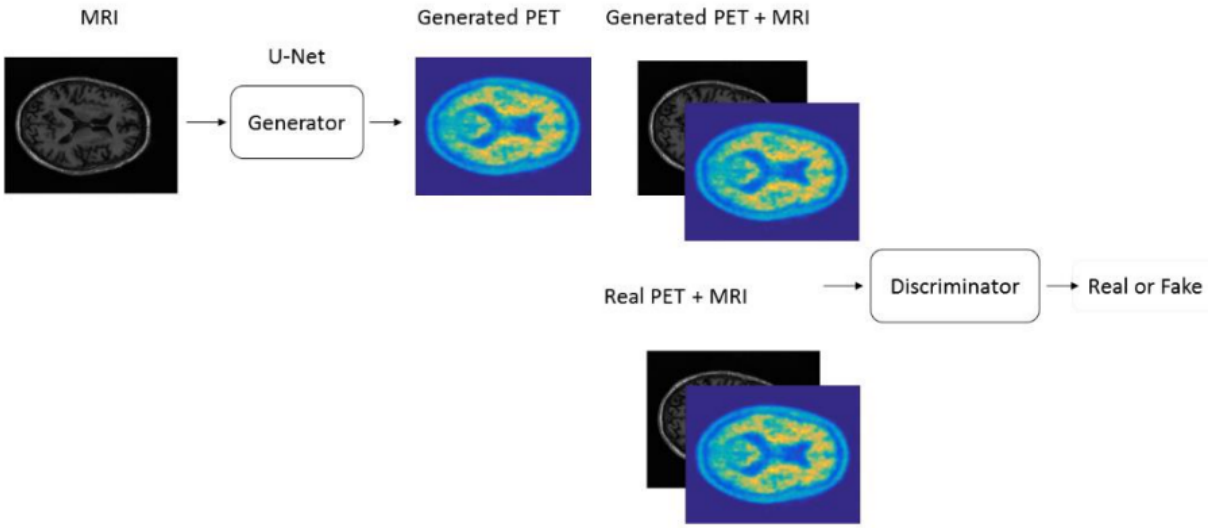


Figure 4.4.1: Schematic of the conditional GANs used in this work. The MRI scan is transformed into its corresponding PET by a U-Net working as Generator, while the Discriminator tries to differentiate between the generated and real PETs, in both cases paired to their corresponding MRI.

5.1 Materials

5.1.1 Data and pre-processing

All image data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI aims to improve clinical trials for the prevention and treatment of Alzheimer's disease. To date, over 1000 scientific publications have used ADNI data. ADNI has been running since 2004 and is currently funded until 2021.

In this study, as mentioned above, the same data as in previous chapters were used: 134 sMCI and 29 pMCI florbetapir images with corresponding T1 MRI. Additionally, a second group of 105 T1 MRI images from a different pMCI group were also downloaded from ADNI and used independently. More details about the use of these datasets in training/validating/testing the cGANs and ResNet are provided in the relevant sections below. All the florbetapir images were pre-processed using same procedure in previous chapters.

5.2 Methods

5.2.1 Amyloid PET generation with conditional adversarial training

GANs are generative networks that learn a mapping from random noise to a meaningful output (in our case, an image.) They are composed of two separate networks, a generator and a discriminator, trained in an adversarial way. The goal of the generator is to generate high-quality synthetic images, while the discriminator evaluates them for authenticity. In conditional GANs, the generator learns a mapping between an input and an output image [87].

In this study, the generator is a 3D U-Net convolutional neural network developed from the original 2D U-Net [147]. The U-Net architecture is based on the fully convolutional network and modified to make it more suited to medical image. While segmentation was the objective of the original U-Net paper, in subsequent studies it has been applied to other image-to-image transformations, as we do in the work presented here. In our work, the U-Net contains three parts: down sampling path (encoder), bottleneck, and up sampling path (decoder).

U-Net is symmetric and the skip connections between the down-sampling path and the up-sampling path apply a concatenation operator instead of a sum. Skip connections are used between each layer i in the encoder and layer $n - i$ in the decoder, where n is the total number of layers. Through these skip connections, the activations in layer i are concatenated with layer $n - i$. Because of its symmetry, the network also has a large number of feature maps in the up-sampling path, which allows the transfer of information at the relevant scales.

Down sampling path

The down sampling path is composed of 4 blocks. Each block is composed of:

- Two 3x3x3 Convolutional Layer + activation function (with batch normalization)
- 2x2x2 Max Pooling

The purpose of this down sampling path is to capture the context of the input image at multiple scales, in order to be able to perform an accurate segmentation (or other image-to-image transformations), with the coarse features capturing the main structures in the images and the fine ones ensuring accuracy at a voxel level. This contextual information is also transferred to the up-sampling path by means of skip connections, which allow direct transmission of information at the same scale.

Bottleneck

For nonlinear dimensionality reduction, we used bottleneck layers between the down sampling and up sampling paths. The bottleneck is built from two 3x3x3 convolutional layers (with batch normalization), with dropout.

Up sampling path

The up sampling path is also composed of 4 blocks. Each block is composed of:

- One 2x2x2 deconvolution layer with stride 2
- Concatenation with the corresponding cropped feature map from the down sampling path
- Two 3x3x3 Convolutional layer + activation function (with batch normalization)

The purpose of this up sampling path is to enable precise localization combined with contextual information from the down sampling path.

The discriminator is a convolutional Markovian discriminator (PatchGAN), which only analyses structure at the scale of image patches. This makes the images more accurate at the level of small patches, making them have a realistic look. This discriminator tries to classify if each patch in an image is real or fake. The smaller PatchGAN is also advantageous due to a lower number of parameters and higher computational speed, compared with discriminators that analyse the image as a whole. We tested the effect of

varying the patch size of our discriminator receptive fields and the variations will be shown in the following discussion section.

During the GAN training process, the generated PET was paired with its corresponding MRI and entered into the discriminator. The loss function of the conditional GAN is:

$$L_{cGAN}(G,D) = E_{x,y}[logD(x,y)] + E_x\left[\log\left(1 - D\left(x, G(x)\right)\right)\right] \quad (15)$$

where x are MRI images and y are PET images. G represents the generator, D refers to the discriminator, E means the expectation operation.

The first term is maximized when D(x, y) =1, and the second is maximized when D(x, G(x)) = 0, while it is minimised when D(x, G(x)) = 1, i.e. when the discriminator is not able to distinguish the generated images and real images. The generator G tries to minimize this objective against an adversarial discriminator D that tries to maximize it. In addition, conditional GANs also add an L1 loss term, which quantifies the difference between the output of the generator and the ground truth image available during training:

$$L_{L1}(G) = E_{x,y}\left(||y - G(x)||\right) \quad (16)$$

Therefore the complete form of the loss function is:

$$L_{total}(G,D) = L_{cGAN}(G,D) + \varepsilon L_{L1}(G) \quad (17)$$

where ε is used to adjust the relative contribution of the L1 loss. We also tested the effect of tuning ε and this will be shown in the following discussion section.

For the implementation of the cGANs architecture we used the Keras framework (TensorFlow backend). The experiments were performed on a computer cluster within the IBME, University of Oxford. The cluster has one headnode and three nodes. Each node has four Nvidia GeForce GTX1080 Ti GPUs, two Intel Xeon E5-2667 v4 3.2GHz CPUs and a storage server (1TB). Node 1 has 512 GB memory while Node 2 and Node 3 have 256 GB memory. All the nodes run on CentOS 7.3. This work was conducted with 3 GPUs using Node 1.

5.2.2 MCI progression classification architecture using 3D ResNet

Deep Residual Network (ResNet) [148] is arguably one of the most important developments in the deep learning area in the last few years. ResNet makes it possible to train up to thousands of layers and still achieves competitive performance with fast convergence. The original paper [148] showed that these residual networks are easier to optimize, and can gain accuracy from significantly increased depth. In 2015, ResNet achieved 3.57% error on the ImageNet test set. This result won the first place on the ILSVRC 2015 classification task. The core concept of ResNet is introducing an identity shortcut connection that skips one or more layers. ResNets have been used successfully for 3D image segmentation as in VoxResNet, where the authors use identity mappings as skip connections [149]. The residual unit is shown in Figure 5.2.1.

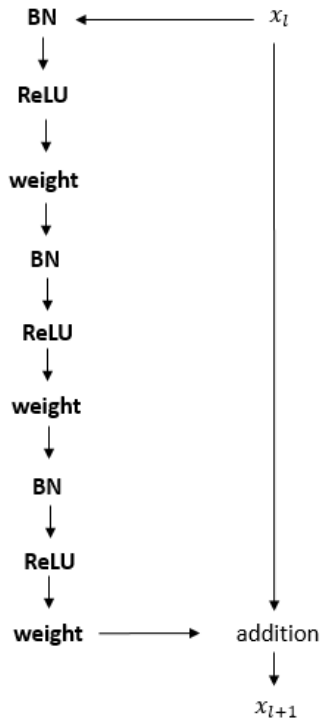


Figure 5.2.1: One residual unit architecture.

Following [148], the computation procedure in one residual unit can be described as:

$$y_l = h(x_l) + F(x_l, W_l) \tag{18}$$

$$x_{l+1} = f(y_l) \tag{19}$$

Where x_l is the input of the l-th residual unit. $W_l = \{W_{l,k} | 1 \leq k \leq K\}$ are the weights and biases of the l-th residual unit, and K refers to the number of layers in a residual unit. F means the residual function and f indicates the ReLU operation. The function h is normally used as an identity mapping:

$$\mathrm{h}(x_l) = x_l \tag{20}$$

If f is also an identity mapping, we have:

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i) \tag{21}$$

for any deeper unit L and any shallower unit l, to achieve an identity mapping $f(y_l) = y_l$, the first ReLU and BN in Figure 5.2.1 are named “pre-activation” of the weight layers.

From the chain rule of the backpropagation, we have:

$$\frac{\partial \varepsilon}{\partial \mathrm{x}_l} = \frac{\partial \varepsilon}{\partial \mathrm{x}_L}\frac{\partial x_L}{\partial \mathrm{x}_l} = \frac{\partial \varepsilon}{\partial \mathrm{x}_L}\left(1 + \frac{\partial}{\partial \mathrm{x}_l}\sum_{i=l}^{L-1} F(x_i, W_i)\right) \tag{22}$$

In this equation, ε is the loss function and $\frac{\partial \varepsilon}{\partial \mathrm{x}_l}$ is divided into two parts: $\frac{\partial \varepsilon}{\partial \mathrm{x}_L}$ is the direct path for the information propagation without passing the weight layers, $\frac{\partial \varepsilon}{\partial \mathrm{x}_L}\left(\frac{\partial}{\partial \mathrm{x}_l}\sum_{i=l}^{L-1} F\right)$ is the term that passes through the weight layers.

In our work, the ResNet architecture extended the work in [150] that applied to 2D images, to work with 3D images. The main difference between our network and the identity mappings version is the number of dimensions of convolutional kernels and pooling. Our ResNet architecture has 152 layers containing 50 3-layer blocks. The three layers are 1x1x1, 3x3x3,

1x1x1 convolutions, where the 1x1x1 layers are responsible for reducing and then increasing dimensions, leaving the 3x3x3 layer a bottleneck with smaller input and output dimensions, as detailed in Table 5.2.2. Down-sampling is performed by conv3_1, conv4_1, conv5_1 with a stride of 2.

To implement the ResNet architecture, we used the Keras framework (TensorFlow backend). The experiments were performed on a computer cluster within IBME, University of Oxford. The cluster has one headnode and three nodes. Each node has four Nvidia GeForce GTX1080 Ti GPUs, two Intel Xeon E5-2667 v4 3.2GHz CPUs and a storage server (1TB). Node 1 has 512 GB memory while Node 2 and Node 3 have 256 GB memory. All the nodes run on CentOS 7.3. This work was conducted with 3 GPUs using Node 1. The complete 152-layer ResNet network architecture is shown in Figure 5.2.2.

