

Classification of 4D fMRI Images Using ML, Focusing on Computational and
Memory Utilization Efficiency

by
Nazanin Beheshti

A dissertation submitted to the Department of Computer Science,
College of Natural Sciences and Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Computer Science

Chair of Committee: Lennart Johnsson

Committee Member: Nikolaos Tsekos

Committee Member: Ricardo Vilalta

Committee Member: Hien Van Nguyen

University of Houston
December 2021

Copyright 2021 Nazanin Beheshti

ACKNOWLEDGMENTS

I will forever be thankful to my advisor, Dr. Lennart Johnsson, who took me when nobody would have given me a second glance, for his endless support and advice throughout my graduate studies at the University of Houston. Even though I had zero knowledge about research, computer architecture, and deep learning, he still gave me the opportunity to grow and learn in this field. I am deeply grateful for your patience, support, and guidance during all these years.

I also want to thank my dissertation committee members, Dr. Nikolaos Tsekos, Dr. Ricardo Vilalta, and Dr. Hein Van Nguyen, for their feedback, and support to make this dissertation better every time. I am also thankful to my colleagues in the ACRL lab for all our group readings, discussion, and learning experience that we had on Fridays during our group meetings.

Furthermore, I was lucky to have had two amazing internships during my Ph.D. I would also like to thank Dr. Hars Vardhan from Samsung Strategy and Innovation Center, and Dr. Valentina Popsecu from Cerebras Systems, who trusted me and gave me the opportunity to learn and grow in the industry. I learned a lot from both of you, and I will be grateful to you for giving me the chance to work with you and learn from you.

I want to thank my parents for their unconditional love, patience, and support for all these years. They always encouraged me to work hard to achieve my goals. I am sincerely grateful to you for being the support and driving force I needed.

I would also like to thank my husband, Mohammad, who always brings love, hope, and energy to my life. Thank you, Mohammad, for believing in me and encouraging me to be the best version of myself in every aspect of my life.

ABSTRACT

Alzheimer's disease is progressively degenerative with a characteristic of memory loss, mood and behavior changes, and deepening confusion about time and place. It is estimated that about 50 million people are affected by AD worldwide. AD's lifetime per patient care cost is estimated to be about \$250k, and the total cost of care of AD patients could exceed \$1 trillion by 2050.

In this research, we use novel data reduction techniques in determining functional brain connectivity from Resting-State fMRI data and show that small Machine Learning models can with good accuracy classify subjects with respect to Alzheimer's disease (AD) or Mild Cognitive Impairment (MCI) or being Cognitive Normal (CN). In fMRI, brain activity is captured from Blood Oxygen Level-Dependent (BOLD) magnetization detected by the MRI scanner. The functional connectivity is inferred from correlations of the observed BOLD signals from typically cubic voxels with sides in the 3 – 4 mm range. The BOLD signals are typically sampled every 2 – 3 seconds for a duration of five to six minutes, generating a data set of 5 – 10 million voxel BOLD signal values per subject. Classification is typically carried out based on signal aggregates for anatomical regions defined in brain atlases to reduce the computational effort. In this research, we use the 90 region Automated Anatomical Labeling atlas, AAL-90, in establishing Regions of Interest, ROIs that are subsets of voxels in the AAL-90 atlas. The functional connectivity is measured by the correlation of BOLD signal aggregates for the ROIs.

In the data reduction step, we represent the 4D data set for a region with a vector that on average reduces the data set for a region from about 100,000 voxel signal values to 100 to 200 values in our spatial representation and in the order of 15,000 – 30,000 in our spatial-temporal representation. We show that a small Convolutional Neural Network (CNN) with a model size of about 168 kiB and a Transformer model of only 37 kiB yields classification accuracies of 80 – 90% for AD, MCI, and CN subject classification. We further show that our region data aggregation technique is more robust to BOLD signal artifacts than the commonly used aggregation technique. The training time for the CNN and Transformer on a data set of 551 subjects required 184 and 23.73 seconds respectively. The experiments are conducted on the Opuntia Cluster using Pytorch.1.5.0, Python 3.7.7, and CUDA 10.1 on a 2.8GHz Intel Xeon E5-2670v2 processor with 2 CPU sockets and 20 cores, and NVIDIA K40 GPU.

