Application of graph convolutional neural networks to Alzheimer's and Parkinson's disease classification

Goli, Haveesh

2021

Goli, H. (2021). Application of graph convolutional neural networks to Alzheimer's and Parkinson's disease classification. Master's thesis, Nanyang Technological University, Singapore. <https://hdl.handle.net/10356/148634>

<https://hdl.handle.net/10356/148634>

<https://doi.org/10.32657/10356/148634>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 19:11:20 SGT

Application of Graph Convolutional Neural Networks to Alzheimer's and Parkinson's disease classification

by

GOLI HAVEESH

G1802705C

School of Computer Science and Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Master of Engineering

January 01, 2021

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

01/01/2021

.....

Date

.....

Goli Haveesh

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

01/01/2021

.....

Date

.....

Prof. Jagath C. Rajapakse

Authorship Attribution Statement

*(A) This thesis **does not** contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author.

01/01/2021

.....

Date

.....

Goli Haveesh

Acknowledgement

I take this opportunity to express my gratitude to everyone who supported in my Master of Engineering course. I am grateful to them for their omnipresent guidance, constructive criticism, and invaluable advice during the course of my work.

Firstly, I would like to express sincere gratitude to my supervisor Professor Jagath Chandana Rajapakse for providing me the opportunity to pursue Masters degree in this prestigious university. I thank him for giving me this opportunity to pursue research in as meaningful and interesting topic as brain imaging. His continuous support, patience and immense knowledge kept me motivated throughout the course. His vocational advice helped me in all the time of research and future career matters.

I would also like to extend my gratitude to Dr. Satya P. Singh and Dr. Sukrit Gupta for sharing their knowledge and expertise with me in the concepts of brain imaging. I thank Dr. Satya for his patience in clarifying my conceptual doubts and guiding me through the techniques that, initially, were totally foreign to me.

Finally, I would like to thank my parents who supported me throughout the course and encouraged me in every manner possible. Without their support, all my efforts would have been futile. My time here at NTU was one that I shall remember and cherish for a long time and would like to thank everyone who made it special.

Contents

Chapter 1: Introduction	11
1.1 Software and Environment	14
1.2 Report organization	14
Chapter 2: Background	14
2.1 Data	16
2.1.1 Structural Magnetic Resonance Imaging (sMRI)	16
2.1.2 Diffusion Tensor Imaging (DTI)	19
2.1.3 Previous work on combining DTI and sMRI scans	23
2.1.4 Transcriptome data	24
2.2 Deep Neural Networks (DNN)	26
2.2.1 Feed Forward Neural Networks (FFN)	26
2.2.2 Convolutional Neural Networks (CNN)	27
2.2.3 Graph Convolutional Neural Networks (GCN)	28
2.2.4 Previous work on application of Machine Learning for Brain Connectome Analysis	29
Chapter 3: Data preparation and pre-processing	34
3.1 Alzheimer's Disease Neuroimaging Initiative (ADNI)	34
3.2 Data pre-processing in DTI	35
3.3 Data processing for DTI scans	36
3.4 Data processing for sMRI scans	41
3.5 Data processing for gene expression data	44
Chapter 4: Ensemble learning of Population based Graph Convolutional Networks (pGCN)	45
4.1 Population graph	45
4.2 Population graph Convolutional Network (pGCN)	45
4.3 Ensemble learning	47
4.4 Experiments and Results	50
4.5 Discussion	54
Chapter 5: Combining of DTI and sMRI scans	55
5.1 DTI features for fusion	56
5.2 sMRI features for fusion	56
5.3 Fusion of features	57
5.4 Experiments and Results	57
5.5 Discussion	59

Chapter 6: Population graphs from transcriptome data	61
6.1 Population graphs for gene expression data	62
6.2 Experiments and Results	63
6.3 Discussion	65
Chapter 7: Conclusion	68
7.1 Summary	68
7.2 Discussion	69
7.3 Future work	70
Bibliography	72

Summary

Identifying connectivity patterns of the human structural connectome plays an important role in diagnosing various neurodegenerative conditions including Alzheimer’s disease (AD) and Parkinson’s disease (PD). The structural connectome is measured by diffusion weighted imaging (DTI) scans and is represented by a graph where the nodes are brain regions of interests and the edges represent the strengths of white matter connections between the regions. Deep neural networks (DNNs) models have been explored for classification of brain diseases by encoding the structural connectome. However, feedforward deep neural networks models operate on vectorized connectivity patterns and convolutional neural networks operate on regular structures of data. Therefore, DNN models have been unable to capitalize on the graphical structure inherent in the connectome data.

In this thesis, we introduce graph convolutional neural network (GCN) to encode the structural connectome in order to classify brain scans. GCNs have been successful in modelling graph-structured data in many applications. In this work, we introduce population based GCN (pGCN) for disease classification using brain connectivity features obtained from structural MRI (sMRI) and diffusion tensor imaging (DTI) brain scans. Information of patient samples are represented in a population graph in pGCN.

Structural connectome features are extracted from DTI scans and grey matter anatomical features are extracted from sMRI images, which were then combined and processed in a pGCN. We show that the fusion of sMRI and DTI to supplement each other for feature representation and improves classification performance. Furthermore, we propose an ensemble learning strategy for training multiple instances of pGCN as random graph embeddings and demonstrate improved classification performance. Ensemble learning increases the diversity of data available for training and simplifies the choices for graph construction.

In order to incorporate genomic features in disease classification, we demonstrate how population graph can be derived from gene expression data. By using large public-domain patient database, we demonstrate that ensemble learning of pGCN combining DTI and sMRI scans and genomic data improves the classification of Alzheimer's disease (AD) and Parkinson's disease (PD).

Author's Manuscripts

- 1) **Haveesh Goli**, Satya P. Singh, Jagath C Rajapakse “Ensemble learning of graph convolutional networks with applications to Alzheimer’s disease and Parkinson’s disease classification”, Under review at 27th *Conference in the Information Processing in Medical Imaging (IPMI)*, 2021.

Chapter 1

Introduction

The Central Nervous System (CNS) consists of two types of tissues: the grey matter that consists of cell bodies and the white matter that are formed by white matter fibres connecting different parts of grey matters.

Neurodegenerative diseases like Alzheimer’s disease (AD) and Parkinson’s disease (PD) cause neurons in the grey matter to lose their structure and function. With such damages in brain cells, the patients experience progressive loss of memory, intelligence, judgment, and other cognitive abilities. Although it has long been established in the science community that Alzheimer’s is a disease that affects the grey matter of the brain, recent studies have expanded their analysis into investigating the connection between Alzheimer’s disease and white matter damage [1][2].

In this work, our focus is encoding the structural connectome for classification of AD and PD patients from DTI and sMRI scans. Deep neural networks (DNNs) models have been explored for classification of brain diseases in DTI and MRI scans. The structural connectome is measured by DTI scans and is represented by a graph where the nodes are brain regions of interests and the edges represent of white matter connections between the regions. However, in previous approaches, the connectivity matrix has been linearized and hence graphical nature of the connectome has been lost. In this thesis, we propose ensemble graph convolutional neural network (GCN) for analysis of the structural connectome.

In order to classify disease, fusion of sMRI and DTI to supplement each other for feature representation was found to be highly desirable. Multimodal integration of sMRI and DTI has improved the ability of classification for other brain disorders such as schizophrenia [3][4]. In [5], the authors argued that potential use of sMRI and DTI modalities as imaging biomarkers can be integrated into monitoring clinical progression in PD. In this paper, we propose to extract features from brain regions from

both sMRI and DTI scans and concatenate them to achieve multimodal fusion of feature representation. Apart from information from images, integration of other information such as subject's gender and age can boost the disease classification.

Neural networks (NN) have seen great success in executing tasks that are challenging when provided with huge amounts of data along with sizable computational resources. Feature extraction and classification are two major tasks that NN models successfully perform. Their capability to simultaneously automate these processes makes them stand out from other machine learning algorithms like regression, decision trees and support vector machine (SVM) at several tasks. Feed forward neural network models operate on vectorized connectivity patterns while convolutional neural networks (CNNs) have seen success in discrete grid structured data representations such as images and videos [6][7]. However, data such as structural connectome of brain and protein graphs can easily be processed into graphs [8][9]. Such unstructured data needs different convolutional operators as the conventional convolutional filters fail to perform on graph data. This led to the research field of Graph CNNs.

Deep neural networks (DNNs) models have been explored for classification of brain diseases using DTI. In [10], to tackle the lack of publicly available data, deep CNN was used with application of transfer learning from MRI to DTI. DNN models have been unable to capitalize on graphical structure inherent in the connectome data. Recently, graph convolutional networks (GCN) have become popular in dealing with graph-structured data in various domains. A Multiview GCN which can directly take brain graphs from multiple views as inputs was presented in [11] for relationship prediction in PD. A few researchers have embarked on using GCN for disease prediction from whole brain scans, for example, for autism and AD classification using resting state fMRI scans [12], [13] and for AD classification from cortical networks derived from MRI images [14],[15].

In this study we classify different stages of cognitive impairment leading up to AD and PD by analysing the structural connectome of the brain and the population traits of the subjects. The process includes- (1)

application of tractography to detect white matter pathways of the brain; (2) generation of white matter connectivity matrices based on tractography results and subjects' structural MRI scan; (3) generation of the population characteristics of the subjects which shall be used in modelling the architecture for disease classification; and (4) use of GCNs to perform classification task. We investigate population based GCN (pGCN) for disease classification from DTI scans. pGCN was introduced in [16] for disease classification using resting-state fMRI scans. One advantage of using pGCN over classical GCN is that it allows a natural mechanism for blending phenotypic information such as age and gender with connectome features [16][17]. Learning weights of pGCN are initialized with a population matrix defined by phenotypic information of subjects. We further proposed how gene expression data is used to generate population graphs.

Furthermore, we present an ensemble learning strategy for pGCN to boost its classification performance. We create an ensemble of pGCN by randomly dropping out [18] a fraction of connections in its weight matrix during learning. Such an ensemble is easy to implement and effectively increases the diversity in learning [19]. We demonstrate that ensemble learning notably improves classification performance. Existing methods that use GCN on ADNI and PPMI datasets for disease classification are limited to using clinical, biomarker and assessment information. They do not make best use of the population, gene and clinical information available in these datasets in graph modelling. Moreover, ensemble models have not been used in modelling GCNs with sMRI and DTI brain scans. In this thesis, we address these limitations and advance the field. The main contributions in this study can be summarized as follows:

- Extraction of sMRI and DTI features from brain scans of subjects and fuse them to achieve multimodal fusion for feature representation
- pGCN for modelling the structural connectome feature derived from DTI and anatomical features from sMRI brain scans.
- Ensemble learning strategy to boost classification performance of pGCN.

- To derive population graph using gene expression data of patients.
- State-of-the-art accuracies for AD and PD classification with ensemble learning of pGCN on large public-domain datasets.

1.1 Software and Environment

In this study, all scripts were written in either Python (3.1.3)¹, Matlab 2018² and processed under Linux OS. Software programs including Diffusion Toolkit³, FSL⁴, Freesurfer⁵ and TrackVis⁶ was used for data pre-processing. Python package called Nipype was used for batch processing. Nipype⁷ allows users to combine processes from different software programs in a single python script. For connectivity matrices generation and machine learning tasks, Python packages including Dipy⁸, Numpy⁹, Pandas¹⁰, NiBabel¹¹, Scipy¹² and Scikit-learn¹³ were also used. TensorFlow 1.0.0¹⁴, also a python library was used to build deep learning models in this study. For data visualization Freesurfer and ‘Aliza Medical Imaging and Dicom Viewer’¹⁵ softwares were used.

1.2 Thesis organization

Chapter 1 gives an overview of the research carried out in this thesis. It states the objectives and contributions made in this work. Chapter 2 describes the background and concepts used in the thesis. Chapter 2 also reviews related work in the area and describes the data used in this study. Chapter 3 describes the data preparation and the pre-processing steps used in the study. Chapter 4 focus on the pGCN framework and the ensemble learning strategy proposed for pGCN. Chapter 5 will elaborate on combining DTI and sMRI modalities to achieve multimodal fusion strategy

¹ <https://www.Python.org/>

² <https://www.mathworks.com/products/matlab.html>

³ <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>

⁴ <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET>

⁵ <https://surfer.nmr.mgh.harvard.edu/>

⁶ <http://trackvis.org/>

⁷ <https://nipype.readthedocs.io/en/latest/>

⁸ <https://dipy.org/>

⁹ <https://numpy.org/>

¹⁰ <https://pandas.pydata.org/>

¹¹ <https://nipy.org/nibabel/>

¹² <https://scikit-learn.org/stable/>

¹³ <https://scikit-learn.org/stable/>

¹⁴ <https://www.tensorflow.org/install/pip>

¹⁵ <https://www.aliza-dicom-viewer.com/>

for a joint feature representation in pGCN. Chapter 6 focuses on building the population graph with gene expression data and describes how genomic data can be incorporated into the framework.

Chapter 2

Background and related work

In this chapter, we describe the various data used in the study. We also review the principles of machine learning for brain connectome in the literature with a focus on GCN. We review various data types used in the study. Our first aim was to use DTI data for disease classification types. We used ensemble pGCN to encode the structural connectome. We later combined the sMRI modality with DTI to achieve multimodal fusion in order to improve the disease classification. We also used the transcriptome data to derive population graphs using gene expression data of patients.