5.2.3 Experiments

5.2.3.1 Data generation setup

In this first experiment, which we refer to as EX 1, we used the 29 paired PET and MRI images from the pMCI group to train the conditional GANs, and subsequently applied the mapping to the unseen 105 pMCI MRI images to generate 105 synthetic PET images, thus enlarging the size of the pMCI dataset. We used SSIM values to test the quality of the images generated from the cGAN, by comparing them to their corresponding real PETs (which were kept fully separated from the training set). The number of datasets from different categories used in EX 1 are shown in Table 5.2.1.

Table 5.2.1: Illustration of data used in EX 1.

EX 1: Generation of synthetic PET images		
	PET (pMCI)	MRI (pMCI)
Data	29 (train)	29 (train) + 105 (test)

Table 5.2.2: 152-ResNet architecture for pMCI and sMCI classification

Layer name	152-layer
Conv1	7x7x7, 64, stride 2
Conv2_x	3x3x3 max pool, stride 2
	$\begin{bmatrix} 1x1x1 & 64 \\ 3x3x3 & 64 \\ 1x1x1 & 256 \end{bmatrix}$ x 3
Conv3_x	$\begin{bmatrix} 1x1x1 & 128 \\ 3x3x3 & 128 \\ 1x1x1 & 512 \end{bmatrix}$ x 8
Conv4_x	$\begin{bmatrix} 1x1x1 & 256 \\ 3x3x3 & 256 \\ 1x1x1 & 1024 \end{bmatrix}$ x 36
Conv5_x	$\begin{bmatrix} 1x1x1 & 512 \\ 3x3x3 & 512 \\ 1x1x1 & 2048 \end{bmatrix}$ x 3
	average pool, 2-d fc, softmax

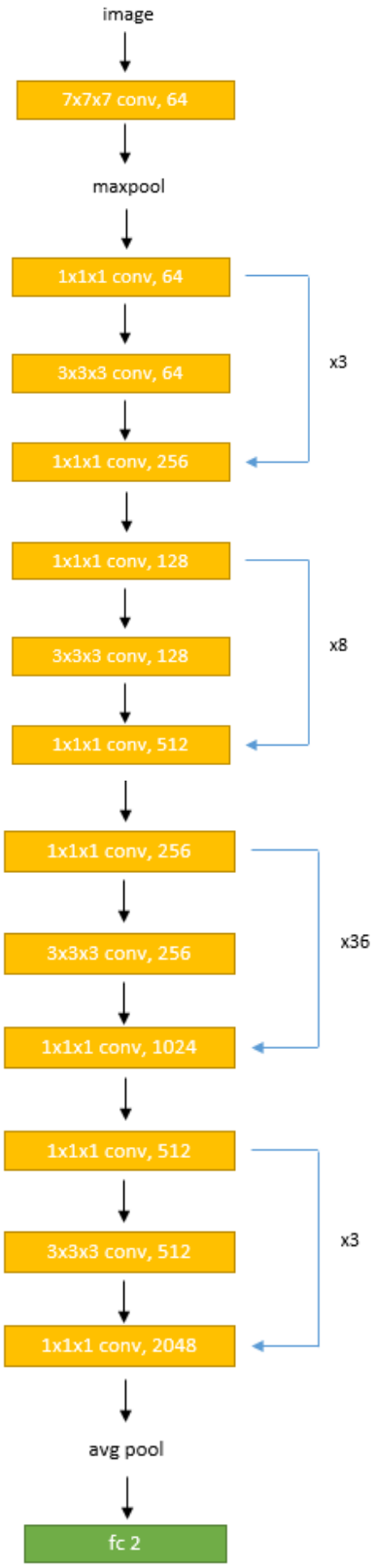


Figure 5.2.2: Complete 152-layer ResNet network architecture (for simplicity, only convolutional layer is shown in each residual block)

5.2.3.2 MCI Classification setup

In this experiment (EX 2) we tested the ability of the network to classify pMCI vs sMCI PET datasets, with or without the inclusion of the generated PET scans in the training. A 10-fold cross-validation was applied to the whole dataset, in both cases. We tested two scenarios. In the first one we used traditional augmentation to generate 105 pMCI PET images. Specifically, we introduced random rotation and flipping, with the random rotation range set to 60 degrees, and images were randomly flipped horizontally and vertically. In the PET synthesis experiment, we used our cGAN augmented dataset, including the additional 105 PET images generated from the MRI scans by the cGAN. Overall there are 214 samples for training, 26 for validation, and 28 for testing (268 in total) for both scenarios. For training we used a batch size of 1 with a learning rate of 0.0001 for 200 epochs. In order to prevent overfitting, we used early stopping, which will ensure the training stops when the chosen performance measure stops improving. The details of EX 2 are shown in Table 5.2.3.

Table 5.2.3: Illustration of data used in EX 2.

EX 2: classification of MCI to AD using amyloid imaging	
Data	Florbetapir: 134 sMCI+29 pMCI+105 generated pMCI (268 in total)

Finally, we performed a multi modal classification experiment (EX 3) using both florbetapir PET and T1 MRI. In chapter 4, we developed a multimodal 3D CNN for MCI classification task, but in this chapter we used two parallel 3D 152-ResNets (one processing the PET, one the MRI) for the same task. The traditional augmentation method in Ex3 is same as Ex2. The data used in EX3 is shown in Table 5.2.4.

Table 5.2.4: Illustration of data used in EX 3.

EX 3: classification of MCI to AD using amyloid imaging and MRI	
Data	Florbetapir: 134 sMCI+29 pMCI+105 generated pMCI (268 in total) MRI: 268 corresponding MRI

5.3 Results and Discussion

5.3.1 Data generation

In this study, we used 105 volumetric MRI images to generate volumetric PET images to enlarge the pMCI group. To evaluate the model performance, we use three measures in the experiments. Specifically, structural similarity index (SSIM), peak signal-to-noise (PSNR) and normalized mean squared error (NMSE) are employed to evaluate the generated image quality.

SSIM combines information about differences in image contrast and overall structure, and has been used routinely in the validation of image generation methods [144,151]. SSIM is defined as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (23)$$

where μ_x , μ_y, σ_x , σ_y , σ_{xy} are the local means, standard deviations, and cross-covariance for images x, y. C_1 and C_2 are regularization constants which are set to small non-zero values. SSIM ranges between 0 and 1, with SSIM=1 meaning that the two images are identical.

PSNR is defined as

$$PSNR = 10log10(\frac{NM^2}{||I-I_G||_2^2}) \quad (24)$$

Where M is the maximum intensity range of the real PET image and the generated PET image I_G, and N refers to the total number of voxels in the image. The metric is used to evaluate the image generation quality of the PET image quantitatively.

The NMSE is defined as

$$NMSE = \frac{||I - I_G||_2^2}{||I||_2^2} \quad (25)$$

This metric is used to measure the voxel-wise intensity differences between the real and generated images. Theoretically, an image with higher SSIM, PSNR and lower NMSE represents higher quality. Recently, some studies have conducted PET synthesis using GANs. In [152], the authors generated synthetic PET from random noise using Deep Convolutional GANs, and they achieved a mean SSIM 77.48 and PSNR 32.83. In [153], the authors produced PET data based on their corresponding MRI data by using 3D Cycle-consistent GANs, with PSNR values of synthetic PET images 24.49 $\pm$ 3.46. In our experiments, the averaged SSIM, PSNR and NMSE for synthetic and real PET are 0.94± 0.03, 33.2$\pm$ 1.6 and 0.012$\pm$ 0.004, which indicate high quality of the generated PET images compared with state of the art results.

Examples of real and generated pMCI PET images in three views are also shown in Figure 5.3.1, Figure 5.3.2 and Figure 5.3.3. From the results we could observe the similarity of the generated pMCI PET images with respect to the real pMCI PET images, which indicates that our model is able to generate high quality PET images from MRI.

5.3.2 Classification results

Classification results for pMCI against sMCI using the 152-ResNet are shown in Table 5.3.1. We computed both the area under the receiver operating characteristic curve (AUC) and the

accuracy (ACC). Two different cases are compared: classification with a network trained using traditional augmentation and using our synthetic images based augmentation method.

Table 5.3.1: Classification ROC AUC and accuracy with 152-ResNet

Amyloid PET		
	AUC (%)	ACC (%)
sMCI vs pMCI (traditional)	78.3	82.7
sMCI vs pMCI (synthetic)	82.6	84.5
Amyloid + MRI		
sMCI vs pMCI (traditional)	80.5	84.2
sMCI vs pMCI (synthetic)	84.4	86.1

In Table 5.3.1, for EX 2, the classification accuracy for sMCI against pMCI using real PET images achieved accuracy 82.7%, with the aid of traditional data augmentation. Meanwhile, our proposed synthetic augmentation method achieved 84.5% accuracy. For EX 3, the classification score for sMCI against pMCI using real PET images achieved accuracy 84.2%, with the aid of traditional data augmentation. As we expected, the highest accuracy was obtained by using multimodal synthetic augmentation method (86.1%), achieving an improvement of 2.3% over the traditional augmentation. The line plot of loss against epoch for EX2 and EX3 is shown in Figure 5.3.4. In EX2 and EX3, the model was saved using early stopping, when the epochs reached 165 and 162 respectively.

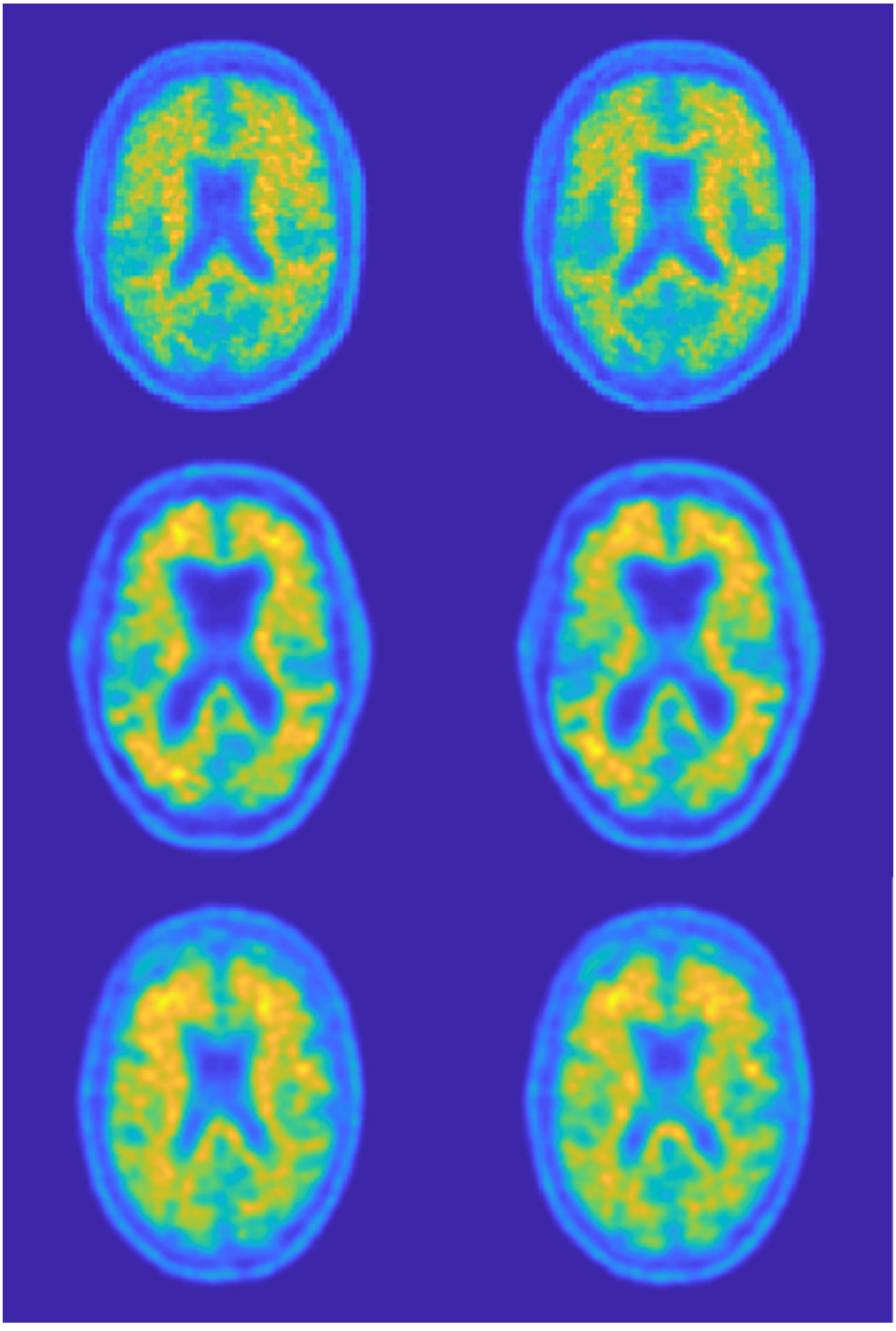
	Synthetic	Real
Axial		

Figure 5.3.1: Randomly selected generated images in axial plane with corresponding real image