Table of Contents

Acknowledgments	iii
ABSTRACT	v
LIST OF TABLES	ix
LIST OF FIGURES	xv
1. Introduction	1
2. Related work	7
3. Bold signal representation for Machine Learning	17
3.1 ROI formation	17
3.2 ROI vector values	19
3.3 Functional connectivity – Pearson correlation	21
4. Data Sets for assessment of ROI image intensity representation and ML classification accuracy and computational efficiency	22
4.1 Data preprocessing	24
5. Machine learning models	29
5.1 CNN models	30
5.2 Transformer model	33
5.3 Comparison of CNN and Transformer models computational and memory requirements	39
6. Data sets for statistical model evaluation	40
7 Training	41
7.1 Training procedure	41
7.2 Training performance	41
7.3 Parameter sparsity	42
8 Experimental Results	45
8.1 AAL-90 region center-voxel based ROIs	45
8.1.1 CNN models	45
8.1.2 The Transformer model.....	53
8.2 Uniformly randomly selected center-voxels for ROI Hilbert curve segments	59
9 Artifact influence on classification accuracy	61
9.1 Artifacts in fMRI	61
9.2 ROI intensity magnitude representation impact on Pearson correlation matrices	62
9.3 Noise impact on Pearson correlation matrices	63

9.4	CNN-B classification accuracy with intra ROI (spatially local) correlated but inter ROI and time uncorrelated noise added to the measured Voxel intensity	66
9.5	CNN-B classification accuracy with intra ROI and temporally uncorrelated but inter ROI correlated noise added to the measured voxel intensity.....	69
9.6	Summary of the artificial noise study	73
Bibliography		74
Appendix.....		82
	An adaptive space-filling curve for MRI based schizophrenia classification.....	82

LIST OF TABLES

Table 1: Comparison of preprocessing steps in our study and related work	8
Table 2: Related work summary	9
Table 3: Identifying patients with Alzheimer’s disease using resting-state fMRI and graph theory [26]	10
Table 4: The identification of Alzheimer's disease using functional connectivity between activity voxels in resting-state fMRI data [36]	10
Table 5: Analysis of Alzheimer’s disease based on random neural network cluster in fMRI [37]	11
Table 6: Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer’s disease [38]	12
Table 7: Gaussian process classification of Alzheimer’s disease and mild cognitive impairment from resting state fMRI [29]	13
Table 8: Summary of methods used in Tables 3-7 for AD classification, ACC, SE, SP, and ML approaches are reported in Table2	14
Table 9: MRI parameters during scanning for the ADNI and OASIS [41] datasets	22
Table 10: Summary of sociodemographic information of subjects in our three data sets.	23
Table 11: Proposed CNN-A architecture, and model characteristics	32
Table 12: Proposed CNN-B architecture, and model characteristics.....	32
Table 13: Computational and memory requirements with parameters represented as FP32 numbers	32
Table 14: Proposed Transformer architecture.....	38
Table 15: Comparison of CNNs and Transformer architecture regarding memory footprint, and number of MACs	39
Table 16: Training time for the CNN-A, CNN-B, and Transformer models for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, MCI-CN- ADNI, and MCI-AD-ADNI classification experiments	42
Table 17: Sparsity level of different layers of the proposed CNN-A model	43
Table 18: Sparsity level of different layers of the proposed CNN-B model.....	44

Table 19: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for AD-CN-OASIS, AD-CN-ADNI, and AD-CN-Mixed subjects for spatial correlation matrices	46
Table 20: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for MCI-CN-ADNI and MCI-AD-ADNI subjects for spatial correlation matrices	46
Table 21: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI classification experiment for spatial correlation matrices	46
Table 22: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, and MCI-AD-ADNI classification experiments for spatial correlation matrices	47
Table 23: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, and AD-CN-ADNI classification experiments for correlation matrices based on spatial-temporal ROI vectors...	48
Table 24: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the MCI-CN-ADNI and MCI-AD-ADNI classification experiments for correlation matrices based on spatial-temporal ROI vectors.....	48
Table 25: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, and MCI-AD-ADNI classification experiments for correlation matrices based on spatial-temporal ROI vectors	48
Table 26: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI classification experiment for correlation matrices based on spatial-temporal ROI vectors	49
Table 27: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI classification experiments for correlation matrices based on ROI temporal vectors	49
Table 28: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the MCI-CN-ADNI and MCI-AD-ADNI classification experiments for correlation matrices based on ROI temporal vectors	50
Table 29: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-	