2.1 Data

2.1.1 Structural Magnetic Resonance Imaging (sMRI)

Structural magnetic resonance imaging (sMRI) translates the local differences in water content into different gray intensities that serve to outline the shapes and sizes of the brain anatomy. MRI scanner delivers a specific radiofrequency that excites hydrogen atoms in the water molecule, which return some of this energy in the form of a characteristic nuclear magnetic resonance signal. Not all protons resonate in that way, but generate enough signals such that the resulting computer-generated image constitutes a highly detailed map of the brain's tissues and structures. Thus, sMRI can be used to discover the presence of abnormal tissue through the changes in tissue density or composition. Scientists examining an sMRI can readily distinguish between gray and white matter and other types of tissue, both normal, such as blood vessels, and abnormal, such as tumors and by their different shading and contrast with surrounding areas.

Such measurements can help scientists and doctors to home in on the regions that are most heavily affected by drugs [20]. Importantly, these initial observations often guide additional investigations, using other research tools and techniques, to determine the reasons for the

structural changes as well as their experiential and behavioral consequences.

Most studies performed measurements on 3T scanners [21]; however, there are no major differences in Voxel Based Morphometry/Surface Based Analysis results when measurements are performed on 1.5T cameras. Most common approaches to the whole brain data evaluation when comparing groups of subjects are briefly introduced in the text below. Clinical scales for rating brain atrophy are useful for evaluating individual subjects in clinical practice.

T1 weighted sequences are part of almost all MRI protocols and are best thought of as the most 'anatomical' of images, resulting in images that most closely approximate the appearances of tissues macroscopically, although even this is gross simplification. T1 sMRI provides information on the shape, size and integrity of brain structures [22]. The dominant signal intensities of different tissues are- fluid (e.g. urine, CSF): low signal intensity (black), muscle: intermediate signal intensity (grey), fat: high signal intensity (white), brain grey matter: intermediate signal intensity (grey), brain white matter: hyperintense compared to grey matter

T2 weighted sequences are part of almost all MRI protocols. Without modification the dominant signal intensities of different tissues are- fluid (e.g. urine, CSF): high signal intensity (white), muscle: intermediate signal intensity (grey), fat: high signal intensity (white), brain grey matter: intermediate signal intensity (grey), white matter: hypointense compared to grey matter (dark) [23].

Similarly, in the brain, we often want to detect parenchymal edema without the glaring high signal from CSF. To do this we suppress CSF. This sequence is called FLAIR [24]. Importantly, at first glance FLAIR images appear similar to T1 (CSF is dark). The best way to tell the two apart is to look at the grey-white matter. T1 sequences will have grey matter being darker than white matter. T2 weighted sequences, whether fluid attenuated or not, will have white matter being darker than grey matter.

Figure 1. Illustration of different types of MRI.

Voxel Based Morphometry is a leading tool reprocessing structural MRI of the brain. It fits different brain regions from a T1-weighted image into a given template enabling users to make intersubject comparisons between brains of different shapes and topography, i.e., in diverse subject populations [25]. Using this non-linear registration in order to achieve voxel-for-voxel correspondence across subjects, brain areas stretch and compress with respect to each other when moving at the matching point in the template image, thus creating a deformation field. The deformed image is segmented into classes according to tissues; gray matter, white matter, and cerebrospinal fluid according to the intensity in the image. The segmented image is then spatially smoothed [25].

In Surface Based Analysis the morphometric measures are derived from geometrical models of the cortical surface. During surface-based analysis, cortex is modelled as a two-dimensional surface model which consists of triangles or polygons, the meeting point of the triangle corners is called a vertex. The parameters of the model are the X, Y and Z coordinates of each vertex. From the known coordinates, the surface can be rendered as a surface embedded in 3D [26]. The coordinates enable measuring morphometric measures such as surface area of cortex, or cortical thickness. The

Freesurfer¹⁶ software has been used for data analysis. Cortical thickness method seems to be more sensitive than voxel based morphometry in depicting early gray matter changes, particularly where there are smaller changes occurring over a larger cortical area [27]. The presumably high sensitivity can be achieved partly through reduced registration errors across different brains.

In this study, from T1 sMRI scans we extract intensity features from brain regions and represent them in population graph in pGCN.

2.1.2 Diffusion Tensor Imaging (DTI)

Diffusion MRI [28], better known as two distinct classes of application, Diffusion Weighted MRI (DWI) and DTI, has shown the ability to detect subtle changes in brain tissue. These methods indirectly identify microscopic aspects that provide measures reflecting the patterns in size, orientation and organization of tissue which are supposed precursors to the final stage of macroscopic tissue atrophy [29]

DWI the brain was first used during the early nineties and found to have immediate utility for the evaluation of suspected acute ischemic stroke. Since then, the technology of diffusion imaging has led to greatly improved image quality and enabled many new clinical applications. DTI and fiber tractography have advanced scientific understanding of many neurologic and psychiatric disorders, having been applied clinically for presurgical mapping of white matter pathways to avoid postoperative injury and proved useful for the diagnosis of dementia [30].

In DWI, each image voxel has an image intensity that reflects a single best measurement of the rate of water diffusion at that location. DWI is most applicable when the tissue of interest is dominated by isotropic water movement and the diffusion rate appears to be the same when measured along any axis [28].

DTI is valuable when a tissue has an internal fiber structure analogous to the anisotropy of some crystals, such as the white

¹⁶ <https://surfer.nmr.mgh.harvard.edu/>

matter fiber tracts in the brain. Water tends to diffuse more rapidly in the direction of the internal structure and more slowly as it moves perpendicular to the passage of least resistance. The measured rate of diffusion differs depending on the direction from which is observed. Each voxel therefore has one or more associated pairs of parameters: a rate of diffusion and a preferred direction of diffusion [28].

The properties of each voxel of a single DTI image is usually calculated by vector or tensor math from six or more different diffusion weighted acquisitions, each obtained with a different orientation of the diffusion sensitizing gradients. In some methods, hundreds of measurements are made to generate a single resulting calculated image data set. The higher information content of a DTI voxel makes it extremely sensitive to subtle pathology in the brain. The directional information can be exploited at a higher level of structure to select and follow neural tracts through the brain, a process called tractography. Detailed review of the physical aspects of DWI and imaging techniques characteristics are described in relevant articles [28][29][30].

DTI provides derived quantitative measures which reveal cellular microstructure and fiber tract integrity. Diffusion tensor imaging (DTI) is an MRI technique that uses anisotropic diffusion to estimate the white matter organization of the brain. It is based on the measurement of Brownian motion of water molecules. DTI provides a quantitative analysis of the magnitude and directionality of water molecules. Within cerebral matter, water molecules tend to diffuse more freely along the direction of white matter fascicles rather than across them as shown in figure 1. Such directional dependence of diffusivity is termed as anisotropy. This direction of maximum diffusivity along the white matter fibers is projected in the final image. The two main parameters derived from DTI data are mean diffusivity (MD), also referred to as apparent diffusion coefficient (ADC), and fractional anisotropy (FA). FA reflects the directionality of molecular displacement by diffusion and varies

between 0 (isotropic diffusion) and 1 (infinite anisotropic diffusion). FA value of CSF is 0. MD reflects the average magnitude of molecular displacement by diffusion. The higher the MD value, the more isotropic the medium.

Figure 2. In DTI, a voxel represents the diffusion properties of several bundles of axons to calculate a diffusion tensor, illustrated as ellipsoid. For each voxel, the tensor information could be parameterized in fractional anisotropy (FA) maps or used for fiber tracking (FT) [31].

Tractography is a 3- dimension (3D) modelling technique used to access neural tracts of the brain using DTI. DTI is commonly used in white matter tractography of the brain because of the unique structure of white matter. White matter consists of bundles of axons which are often organised in a single direction such that the barrier of diffusion is set up. The DTI images used in this project were taken in 47 directions. The gradient information was stored in two files named *bval*, which contains the b-value, and *bvec*, which contains the b-vector. *bval* is a 47 x 3 matrix which represents the gradient direction taken for all 47 directions and *bvec* represents the strength of the magnetic field. These parameters are pre-determined by scanner and hence they stay consistent for all DTI images. In this project, deterministic tractography was used, where tracking is focused on the dominant direction of each of the voxels. A greedy algorithm Fiber Assessment by Continuous Tracking (FACT) [32] was used to track white matter fibres by following the principal diffusion direction of each voxel, from a seed location to a destination voxel.

Figure 3. Illustration of diffusion tensor imaging (DTI), each color represents a different white matter pathway.

In this study DTI images used where eddy-current corrected images. The feature information extracted from DTI was incorporated in the population graph in pGCN.

2.1.3 Previous work on combining DTI and sMRI scans

In [33], authors combined DTI and sMRI modalities for detection of AD by using SVM on a large European multicentre dataset. In [34] authors use a fusion of sMRI and PET images together with canonical correlation analysis (CCA). After pre-processing and aligning images of 2 modalities given the covariance data of sMRI and PET image they find the projection matrices by maximizing the correlation between projected features. Ahmed et al. in [35] demonstrated the efficiency of using the amount of cerebrospinal fluid (CSF) in the hippocampal area calculated by an adaptive Otsu's thresholding method as an additional feature for AD diagnostics. In [36], they further improved the result of [37] by combining visual features derived from sMRI and DTI MD maps with a multiple kernel learning scheme. Similar to [37] they selected hippocampus ROIs on the axial, sagittal and coronal projections and described them using Gauss-Laguerre Harmonic Functions.

Another way is to use more input data e.g. consider several ROIs instead of one. So Liu et al. in [38] first identified discriminative 50 anatomical landmarks from MR images in a data driven manner, and then extract multiple image patches around these detected landmarks. After that they use a deep multi-task multi-channel convolutional neural network for disease classification. The authors addressed the problem of classification of patients into NC, stable MCI (sMCI) and progressive MCI (pMCI). The authors used MRI images from ADNI database containing in total 1396 images and achieved 0.518 accuracy in four-class (NC/sMCI/pMCI/AD) classification. A more simplified idea was proposed by Cheng et al. in [39] as they used a number of 3D convolutional neural networks with 4 layers together with late fusion. With the subset of ADNI database of 428 sMRI images the authors achieved an accuracy value of 0.872 for AD/NC classification. In [5], they proved that potential use of sMRI and DTI modalities as imaging biomarkers can be integrated into prediction of clinical progression in PD. However, combining sMRI and DTI using GCNs has been unexplored yet.

2.1.4 Transcriptome data

The human genome is made of DNA (deoxyribonucleic acid), a long, winding molecule that contains the instructions needed to build and maintain cells. These instructions are spelled out in the form of "base pairs" of four different chemicals, organized into 20,000 to 25,000 genes. For the instructions to be carried out, DNA must be "read" and transcribed, in other words, copied into RNA (ribonucleic acid). These gene readouts are called transcripts, a transcriptome is a collection of all the gene readouts present in a cell.

There are various kinds of RNA. The major type, called messenger RNA (mRNA), plays a vital role in making proteins. In this process: mRNA is transcribed from genes; then the mRNA transcripts are delivered to ribosomes, the molecular machines located in the cell's cytoplasm; then the ribosomes read, or "translate," the sequence of chemical letters in the mRNA and assemble building blocks called amino acids into proteins.

An RNA sequence mirror is the sequence of the DNA from which it was transcribed. Consequently, by analysing the entire collection of RNA sequences in a cell (the transcriptome) researchers can determine when and where each gene is turned on or off in the cells and tissues of an organism. Depending on the technique used, it is often possible to count the number of transcripts to determine the amount of gene activity - also called gene expression, in a certain cell or tissue type. In human and other organisms, nearly every cell contains the same genes, but different cells show different patterns of gene expression. These differences are responsible for the many different properties and behaviours of various cells and tissues, both in health and disease.

The function of most genes is not yet known. A search of a transcriptome database can give researchers a list of all the tissues in which a gene is expressed, providing clues about its possible function. Human brain tissue is a valuable resource for determining the role of gene expression in neuropsychiatric diseases, and establishing a

connection with experimental models [40]. Whole-genome profiling, primarily using microarrays, has identified subtle alterations in gene expression within discrete brain regions related to dementia [41]. High-throughput sequencing technologies (RNA-Seq) offer novel insights into transcriptome complexity, including non-coding RNAs (ncRNAs) and alternatively spliced variants, across different tissues and disease states [42].

Recent advances in biotechnology have led to full-scale analyses of the genome, transcriptome, and epigenome rather than focusing on a few biomarkers. A large-scale genome-wide association study (GWAS) of individuals with AD and normal controls (NC) was presented in 2009 and it identified variants at *CLU* and *CR1* genes, which were associated with AD [43]. Additionally, a meta-analysis of four previously reported GWAS datasets (17,008 AD cases, 37,154 controls) yielded 11 new loci of susceptibility to AD [44]. Recently, Xu et al. constructed an AlzData database integrating data from GWAS, eQTL, interactome, and laboratory experiments, which provides all human genes with scores for association with AD, called the Convergent Functional Genomics (CFG) score [45].

In recent years, two large multi-center studies were conducted to identify biomarkers for early AD diagnosis and MCI progression to AD: the Europe-based ANM and US-based Alzheimer's Disease Neuroimaging Initiative (ADNI) [46][47]. Furthermore, a large amount of publicly available gene expression data on AD have been provided in the NCBI GEO [47]. As a result, various studies, especially gene expression-based studies, have been published to uncover the informative genes associated with AD. The BrainNet study analyzed 113 samples of well characterized postmortem brain tissues, yielding 21 genes dysregulated in AD cases [47]. A study by Liang et al.[48] consisting of 87 brain tissues samples revealed that in brain tissues of AD cases, the genes encoding subunits of the mitochondrial components showed significantly lower expression. By analyzing RNA expression from brain tissues of AD patients, Xu et al. [48] demonstrated that an early alteration of *YAP1* could promote AD.