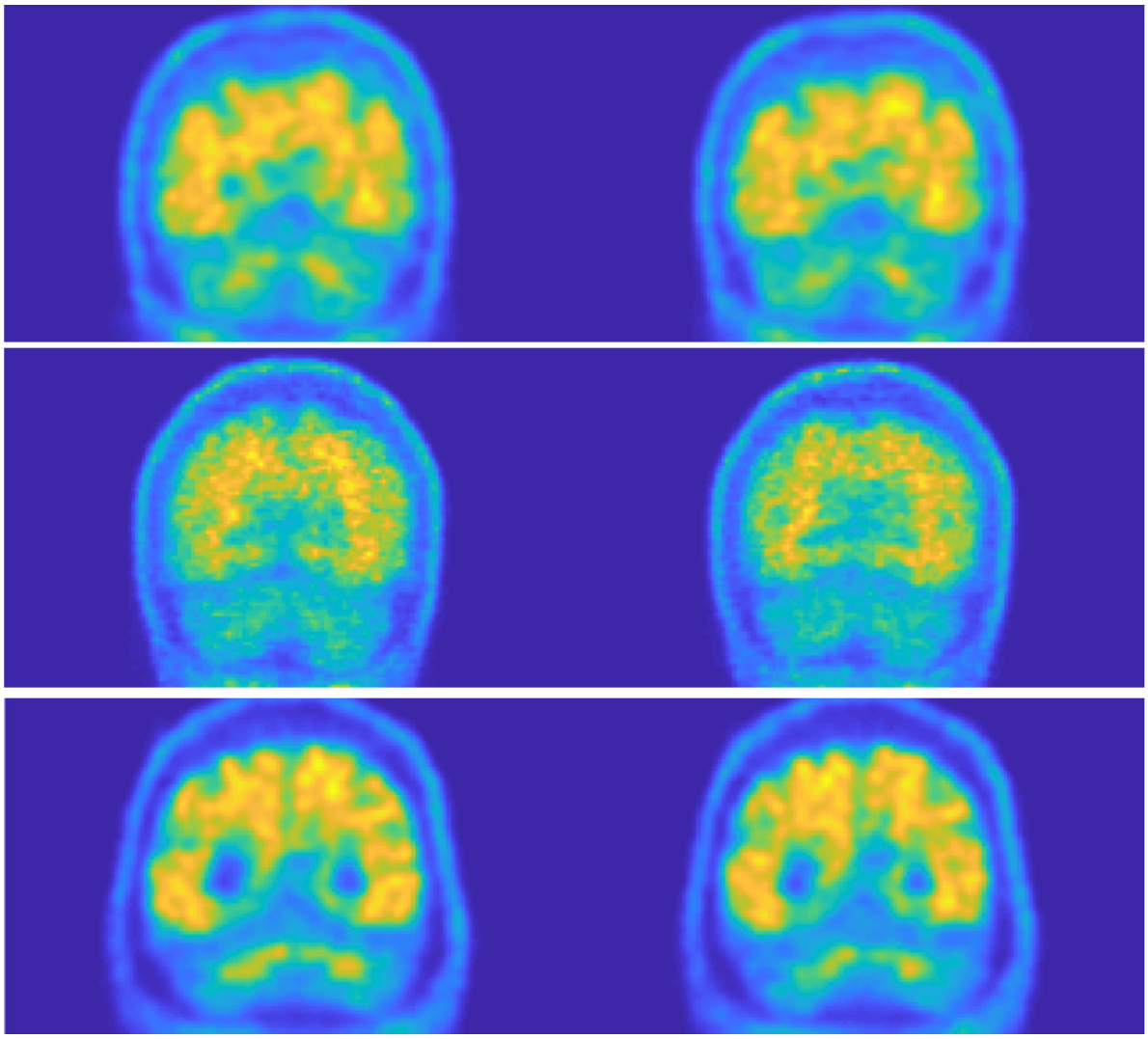
	Synthetic	Real
Coronal		

Figure 5.3.2: Randomly selected generated images in coronal plane with corresponding real image

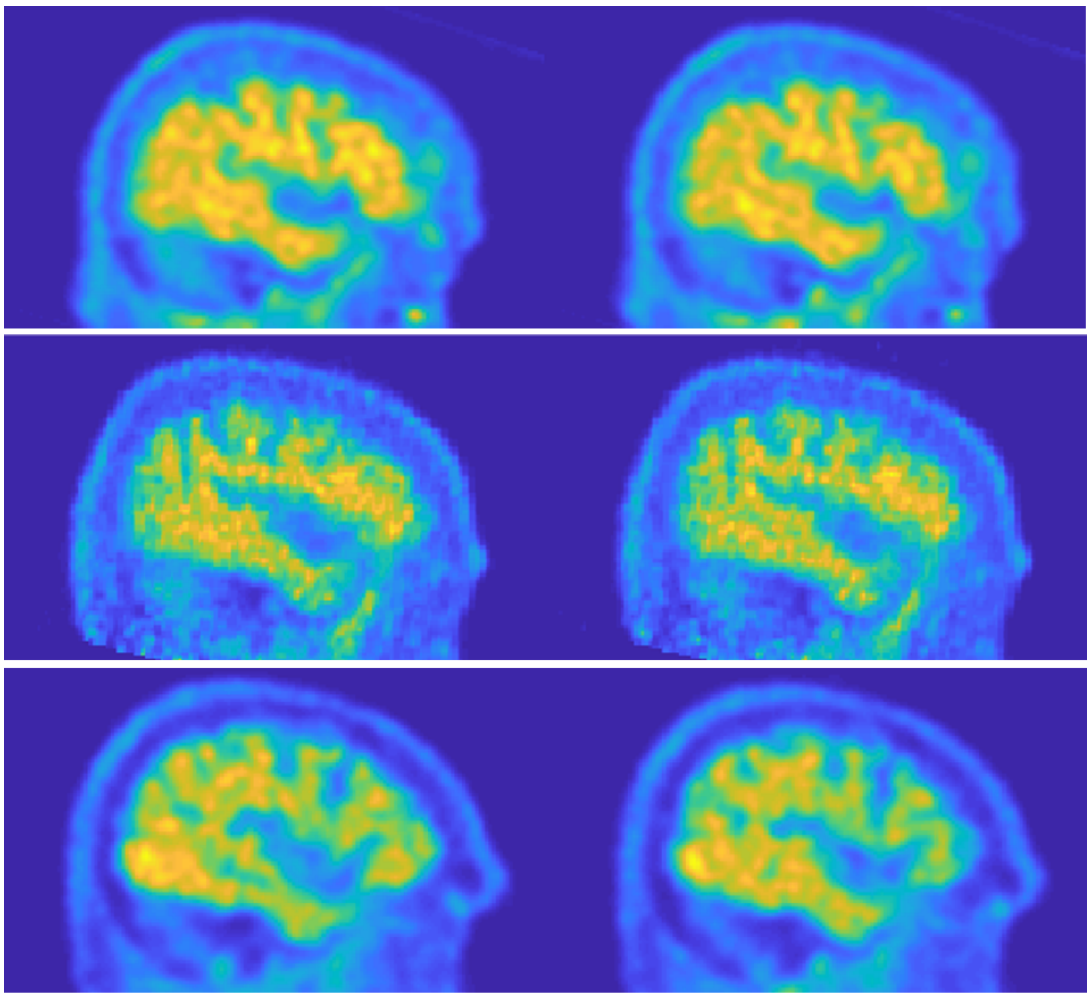
	Synthetic	Real
Sagittal		

Figure 5.3.3: Randomly selected generated images in sagittal plane with corresponding real image

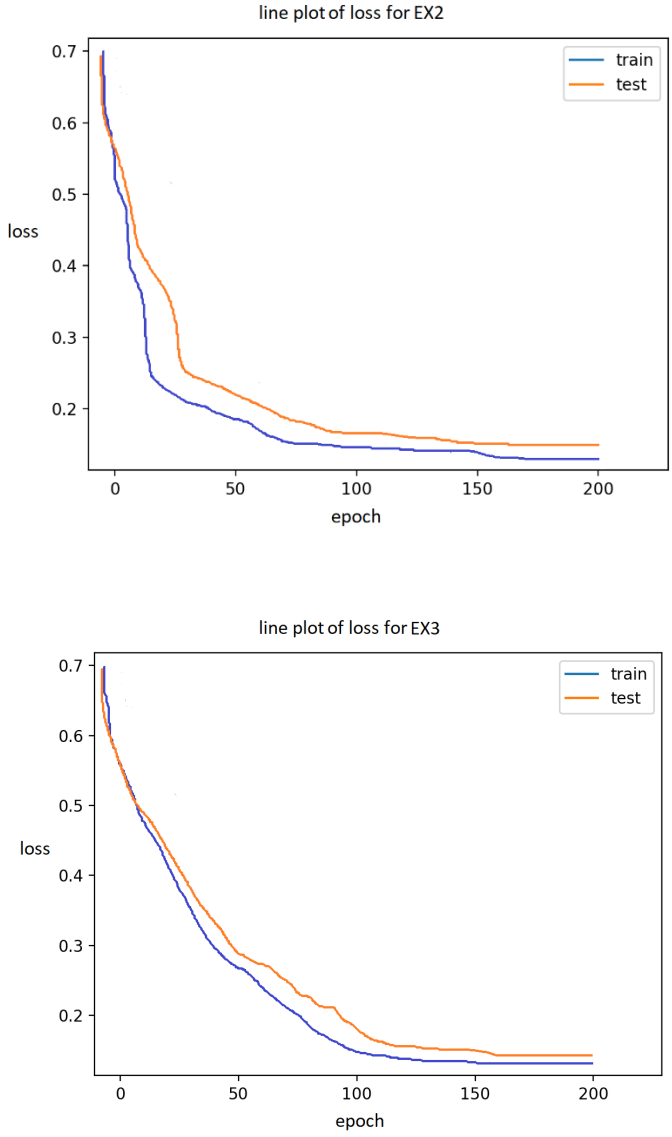
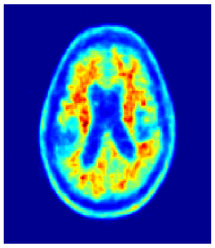
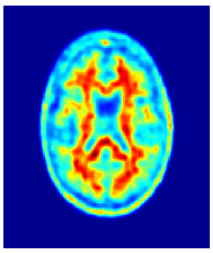
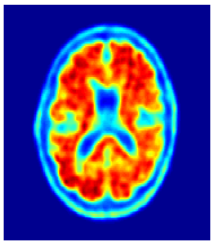
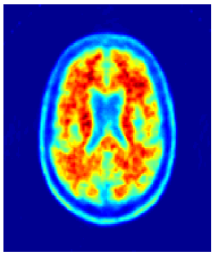
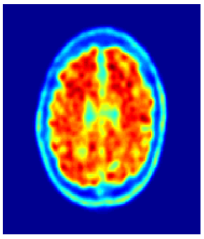
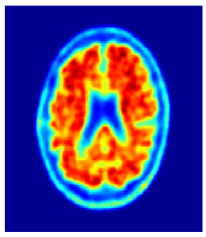
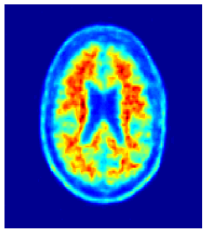
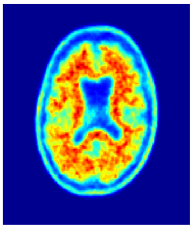


Figure 5.3.4: Line plot loss for EX2 and EX 3 using early stopping

For EX 3, some example PET images that are correctly classified as well as PET images that the algorithm has failed to classify are shown in Figure 5.3.5. Similar to the previous chapter, the main reason that the classification failed in this work is due to the fact that these images showed different amyloid deposition in gray matter area within the same class. For instance, the majority of sMCI subjects that are classified incorrectly show a heavy amount of amyloid depositions within the white matter and gray matter, and vice versa for the majority of pMCI subjects.

	correctly classified		not correctly classified	
sMCI				
pMCI				

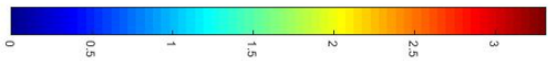


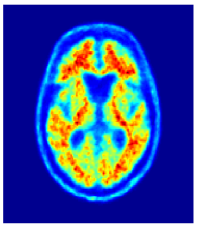
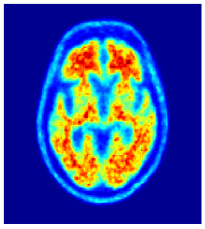
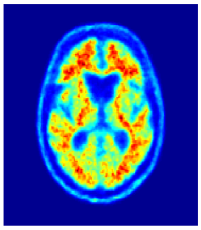
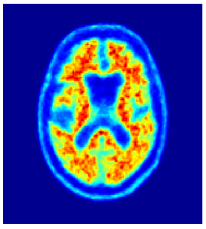
Figure 5.3.5: In EX 3, example PET images (left two columns--correctly classified, right two columns-- incorrectly classified)

5.3.3 Tuning Hyperparameters

We tested the effect of varying regularization term ε and then tested the effect of different patch sizes in our discriminator. We calculated the average SSIM scores for using both different ε and patch sizes which is shown in Table 5.3.2 and Table 5.3.3. The example PET images using different ε and patch size are shown in Figure 5.3.6 and Figure 5.3.7.

Table 5.3.2: SSIM scores between generated and real PET for different ε (patch size 32). Bold text refers to the chosen regularization term ε

Regularization term ε	Average SSIM scores
10	0.785
100	**0.924**
500	0.792

Ground truth	ε =10	ε =100	ε =500
			

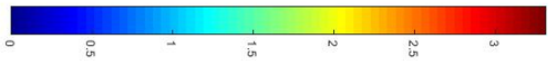
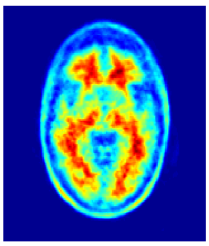
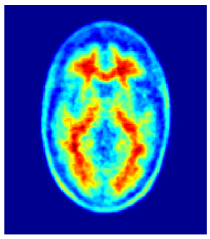
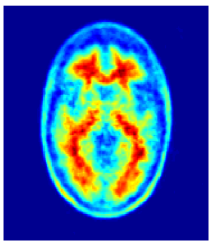
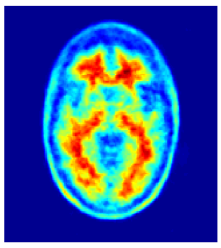


Figure 5.3.6: The example PET images using different ε with corresponding ground truth (patch size 32)

Table 5.3.3: SSIM scores between generated and real PET for different patch sizes in discriminator (ε=100). Bold text refers to the chosen patch size in discriminator

Patch size in discriminator	Average SSIM scores
16	0.904
32	**0.924**
64	0.931

Ground truth	16	32	64
			

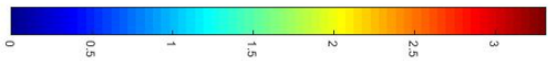


Figure 5.3.7: The example PET images using different patch size with corresponding ground truth (ε=100)

Based on Figure 5.3.6, the visual difference between generated PET and corresponding ground truth using different loss ratio value ε can be clearly observed in some areas of the brain. Meanwhile, in Figure 5.3.7, the visual difference between generated PET and the corresponding ground truth using different patch sizes is not easily detectable.

In Table 5.3.2 and Table 5.3.3, ε size 100 and patch size 64 achieved the highest SSIM score, though the difference compared with patch size 32 was very small. In this case, we use ε=100 to adjust the weight of L1 loss, and to reduce the computational cost, we decided to use PatchGANs with patch size 32 on all subsequent experiments.

5.3.4 Comparison with previous results

In chapters 3 and 4, we proposed a SVM classification model using novel features (fractal dimension and Shannon entropy) as well as a multi-modality 3D CNN classification model respectively. The florbetapir data used in this chapter is the same as the florbetapir data

used in the two previous chapters. This allows us to compare the ResNet classification model in this chapter with the two previous chapters. This is shown in the table below:

Table 5.3.4: Comparison between different chapters for MCI prediction performance (tra: traditional augmentation method, syn: synthetic augmentation method)

Accuracy (%)	Chapter 3 (E0) (134sMCI+29pMCI)	Chapter 4 (E2) (134sMCI+29pMCI+36pAD)	Chapter 5 (EX2 & EX3) (134 sMCI+29 pMCI+105 synthetic pMCI)
MRI	71.8	73.7	n/a
amyloid	80.4	81.8	82.7 (EX2 tra) 84.5 (EX2 syn)
MRI+amyloid	82.8	83.3	84.2 (EX3 tra) 86.1 (EX3 syn)

Based on Table 5.3.4, for the multi-modality experiment, we can see that the 152-ResNet model using synthesis data augmentation method has achieved the highest accuracy (86.1%) compared with chapter 3 (handcrafted feature method) and chapter 4 (multimodal CNN). This results further indicate the competitive classification capacity of ResNet and the significant gain in classification performance using data synthesis as an augmentation method. We suggest that the mapping from T1 MRI to amyloid PET brings new feature information that cannot be generated by PET or MRI alone. In addition, the traditional augmentation in EX 2 and EX 3 gives 82.7% and 84.2% accuracy respectively, which outperformed the results in chapter 3 and chapter 4 as well. This is possibly due to the fact that ResNet is better than SVM and regular CNN when the data set is sufficiently large for classification tasks: 163 scans for chapter 3, 199 for chapter 4, and finally 268 in this chapter. Meanwhile, it is worth mentioning that the results we achieved in this chapter also outperformed the results from the most recent literature [108] (MCI to AD prediction accuracy 84.2%).

Among the three research chapters, flutemetamol achieved the highest MCI prediction accuracy (88.9%). The possible reason for this has been discussed in chapter 3: the structural similarities to Thioflavin-S. The reason that flutemetamol data is not used in chapter 4 and chapter 5 is its very limited availability (just 9 pMCI and sMCI cases each), which is not compatible with deep learning methods such as the ones proposed here, even with sophisticated data augmentation. For such small datasets, we believe the use of handcrafted features is advisable. More specifically, traditional machine learning methods with handcrafted features such as FD and SE can be superior compared with deep learning methods when the dataset is small. However, one of the main advantages of using deep learning is that it reduces the need for feature engineering, one of the most time-consuming parts of machine learning practice.

5.4 Conclusion

We developed a model for generating florbetapir PET from structural MR using deep generative networks, with generated data showing a high similarity to real corresponding PET. The generated data were then used for data augmentation for MCI classification on a limited dataset. We compared the synthetic augmentation method with a traditional augmentation method, using both MRI and amyloid, and the synthetic augmentation outperformed the traditional augmentation. In addition, the works present in this chapter showed the highest classification accuracy for MCI to AD conversion compare to the two previous chapters, when using florbetapir. Future work will focus on other 3D CNN variants and novel PET imaging such as Tau imaging.

6 Conclusions

6.1 Summary 119
6.2 Future work 121
6.2.1 Tau imaging 121
6.2.2 Other variants of CNN 123
6.2.3 Other possibilities for future research 125

This chapter summaries the contributions of the thesis and then proposes future works that could be possibly addressed in the near future.

6.1 Summary

This thesis has presented and discussed methods addressing the challenging problem of early diagnosis of Alzheimer's disease (AD) and particularly the intermediate stage (MCI) to AD prediction, using both traditional machine learning and deep learning methods. The methods presented used handcrafted novel features as well as automatically extracted features. The proposed approach can predict AD conversion two years prior to the confirmation of probable AD. This thesis has discussed, in detail, several contributions to the medical imaging field, which will be briefly summarised in this chapter.

The first contribution of this thesis is the application of traditional machine learning methods using novel handcrafted features for early diagnosis of AD. In this method we have applied fractal dimension and Shannon entropy, two complexity descriptors, to measure the complexity difference between the healthy subjects and the diseased patients. More specifically, we found that the complexity of the gray-white matter and gray matter-CSF boundaries can be used to distinguish different stages in AD, with the aid of fractal dimension and Shannon entropy. Fractal dimension and Shannon entropy have been previously applied to T1 MRI and DTI while the application to PET and amyloid imaging in

particular is novel. We further introduced a nonlinear SVM using a radial basis function kernel for the AD conversion prediction. The method was applied to both volumetric MRI and volumetric PET, where PET included two different radiotracers: F18-flutemetamol and F18-florbetapir. The applications on these two radiotracers are also novel and we therefore use them to address the most important classification problem in AD: MCI to AD conversion prediction. We compared our method with other methods in the field and showed that ours performs favourably in terms of MCI classification accuracy.

The second contribution is the introduction of a multimodality deep learning approach for MCI to AD prediction. The method used two volumetric convolutional neural networks in parallel, with the aid of multi-modality data: T1 MRI, FDG PET and amyloid PET. We made five different combinations using different imaging modality combinations: 1) MRI 2) Amyloid 3) FDG 4) MRI+FDG 5) MRI+Amyloid. In addition, we proposed two more experiments which augmented the original training data with the most "similar and relevant" data: add pAD to pMCI group and afterwards add HV to sMCI group, which will lead to two extra classification tasks: (pAD+pMCI) against sMCI as well as (pAD+pMCI) against (sMCI+HV). We made a comprehensive discussion of the results that have been achieved and we found that the results achieved by adding pAD to pMCI has the highest accuracy, and outperformed the results in chapter 3. Our results were also compared with several other traditional machine learning and deep leaning methods that used for MCI classification and achieved competitive accuracy.

The third contribution in this thesis is to address the results and conclusions from the two previous chapters, and develop a more complete network for MCI classification. First, we address the problem encountered in previous chapters: the datasets for diseased patients are relatively small, and the data is therefore imbalanced. We developed a conditional generative adversarial network in order to generate amyloid PET (pMCI) from corresponding T1 MRI. Subsequently, the generated PET was added to the original pMCI group to enlarge the dataset. We used the structural similarity index to test the quality of the generated

images and it showed that the generated images have very close image textures when compared with the real images. Based on chapter 4, we further developed a deep learning classification model using 152 layers residual deep learning networks, and the results using both T1 MRI and amyloid PET again reached the highest accuracy, compared with the two previous chapters.

In summary, we can conclude that 1) for single modality, amyloid PET outperformed MRI for MCI classification 2) adding pAD to pMCI group can improve the classification performance 3) Multimodality analysis can provide complementary information and the results outperformed any of the single modality alone. The works in thesis may have a great potential to lead to a new perspective for computer aided diagnosis in Alzheimer's disease, especially when the cure of Alzheimer's disease has been found.

6.2 Future work

6.2.1 Tau imaging

Abnormal aggregation of tau has become an important hallmark for different neurodegenerative diseases such as AD. The effect of tau phosphorylation in AD remains unclear. Recently, researchers have made substantial efforts to determine how to accurately target the tau deposits in vivo in the brain [154]. Thanks to the development of molecular imaging in recent years, several tau specific tracers has been produced for PET, such as THK5317, PBB3 and AV-1451 [155]. These tracers have already been utilized in a number of clinical studies. The patterns in tau depositions may lead to the early diagnosis of AD and other neurodegenerative diseases. The progression of tau depositions in AD according to Braak staging are shown in Figure 6.2.1.