2.2 Deep Neural Networks (DNN)

Deep learning extracts patterns from data using neural networks. Deep learning refers to artificial neural networks containing multiple interconnected layers of artificial neurons. Deep neural network architecture has provided breakthrough results in many fields such as image classification, text generation and language translation [6][49][50]. Deep Learning network's ability to extract its own features for classification has provided it an edge over other machine learning algorithms like decision trees and SVMs. An artificial neuron by analogy to a biological neuron is something that takes multiple inputs, performs a simple computation, and produces an output. This simple of a linear function of the inputs followed by an activation function (usually non-linear) denoted by $f(\cdot)$. Examples of some commonly used non-linear activation functions are the hyperbolic tangent (Tanh), sigmoid transformation and the rectified linear unit (ReLU) (gives a maximum of 0 and the input) [19]. Deep learning is essentially a reincarnation of the artificial neural network where we stack layer upon layer of artificial neurons. Using the outputs of the terminal layers built on the outputs of previous layers, one can start to describe arbitrarily complex patterns. Many such merits of DNNs have resulted in deep learning methods being the choice for automated detection and diagnosis of diseases.

2.2.1 Feed Forward Neural Networks (FFN)

Feedforward neural networks were among the first and most successful learning algorithms. They are also called deep networks, multi-layer perceptron (MLP), or simply neural networks. As data travels through the network's artificial mesh, each layer processes an aspect of the data, filters outliers, spots familiar entities and produces the final output.

The input layer consists of the neurons that receive inputs and pass them on to the other layers. The number of neurons in the input layer should be equal to the attributes or features in the dataset. The output layer is the predicted feature and depends on the type of model you're building.

In between the input and output layer, there are hidden layers based on the type of model. Hidden layers contain a vast number of neurons which apply transformations to the inputs before passing them. As the network is trained, the weights are updated to be more predictive. Weights refer to the strength or amplitude of a connection between two neurons. If you are familiar with linear regression, you can compare weights on inputs like coefficients. Weights are often initialized to small random values, such as values in the range 0 to 1.

Figure 4. 2-layer feed forward neural network (FFN) architecture

2.2.2 Convolutional Neural Networks (CNN)

The success of deep learning can majorly be attributed to Convolutional Neural networks (CNNs). CNNs are a breakthrough in computer vision. The change in paradigm from handcrafted features (SIFT) [51] to learned features (end-to-end systems) has started with the invention of CNNs. CNN architectures are powerful to solve high dimensional learning problems. They have been exceptional not just in tasks related to image and video data interpretation and classification [52], but also in speech recognition and language translation [49]. CNNs further build on the

capabilities of multilayer perceptrons (MLPs) while not altering the number of parameters. They leverage compositionality structure to extract the compositional features and feed them to a classifiers, regression systems, recommender systems etc. Essentially, in CNN, features are generated within the network by convolving a kernel with the input images. These features are generally in the form of shapes, curves edges, et cetera.

A CNN comprises of four basic components: 1) Local receptive field, 2) Sharing weights, 3) Pooling, and 4) Fully connected (FC) layers. Deep CNN architecture is constructed by stacking several convolutional layers and pooling layers and one or so fully connected layers at the end of the network.

The key for the success of CNNs is that the common set of parameters are shared over different inputs, but the limitation here is that the input needs to be gridded with fixed neighbours. Graph Convolutional Networks address this limitation using graph-based convolutional operation.

2.2.3 Graph Convolutional Neural Networks (GCN)

Graph convolutional networks (GCNs) have become popular in dealing with graph structured data in various domains. A graph G is represented using a tuple (V, A) where V are the vertices and A is the adjacency matrix. GCNs can broadly be divided into two categories based on defining the convolutional operation on graphs: spectral based and spatial based methods.

Spectral methods are based on spectrum of graph Laplacian and convolutional kernel in the spectral domain. ChebyNet [53] introduces a fast localized convolutional filter on graphs based on Chebysev's polynomial approximation. Kipf's and Welling's Vanilla GCN [53] further extends the spectral graph convolutions to large scale networks by several simplifications. Spectral graph theory works on the construction of a filter based on the eigen vector decomposition of Laplacian $\mathcal{L}$ defined as:

$$L = D - A \quad (3)$$

which is unnormalized Laplacian for continuous domain. $\mathcal{L}$ can be normalized as shown in equation (4).

$$L = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \quad (4)$$

where A is the adjacency matrix. I and D are the identity and diagonal degree matrices respectively. L or Laplacian is the core operator in spectral graph theory and is used to calculate the eigen vectors and eigen values U which is similar to the discrete fourier transform (DFT). The idea is to build the smoothest orthogonal basis on the graph. Good measure of smoothness is Laplacian. A graph signal x is transformed into spectral domain as follows:

$$x * h = U^T.(Ux \odot h) \quad (5)$$

where, $\odot$ represents the elementwise product and ‘.’ is matrix multiplication. Each frequency is multiplied by a filter h to get its filtered output in spectral domain. The filter coefficients are then multiplied with the graph signals that are spectrally transformed. This is essentially the filtering operation in spectral domain. Many researchers have proposed GCNs based on this method of filtering [54][55]. Spectral methods have the necessity of having input graph samples to be homogeneous. This is necessary as the eigen decomposition of the graph Laplacian is unique for every graph structure and needs to be solved separately. Spectral based GCNs are usually used for the graph structured data where there is a single common graph structure across all the samples. We focus on spectral based GCNs in our work as our input graphs formed by the connectome data are homogenous. Spatial approaches can filter graphs without the graphs being homogenous.

2.2.4 Previous work on application of Machine Learning for Brain Connectome Analysis

Machine learning has previously been predominantly been tested on MRI and DTI data. The k-nearest neighbor [56] is a simple instance-based learning algorithm, which employs the distance metric criterion for discrimination using its k-nearest neighbors. Bagging [57] is a meta

algorithm, which reduces variance and helps to avoid overfitting. Classification via regression classifies the data using regression methods. In this method each class is binarized and one regression model is built for each class. Recently, Zhang et al. [58] have used a L2 regularized logistic regression classifier on fMRI data for MCI classification. Decision tree [59] is a classifier for generating a pruned or unpruned decision tree and it is a mapping of observations to classification.

Support vector Machine (SVM) [60] is a supervised machine learning algorithm that can be employed for both classification and regression purpose. SVMs are based on the idea of a hyperplane that best divides a dataset into two classes. SVMs have been used for text classification, sentiment analysis and image recognition tasks particularly well in aspect -based recognition and color-based classification [61],[62]. Clinical data is often not linearly separable. To deal with non-separable data, [63] has introduced optimal soft-margin hyperplane which introduces slack variables to penalize classification errors based on some predetermined weights. However, in this work, the classification performance was not consistent. In [64], SVM was used to classify PD patients from using a method to find the best possible accuracy on different kernel values. In [65], they have investigated how multiple indices of diffusion in DTI can be used in conjunction with SVMs for the classification of NC and MCI subjects. In [66], a framework is presented to introduce spatial and anatomical priors in SVM for brain image analysis based on regularization operators. SVM [67], univariate [68], and k-best [69], are among the methods that have successfully been utilized for classifying various stages of dementia: AD, MCI, and NC. Mateos-Pérez et al. [70] have done an extensive review of different variants of SVM used on the structural connectome for disease prediction. The major drawback of SVM is that, it fails to perform on large and noisier datasets as it is prone to overfitting if the number of features is much greater than the number of samples. Data used in this work falls into this category of high imbalance between the sample size and the number of features and hence SVMs are ideally not suitable for our data. Data acquisition and pre-processing shall be discussed in depth in sections 2.5.1 and 2.5.2. Also, SVM needs several key parameters that need to be set correctly to achieve

the best classification results for any given problem as opposed to neural network models.

Artificial neural networks (ANNs) were originally inspired by the human brain and mimic the information transferring among distributed neurons in the brain [71]. With the development of computational power and big data analytics, DNNs achieved promising results in various pattern recognition tasks.

The connectome helps to explain the structure of the brain. It is quantified and analysed from neuroimaging data, including MRI and DTI. Brain networks offer advantages when describing relationships among brain regions and revealing progression of diseases. Since several neurodegenerative conditions are difficult to diagnose at an early stage, detection of brain disorders becomes more important early. Due to the unusual structured and high-dimensional characteristics of the networks, the application of deep learning on network structured data is challenged by extracting meaningful information from networks. Specific attributes of brain networks include node attributes, edge attributes, and graph attributes. The first deep neural network on connectome is BrainNetCNN, proposed by Kawahara et al in 2017 [72]. They specifically designed edge-to-edge, edge-to-node and node-to-graph convolutional layers to detect the properties of the connectome. Inputs to BrainNetCNN is adjacency matrix derived from DTI and the output is prediction scores of cognitive and motor developmental outcomes. The drawback of this model is that it does not take node attributes and graph attributes into account.

Figure 5. Schematic representation of the BrainNetCNN architecture [73].

Each block represents the input and/or output of the numbered filter layers. The 3rd dimension of each block (i.e., along vector m) represents the number of feature maps, M , at that stage. The

brain network adjacency matrix (leftmost block) is first convolved with one or more (two in the figure) E2E filters which weight edges of adjacent brain regions. The response is convolved with an E2N filter which assigns each brain region a weighted sum of its edges. The N2G assigns a single response based on all the weighted nodes. Finally, fully connected (FC) layers reduce the number of features down to two output score predictions.

Another popular deep learning approach on the connectome is through spectral graph convolution or spectral GCN. Parisot et al. [74] proposed spectral GCN for brain disease prediction. Essentially, their model uses graph Laplacian operator defined as:

$$L = I_N - D^{-\frac{1}{2}}WD^{-\frac{1}{2}} \quad (6)$$

where I_N and D are respectively identity and diagonal degree matrices. Their eigen decompositions, $L = U\Lambda U^T$ gives a set of orthonormal eigenvectors $U \in R^{N \times N}$ with associated real, non-negative eigen values $\Lambda \in R^{N \times N}$. Spectral convolutions over signal x with a filter $g_\theta = \text{diag}(\theta)$ are defined as $g_\theta * x = g_\theta(L)x = g_\theta(U\Lambda U^T)x$ where $\theta \in R^{N \times N}$ is a vector of Fourier coefficients. They use population traits from the data into predictive modelling using graph connectivity. The architecture of their model has been shown in figure 6.

Figure 6. Spectral GCN for population-based disease classification [16].

Deep learning has the ability to discover hidden features and predict disease from learning. Deep learning on connectome has been used to early predict Alzheimer's, Autism and other neurogenerative disorders [74][75]. A spectral GCN was employed for AD and MCI classification using the

cortical thickness data from T1-MRI [76]. They used transfer learning to determine the best performing model on two different datasets.

We summarize the literature about deep learning on brain connectome in Table 1.

Reference	Approach	Description
C.Y.We et al. (2019) [76]	GCN	Transfer learning to train the model; classification using cortical thickness data from T1-MRI scans.
Ktena et al. (2018) <i>Population-based GCN</i> [74]	GCN	Learning similarity of brain connectivity networks using population traits; spectral GCN to learn features in graph spectral domain.
He et al. (2018) [77]	auto-encoder	Prediction of cognitive deficits in infants; use auto-encoder to extract features from functional connectivity of brain.
Wang et al. (2018) [78]	recurrent neural network (RNN)	Combining RNN with dynamic causal modelling to estimate effective connectivity of brain
Saman Saraf et al. (2017)	CNN	Functional and structural MRI data pre-processing pipelines for classification of AD
Kawahara et al. (2017) <i>BrainNetCNN</i> [72]	CNN	Development of edge-to-edge, edge-to-node and node-to-graph filters to leverage network structure.
Suk et al. (2016) [79]	deep auto-encoder	to estimate functional dynamics; to encode functional brain networks into embedding space

Table 1. Deep learning techniques in literature for brain connectome analysis.

Chapter 3

Data processing steps

In the previous chapter, we have discussed the data and the previous works in literature related to sMRI, DTI and transcriptome data. In this chapter we discuss the data acquisition and the processing steps for DTI and sMRI scans, and transcriptome data.

The DTI and sMRI data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI)¹⁷ and Parkinson's Progression Markers Initiative (PPMI)¹⁸. The gene expression data of the patients was obtained from ADNI database. ADNI and PPMI are clinical studies that seek to build data driven approaches for early diagnosis of AD and PD, respectively. In this study, only the participants who had both structural MRI (sMRI) and DTI images were selected.

3.1 Alzheimer's disease Neuroimaging Initiative (ADNI)

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's disease (AD). Since its launch more than a decade ago, the landmark public-private partnership has made major contributions to AD research, enabling the sharing of data between researchers around the world. One defining characteristic of ADNI is the commitment by all cores to share data without embargo and within hours of collection. All data generated by the ADNI study investigators are entered into the data repository hosted at the Laboratory of Neuroimaging (LONI) at the University of Southern California, the LONI Image & Data Archive (IDA). Qualified researchers worldwide can submit an online data access request and generally begin using ADNI data including cognitive/neuropsychological, image, biofluid and genetic data sets within a few days of request submission. Since its launch in 2004, the

¹⁷ <http://adni.loni.usc.edu/>

¹⁸ <https://www.ppmi-info.org/>

overarching aim of the ADNI study has been realized in informing the design of therapeutic trials in AD. ADNI3 continues the previously funded ADNI1, ADNI-GO, and ADNI2 studies that have combined public/private collaborations between academia and industry to determine the relationships between the clinical, cognitive, imaging, genetic and biochemical biomarker characteristics of the entire spectrum of sporadic late onset AD.