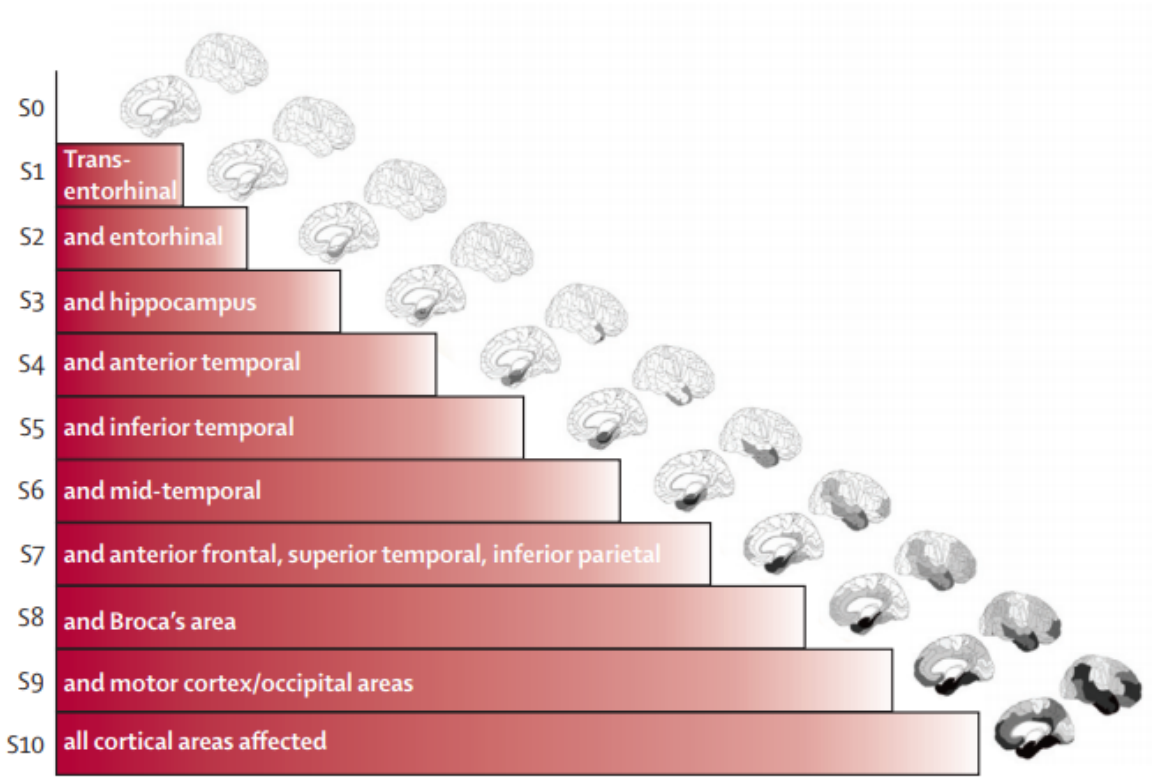


Figure 6.2.1: Tau deposition progression based on Braak staging (from[156])

The availability of data (and, in particular, annotated data) is one of the biggest concerns in the medical imaging field. Fortunately, the latest data collection launch in ADNI, named ADNI 3, includes longitudinal tau PET scans with AV1451.

Currently, the applications of using tau PET for early diagnosis of AD are very limited. In 2018, Jack et al. proposed that Tau PET SUVR is useful for differentiating stages of AD: cognitively impaired abnormal amyloid group, cognitively unimpaired abnormal amyloid group, and cognitively unimpaired normal amyloid group [157].

In the near future, due to the failure of several therapies aiming to reduce amyloid deposition, the inhibition of tau aggregation has drawn attention for clinical trials. The combination of modern machine learning or deep learning algorithms with tau PET may lead to more promising results for early diagnosis of AD, compared to amyloid PET. In this case, the contributions in this thesis may be used with the combination of tau imaging, because the tau imaging dataset size is small and data augmentation is necessary.

6.2.2 Other variants of CNN

The deep learning field is evolving at a very high speed, and network architectures are rapidly replaced by new ones that achieve superior results. While the timing did not allow us to include them in our research, they could be the topic of future work. Some ideas are provided in the following.

DenseNet

Deep convolutional networks can be difficult to train and suffer from gradient vanishing. The early convolutional networks that were developed such as LeNet and VGG only had 5 layers and 19 layers. In this case, the development of ResNet and Dense Convolutional Network (DenseNet) has made very deep convolutional networks possible. We used ResNet in the work presented here. A potentially superior alternative is DenseNet, which could be immediately applied to the same dataset used in chapter 5 and compare the results with those that ResNet has achieved [158].

DenseNet simplifies the connectivity pattern between layers, to ensure the maximum information and gradient flow between layers. To do it and tackle the training difficulty in deep networks, it simply connects every layer directly with each other, and thus each layer has direct access to the gradients of the loss function and the original input. Instead of making the network architecture deeper or wider, the network design focuses on the feature reuse. By connecting this way, DenseNet requires fewer parameters than an equivalent traditional CNN [158], and thus it should be possible to train it with smaller datasets. The layout of a dense block in DenseNet is shown in Figure 6.2.2.

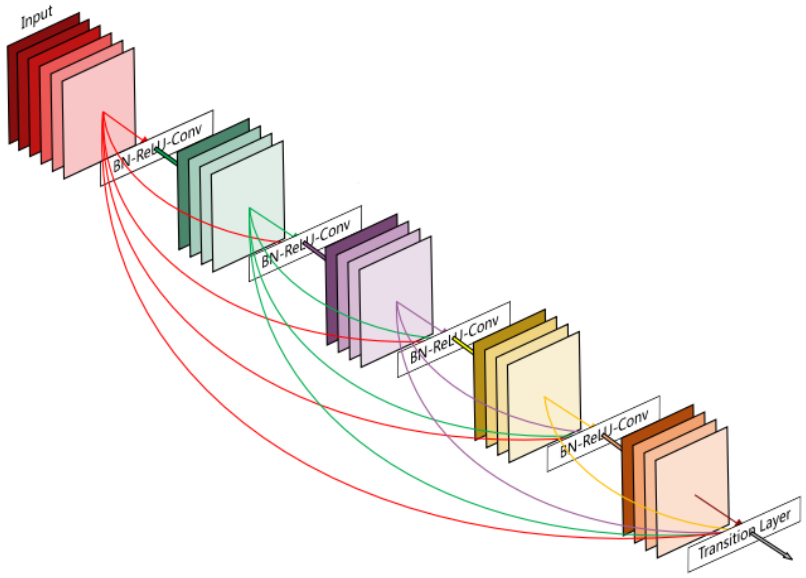


Figure 6.2.2: A 5-layer dense block in DenseNet, each layer takes all preceding feature maps as input (from [158]*)*

FractalNet

As can be seen by the ResNet and DenseNet architectures, modern deep networks include a combination of direct and skip connections. Determining the right number and configuration of these connections is challenging, and involves multiple decisions by the designer, many of them interrelated.

In 2017, Larsson et al. developed another ultra-deep neural network named FractalNet. FractalNet proved that the residual block may not be essential for ultra-deep networks [159] .

The design of FractalNet architecture is based on self-similarity and its structural layouts can be seen as truncated fractals. The fractal architecture is shown in Figure 6.2.3. In this diagram, C refers to the number of columns. The FractalNet structure simplifies its design, by reducing the number of decisions. An additional important contribution of this work is not the architecture but the training procedure called drop-path. Drop-path is an extension of dropout, which randomly disables the path between different layers while still guarantee at least one path enabled [159]. With drop-path, regularization of extremely deep fractal

networks is intuitive and effective. This mechanism encourages the development of individual columns as a subnetwork which can perform as well as the whole network.

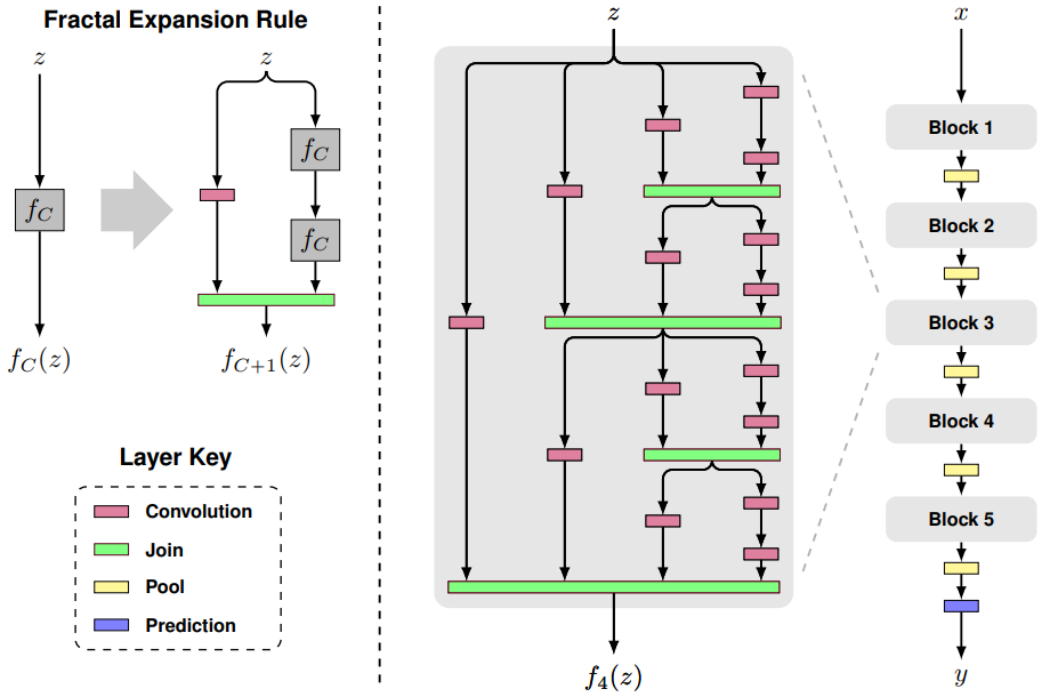


Figure 6.2.3: The fractal architecture of the proposed network (from [159] *)*

6.2.3 Other possibilities for future research

Feature visualization

Visualization and interpretation of network features has been a difficult but promising direction for researchers in deep learning. For instance, Baumgartner et al. proposed a Wasserstein Generative Adversarial Networks (WGAN) based method for visual attribution, where the generator is used to produce effect maps that can highlight the most affected regions for Alzheimer's disease. They evaluated the proposed method using ADNI 3D brain MRI data and have shown that this method can capture most category specific effects [160].

Recently, a new research direction named graph convolutional networks (GCNs) has attracted wide attention, which is an extension of CNN in graph domain [161] [162]. The main idea behind GCNs is to learn how to iteratively aggregate feature information from local

graph neighbourhoods using neural networks. The architecture of GCNs with first order filters is shown in Figure 6.2.4.

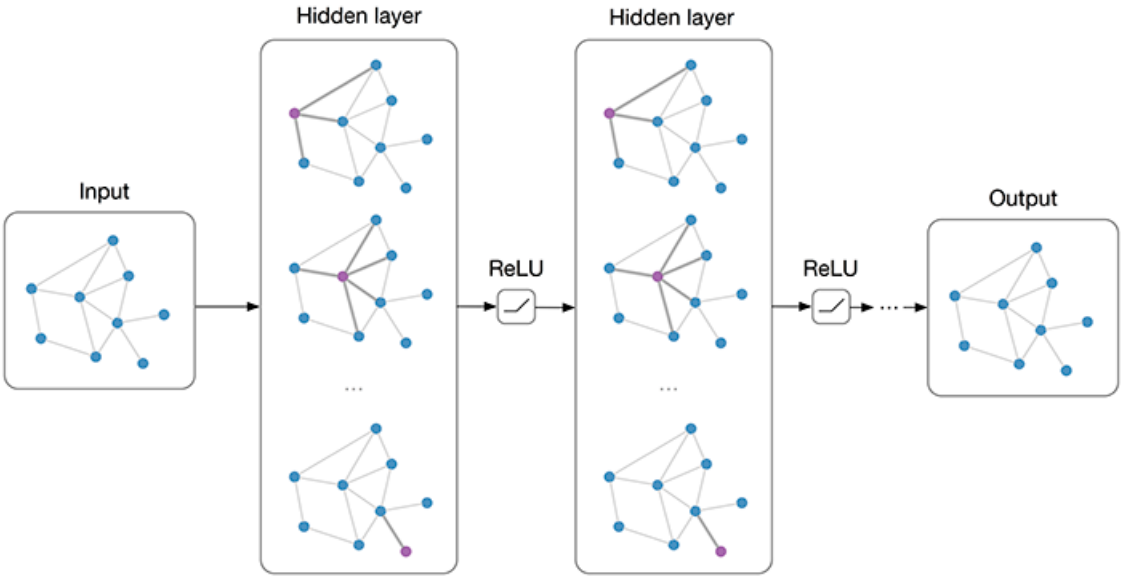


Figure 6.2.4: Multi-layer Graph Convolutional Network (GCN) with first-order filters (from[163]*)*

In this case, two possible extensions of this thesis could include 1) visualization and interpretation of 3D ResNet features 2) development of deep learning techniques for analysis of graphs (such as MRI cortical mesh data) using graph convolutions.

Pseudo Healthy Synthesis

Another possible extension is to generate pseudo healthy synthesis from diseased patients. Generating such images can be valuable both in research and clinical applications. For instance, Yang et al. [164] used a Variational-auto-encoder (VAEs) to learn a mapping from diseased images to pseudo healthy images to improve atlas-to-image registration accuracy. Afterwards, Sun et al. [165] developed a CycleGAN based method to perform pseudo healthy synthesis to aid experts with additional diagnostic information.

AD classification and MCI prediction using machine learning and deep learning algorithms is a dynamic field and in this section we proposed a series of deep learning alternatives which can be applied to our dataset. The new tracers such as Tau imaging also play an important role and will possibly lead to a new stage for AD prediction tasks. In addition, the continuation of the ADNI study will mean that further clinical follow-up information will be available.

Reference

[1] S. Todd, P. Passmore, Alzheimer's Disease – The Importance of Early Detection, Eur. Neurol. Rev. 2008;3(2):18-21 https://doi.org/10.17925/ENR.2008.03.02.18.

[2] Rathore, Saima et al. "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages." NeuroImage vol. 155 (2017): 530-548.

[3] Scatliff, James H, and Peter J Morris. "From Roentgen to magnetic resonance imaging: the history of medical imaging." North Carolina medical journal vol. 75,2 (2014): 111-3.

[4] A El-Gamal, Fatma El-Zahraa et al. "Medical imaging diagnosis of early Alzheimer's disease." Frontiers in bioscience (Landmark edition) vol. 23 671-725. 1 Jan. 2018, doi:10.2741/4612

[5] Erickson, Bradley J et al. "Machine Learning for Medical Imaging." Radiographics : a review publication of the Radiological Society of North America, Inc vol. 37,2 (2017): 505-515.

[6] M.I. Razzak, S. Naz, A. Zaib, Deep learning for medical image processing: Overview, challenges and the future, Lect. Notes Comput. Vis. Biomech. 26 (2018) 323–350. https://doi.org/10.1007/978-3-319-65981-7_12.

[7] McKhann, Guy M., et al. "The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging‐Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease." Alzheimer's & dementia 7.3 (2011): 263-269.

[8] Albert, Marilyn S., et al. "The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease." Focus 11.1 (2013): 96-106.

[9] Anand, R., Kiran Dip Gill, and Abbas Ali Mahdi. "Therapeutics of Alzheimer's disease: Past, present and future." Neuropharmacology 76 (2014): 27-50. https://doi.org/10.1016/j.neuropharm.2013.07.004.

[10] A. Association, others, 2017 Alzheimer's disease facts and figures, Alzheimer's Dement. 13 (2017) 325–373.

[11] M.G. M. Knapp, R. M. McCrone, P. Prina, M. Comas-Herrera, A. Wittenberg, D.S. Adelaja, B. Hu, B. King, D. Rehill, Dementia UK: Update, 2014.

[12] A.M. Clarfield, The decreasing prevalence of reversible dementias: An updated meta-analysis, Arch. Intern. Med. 163 (2003) 2219–2229. https://doi.org/10.1001/archinte.163.18.2219.

[13] D.P. Perl, Neuropathology of Alzheimer's Disease, Mt Sinai J Med. 77 (2010) 32–42. https://doi.org/10.1002/msj.20157.Neuropathology.

[14] G.S. Bloom, Amyloid-beta and tau: The trigger and bullet in Alzheimer disease pathogenesis, JAMA Neurol. 71 (2014) 505–508. https://doi.org/10.1001/jamaneurol.2013.5847.

[15] C.L. Masters, R. Bateman, K. Blennow, C.C. Rowe, R.A. Sperling, J.L. Cummings, Alzheimer's disease, Nat. Rev. Dis. Prim. 1 (2015) 150-156. http://dx.doi.org/10.1038/nrdp.2015.56%0Ahttp://10.0.4.14/nrdp.2015.56.

[16] O. Preische, S.A. Schultz, A. Apel, J. Kuhle, S.A. Kaeser, C. Barro, S. Gräber, E. Kuder-Buletta, C. LaFougere, C. Laske, others, Serum neurofilament dynamics

predicts neurodegeneration and clinical progression in presymptomatic Alzheimer's disease, Nat. Med. 25 (2019) 277–283.

[17] Bekris, Lynn M., et al. "Genetics of Alzheimer disease." Journal of geriatric psychiatry and neurology 23.4 (2010): 213-227.

[18] Liu, Chia-Chen, et al. "Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy." Nature Reviews Neurology 9.2 (2013): 106.

[19] Ferreira, Luiz Kobuti, and Geraldo F. Busatto. "Neuroimaging in Alzheimer's disease: current role in clinical practice and potential future applications." Clinics 66 (2011): 19-24.

[20] Varghese, Tinu et al. "A review of neuroimaging biomarkers of Alzheimer's disease." Neurology Asia vol. 18,3 (2013): 239-248.

[21] Guptha, Sunku H., Eric Holroyd, and Gordon Campbell. "Progressive lateral ventricular enlargement as a clue to Alzheimer's disease." The Lancet 359.9322 (2002): 2040.

[22] Johnson, Keith A et al. "Brain imaging in Alzheimer disease." Cold Spring Harbor perspectives in medicine vol. 2,4 (2012): a006213. doi:10.1101/cshperspect.a006213

[23] P.M. Frisoni, Giovanni B and Fox, Nick C and Jack, Clifford R and Scheltens, Philip and Thompson, The clinical use of structural MRI in Alzheimer disease, Nat. Rev. Neurol. 6 (2010) 67–77.

[24] Risacher, Shannon L et al. "Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort." Neurobiology of aging vol. 31,8 (2010): 1401-18. doi:10.1016/j.neurobiolaging.2010.04.029

[25] López-Gómez, Carlos et al. “ALTEA: A Software Tool for the Evaluation of New Biomarkers for Alzheimer's Disease by Means of Textures Analysis on Magnetic Resonance Images.” Diagnostics (Basel, Switzerland) vol. 8,3 47. 19 Jul. 2018

[26] Tripoliti, Evanthia E et al. “A six stage approach for the diagnosis of the Alzheimer's disease based on fMRI data.” Journal of biomedical informatics vol. 43,2 (2010): 307-20. doi:10.1016/j.jbi.2009.10.004

[27] R. Sperling, Functional MRI studies of associative encoding in normal aging, mild cognitive impairment, and Alzheimer’s disease, Ann. N. Y. Acad. Sci. 1097 (2007) 146–155.

[28] He, Yong et al. “Regional coherence changes in the early stages of Alzheimer's disease: a combined structural and resting-state functional MRI study.” NeuroImage vol. 35,2 (2007): 488-500. doi:10.1016/j.neuroimage.2006.11.042

[29] Damoiseaux, Jessica S. "Resting-state fMRI as a biomarker for Alzheimer's disease?." Alzheimer's research & therapy 4.2 (2012): 8.

[30] R. Sperling, The potential of functional MRI as a biomarker in early Alzheimer’s disease, Neurobiol. Aging. 32 (2011) S37–S43.

[31] Mosconi, Lisa, et al. "Pre-clinical detection of Alzheimer's disease using FDG-PET, with or without amyloid imaging." Journal of Alzheimer's Disease 20.3 (2010): 843-854.

[32] Mueller, Susanne G., et al. "Ways toward an early diagnosis in Alzheimer’s disease: the Alzheimer’s Disease Neuroimaging Initiative (ADNI)." Alzheimer's & Dementia 1.1 (2005): 55-66.

[33] Mosconi, L., and P. F. McHugh. "FDG-and amyloid-PET in Alzheimer's disease: Is the whole greater than the sum of the parts?." The Quarterly Journal of Nuclear Medicine and Molecular Imaging 55.3 (2011): 250.

[34] Marcus, Charles, Esther Mena, and Rathan M. Subramaniam. "Brain PET in the diagnosis of Alzheimer's disease." Clinical nuclear medicine 39.10 (2014): e413.

[35] Klunk, William E et al. "Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B." Annals of neurology vol. 55,3 (2004): 306-19. doi:10.1002/ana.20009

[36] FDA approves 18F-florbetapir PET agent, J. Nucl. Med. 53 (2012) 15N.

[37] V. Kepe, M.C. Moghbel, B. Långströmd, H. Zaidi, H. V. Vinters, S.C. Huang, N. Satyamurthy, D. Doudet, E. Mishani, R.M. Cohen, P.F. Høilund-Carlsen, A. Alavi, J.R. Barrio, Amyloid-β positron emission tomography imaging probes: A critical review, J. Alzheimer's Dis. 36 (2013) 613–631. https://doi.org/10.3233/JAD-130485.

[38] B. Stankoff, L. Freeman, M.S. Aigrot, A. Chardain, F. Dollé, A. Williams, D. Galanaud, L. Armand, S. Lehericy, C. Lubetzki, B. Zalc, M. Bottlaender, Imaging central nervous system myelin by positron emission tomography in multiple sclerosis using [methyl-11C]-2-(4-methylaminophenyl)- 6-hydroxybenzothiazole, Ann. Neurol. 69 (2011) 673–680. https://doi.org/10.1002/ana.22320.

[39] S.A. Lipton, Paradigm shift in NMDA receptor antagonist drug development: molecular mechanism of uncompetitive inhibition by memantine in the treatment of Alzheimer's disease and other neurologic disorders, J. Alzheimer's Dis. 6 (2004) S61–S74.

[40] Stella, Florindo et al. "Anti-dementia medications: current prescriptions in clinical practice and new agents in progress." Therapeutic advances in drug safety vol. 6,4 (2015): 151-65. doi:10.1177/2042098615592116

[41] Jeffrey Cummings, Garam Lee , Travis Mortsdorf , Aaron Ritter, Kate Zhong Alzheimer's disease drug development pipeline: 2017, Alzheimer's Dement. Transl. Res. Clin. Interv. 3 (2017) 367–384.

[42] Grutzendler, J, and J C Morris. "Cholinesterase inhibitors for Alzheimer's disease." Drugs vol. 61,1 (2001): 41-52.

[43] Schneider, Lon S et al. "Lack of evidence for the efficacy of memantine in mild Alzheimer disease." Archives of neurology vol. 68,8 (2011): 991-8. doi:10.1001/archneurol.2011.69

[44] Pereira, A C et al. "Age and Alzheimer's disease gene expression profiles reversed by the glutamate modulator riluzole." Molecular psychiatry vol. 22,2 (2017): 296-305. doi:10.1038/mp.2016.33

[45] Wang, Rui, and P Hemachandra Reddy. "Role of Glutamate and NMDA Receptors in Alzheimer's Disease." Journal of Alzheimer's disease : JAD vol. 57,4 (2017): 1041-1048. doi:10.3233/JAD-160763

[46] J.L. Cummings, K. Zhong, Chapter 17 - Symptomatic Cognitive Enhancing Agents, in: M.S. Wolfe (Ed.), Dev. Ther. Alzheimer's Dis., Academic Press, Boston, 2016: pp. 459–475. https://doi.org/https://doi.org/10.1016/B978-0-12-802173-6.00017-4.

[47] Lopresti, Brian J et al. "Simplified quantification of Pittsburgh Compound B amyloid imaging PET studies: a comparative analysis." Journal of nuclear medicine : official publication, Society of Nuclear Medicine vol. 46,12 (2005): 1959-72.