We downloaded 56 AD, 93 MCI (Mild Cognitive Impairment), and 67 NC (Normal Controls) subjects specifically from ADNI2 and ADNI-GO baseline studies. A subject may have several visits and we use all visits in our experiments. Therefore, in total, we prepared 155 samples of AD, 221 samples of MCI, and 155 samples of NC. During data split for test, train and validation, we have taken care to not mix the same subject's scans with others.

3.2 Parkinson's Progression Markers Initiative (PPMI)

Parkinson's disease is a chronic, degenerative neurological disorder. There is no objective test, or biomarker, for Parkinson's so the rate of misdiagnosis can be relatively high, especially when the diagnosis is made by a non-specialist. Therefore, estimates of the number of people living with the disease vary.

The Parkinson's Progression Markers Initiative (PPMI) is a landmark observational clinical study to comprehensively evaluate cohorts of significant interest using advanced imaging, biologic sampling and clinical and behavioural assessments to identify biomarkers of Parkinson's disease progression. PPMI is taking place at clinical sites in the United States, Europe, Israel, and Australia. Data and samples acquired from study participants will enable the development of a comprehensive Parkinson's database and biorepository, which is currently available to the scientific community to conduct field-changing research. PPMI is made possible by the concerted efforts of a number of collaborators.

A primary goal of PPMI is to develop a large and comprehensive set of correlated clinical data and biospecimens that will be made

available to the entire scientific community to help accelerate biomarker verification research. PPMI data have been aggregated into the PPMI study database, which is managed by the PPMI Bioinformatics Core, the Laboratory of NeuroImaging (LONI) at the University of Southern California.

We downloaded 185 PD and 72 NC subjects from PPMI database. We did not include SWEDD, genetic cohorts, and prodromal subjects in this study. Similar to ADNI dataset, we use all visits of subjects in our experiments with all due care that same subject’s visits are not mixed with others during cross validation. Therefore, in total we have 655 samples of PD and 172 samples of NC. Table 2 summarizes the demographic information of subjects in ADNI and PPMI datasets.

Database	Class	Participant	Sample size	Sex	Age (years)
ADNI	AD	56	155	120M 53F	76.15 ± 8.3
	MCI	93	221	153M 68F	73.84 ± 7.8
	NC	67	155	93M 62F	74.50 ± 6.5
PPMI	PD	185	655	412M 243F	62.2 ± 9.3
	NC	72	172	107M 65F	59.7 ± 11.1

Table 2. Details of subjects in ADNI and PPMI databases. Samples of participants who had both structural MRI (sMRI) and DTI images were only selected

3.3 Data processing for DTI scans

Pre-processing of DTI scans converts raw data and signals into data representation suitable as inputs to the population in pGCN model. The objectives of pre-processing include size reduction of the input space, smoother relationships, data normalization, noise reduction, and feature extraction. The structural MRI (sMRI) scans used were T1-weighted images that had been spatially normalized, masked and N3 corrected. The data pre-processing step was done using Free-surfer brain imaging software. DTI images used were eddy-current corrected images. We use the same pre-processing pipeline for ADNI and PPMI data.

3.3.1 Tractography

The brain extraction tool of FSL was applied to remove extracerebral tissues from diffusion-weighted images. The eddy corrected images, *bvec* and *bval*, were then used as input for diffusion tensor computation. Deterministic tractography was performed on the generated brain mask and diffusion tensor. The pipeline is illustrated in flowchart 1

Flowchart 1. Illustration of processing pipeline used in tractography.

3.3.2 Region of interest extraction and Brain mask generation

For ROI extraction, FSL's function was used to extract the first volume (i.e. the image with $b = 0$). This image is known as *b0* image. FSL's function were then used to obtain a mask of the brain. The parameter was

used to determine where the edge of the final segmented brain is located. These two steps are essential in preparing the data for deterministic tractography.

3.3.3 Eddy Current Correction

Although images used in this project are eddy current corrected, eddy current correction is an essential process as it ensures that brain images used for diffusion tensor computation are not distorted. Hence, if images used are not eddy current corrected, eddy current correction should be performed. FSL's function can be used to perform eddy current correction

3.3.4 Diffusion Tensor Computation

The Eddy corrected images, as well as the *bvec* and *bval*, will then be used as input for diffusion tensor computation. The computation can be done by using Diffusion Tool Kit's function.

3.3.5 Deterministic Tractography

The step following diffusion tensor computation is deterministic tractography where the diffusion tensor and ROI mask were used to find the tracts. Diffusion Tool Kit's was used to perform deterministic tractography.

A track file was generated from each DTI image. The white matter pathways were then visualized using software TrackVis. Figure 7 below shows four tractography results obtained from two DTI images.

Figure 7. Results of tractography obtained from DTI images

3.3.6 Cortical parcellation

DTI images were linearly aligned and elastically registered to their respective baseline T-1 weighted structures. Then, inverse transformation of T1 to b_0 space was calculated, and transformation matrices were obtained. Finally, we warped the transformation matrices with automated anatomical labelling (AAL) to b_0 . The pipeline is

Flowchart 2. Pipeline architecture used in cortical parcellation

illustrated in flowchart 2.

3.3.7 Computation of connectivity matrices

After performing cortical parcellation, connectivity matrices were then generated using both parcellated brain and white matter tracts. The non-weighted connectivity matrix shows how different regions of the brain are connected in terms of number of white matter fibre tracts connecting two regions of the brain. The AAL atlas divided the brain into 116 regions, therefore, the connectivity matrix generated is a matrix with dimension

116 x 116 where the value of each element represents the number of tracts connectivity the two regions. The two main steps for generation of connectivity matrices are, first, to transform streamlines into parcellated brain space. The second step is to generate the matrices.

The resulting output from the above steps are connectivity matrices, each with a dimension of 116 x 116. Each column and each row of the matrix represents one region of the brain. The weighted connectivity matrix was then computed by doing element wise multiplication between the matrix representing average FA of streamlines and the non-weighted connectivity matrix of the sample.

Figure 8. (a) An example of a weighted connectivity matrix generated, (b) Logarithm scaling of the weighted connectivity matrix

3.3.8 Feature vector generation

Symmetric connectivity matrices were generated using both parcellated brain and white matter tracts. The AAL atlas divided the brain into 116 regions thereby generating a connectivity matrix with a dimension 116 x 116. Since the connectivity matrices are symmetric in nature, we further eliminated the elements in the lower half triangles of the matrices thereby reducing the number of features to $M = 6786$. The connectivity matrices were then linearized into feature vectors where each subject was represented as an input node on the graph.

3.4 Data processing for sMRI scans

The feature extraction of sMRI was done using the Freesurfer¹⁹ software suite. ‘Freesurfer’ is an open-software suite involving the preprocessing stages of brain MRIs. The rsurfer²⁰ package contains functions to aid the importing and manipulation of data generated by Freesurfer.

All of the T1 images were first normalized to the standard Montreal Neurological Institute (MNI) space, and the images were segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) partitions using the tissue prior free segmentation routine of the “Estimate and write” toolbox in VBM8 [80]. Then, we checked the segment and normalization quality.

3.4.1 Intracranial Volume Normalisation

There are several intracranial volume (ICV) normalisation techniques. In this section we have explained the techniques we followed.

— *Proportional Intracranial Volume Normalisation*

This is the most commonly used method, and is the easiest to implement. the aim is to express a volume as the proportion of the brain it occupies and it is computed by:

$$v' = \frac{v}{V} \quad (7)$$

Where v is the unnormalised volume of the patient, V is the intracranial volume of the patient and v' is the normalised volume of the patient. This method is completely independent of any other subjects.

— *Residual Intracranial Volume Normalisation*

The residual method assumes a linear relation between the volumes and the subject's V and estimates the relationship such that:

$$v' = v - w_1(V - \bar{V}) \quad (8)$$

¹⁹ <https://surfer.nmr.mgh.harvard.edu/>

²⁰ <https://cran.r-project.org/web/packages/rsurfer/index.html>

where $\bar{V}$ is the average intra cranial volume across all subjects being considered, and w_1 is a coefficient computed by solving the linear regression problem of

$$v' = w_1 V + w_0 \quad (9)$$

— *Power Proportional Intracranial Volume Normalisation*

The power proportion method computes the normalised volume as

$$v' = \frac{v}{V^{w_1}} \quad (10)$$

where w_1 is computed by the power regression problem: $v' = w_0 V^{w_1}$. Note that this power regression problem can be converted to a linear regression problem by taking the natural logarithm of both sides to create

$$\ln(v') = w_1 \ln(V) + \ln(w_0) \quad (11)$$

After normalization, the GM images without modulation were smoothed with a Gaussian filter with an 8mm full-width half-maximum kernel. The detailing flowchart of the structural MRI preprocessing is shown in Flowchart 3.

Flowchart 3. Flowchart of sMRI pre-processing steps

3.5 Data processing of gene expression data

Gene expression profiles of all the patients were acquired from the ADNI database. Table 3 summarizes the ADNI dataset.

Class	Patients with sMRI	Patients with DTI	Patients with DTI and sMRI	Patients with gene expression	Patients with DTI and sMRI and gene expression
AD	56	68	56	53	53
MCI	93	104	93	27	27
NC	67	80	67	35	35
Total	216	252	216	115	115

Table 3. Details of subjects in ADNI database. Patients who had all three i.e. sMRI scans, DTI scans and gene expression profiles were selected.

The data processing is done in three steps. Our first strategy is based on Wilcoxon analysis that identifies the features that are most correlated with each clinical outcome using only the training data (Wilcoxon $p < 0.05$). When this analysis did not return any feature, the top 5% features (computed from the Wilcoxon test statistic) were used regardless of their p -values. After dimension reduction, there are between 2478 and 2553 features left depending on the clinical outcome. Another feature selection method experimented with involves computing the variance of each feature (across the dataset) and keeping the top 5% most varying features. This gives 2470 features.

Chapter 4

Ensemble learning of Population based Graph Convolutional Networks (pGCN) using DTI data

In chapter 3, we have discussed the data processing steps. In this chapter we discuss the methodology, experiments and results of ensemble pGCN using DTI data.

4.1 Population graph

A *population graph* is a graph where nodes represent feature vector set extracted from subjects and edges represents similarities or interactions between them and incorporates the phenotypic information. We used graph construction strategy proposed in [16] to construct the population graph. Let the population graph be represented by matrix $W = (w_{pq}) \in \mathbf{R}^{D \times D}$ where D is the number of subjects and weight w_{pq} between subject p and q is measured by the similarity between their phenotypic vectors ϕ_p and ϕ_q , respectively. That is

$$w_{pq} = \text{sim}(\phi_p, \phi_q) \quad (12)$$

Age and gender were used as phenotypic measures of subjects. A similarity equal to 1 was given if the genders were different else a greater value was assigned. Similarly, subjects were given a score equal 1 if ages were different else a greater value was assigned if ages were different beyond a threshold.

The population graph is implemented as weight matrix W of pGCN where the input layer nodes represent subjects and input samples corresponds to structural connectivity features derived from DTI scans.

4.2 Population Graph Convolutional Network (pGCN)

The pGCN architecture consists of a few alternate layers of graph convolution and pooling, followed by a fully connected layer with hidden

layers using the rectified Linear Unit (ReLU) function. We used the upper triangular matrices of the structural connectivity matrices as imaging features to the pGCN. The output layer is followed by a softmax function. Figure 9 illustrates the architecture.

Figure 9. Architecture of pGCN

Let $x_i \in \mathbf{R}^3$ be the features from the imaging data which are the input signals to the pGCN. Inputs of pGCN are convolved with learnable filters and pass through nonlinear activation functions. Graph coarsening is achieved using an efficient pooling strategy on the graph after rearrangement of the vertices as a binary tree structure. GCN formulation was derived from spectral graph theory [81]. Given a graph with a population matrix W , the graph convolution y on spatial signal x is defined as

$$y = g_{\theta}(L)x = Ug_{\theta}(\Lambda)U^T x \quad (13)$$

where $L = D^{-1/2}WD^{1/2}$ is the graph Laplacian, D is the diagonal degree matrix, U is the set of orthonormal eigenvectors derived from eigen decomposition of L , and $\Lambda = (\lambda_k)$ are non-negative eigen values. In order to derive more efficient implementation of trainable filters $g_{\theta}(\Lambda)$, a Chebyshev polynomial expansion was introduced in [82]:

$$g_{\theta}(L) = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L}) \quad (14)$$

where $T_k(\tilde{L})$ is k th order Chebyshev polynomial evaluated at $\tilde{L} \in \mathbf{R}^{N \times N} = \frac{2L}{\lambda_{max}} - I$ and $\theta = (\theta_k \in \mathbf{R}^{K \times N})$ are the Chebyshev coefficients. Then

$$z = \sigma \left(\sum_{k=0}^{K-1} \theta_k T_k(\tilde{L}) x \right) \quad (15)$$

which can be recursively implemented as $y = \tilde{x}\theta$ where $\tilde{x} = (\tilde{x}_k)$ are evaluated recursively as $\tilde{x}_k = T_k(L)x = 2L\tilde{x}_{k-1} - \tilde{x}_{k-2}$ with $\tilde{x}_0 = x$ and $\tilde{x}_1 = Lx$. The polynomial order K ensures that the filters are strictly localized within K hops from the central node and σ is the non-linear activation function. The final output y is given by

$$y = \text{softmax}(Vz + b) \quad (16)$$

where b is the bias, Vz is the dot product of V , a vector which component is the weights, and z , a vector consisting of the inputs and softmax is the activation.