[48] J.A. Thie, Understanding the standardized uptake value, its methods, and implications for usage, J. Nucl. Med. 45 (2004) 1431–1434.

[49] Zasadny, K R, and R L Wahl. "Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a

method for correction.” Radiology vol. 189,3 (1993): 847-50. doi:10.1148/radiology.189.3.8234714

[50] Joshi, Abhinay D et al. “Performance characteristics of amyloid PET with florbetapir F 18 in patients with alzheimer's disease and cognitively normal subjects.” Journal of nuclear medicine : official publication, Society of Nuclear Medicine vol. 53,3 (2012): 378-84. doi:10.2967/jnumed.111.090340

[51] Barthel, Henryk et al. “Cerebral amyloid-β PET with florbetaben (18F) in patients with Alzheimer's disease and healthy controls: a multicentre phase 2 diagnostic study.” The Lancet. Neurology vol. 10,5 (2011): 424-35. doi:10.1016/S1474-4422(11)70077-1

[52] Blautzik, Janusch et al. “Reference region selection and the association between the rate of amyloid accumulation over time and the baseline amyloid burden.” European journal of nuclear medicine and molecular imaging vol. 44,8 (2017): 1364-1374. doi:10.1007/s00259-017-3666-8

[53] V.W. Pike, PET radiotracers: crossing the blood--brain barrier and surviving metabolism, Trends Pharmacol. Sci. 30 (2009) 431–440.

[54] Pontecorvo, Michael, et al. "Can incorporation of a quantitative estimate of cortical to cerebellar SUVr as an adjunct to visual interpretation improve the accuracy and reliability of florbetapir PET scan interpretation?." Journal of Nuclear Medicine 55.supplement 1 (2014): 245-245.

[55] Wong, Dean F., et al. "In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (flobetapir F 18)." Journal of Nuclear Medicine 51.6 (2010): 913-920.

[56] Sabri, Osama, et al. "Florbetaben PET imaging to detect amyloid beta plaques in Alzheimer's disease: phase 3 study." Alzheimer's & Dementia 11.8 (2015): 964-974.

[57] de Bruijne, Marleen. “Machine learning approaches in medical image analysis: From detection to diagnosis.” Medical image analysis vol. 33 (2016): 94-97.

[58] T. Lindeberg, Scale Invariant Feature Transform, Scholarpedia. 7 (2012) 10491. https://doi.org/10.4249/scholarpedia.10491.

[59] Gómez, Carlos, and Roberto Hornero. "Entropy and complexity analyses in Alzheimer’s disease: an MEG study." The open biomedical engineering journal 4 (2010): 223.

[60] McBride, Joseph, et al. "EEG multiscale entropy dynamics in mild cognitive impairment and early Alzheimer's disease." Proceedings of the 2014 Biomedical Sciences and Engineering Conference. IEEE, 2014.

[61] Mandelbrot, Benoit B. The fractal geometry of nature. Vol. 173. New York: WH freeman, 1983.

[62] Bullmore, Edward, et al. "Fractal analysis of the boundary between white matter and cerebral cortex in magnetic resonance images: a controlled study of schizophrenic and manic-depressive patients." Psychological medicine 24.3 (1994): 771-781.

[63] King, Richard D., et al. "Fractal dimension analysis of the cortical ribbon in mild Alzheimer's disease." Neuroimage 53.2 (2010): 471-479.

[64] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423. https://doi.org/10.1145/584091.584093.

[65] Rényi, Alfréd. "On measures of entropy and information." Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics. The Regents of the University of California, 1961.

[66] Aczél, János, Bruno Forte, and Che Tat Ng. "Why the Shannon and Hartley entropies are ‘natural’." Advances in applied probability 6.1 (1974): 131-146.

[67] R. S. Varma, Generalization of Renyi's entropy of order α, J. Math. Sci. 1 (1966) 34–48.

[68] Delic J, Alhilali LM, Hughes MA, Gumus S, Fakhran S. White Matter Injuries in Mild Traumatic Brain Injury and Posttraumatic Migraines: Diffusion Entropy Analysis. Radiology. 2016;279(3):859 - 866. doi:10.1148/radiol.2015151388

[69] Young, Karl, and Norbert Schuff. "Measuring structural complexity in brain images." NeuroImage vol. 39,4 (2008): 1721-30. doi:10.1016/j.neuroimage.2007.10.043

[70] Z. Camlica, H.R. Tizhoosh, F. Khalvati, Medical image classification via SVM using LBP features from saliency-based folded data, Proc. - 2015 IEEE 14th Int. Conf. Mach. Learn. Appl. ICMLA 2015. (2016) 128–132. https://doi.org/10.1109/ICMLA.2015.131.

[71] Cristianini, Nello, and John Shawe-Taylor. An introduction to support vector machines and other kernel-based learning methods. Cambridge university press, 2000.

[72] Magnin, Benoît, et al. "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI." Neuroradiology 51.2 (2009): 73-83.

[73] Katako, Audrey et al. "Machine learning identified an Alzheimer's disease-related FDG-PET pattern which is also expressed in Lewy body dementia and Parkinson's disease dementia." Scientific reports vol. 8,1 13236. 5 Sep. 2018, doi:10.1038/s41598-018-31653-6

[74] Tang, An et al. "Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology." Canadian Association of Radiologists journal = Journal l'Association canadienne des radiologistes vol. 69,2 (2018): 120-135. doi:10.1016/j.carj.2018.02.002

[75] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297. https://doi.org/10.1007/BF00994018.

[76] H. Drucker, D. Wu, V.N. Vapnik, Support vector machines for spam categorization, IEEE Trans. Neural Networks. 10 (1999) 1048–1054.

[77] Dechter, Rina. Learning while searching in constraint-satisfaction problems. University of California, Computer Science Department, Cognitive Systems Laboratory, 1986.

[78] Aizenberg, Igor, Naum N. Aizenberg, and Joos PL Vandewalle. Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications. Springer Science & Business Media, 2013.

[79] D. Ciregan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2012) 3642–3649. https://doi.org/10.1109/CVPR.2012.6248110.

[80] A. Krizhevsky, I. Sutskever, G.E. Hinton., Imagenet classification with deep convolutional neural networks. In Advances in neural information, Adv. Neural Inf. Process. Syst. (2012) 1097–1105. https://doi.org/10.1016/B978-008046518-0.00119-7.

[81] Lo, S-CB, et al. "Artificial convolution neural network techniques and applications for lung nodule detection." IEEE transactions on medical imaging 14.4 (1995): 711-718.

[82] F. Agostinelli, M. Hoffman, P. Sadowski, P. Baldi, Learning activation functions to improve deep neural networks, ArXiv Prepr. ArXiv1412.6830. (2014).

[83] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, ArXiv Prepr. ArXiv1412.6980. (2014).

[84] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A. A. Bharath, "Generative Adversarial Networks: An Overview," in IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 53-65, Jan. 2018, doi: 10.1109/MSP.2017.2765202.

[85] A. Radford, L. Metz, S. Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, ArXiv. (2015). https://doi.org/10.1051/0004-6361/201527329.

[86] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. InAdvances in neural information processing systems 2014 (pp. 2672-2680).

[87] P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632.

[88] Sabuncu, Mert R., Ender Konukoglu, and Alzheimer's Disease Neuroimaging Initiative. "Clinical prediction from structural brain MRI scans: a large-scale empirical study." Neuroinformatics 13.1 (2015): 31-46.

[89] Dickerson, B. C., et al. "Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults." Neurology 76.16 (2011): 1395-1402.

[90] Dickerson, Bradford C., et al. "The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals." Cerebral cortex 19.3 (2009): 497-510.

[91] Li, Shuyu, et al. "Abnormal changes of multidimensional surface features using multivariate pattern classification in amnestic mild cognitive impairment patients." Journal of Neuroscience 34.32 (2014): 10541-10553.

[92] Baron, J. C., et al. "In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease." Neuroimage 14.2 (2001): 298-309.

[93] Braak, Heiko, and Eva Braak. "Neuropathological stageing of Alzheimer-related changes." Acta neuropathologica 82.4 (1991): 239-259.

[94] Mu, Yangling, and Fred H. Gage. "Adult hippocampal neurogenesis and its role in Alzheimer's disease." Molecular neurodegeneration 6.1 (2011): 85.

[95] Sørensen, Lauge et al. “Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry.” NeuroImage. Clinical vol. 13 470-482. 7 Dec. 2016, doi:10.1016/j.nicl.2016.11.025

[96] J. Pacák, Z. Toík, M. Erný, Synthesis of 2-deoxy-2-fluoro-D-glucose, J. Chem. Soc. D Chem. Commun. 77-77 (1969). https://doi.org/10.1039/C29690000077.

[97] C. Cabral, M. Silveira, Classification of Alzheimer’s disease from FDG-PET images using favourite class ensembles, in: Eng. Med. Biol. Soc. (EMBC), 2013 35th Annu. Int. Conf. IEEE, 2013: pp. 2477–2480.

[98] Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Recent publications from the Alzheimer’s Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials., Alzheimer’s Dement. J. Alzheimer’s Assoc. 13 (2017) 1–85.

[99] Vandenberghe, Rik, et al. "Binary classification of 18F-flutemetamol PET using machine learning: comparison with visual reads and structural MRI." NeuroImage 64 (2013): 517-525.

[100] B. Cheng, M. Liu, H.-I. Suk, D. Shen, D. Zhang, Multimodal manifold-regularized transfer learning for MCI conversion prediction., Brain Imaging Behav. 9 (2015) 913–926. https://doi.org/10.1007/s11682-015-9356-x.

[101] Zhang, Daoqiang, et al. "Multimodal classification of Alzheimer's disease and mild cognitive impairment." Neuroimage 55.3 (2011): 856-867.

[102] Hosseini-Asl, Ehsan et al. “Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network.” Frontiers in bioscience (Landmark edition) vol. 23 584-596. 1 Jan. 2018, doi:10.2741/4606

[103] Payan, Adrien, and Giovanni Montana. "Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks." arXiv preprint arXiv:1502.02506 (2015).

[104] B. Shi, Y. Chen, P. Zhang, C.D. Smith, J. Liu, Erratum to “Nonlinear feature transformation and deep fusion for Alzheimer’s disease staging analysis” [Pattern Recognition 63 (2017) 487–498] (S0031320316302916)(10.1016/j.patcog.2016.09.032), Pattern Recognit. 66 (2017) 447. https://doi.org/10.1016/j.patcog.2017.01.004.

[105] S. Singh, A. Srivastava, L. Mi, R.J. Caselli, K. Chen, D. Goradia, E.M. Reiman, Y. Wang, Deep-learning-based classification of FDG-PET data for Alzheimer’s disease categories, in: 13th Int. Conf. Med. Inf. Process. Anal., 2017: p. 105720J.

[106] Ding, Yiming et al. “A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain.” Radiology vol. 290,2 (2019): 456-464.

[107] Liu, Manhua et al. “Classification of Alzheimer's Disease by Combination of Convolutional and Recurrent Neural Networks Using FDG-PET Images.” Frontiers in neuroinformatics vol. 12 35. 19 Jun. 2018, doi:10.3389/fninf.2018.00035

[108] H. Choi, K.H. Jin, A.D.N. Initiative, others, Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging, Behav. Brain Res. 344 (2018) 103–109.

[109] D. Lu, K. Popuri, W. Ding, R. Balachandar, M.F. Beg, Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer’s Disease using structural MR and FDG-PET images, (2017) 1–13. https://doi.org/10.1038/s41598-018-22871-z.

[110] H.-I. Suk, S.-W. Lee, D. Shen, T.A.D.N. Initiative, Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis, Brain Struct. Funct. 221 (2016) 2569–2587. https://doi.org/10.1007/s00429-015-1059-y.

[111] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M.J. Fulham, Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease, IEEE Trans. Biomed. Eng. 62 (2015) 1132–1140. https://doi.org/10.1109/TBME.2014.2372011.

[112] S. Korolev, A. Safiullin, M. Belyaev, Y. Dodonova, Residual and plain convolutional neural networks for 3D brain MRI classification, in: Proc. - Int. Symp. Biomed. Imaging, 2017. https://doi.org/10.1109/ISBI.2017.7950647.