4.3 Ensemble learning

Ensemble learning methods train several baseline models and use some rules to combine them together to make predictions. The ensemble learning methods have gained popularity because of their superior prediction performance in practice. The performance of a particular learner depends on how effective its searching strategy is in approximating the optimal predictor defined by the true data generating mechanism [83]. The performance of a particular learner depends on how

effective its searching strategy is in approximating the optimal predictor defined by the true data generating distribution [84]. In theory, the relative performance of various learners will depend on the model assumptions and the data distribution. In practice, the performance of the learners will depend on the sample size, dimensionality, parameters etc. Thus, it is generally impossible to know a priori which learner would perform best given the dataset and prediction problem. The idea of ensemble learning, which combines classifiers instead of selecting a single classifier, is well studied in the literature [85][86]. Recently the Super Learner Selector has shown success in areas including precision medicine [87]. In addition, He et. Al., 2015 [6] also won the ImageNet detection task with ensemble of 3 residual network models. We employ a similar idea in building our model with graph construction that can lead to remarkably different classifiers.

We built an ensemble with Q number of pGCN having weights $\{W^q\}$ where W^q denotes the weight of the q th network. Our approach is illustrated in figure 9. During training, the population graph was trained in individual networks by dropping edges uniformly at edge dropout rate γ . At each training iteration, pGCN instantiation corresponds to a network invoked by a randomized graph ensemble. Note that we dropped only edges from the population graph and did not sub-sample data like other conventional methods for ensemble learning [88] because dropping edges is a novel and flexible technique to alleviate over-smoothing and overfitting issue in deep graph convolutional networks. Conceptually, this is similar to the idea of using multiple attention heads in the recently successful attention models in deep learning. In order to find a consensus among the predictions $\{y^q\}$, we computed the average of the class conditional probabilities of component networks. The disease class label c^* is given by

$$c^* = \operatorname{argmax}_c \left\{ \sum_{q=1}^Q p(y^q = c | x, W^q) \right\} \quad (17)$$

As the literature suggests [89] unweighted averaging might be reasonable ensemble for similar base learners with comparable performances.

Figure 10. Ensemble learning of pGCN for classification of brain diseases. Q population graphs were generated independently by dropping edges at a rate γ and a consensus combination of predictions was obtained.

Algorithm for the proposed model:

L is a normalization of weighted graph $G = \{X, W\}$

1. **function NETWORK** (GRAPH MODEL FN, W, X, L)
 2. $Z \leftarrow X$
 3. GraphModels $\leftarrow []$
 4. for $q = 1$ to Q do
 5. GraphModels.append (Graph Model FN, W, X, L)
 6. Output tensors.concatenate
 7. $Z \leftarrow LX$
 8. $Z \rightarrow \text{activation (softmax)}$
 9. **return Z**
 10. final predictions $[]$
 11. **Output y :** mean (final predictions)
-

Table 3. Algorithm for ensemble learning of pGCN.

4.4 Experiments and Results

We evaluated ensemble learning approach to pGCN for classification of AD and PD with data gathered in ADNI and PPMI databases, respectively. We had sample sizes $N = 531$ from ADNI and $N = 827$ from PPMI. The number of features $M = 6786$ available as samples for training. The datasets were split into the ratio of 60:20:20 and 10-fold cross validation was performed. We performed and evaluated on binary classification of subjects: AD vs NC, MCI vs NC, AD vs MCI, and PD vs NC.

We implemented the pGCN architecture as described in [16] which in turn utilized the original implementation from [55]. The training set consists of labelled subset of graph nodes. The test set consists of the remaining unlabelled graph nodes and influences the convolutions of labelled samples making it a semi-supervised classification scheme. After training the pGCN, the softmax activations are computed on the test set and the unlabelled nodes are assigned labels. The network was a fully connected GCN with 5 hidden layers and Chebyshev's polynomial of degree $K = 3$ to approximate the graph Fourier transform. The learning parameters were slightly different for ADNI and PPMI and the best parameters were found empirically. For ADNI, dropout = 0.02 and learning rate = 0.01 while for PPMI, dropout = 0.03 and learning rate = 0.005. Both the networks are trained for 200 epochs each. To build the population graph, threshold = 2 was used to determine similar ages.

Figures 11-14 show variation of the ensemble pGCN accuracy with hyperparameters for all the binary classification tasks. It is observed that there are several combinations of hyperparameters that could boost the performance of pGCN. For AD vs NC and AD vs MCI, it is observed that the model performs at its best with an ensemble of 25 graphs while for MCI vs NC and PD vs NC, the highest accuracy is achieved with 20 graphs in the ensemble. A simple optimization without the use of any additional information is able to boost the performance of the model.

The classification results for AD and PD are reported in table 4. The best accuracies reported for ensemble pGCN were achieved by varying the number Q of networks and edge dropout rate γ , empirically looking for

best accuracies. The corresponding parameter values for best performing ensembles are reported in table 4. We compared our results with RBF SVM. SVM was implemented with radial basis function (RBF) for kernel. During the training process, best parameters (kernel, penalty parameter (C), and gamma) were selected using grid search approach. We selected gamma and C; 5, 0.52 for AD vs NC, 5, 0.50 for NC vs MCI, 5, 0.51 for MCI vs AD, and C; 5, 0.52 for PD vs NC. For further assessment, we also compared our results with a modified version of the BrainNetCNN [65] trained on connectivity matrices. The architecture is formed from an E2E layer followed by an E2N layer and two fully connected (FC) layers. Note that for better convergence during training, a batchnorm operation was added at the end of each layer. These layers are then followed by two FC layers and then the output layer. To obtain a binary classification (e.g. AD vs NC), a sigmoid function is added to the output layer on the connectivity matrices. In order to ensure that configurations with best performance were retained, 10-fold cross validation was utilized. The loss function is the well-known cross-entropy which we minimize with the Adam optimizer. For the learning procedure, along with the 10-fold cross-validation, the dataset was split into three sets: training (70%), validation (15%) and test (15%).

Ensemble pGCN improves the performance of pGCN by 2.7% for AD vs NC and 2.9% for PD vs NC classification. As seen, pGCN performs better with ensemble learning and outperforms BrainNetCNN and SVM notable in all classification tasks.

Figure 11. Analysis of performance measure with variation of hyperparameters: size of ensemble (Q) and edge dropout rate γ for AD vs NC classification.

Figure 12. Analysis of performance measure with variation of hyperparameters: size of ensemble (Q) and edge dropout rate γ for MCI vs NC classification.

Figure 13. Analysis of performance measure with variation of hyperparameters: size of ensemble (Q) and edge dropout rate γ for AD vs MCI classification.

Figure 14. Analysis of performance measure with variation of hyperparameters: size of ensemble (Q) and edge dropout rate γ for PD vs NC classification

Model	AD vs NC	MCI vs NC	AD vs MCI	PD vs NC
RBF SVM	80.4 ± 2.3	79.4 ± 3.1	77.8 ± 1.5	78.3 ± 2.1
BrainNetCNN	80.7 ± 1.1	76.3 ± 2.3	78.8 ± 1.1	78.6 ± 1.3
pGCN	82.0 ± 1.6	80.2 ± 0.4	79.7 ± 2.2	80.7 ± 1.5
Ensemble pGCN	84.7 ± 1.7	81.3 ± 1.6	80.7 ± 2.7	83.6 ± 2.3
(Q, γ)	(25, 0.35)	(20, 0.4)	(25, 0.35)	(20, 0.35)

Table 4. Comparative analysis of performance measure (accuracy %) of pGCN with ensemble learning for DTI

4.5 Discussion

In this chapter, we proposed a method for group-level population diagnosis that exploits the novel concept of spectral graph convolutions. We modelled population graphs incorporating subject-specific imaging data from DTI and pairwise interactions described using phenotypic and demographic information. The population graph is used to train a GCN model (pGCN) for node classification, learning on a subset of labelled nodes and evaluating on the rest. We show how our population graph formulation yields more accurate results. Our experiments on two large and challenging databases (ADNI and PPMI) confirm our initial hypothesis about the importance of population graph formulation and contextual pairwise information for classification, as we obtain state of the art performance with 82% (ADNI) and 80.7% (PPMI) accuracy. We used ensemble learning approach by combining multiple pGCN classifiers invoked by a randomized graph ensemble and observed that ensemble pGCN improves the performance of pGCN by 2.7% for AD vs NC and 2.9% for PD vs NC classification. We obtained 84.7% (ADNI) and 83.6% (PPMI) accuracy, corresponding to increases of 4.3% and 5.3% with respect to SVM classifier.

In addition, we perform significance test on the classification performances between the models by using the standard paired t -test under the significance level of 95 %. Table 5 shows the results of t -test between pGCN and RBF SVM; ensemble pGCN and pGCN.

Compared model	$p - value$			
	AD vs NC	MCI vs NC	AD vs MCI	PD vs NC
pGCN	1.3×10^{-5}	1.1×10^{-5}	1.4×10^{-5}	1.3×10^{-5}
Ensemble pGCN	1.7×10^{-5}	5×10^{-5}	7×10^{-5}	1.5×10^{-5}

Table 5. Significance test on the classification accuracies between pGCN and SVM; pGCN and ensemble pGCN

Ensemble pGCN is significantly different and better than pGCN and likewise pGCN is better than RBF SVM, which justifies our approach.

Chapter 5

Combining DTI and sMRI scans

In chapter 4, we have discussed the implementation of ensemble pGCN using DTI data. In this chapter, we present the methodology to fuse DTI and sMRI data. We discuss the experiments and results of ensemble pGCN using the combination of DTI and sMRI data.

“Multimodality” is a widely used phrase in the context of brain imaging studies as multiple modalities of brains are acquired to characterize brain tissues. Collecting multiple modalities of brain imaging data from the same individual has been common in brain imaging studies as the information of a single modality is often insufficient for making accurate inferences. There is increasing evidence that multimodal brain imaging studies can help provide a more complete understanding of the brain and its disorders. For example, multimodal images can accurately inform us about how brain structure shapes brain function, in which way they are impacted by psychopathology and which functional or structural aspects of physiology could drive human behavior and cognition.

Previous studies in literature have shown that superior classification results for sMRI measures when compared to DTI measures for disease classification [90]. In contrast, in the regions of interest (ROIs) for early stages of AD such as MCI, hippocampal separation showed is more evident using DTI markers. Until today, the question remains unresolved as to what extent the two imaging techniques are beneficial in disease classification and how multiple modalities are effectively combined.

In this chapter we present the fusion of sMRI and DTI to supplement each other for feature representation via the local weighted clustering coefficients (LWCC) [91]. The data processing of DTI and sMRI scans is explained in chapter 3. We extract features from brain regions of both sMRI and DTI from one subject and concatenate them via LWCC to achieve multimodal fusion for feature representation.

5.1 DTI features for fusion

Most of the existing researches use the upper half feature representation matrix of features of the subject as the feature vector. However, such an operation will bring the high feature dimension and destroy the original structure of the network. To address this problem, for connectivity matrix $S = (S_{I,J}) \in \mathbf{R}^{N \times N}$ where $S_{I,J}$ denotes structural connectivity between brain regions I and J , we extracted the clustering coefficients c_I of brain region I :

$$c_I = \frac{3\sum_{J \in \varepsilon_I} (m_{I,J})^{\frac{1}{3}}}{|\varepsilon_I|(|\varepsilon_I| - 1)} \quad (18)$$

where, ε_I denotes the set of vertices directly connected to vertex I , and $|\varepsilon_I|$ is the number of elements in ε_I . By using clustering coefficient as the nodal feature, we not only reduce the dimensionality of connectivity features but also transform the connectivity feature into nodal features.

5.2 sMRI features for fusion

From sMRI, we extracted intensity features, mean intensity μ_I and standard deviation σ_I for each brain region I . Volumetric Brain Mapping (VBM) [80] has been applied to study volumetric atrophy of the grey matter in brain scans against normal controls (NC). The procedure involve the spatial normalization of subject images into a standard space, segmentation of tissue classes using a priori probability maps, smoothing to correct noise and small variations, and voxel-wise statistical tests. Statistical analysis is based on the general linear model (GLM) to describe the data in terms of experimental and confounding effects, and residual variability. Classical statistical inference is used to test hypotheses that are expressed in terms of GLM estimated regression parameters. This computation of a given contrast provides a statistical parametric map, which is thresholded according to the Random Field theory [92] to obtain the clusters of voxels where the null hypothesis of no effect cannot be accepted. These voxels represent the regions of the brain maximally affected by the disease, and we assume that their associated information will be discriminant of AD and PD patients and control subjects.

5.3 Fusion of features

The first feature extraction process computes the mean and standard deviation of grey matter intensity values of each voxel location cluster, we denote these features as (μ_I, σ_I) where μ_I is the mean intensity and σ_I is the standard deviation. Hence, we have one structural connectivity marker and two sMRI markers. For each region I , three features are input to the pGCN as a vector (c_I, μ_I, σ_I) .

Figure 15. Graph construction approach using combination of sMRI and DTI imaging features and non-imaging phenotypic measures of the subjects

In chapter 4 we discussed the strategy to build the population graph W using age and similarity measures of the subjects. The same strategy has been used to build the population graph here.

5.4 Experiments and Results

We performed the same experiments as done in Chapter 3 with different inputs. We evaluated ensemble learning approach to pGCN for classification of AD and PD with data gathered in ADNI and PPMI databases, respectively. We had sample sizes $N = 531$ from ADNI and $N = 827$ from PPMI. The number of features $M = 6786$ available as samples for training. The datasets were split into the ratio of 60:20:20 and 10-fold cross validation was performed. We performed and evaluated on binary classification of subjects: AD vs NC, MCI vs NC, AD vs MCI, and PD vs NC.

Figure 16 shows variation of ensemble pGCN (DTI + sMRI) accuracy with hyperparameters for the combination of sMRI and DTI features for all the binary classification tasks. When compared to Figures 11-14 in chapter 4, it is observed that, the performance measures for all the ensembles across all the tasks have consistently improved with the combination of features which validates our approach.

The classification results for AD and PD are reported in Table 6. The best accuracies reported for ensemble pGCN were achieved by varying the number Q of networks and edge dropout rate γ , empirically looking for best accuracies. The corresponding parameter values for best performing ensembles are reported in Table 6. To compare our results, RBF SVM and BrainNetCNN baseline architectures were implemented in a similar manner as described in sec 4.4.

Input	Model	AD vs NC	MCI vs NC	AD vs MCI	PD vs NC
DTI	RBF SVM	80.4 ± 2.3	79.4 ± 3.1	77.8 ± 1.5	78.3 ± 2.1
	BrainNetCNN	80.7 ± 1.1	76.3 ± 2.3	78.8 ± 1.1	78.6 ± 1.3
	pGCN	82.0 ± 1.6	80.2 ± 0.4	79.7 ± 2.2	80.7 ± 1.5
	Ensemble pGCN	84.7 ± 1.7	81.3 ± 1.6	80.7 ± 2.7	83.6 ± 2.3
	(Q, γ)	(25, 0.35)	(20, 0.4)	(25, 0.35)	(20, 0.35)
sMRI	RBF SVM	68.0 ± 0.2	64.4 ± 1.1	65.3 ± 1.5	69.8 ± 1.8
	BrainNetCNN	70.1 ± 1.1	69.3 ± 2.1	66.3 ± 1.1	67.2 ± 1.5
	pGCN	64.0 ± 0.6	68.2 ± 0.4	60.7 ± 0.3	65.8 ± 2.2
	Ensemble pGCN	65.0 ± 0.7	69.3 ± 1.6	63.4 ± 1.2	67.6 ± 0.4
	(Q, γ)	(25, 0.35)	(20, 0.4)	(25, 0.35)	(20, 0.35)
DTI + sMRI	RBF SVM	81.8 ± 0.3	80.5 ± 0.9	79.9 ± 0.7	80.2 ± 0.2
	BrainNetCNN	79.9 ± 1.1	79.1 ± 1.3	79.4 ± 2.1	80.0 ± 2.3
	pGCN	86.0 ± 1.2	83.2 ± 0.5	81.3 ± 1.1	84.8 ± 0.7
	Ensemble pGCN	87.8 ± 1.9	84.4 ± 2.2	83.0 ± 2.1	88.0 ± 1.9
	(Q, γ)	(25, 0.35)	(20, 0.4)	(25, 0.35)	(20, 0.35)

Table 6. Comparative analysis of performance measure (accuracy %) of the proposed model

Figure 16. Comparative analysis of performance measures of pGCN combining DTI and sMRI features with variation of hyperparameters: size of ensemble (Q) and edge dropout rate γ . (a) AD vs NC; (b) MCI vs NC (c) AD vs MCI (d) PD vs NC

5.5 Discussion

In this chapter, we proposed a method for combining DTI and sMRI scans as image information into the pGCN. We performed experiments as similar to the ones in the previous chapter but with combined input features from DTI and sMRI. We compared the results of pGCN with only DTI input and observed that ensemble pGCN with fusion of DTI and sMRI performed at 87.8% (ADNI) and 88% (PPMI) accuracy, corresponding to increases of 3% and 4.5% with respect to ensemble pGCN using only DTI data. The improvement in performance was consistent across all classification tasks.

The complexity of the human brain coupled with the incomplete measurement provided by existing imaging technology makes multimodal fusion essential in order to provide a key to finding the missing links in

complex mental illness like AD and PD. The fusion of grey matter features from sMRI and white matter connectivity features from DTI has boosted the classification performance.

In addition, we perform significance test on accuracies of classification tasks between the models with different imaging inputs by using the standard paired t -test under significance level of 95 %. Table 7 shows the results of t -test of ensemble pGCN between only DTI input and the fusion of DTI and sMRI; only DTI input and only sMRI input.

Ensemble pGCN model	$p - value$			
	AD vs NC	MCI vs NC	AD vs MCI	PD vs NC
DTI only	1.1×10^{-5}	1.2×10^{-5}	1.7×10^{-5}	1.1×10^{-5}
DTI + sMRI	3×10^{-5}	1.7×10^{-4}	1.2×10^{-4}	5×10^{-5}

Table 7. Significance test on the classification accuracies between pGCN and SVM; pGCN and ensemble pGCN

As we can see from Table 7, classification performance of ensemble pGCN model with multimodal fusion of sMRI and DTI is significantly different than ensemble pGCN with only DTI input and performs better across all classification tasks which justifies our method.

Chapter 6

Population graphs from Transcriptome data

In chapters 4 and 5, we have seen the population graph built from demographic and age information of the patients. In this chapter we discuss the population graphs built from the transcriptome data and compare the results of ensemble pGCN using gene expression with that of ensemble pGCN built using demographic information.

Biological data are often represented as networks, where nodes represent biologically relevant entities (typically genes or proteins) and edges represent relationships between these entities (e.g., regulation, interaction). Although most biological networks represent gene or protein networks it is often relevant to represent the data as Patient Similarity networks (PSN)[93]. In these networks, nodes represent patients and edges represent similarities between patients' profiles. These networks can be used to group patients and to associate these groups with distinct clinical features. It was observed for instance that, within a network obtained by integrating multiple omics data, cancer patient clusters had different clinical outcomes, including different overall survival [94]. Similarly, a network topology-based analysis of diabetes patient genotypes revealed that patients can be clustered into three groups and that these groups have distinct clinical features, including different comorbidities [95].

In the current study, we hypothesize that clinically relevant information is encoded within PSN built from genomics data. To investigate whether we can use this topological information to predict patient clinical outcome, we incorporate the gene expression of patients as the population graph G . In chapters 4 and 5, age and gender phenotypic measures were used to construct the population graph. In this chapter we present the graph construction of pGCN model using the gene expression data of ADNI patients. In this chapter, the combination of

both DTI and sMRI data were used in modelling. We compare the model performance of pGCN using gene expression with that of the approach used in chapters 4 and 5. We also analyse the correlation of the gene expression and the phenotypic measures of the patients.

In our approach, we transform the gene expression data into networks and then train pGCN with network topological data. Our results indicate that the performance of classifiers trained with topological data is slightly better to the performance of the models built on demographic data. Our approach represents a complementary strategy to analyze and integrate large collections of gene expression data with brain imaging data. The data processing of gene expression data has been discussed in section 3.5.

6.1 Population graphs for gene expression data

Top 5% features were selected across the dataset. We used the reduced datasets to infer Patient Similarity Networks (PSN), graphs in which a node represents a patient and an edge between two nodes represents the similarity between the two profiles of the corresponding patients. These graphs are built by computing the Pearson's correlation coefficients between all profiles pairwise. Similarity between two patients p and q , are represented by the weights $w_{p,q}$ between the nodes representing two patients and is evaluated with a distance measure between the features:

$$w_{(p,q)} = distance(r_p, r_q) \quad (19)$$

where r_p and r_q denotes the set of transcriptomic features of patients p and q respectively. The PSN of a set $P = \{p\}_{p=1}^N$ of N patients is represented by the graph $G = (P, W)$ where $W = (w_{(p,q)})_{p,q \in P}$ denotes the similarity matrix of patients. The distances between features of the dataset were determined by computing the Pearson's correlation coefficient between the patients:

$$distance(r_p, r_q) = \frac{N(\Sigma_p \Sigma_q r_p r_q) - (\Sigma_p r_p)(\Sigma_q r_q)}{\sqrt{(N\Sigma_p r_p^2 - (\Sigma_p r_p)^2)(N\Sigma_q r_q^2 - (\Sigma_q r_q)^2)}} \quad (20)$$

The final strategy was to normalize and rescale the correlation values to represent positive edge weights using Weighted Correlation Network Analysis (WGCNA) algorithm [96]. The WGCNA enforces scale-freeness of the PSN by making its nodal degree distribution follow a power law, or at least asymptotically, and is more robust to noise. WGCNA has a parameter β which has to be tuned. The search space used was between 1 and 10, with intervals of 0.01. The smallest β that gives a scale-free index of at least 0.9 was used.

Figure 17. Graph construction approach using gene expression data of the patients

6.2 Experiments and Results

We evaluated ensemble learning approach to pGCN for classification of AD with data gathered from ADNI database. We had sample sizes $N = 115$ from ADNI. The variance method experimented in data pre-processing performed better than wilcoxon analysis. So, we extracted $M = 2470$ features using variance method. These are available as samples for training. The datasets were split into the ratio of 60:20:20 and 10-fold cross validation was performed. We performed and evaluated on binary classification of subjects: AD vs NC, MCI vs NC and AD vs MCI.

We implemented the pGCN architecture as described in [16] which in turn utilized the original implementation from [55]. The results reported are for a reduced dataset of 115 samples as opposed to 531 samples of data experimented in previous chapters because gene expression profile of a patient is the same across different samples and the transcriptome data was not available for all the patients in ADNI database. The training set consists of labelled subset of graph nodes. The test set consists of the remaining unlabelled graph nodes and influences the convolutions of labelled samples making it a semi-supervised classification scheme. After training the pGCN, the softmax activations are computed on the test set and the unlabelled nodes are assigned labels. The network was a fully connected GCN with 5 hidden layers and Chebyshev’s polynomial of degree $K = 3$ to approximate the graph Fourier transform. The learning parameters for ADNI were found empirically. The dropout = 0.01 and learning rate = 0.03. The network was trained for 200 epochs each. The classification results for AD are reported in Table 8. To compare our results, RBF SVM and BrainNetCNN architectures were implemented in a similar manner as described in sec 4.4.

Input	Model	AD vs NC	MCI vs NC	AD vs MCI
DTI + sMRI	RBF SVM	76.8 ± 1.3	75.3 ± 1.4	75.0 ± 2.1
	BrainNetCNN	77.1 ± 1.1	75.2 ± 2.0	73.0 ± 1.5
	Ensemble pGCN	80.4 ± 2.1	78.5 ± 1.7	79.8 ± 2.3
	(demo info)			
	(Q, γ)	(25, 0.35)	(20, 0.4)	(25, 0.35)
	Ensemble pGCN	83.1 ± 1.3	79.2 ± 1.5	81.2 ± 2.0
	(gene info)			
	(Q, γ)	(20, 0.4)	(20, 0.4)	(25, 0.35)

Table 8. Comparative analysis of performance measure (accuracy %) of ensemble pGCN model using gene expression data.

Figure 18 shows variation of ensemble pGCN (DTI + sMRI) accuracy with hyperparameters for gene expression data across all the binary classification tasks.

Figure 18. Comparative analysis of performance measures of Ensemble pGCN with gene info DTI and sMRI features with variation of hyperparameters: size of ensemble (Q) and edge dropout rate γ . (a) AD vs NC; (b) MCI vs NC (c) AD vs MCI

6.3 Discussion

In this chapter, we incorporated patients' gene expression data into population graph instead of using their phenotypic measures and demographic information. We modelled the new population graph G into pGCN and performed experiments with ensemble pGCN using both DTI and sMRI inputs. We exploited the maximum information available from the ADNI database to the best use into graph modelling. As observed, the

ensemble pGCN built with gene expression information of patients performs slightly better than ensemble pGCN model using demographic information of the patients across all classification tasks. It shows that patients' transcriptome data can be exploited to achieve better classification results even with simple and small size datasets.

In addition, we performed significance test on accuracies of classification tasks between the models with different imaging inputs by using the standard paired t -test under the significance level of 95 %. Table 9 shows the results of t -test of ensemble pGCN between gene expression data and demographic data.

Ensemble pGCN (DTI + sMRI)	$p - value$		
	AD vs NC	MCI vs NC	AD vs MCI
Gene expression data	3×10^{-4}	1.7×10^{-4}	1.2×10^{-3}

Table 9. Significance test on the classification accuracies between ensemble pGCN with gene expression data and ensemble pGCN with demographic data.

Although the difference in accuracies between both the models is less, ensemble pGCN model using gene expression data is significantly different than ensemble pGCN model using demographic information of patients. The $p - values$ reported indicate that the classification performance is stable across all classification tasks.

Figure 19. Heatmaps of the two population graphs. (a) population graph W of demographic data; (b) population graph G of gene expression data

Figure 19 shows the comparison of the population graphs built using demographic data and gene expression data. In figure 19(a), we can observe the correlation between subjects based on age and gender groups which is the demographic information. Figure 19(b) shows the correlation between subjects based on their gene expression. We can observe that population graph W is more correlated than G .