[113] A. Khvostikov, K. Aderghal, J. Benois-Pineau, A. Krylov, G. Catheline, 3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies, ArXiv Prepr. ArXiv1801.05968. (2018).

[114] M. Tang, H.N. Wang, Feature analysis of brain MRI images based on fractal dimension, in: Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc., 2005. https://doi.org/10.1109/iembs.2005.1617168.

[115] Reishofer, Gernot et al. "Age is reflected in the Fractal Dimensionality of MRI Diffusion Based Tractography." Scientific reports vol. 8,1 5431. 3 Apr. 2018, doi:10.1038/s41598-018-23769-6

[116] M.W. Weiner, D.P. Veitch, P.S. Aisen, L.A. Beckett, N.J. Cairns, J. Cedarbaum, R.C. Green, D. Harvey, C.R. Jack, W. Jagust, J. Luthman, J.C. Morris, R.C. Petersen, A.J. Saykin, L. Shaw, L. Shen, A. Schwarz, A.W. Toga, J.Q. Trojanowski, 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception, Alzheimer's Dement. 11 (2015) e1–e120. https://doi.org/10.1016/j.jalz.2014.11.001.

[117] Medina-Carnicer, R., et al. "Determining hysteresis thresholds for edge detection by combining the advantages and disadvantages of thresholding methods." IEEE transactions on image processing 19.1 (2009): 165-173.

[118] Penny, William D., et al., eds. Statistical parametric mapping: the analysis of functional brain images. Elsevier, 2011.

[119] M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images, Neuroimage. 17 (2002) 825–841. https://doi.org/10.1016/S1053-8119(02)91132-8.

[120] Dice, Lee R. "Measures of the amount of ecologic association between species." Ecology 26.3 (1945): 297-302.

[121] C.R. Madan, E.A. Kensinger, Cortical complexity as a measure of age-related brain atrophy, Neuroimage. 134 (2016) 617–629. https://doi.org/10.1016/j.neuroimage.2016.04.029.

[122] I. Bhakta, S. Phadikar, Evaluation of Entropy-Based Segmentation Techniques for Automated Skin Disease Detection, in: R. Singh, S. Choudhury (Eds.), Proceeding Int. Conf. Intell. Commun. Control Devices, Springer Singapore, Singapore, 2017: pp. 335–341.

[123] C. Chang, C. Lin, LIBSVM : A Library for Support Vector Machines, ACM Trans. Intell. Syst. Technol. 2 (2013) 1–39. https://doi.org/10.1145/1961189.1961199.

[124] A.M. Dale, B. Fischl, M.I. Sereno, Cortical surface-based analysis: I. Segmentation and surface reconstruction, Neuroimage. 9 (1999) 179–194. https://doi.org/10.1006/nimg.1998.0395.

[125] recon-all, https://surfer.nmr.mgh.harvard.edu/fswiki/recon-all (accessed April 4, 2020).

[126] B. Fischl, A.M. Dale, Measuring the thickness of the human cerebral cortex from magnetic resonance images, Proc. Natl. Acad. Sci. 97 (2000) 11050–11055. https://doi.org/10.1073/pnas.200033797.

[127] C. Lange, P. Suppa, L. Frings, W. Brenner, L. Spies, R. Buchert, Optimization of statistical single subject analysis of brain FDG PET for the prognosis of mild cognitive impairment-to-Alzheimer's disease conversion, J. Alzheimer's Dis. 49 (2015) 945–959. https://doi.org/10.3233/JAD-150814.

[128] R. Vandenberghe, N. Nelissen, E. Salmon, A. Ivanoiu, S. Hasselbalch, A. Andersen, A. Korner, L. Minthon, D.J. Brooks, K. Van Laere, P. Dupont, Binary classification of18F-flutemetamol PET using machine learning: Comparison with visual reads and structural MRI, Neuroimage. 64 (2013) 517–525. https://doi.org/10.1016/j.neuroimage.2012.09.015.

[129] C. Zu, B. Jie, M. Liu, S. Chen, D. Shen, D. Zhang, the Alzheimer's Disease Neuroimaging Initiative, Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment, Brain Imaging Behav. 10 (2016) 1148–1159. https://doi.org/10.1007/s11682-015-9480-7.

[130] K. Anand, M. Sabbagh, Amyloid imaging: poised for integration into medical practice, Neurotherapeutics. 14 (2017) 54–61.

[131] S.G. Mueller, M.W. Weiner, L.J. Thal, R.C. Petersen, C. Jack, W. Jagust, J.Q. Trojanowski, A.W. Toga, L. Beckett, The Alzheimer's disease neuroimaging initiative, Neuroimaging Clin. 15 (2005) 869–877.

[132] Y. LeCun, Y. Bengio, Convolutional networks for images, speech, and time series, Handb. Brain Theory Neural Networks. 3361 (1995) 255–258. https://doi.org/10.1109/IJCNN.2004.1381049.

[133] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[134] Q. Li, W. Cai, X. Wang, Y. Zhou, D.D. Feng, M. Chen, Medical image classification with convolutional neural network, in: Control Autom. Robot. Vis. (ICARCV), 2014 13th Int. Conf., 2014: pp. 844–848.

[135] A. Kumar, J. Kim, D. Lyndon, M. Fulham, D. Feng, An ensemble of fine-tuned convolutional neural networks for medical image classification, IEEE J. Biomed. Heal. Informatics. 21 (2017) 31–40.

[136] Maioli, F., et al. "Conversion of mild cognitive impairment to dementia in elderly subjects: a preliminary study in a memory and cognitive disorder unit." Archives of gerontology and geriatrics 44 (2007): 233-241.

[137] M. Ganguli, Y. Jia, T.F. Hughes, B.E. Snitz, C.-C.H. Chang, S.B. Berman, K.J. Sullivan, M.I. Kamboh, Mild Cognitive Impairment that Does Not Progress to Dementia: A Population-Based Study., J. Am. Geriatr. Soc. 67 (2019) 232–238. https://doi.org/10.1111/jgs.15642.

[138] Gao, Qi et al. “Mild Cognitive Impairment Reversion and Progression: Rates and Predictors in Community-Living Older Persons in the Singapore Longitudinal Ageing Studies Cohort.” Dementia and geriatric cognitive disorders extra vol. 8,2 226-237. 19 Jun. 2018, doi:10.1159/000488936

[139] Matthews, Fiona E., et al. "Two-year progression from mild cognitive impairment to dementia: to what extent do different definitions agree?." Journal of the American Geriatrics Society 56.8 (2008): 1424-1433.

[140] J. Young, M. Modat, M.J. Cardoso, A. Mendelson, D. Cash, S. Ourselin, A.D.N. Initiative, others, Accurate multimodal probabilistic prediction of conversion to

Alzheimer's disease in patients with mild cognitive impairment, NeuroImage Clin. 2 (2013) 735–745.

[141] X. Zhu, H.-I. Suk, L. Wang, S.-W. Lee, D. Shen, A.D.N. Initiative, others, A novel relational regularization feature selection method for joint regression and classification in AD diagnosis, Med. Image Anal. 38 (2017) 205–214. DOI:10.1016/j.media.2015.10.008

[142] H.-I. Suk, S.-W. Lee, D. Shen, A.D.N. Initiative, others, Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis, Neuroimage. 101 (2014) 569–582.

[143] Korolev, Igor O et al. "Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification." PloS one vol. 11,2 e0138866. 22 Feb. 2016,

[144] H. Choi, D.S. Lee, Generation of Structural MR Images from Amyloid PET: Application to MR-Less Quantification, J. Nucl. Med. 59 (2018) 1111–1117. https://doi.org/10.2967/jnumed.117.199414.

[145] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification, (2018) 1–10. http://arxiv.org/abs/1803.01229.

[146] Ali Madani, Mehdi Moradi, Alexandros Karargyris, and Tanveer Syeda-Mahmood "Chest x-ray generation and data augmentation for cardiovascular abnormality classification", Proc. SPIE 10574, Medical Imaging 2018: Image Processing, 105741M (2018); https://doi.org/10.1117/12.2293971

[147] O. Ronneberger, P. Fischer, B. Thomas, U-Net: Convolutional Networks for Biomedical Image Segmentation, Med. Image Comput. Comput. Interv. 9351 (2015) 234–241. http://arxiv.org/abs/1505.04597.

[148] He, Kaiming, et al. "Deep residual learning for image recognition. CoRR abs/1512.03385 (2015)." (2015): 646-661.

[149] Chen, Hao et al. “VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images.” NeuroImage vol. 170 (2018): 446-455. doi:10.1016/j.neuroimage.2017.04.041

[150] C. Gulcehre, S. Chandar, K. Cho, Y. Bengio, Identity Mappings in Deep Residual Networks, Neural Comput. 30 (2018) 857–884. https://doi.org/10.1162/NECO_a_01060.

[151] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612. https://doi.org/10.1109/TIP.2003.819861.

[152] J. Islam, Y. Zhang, GAN-based synthetic brain PET image generation, Brain Informatics. 7 (2020) 1–12. DOI: 10.1186/s40708-020-00104-2

[153] Pan Y., Liu M., Lian C., Zhou T., Xia Y., Shen D. (2018) Synthesizing Missing PET from MRI with Cycle-consistent Generative Adversarial Networks for Alzheimer’s Disease Diagnosis. In: Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, vol 11072. Springer, Cham

[154] N. Okamura, R. Harada, A. Ishiki, A. Kikuchi, T. Nakamura, Y. Kudo, The development and validation of tau PET tracers: current status and future directions, Clin. Transl. Imaging. 6 (2018) 305–316.

[155] L. Saint-Aubert, L. Lemoine, K. Chiotis, A. Leuzy, E. Rodriguez-Vieitez, A. Nordberg, Tau PET imaging: present and future directions, Mol. Neurodegener. 12 (2017) 19. DOI: 10.1186/s13024-017-0162-3

[156] V.L. Villemagne, M.T. Fodero-Tavoletti, C.L. Masters, C.C. Rowe, Tau imaging: early progress and future directions, Lancet Neurol. 14 (2015) 114–124. DOI: 10.1016/S1474-4422(14)70252-2

[157] Clifford R Jack, Jr, Heather J Wiste, Christopher G Schwarz, Val J Lowe, Matthew L Senjem, Prashanthi Vemuri, Stephen D Weigand, Terry M Therneau, Dave S Knopman, Jonathan Graff-Radford, Kejal Kantarci, Rosebud O Roberts, Michelle M Mielke, Mary M Machulda, Ronald C Petersen, Longitudinal tau PET in ageing and Alzheimer's disease, Brain, Volume 141, Issue 5, May 2018, 1517–1528, https://doi.org/10.1093/brain/awy059

[158] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017: pp. 4700–4708.

[159] G. Larsson, M. Maire, G. Shakhnarovich, Fractalnet: Ultra-deep neural networks without residuals, ArXiv Prepr. ArXiv1605.07648. (2016).

[160] C.F. Baumgartner, L.M. Koch, K. Can Tezcan, J. Xi Ang, E. Konukoglu, Visual feature attribution using wasserstein gans, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018: pp. 8309–8319.

[161] Y. Zhang, S. Pal, M. Coates, D. Üstebay, Bayesian graph convolutional neural networks for semi-supervised classification, ArXiv Prepr. ArXiv1811.11103. (2018).

[162] Yao, Liang, Chengsheng Mao, and Yuan Luo. "Graph convolutional networks for text classification." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. 2019.

[163] T. KIPF, GRAPH CONVOLUTIONAL NETWORKS, https://tkipf.github.io/graph-convolutional-networks/.

[164] Yang X., Han X., Park E., Aylward S., Kwitt R., Niethammer M. (2016) Registration of Pathological Images. In: Tsaftaris S., Gooya A., Frangi A., Prince J. (eds) Simulation and Synthesis in Medical Imaging. SASHIMI 2016. Lecture Notes in Computer Science, vol 9968. Springer, Cham

[165] L. Sun, J. Wang, X. Ding, Y. Huang, J. Paisley, An Adversarial Learning Approach to Medical Image Synthesis for Lesion Removal, ArXiv Prepr. ArXiv1810.10850. (2018).