Chapter 7

Conclusion

7.1 Summary

In this study, we implemented population based GCNs using the graphical structure of population graph for AD and PD classification. We used the natural mechanism of GCN for blending phenotypic information such as age and gender with connectome features. We showed that learning pGCNs with a simple ensemble strategy has improved brain disease classification using structural connectome and anatomical features. Machine learning method using RBF SVM performed poorly on both the datasets (ADNI and PPMI).

Using pGCN with largest ensemble of 20-25 graphs, the accuracies of AD and PD classification were notably improved. Though learning of individual network may be poor, consensus decisions over individual weak learners improved overall accuracy. We observed that dropping 10 to 20% of the edges was effective for graphs performing better individually while for the poor learning graphs, dropping 3 to 5% was enough to boost the performance.

We combined the structural connectome features of DTI and anatomical features of sMRI to achieve multimodal fusion for better feature representation. We observed improved performances across all classification task with combined features from two modalities. We demonstrated ensemble pGCN on two disease datasets: ADNI and PPMI for AD and PD classification, respectively.

We also incorporated gene expression of patients from ADNI into graph construction and observed that ensemble pGCN performs slightly better in classification when compared to the ensemble pGCN using demographic data. We inferred that age and gender information of patients is more desirable than their gene expression information for AD and PD classification.

7.2 Discussion

In this thesis, we modelled population graphs incorporating subject-specific imaging data from brain imaging data and pairwise interactions described using phenotypic and demographic information. The population graph was used to train a GCN model (pGCN) for node classification, learning on a subset of labelled nodes and evaluating on the rest. We showed how our population graph formulation yielded more accurate results. The classification results using pGCN in chapter 4 showed substantial improvement when compared to the RBF SVM baseline method. Our experiments on two large and challenging databases (ADNI and PPMI) confirm our initial hypothesis about the importance of population graph formulation and contextual pairwise information for classification, as we obtain state of the art performance with 82% (ADNI) and 80.7% (PPMI) accuracy.

While recent graph-based predictive models have shown that incorporating contextual non-imaging features such as patient sex, location of the scan etc. can improve performance, they end up being extremely sensitive to the choice of the population graph. To circumvent this challenge, we used ensemble learning approach by combining multiple pGCN classifiers invoked by a randomized graph ensemble and observed that ensemble pGCN boosted the performance of pGCN by 2.7% for AD vs NC and 2.9% for PD vs NC classification. We obtained 84.7% (ADNI) and 83.6% (PPMI) accuracy, corresponding to increases of 4.3% and 5.3% with respect to SVM classifier.

We proposed a method for combining DTI and sMRI scans as image information into population graph. In chapter 5, we performed experiments with different inputs for. We compared the results of pGCN with only DTI input and observed that ensemble pGCN with fusion of DTI and sMRI performed at 87.8% (ADNI) and 88% (PPMI) accuracy, corresponding to increases of 3% and 4.5% with respect to ensemble pGCN using only DTI data. The improvement in performance is consistent across all classification tasks. The complexity of the human brain coupled with the incomplete measurement provided by existing imaging technology makes multimodal fusion essential in order to provide a key

to finding the missing links in complex mental illness like AD and PD. The advantage of fusing grey matter features from sMRI and white matter connectivity features from DTI has improved the classification performance.

We incorporated patients' gene expression data into population graph instead of using their demographic information. Experiments proved that gene expression data of patients helps in classification of AD and PD. We observed, the ensemble pGCN built with gene expression information of patients performs slightly better than ensemble pGCN model using demographic information of the patients across all classification tasks. It shows that patients' transcriptome data can be exploited to achieve better classification results even with simple and small size datasets. We studied the correlation of subjects in the population graphs built using demographic information W and gene expression information G and observed that population graph with demographic information of patients is more correlated than the population graph with genetic information of patients which suggests that age and gender information of patients is more desirable than their gene expression information for AD and PD classification.

7.3 Future work

This study opens several new avenues of future work. AD and PD are the two prominent brain diseases and we have implemented our work to classify these diseases. However, this work can be extended to other dementias like schizophrenia. Any kind of patient data that has graph structure can be well exploited and studied using GCNs. It would be interesting to combine genetic information and the demographic information of patients into graph modelling for classification. It would also be interesting to introduce functional connectome (fMRI) into pGCN modelling in future. It will be interesting to study the ensemble pGCN classifier by representing the features from both structural and functional connectome of brain i.e. fMRI and DTI brain scans along with anatomical brain features (sMRI). This work could also be extendable for other datasets and it would be interesting to see theoretical analysis of our approach.

Low-dimensional vector embeddings of nodes in large graphs are inherently transductive and can only generate embeddings for a single fixed graph. These transductive approaches do not efficiently generalize to unseen nodes (e.g., in evolving graphs), and these approaches cannot learn to generalize across different graphs. In contrast, Graph Neural Network (GNN) framework leverages node attribute information to efficiently generate representations on previously unseen data. In future it would be interesting to study how GNNs can be used to leverage the available data towards improving classification results in highly class imbalanced problems.

Bibliography

- [1] P. S. Sachdev, L. Zhuang, N. Braidy, and W. Wen, “Is Alzheimer’s a disease of the white matter?,” *Current Opinion in Psychiatry*, vol. 26, no. 3, pp. 244–251, May-2013, doi: 10.1097/YCO.0b013e32835ed6e8.
- [2] F. Caso *et al.*, “White Matter Degeneration in Atypical Alzheimer Disease,” *Radiology*, vol. 277, no. 1, pp. 162–172, Oct. 2015, doi: 10.1148/radiol.2015142766.
- [3] S. Guo *et al.*, “Combining multi-modality data for searching biomarkers in schizophrenia,” *PLoS One*, vol. 13, no. 2, p. e0191202, Feb. 2018, doi: 10.1371/journal.pone.0191202.
- [4] J. Sui *et al.*, “Combination of Resting State fMRI, DTI, and sMRI Data to Discriminate Schizophrenia by N-way MCCA + jICA,” *Front. Hum. Neurosci.*, vol. 7, p. 235, May 2013, doi: 10.3389/fnhum.2013.00235.
- [5] S. Lorio, F. Sambataro, A. Bertolino, B. Draganski, and J. Dukart, “The Combination of DAT-SPECT, Structural and Diffusion MRI Predicts Clinical Progression in Parkinson’s Disease,” *Front. Aging Neurosci.*, vol. 11, p. 57, Mar. 2019, doi: 10.3389/fnagi.2019.00057.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [7] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2261–2269, Aug. 2016.
- [8] K. M. Borgwardt, C. Soon Ong, S. Schönauer, S. V N Vishwanathan, A. J. Smola, and H.-P. Kriegel, “Protein function prediction via graph kernels,” *BIOINFORMATICS*, vol. 21, pp.

- 47–56, 2005, doi: 10.1093/bioinformatics/bti1007.
- [9] P. Schweitzer PASCAL *et al.*, “Weisfeiler-Lehman Graph Kernels Nino Shervashidze Kurt Mehlhorn Karsten M. Borgwardt,” 2011.
 - [10] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, “Classification of Alzheimer Disease on Imaging Modalities with Deep CNNs Using Cross-Modal Transfer Learning,” in *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, 2018, vol. 2018-June, pp. 345–350, doi: 10.1109/CBMS.2018.00067.
 - [11] X. Zhang, L. He, K. Chen, Y. Luo, J. Zhou, and F. Wang, “Multi-View Graph Convolutional Network and Its Applications on Neuroimage Analysis for Parkinson’s Disease,” *Annu. Symp. proceedings. AMIA Symp.*, vol. 2018, pp. 1147–1156, 2018.
 - [12] D. R. Ktena, Sofia Ira, Sarah Parisot, Enzo Ferrante, Martin Rajchl, Matthew Lee, Ben Glocker, “Metric learning with spectral graph convolutions on brain connectivity networks,” *Neuroimage*, vol. 169, pp. 431–442, 2018, doi: 10.1016/j.neuroimage.2017.12.052.
 - [13] S. Parisot, S. I. Ktena, and E. Al., “Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer’s disease,” *Med. Image Anal.*, vol. 48, pp. 117–130, 2018, doi: 10.1016/j.media.2018.06.001.
 - [14] W. Lin *et al.*, “Convolutional neural networks-based MRI image analysis for the Alzheimer’s disease prediction from mild cognitive impairment,” *Front. Neurosci.*, vol. 12, no. NOV, Nov. 2018, doi: 10.3389/fnins.2018.00777.
 - [15] C. Y. Wee, C. Liu, A. Lee, J. S. Poh, H. Ji, and A. Qiu, “Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations,” *NeuroImage Clin.*, vol. 23, p. 101929, Jan. 2019, doi: 10.1016/j.nicl.2019.101929.
 - [16] S. Parisot *et al.*, “Spectral graph convolutions for population-based disease prediction,” *Lect. Notes Comput. Sci. (including*

- Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics*), vol. 10435 LNCS, no. 319456, pp. 177–185, 2017, doi: 10.1007/978-3-319-66179-7_21.
- [17] Y. Taki *et al.*, “Correlations among brain gray matter volumes, age, gender, and hemisphere in healthy individuals,” *PLoS One*, vol. 6, no. 7, p. e22734, Jul. 2011, doi: 10.1371/journal.pone.0022734.
 - [18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, 2014.
 - [19] H. Yang, C. Yuan, J. Xing, and W. Hu, “Diversity encouraging ensemble of convolutional networks for high performance action recognition,” in *Proceedings - International Conference on Image Processing, ICIP*, 2018, doi: 10.1109/ICIP.2017.8296802.
 - [20] P. Seeman and B. Madras, *Imaging of the Human Brain in Health and Disease*. Elsevier Inc., 2013.
 - [21] Z. Balážová, M. Nováková, A. Minsterová, and I. Rektorová, “Structural and Functional Magnetic Resonance Imaging of Dementia With Lewy Bodies,” in *International Review of Neurobiology*, vol. 144, Academic Press Inc., 2019, pp. 95–141.
 - [22] “T1-weighted MRI Benchmark (Brain Image Segmentation) | Papers With Code.” [Online]. Available: <https://paperswithcode.com/sota/brain-image-segmentation-on-t1-weighted-mri>. [Accessed: 16-Dec-2020].
 - [23] Q. Lyu, G. Wang, C. / Bme, and / Soe, “Quantitative MRI- Absolute T1, T2 and Proton Density Parameters from Deep Learning.”
 - [24] P. Sati, I. C. George, C. D. Shea, M. I. Gaitán, and D. S. Reich, “FLAIR*: A combined MR contrast technique for visualizing white matter lesions and parenchymal veins,” *Radiology*, vol. 265, no. 3, pp. 926–932, 2012, doi: 10.1148/radiol.12120208.
 - [25] A. M. Michael, E. Evans, and G. J. Moore, “Influence of group on

- individual subject maps in SPM voxel based morphometry,” *Front. Neurosci.*, vol. 10, no. DEC, 2016, doi: 10.3389/fnins.2016.00522.
- [26] “Post-processing of structural MRI for individualized diagnostics - Martin - Quantitative Imaging in Medicine and Surgery.” [Online]. Available: <http://qims.amegroups.com/article/view/5893>. [Accessed: 16-Dec-2020].
- [27] “Structural Magnetic Resonance Imaging - an overview | ScienceDirect Topics.” [Online]. Available: <https://www.sciencedirect.com/topics/neuroscience/structural-magnetic-resonance-imaging>. [Accessed: 16-Dec-2020].
- [28] A. Filler, “The History, Development and Impact of Computed Imaging in Neurological Diagnosis and Neurosurgery: CT, MRI, and DTI,” *Nat. Preced.*, pp. 1–76, 2009, doi: 10.1038/npre.2009.3267.4.
- [29] M. Bozzali, A. C.-M. resonance imaging, and undefined 2007, “Diffusion tensor MRI to investigate dementias: a brief review,” *Elsevier*.
- [30] P. Mukherjee, J. I. Berman, S. W. Chung, C. P. Hess, and R. G. Henry, “Diffusion Tensor MR Imaging and Fiber Tractography: Theoretic Underpinnings,” *Am Soc Neuroradiol.*, doi: 10.3174/ajnr.A1051.
- [31] H. P. Müller and J. Kassubek, “Diffusion tensor magnetic resonance imaging in the analysis of neurodegenerative diseases,” *J. Vis. Exp.*, no. 77, p. 50427, Jul. 2013, doi: 10.3791/50427.
- [32] S. Mori, B. J. Crain, V. P. Chacko, and P. C. M. Van Zijl, “Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging,” *Ann. Neurol.*, vol. 45, no. 2, pp. 265–269, Feb. 1999, doi: 10.1002/1531-8249(199902)45:2<265::AID-ANA21>3.0.CO;2-3.
- [33] M. Dyrba *et al.*, “Combining DTI and MRI for the automated detection of Alzheimer’s disease using a large European

- multicenter dataset,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7509 LNCS, pp. 18–28, doi: 10.1007/978-3-642-33530-3_2.
- [34] B. Lei, S. Chen, D. Ni, and T. Wang, “Discriminative learning for Alzheimer’s disease diagnosis via canonical correlation analysis and multimodal fusion,” *Front. Aging Neurosci.*, vol. 8, no. MAY, 2016, doi: 10.3389/fnagi.2016.00077.
- [35] O. Ben Ahmed, J. Benois-Pineau, M. Allard, C. Ben Amar, and G. Catheline, “Classification of Alzheimer’s disease subjects from MRI using hippocampal visual features,” *Multimed. Tools Appl.*, vol. 74, no. 4, pp. 1249–1266, 2014, doi: 10.1007/s11042-014-2123-y.
- [36] O. Ben Ahmed *et al.*, “Recognition of Alzheimer’s disease and Mild Cognitive Impairment with multimodal image-derived biomarkers and Multiple Kernel Learning,” 2016, doi: 10.1016/j.neucom.2016.08.041.
- [37] O. Ben Ahmed *et al.*, “Alzheimer’s disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex,” *Comput. Med. Imaging Graph.*, vol. 44, pp. 13–25, 2015, doi: 10.1016/j.compmedimag.2015.04.007.
- [38] M. Liu, J. Zhang, E. Adeli, and D. Shen, “Deep multi-task multi-channel learning for joint classification and regression of brain status,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10435 LNCS, pp. 3–11, doi: 10.1007/978-3-319-66179-7_1.
- [39] D. Cheng, M. Liu, J. Fu, ... Y. W.-I. C. on, and undefined 2017, “Classification of MR brain images by combination of multi-CNNs for AD diagnosis,” *spiedigitallibrary.org*.
- [40] G. T. Sutherland, D. Sheedy, and J. J. Kril, “Using Autopsy Brain Tissue to Study Alcohol-Related Brain Damage in the Genomic

- Age,” *Alcohol. Clin. Exp. Res.*, vol. 38, no. 1, pp. 1–8, Jan. 2014, doi: 10.1111/acer.12243.
- [41] C. Contet, “Gene Expression Under the Influence: Transcriptional Profiling of Ethanol in the Brain,” *Curr. Psychopharmacol.*, vol. 1, no. 4, pp. 301–314, Sep. 2012, doi: 10.2174/2211556011201040301.
- [42] R. Hitzemann *et al.*, “Genes, behavior and next-generation RNA sequencing,” *Genes, Brain Behav.*, vol. 12, no. 1, pp. 1–12, Feb. 2013, doi: 10.1111/gbb.12007.
- [43] D. Harold *et al.*, “Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer’s disease,” *nature.com*, 2009, doi: 10.1038/ng.440.
- [44] J. C. Lambert *et al.*, “Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease,” *Nat. Genet.*, vol. 45, no. 12, pp. 1452–1458, Dec. 2013, doi: 10.1038/ng.2802.
- [45] A. B. Niculescu and H. Le-Niculescu, “Convergent Functional Genomics: what we have learned and can learn about genes, pathways, and mechanisms,” *Neuropsychopharmacology*, vol. 35, no. 1, pp. 355–356, Dec. 2010, doi: 10.1038/npp.2009.107.
- [46] S. G. Mueller *et al.*, “Ways toward an early diagnosis in Alzheimer’s disease: The Alzheimer’s Disease Neuroimaging Initiative (ADNI),” *Alzheimer’s Dement.*, vol. 1, no. 1, pp. 55–66, Jul. 2005, doi: 10.1016/j.jalz.2005.06.003.
- [47] S. Lovestone *et al.*, “AddNeuroMed-The European Collaboration for the Discovery of Novel Biomarkers for Alzheimer’s Disease,” *Ann. N. Y. Acad. Sci.*, vol. 1180, no. 1, pp. 36–46, Oct. 2009, doi: 10.1111/j.1749-6632.2009.05064.x.
- [48] W. S. Liang *et al.*, “Alzheimer’s disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 11, pp. 4441–4446, Mar. 2008, doi: 10.1073/pnas.0709259105.

- [49] D. Dong, H. Wu, W. He, D. Yu, and H. Wang, “Multi-Task Learning for Multiple Language Translation.”
- [50] J. Liu, W. C. Chang, Y. Wu, and Y. Yang, “Deep learning for extreme multi-label text classification,” in *SIGIR 2017 - Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017, pp. 115–124, doi: 10.1145/3077136.3080834.
- [51] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.
- [52] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, “Large-scale Video Classification with Convolutional Neural Networks.”
- [53] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering,” *Adv. Neural Inf. Process. Syst.*, pp. 3844–3852, Jun. 2016.
- [54] M. Henaff, J. Bruna, and Y. LeCun, “Deep Convolutional Networks on Graph-Structured Data,” Jun. 2015.
- [55] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2019.
- [56] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-Based Learning Algorithms,” *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, 1991, doi: 10.1023/A:1022689900470.
- [57] D. Opitz and R. Maclin, “Popular Ensemble Methods: An Empirical Study,” 1999.
- [58] X. Zhang, B. Hu, X. Ma, and L. Xu, “Resting-state whole-brain functional connectivity networks for MCI classification using L2-regularized logistic regression,” *IEEE Trans. Nanobioscience*, vol. 14, no. 2, pp. 237–247, Mar. 2015, doi:

10.1109/TNB.2015.2403274.

- [59] P. L. Tu and J. Y. Chung, “A new decision-tree classification algorithm for machine learning,” in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 1992, vol. 1992-November, pp. 370–377, doi: 10.1109/TAI.1992.246431.
- [60] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “Training algorithm for optimal margin classifiers,” in *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, 1992, pp. 144–152, doi: 10.1145/130385.130401.
- [61] S. Tong and D. Koller, “Support Vector Machine Active Learning with Applications to Text Classification,” 2001.
- [62] S. Boughorbel, J. P. Tarel, and N. Boujemaa, “Conditionally positive definite kernels for SVM based image recognition,” in *IEEE International Conference on Multimedia and Expo, ICME 2005*, 2005, vol. 2005, pp. 113–116, doi: 10.1109/ICME.2005.1521373.
- [63] E. Reynolds, B. Callaghan, and M. Banerjee, “SVM–CART for disease classification,” *J. Appl. Stat.*, vol. 46, no. 16, pp. 2987–3007, Dec. 2019, doi: 10.1080/02664763.2019.1625876.
- [64] I. Bhattacharya and M. P. S. Bhatia, “SVM classification to distinguish Parkinson disease patients,” in *Proceedings of the 1st Amrita ACM-W Celebration of Women in Computing in India, A2CWIC’10*, 2010, doi: 10.1145/1858378.1858392.
- [65] L. O’Dwyer *et al.*, “Using Support Vector Machines with Multiple Indices of Diffusion for Automated Classification of Mild Cognitive Impairment,” *PLoS One*, vol. 7, no. 2, p. e32441, Feb. 2012, doi: 10.1371/journal.pone.0032441.
- [66] R. Cuingnet, J. A. Glaunés, M. Chupin, H. Benali, and O. Colliot, “Spatial and anatomical regularization of SVM: A general framework for neuroimaging data,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 682–696, Mar. 2013, doi: 10.1109/TPAMI.2012.142.

- [67] O. Ben Ahmed, J. Benois-Pineau, M. Allard, G. Catheline, and C. Ben Amar, “Recognition of Alzheimer’s disease and Mild Cognitive Impairment with multimodal image-derived biomarkers and Multiple Kernel Learning,” *Neurocomputing*, vol. 220, pp. 98–110, Jan. 2017, doi: 10.1016/j.neucom.2016.08.041.
- [68] and J. J. Xie, Yunyan, Zaixu Cui, Zhongmin Zhang, Yu Sun, Can Sheng, Kuncheng Li, Gaolang Gong, Ying Han, “Identification of amnesic mild cognitive impairment using multi-modal brain features: a combined structural MRI and diffusion tensor imaging study,” *J. Alzheimer’s Dis.*, vol. 47, no. 2, pp. 509–522, 2015.
- [69] A. Ebadi *et al.*, “Ensemble Classification of Alzheimer’s Disease and Mild Cognitive Impairment Based on Complex Graph Measures from Diffusion Tensor Images,” *Front. Neurosci.*, vol. 11, Feb. 2017, doi: 10.3389/fnins.2017.00056.
- [70] J. M. Mateos-Pérez, M. Dadar, M. Lacalle-Aurioles, Y. Iturria-Medina, Y. Zeighami, and A. C. Evans, “Structural neuroimaging as clinical predictor: A review of machine learning applications,” *NeuroImage: Clinical*, vol. 20. Elsevier Inc., pp. 506–522, 01-Jan-2018, doi: 10.1016/j.nicl.2018.08.019.
- [71] W. Pitts, “Some observations on the simple neuron circuit,” *Bull. Math. Biophys.*, vol. 4, no. 3, pp. 121–129, Sep. 1942, doi: 10.1007/BF02477942.
- [72] J. Kawahara *et al.*, “BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment,” *Neuroimage*, vol. 146, pp. 1038–1049, Feb. 2017, doi: 10.1016/j.neuroimage.2016.09.046.
- [73] T. N. Kipf and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*, Sep. 2016.
- [74] S. Parisot *et al.*, “Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer’s disease,” *Med. Image Anal.*, vol. 48, pp. 117–130, Aug. 2018, doi: 10.1016/j.media.2018.06.001.

- [75] A. Rushil and J. Thiagarajan, “BOOTSTRAPPING GRAPH CONVOLUTIONAL NEURAL NETWORKS FOR AUTISM Rushil Anirudh and Jayaraman J . Thiagarajan Lawrence Livermore National Laboratory,” pp. 3197–3201, 2019.
- [76] C. Y. Wee, C. Liu, A. Lee, J. S. Poh, H. Ji, and A. Qiu, “Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations,” *NeuroImage Clin.*, vol. 23, Jan. 2019, doi: 10.1016/j.nicl.2019.101929.
- [77] L. He, H. Li, S. K. Holland, W. Yuan, M. Altaye, and N. A. Parikh, “Early prediction of cognitive deficits in very preterm infants using functional connectome data in an artificial neural network framework,” *NeuroImage Clin.*, vol. 18, pp. 290–297, 2018, doi: 10.1016/j.nicl.2018.01.032.
- [78] Y. Wang, Y. Wang, and Y. W. Lui, “Generalized Recurrent Neural Network accommodating Dynamic Causal Modeling for functional MRI analysis,” *Neuroimage*, vol. 178, pp. 385–402, Sep. 2018, doi: 10.1016/j.neuroimage.2018.05.042.
- [79] H. Il Suk, C. Y. Wee, S. W. Lee, and D. Shen, “State-space model with deep learning for functional dynamics estimation in resting-state fMRI,” *Neuroimage*, vol. 129, pp. 292–307, Apr. 2016, doi: 10.1016/j.neuroimage.2016.01.005.
- [80] J. Ashburner and K. J. Friston, “Voxel-based morphometry - The methods,” *Neuroimage*, vol. 11, no. 6 I, pp. 805–821, 2000, doi: 10.1006/nimg.2000.0582.
- [81] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, “Spectral Networks and Locally Connected Networks on Graphs,” *arXiv*, Dec. 2013.
- [82] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional neural networks on graphs with fast localized spectral filtering,” in *Advances in Neural Information Processing Systems*, 2016, pp. 3844–3852.
- [83] C. Ju, A. Bibaut, and M. J. Van Der Laan, “The Relative Performance of Ensemble Methods with Deep Convolutional

Neural Networks for Image Classification.”

- [84] M. J. Van Der Laan, E. C. Polley, and A. E. Hubbard, “Super learner,” *Stat. Appl. Genet. Mol. Biol.*, vol. 6, no. 1, Sep. 2007, doi: 10.2202/1544-6115.1309.
- [85] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [86] S. Tao, “Deep Neural Network Ensembles,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11943 LNCS, pp. 1–12, Apr. 2019.
- [87] A. R. Luedtke and M. J. Van Der Laan, “Super-Learning of an Optimal Dynamic Treatment Rule,” *Int. J. Biostat.*, vol. 12, no. 1, pp. 305–332, May 2016, doi: 10.1515/ijb-2015-0052.
- [88] T. G. Dietterich, “Ensemble methods in machine learning,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 1857 LNCS, pp. 1–15, 2000, doi: 10.1007/3-540-45014-9_1.
- [89] C. Ju, A. Bibaut, and M. van der Laan, “The relative performance of ensemble methods with deep convolutional neural networks for image classification,” *J. Appl. Stat.*, vol. 45, no. 15, pp. 2800–2818, 2018, doi: 10.1080/02664763.2018.1441383.
- [90] E. PMC, “- {Abstract} - {Europe} {PMC}.” 2016.
- [91] Y. Wang, E. Ghumare, R. Vandenberghe, and P. Dupont, “Comparison of different generalizations of clustering coefficient and local efficiency for weighted undirected graphs,” *Neural Computation*, vol. 29, no. 2. MIT Press Journals, pp. 313–331, 01-Feb-2017, doi: 10.1162/NECO_a_00914.
- [92] M. Brett, W. Penny, and S. Kiebel, “An Introduction to Random Field Theory,” 2003.
- [93] L. C. Tranchevent *et al.*, “Predicting clinical outcome of neuroblastoma patients using an integrative network-based approach,” *Biol. Direct*, vol. 13, no. 1, pp. 1–13, 2018, doi: 10.1186/s13062-018-0214-9.

- [94] B. Wang *et al.*, “Similarity network fusion for aggregating data types on a genomic scale,” *Nat. Methods*, vol. 11, no. 3, pp. 333–337, Mar. 2014, doi: 10.1038/nmeth.2810.
- [95] S. Pimm, C. Jenkins, R. Abell, ... T. B.-, and undefined 2014, “The biodiversity of species and their rates of extinction, distribution, and protection,” *science.sciencemag.org*, 2014, doi: 10.1126/science.1246752.
- [96] B. Zhang and S. Horvath, “A general framework for weighted gene co-expression network analysis,” *Stat. Appl. Genet. Mol. Biol.*, vol. 4, no. 1, 2005, doi: 10.2202/1544-6115.1128.