CN-ADNI, and MCI-AD-ADNI classification experiments for correlation matrices based on ROI temporal vectors	50
Table 30: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI classification experiment for correlation matrices based on ROI temporal vectors.	50
Table 31: smROI, Summary of CNN model and ROI size for best accuracy, sensitivity, and specificity for the three types of ROI vectors for AD-CN-OASIS, AD-CN-ADNI, AD-CN-Mixed.....	51
Table 32: smROI, Summary of CNN model and ROI size for best accuracy, sensitivity, and specificity for the three types of ROI vectors for MCI-CN-ADNI, MCI-AD-ADNI.....	51
Table 33: smROI-2, Summary of best accuracy, sensitivity, and specificity for the three types of ROI vectors and the two CNN models for AD-CN-OASIS, AD-CN-ADNI, AD-CN-Mixed.....	51
Table 34: smROI-2, Summary of best accuracy, sensitivity and specificity for the three types of ROI vectors and the two CNN models for MCI-CN-ADNI, MCI-AD-ADNI.....	52
Table 35: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI classification experiments for correlation matrices based on ROI spatial vectors with Transformer model.....	53
Table 36: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the MCI-CN-ADNI, MCI-AD-ADNI classification experiments for correlation matrices based on ROI spatial vectors with Transformer model ..	53
Table 37: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, and MCI-AD-ADNI classification experiments for correlation matrices based on ROI spatial vectors with Transformer model.....	54
Table 38: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI classification experiments for correlation matrices based on ROI spatial vectors with Transformer model	54
Table 39: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI classification experiments for correlation matrices based on ROI spatial-temporal vectors with Transformer model.....	55

Table 40: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-MCI-CN-ADNI, MCI-AD-ADNI classification experiments for correlation matrices based on ROI spatial-temporal vectors with Transformer model..... 55

Table 41: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, and MCI-AD-ADNI classification experiments for correlation matrices based on ROI spatial-temporal vectors with Transformer model..... 55

Table 42: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI classification experiments for correlation matrices based on ROI spatial-temporal vectors with Transformer model 56

Table 43: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI classification experiments for correlation matrices based on ROI temporal vectors with Transformer model..... 56

Table 44: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the MCI-CN-ADNI, MCI-AD-ADNI classification experiments for correlation matrices based on ROI temporal vectors with Transformer model..... 56

Table 45: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, and MCI-AD-ADNI classification experiments for correlation matrices based on ROI temporal vectors with Transformer model..... 57

Table 46: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for MCI-CN-ADNI classification experiments for correlation matrices based on ROI temporal vectors 57

Table 47: smROI, Summary ROI size for best accuracy, sensitivity, and specificity for the two types of ROI vectors with Transformer model for AD-CN-OASIS, AD-CN-ADNI, AD-CN-Mixed 57

Table 48: smROI, Summary ROI size for best accuracy, sensitivity, and specificity for the two types of ROI vectors with Transformer model for MCI-CN-ADNI, MCI-AD-ADNI 58

Table 49: smROI-2, Summary of best accuracy, sensitivity, and specificity for the two types of ROI vectors with the transformer model for AD-CN-OASIS, AD-CN-ADNI, AD-CN-Mixed..... 58

Table 50: smROI-2, Summary of best accuracy, sensitivity, and specificity for the two types of ROI vectors and the transformer model for MCI-CN-ADNI, MCI-AD-ADNI.....	58
Table 51: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI classification experiments for groups of subjects for two different seed voxels.....	60
Table 52: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the MCI-CN-ADNI and MCI-AD-ADNI classification experiments	60
Table 53: Average, standard deviation, min, max of condition numbers for correlation matrices based on ROI vectors of type temporal and spatial-temporal for all subjects in the MCI-CN-ADNI and MCI-AD-ADNI classification experiments	63
Table 54: Average, standard deviation, min, max of condition numbers for correlation matrices based on ROI vectors of type temporal and spatial-temporal for all subjects in the AD-CN-OASIS and AD-CN-ADNI classification experiments	63
Table 55: Impact on voxel intensity statistics of artificial noise correlated within ROIs but uncorrelated between ROIs	64
Table 56: Impact on voxel intensity statistics of artificial noise uncorrelated within ROIs but correlated between ROIs	65
Table 57: Peak artificial noise magnitude relative to the mean and max voxel intensities for the ADNI and OASIS data sets	65
Table 58: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-OASIS, AD-CN-ADNI, MCI-CN-ADNI, and MCI-AD-ADNI classification experiments for intra ROI correlated and inter ROI uncorrelated artificial noise with scale factors 150 and 450 for 201 ROI length In the table ST stands for spatial-temporal ROI vectors being used for correlation matrices.....	67
Table 59: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-OASIS, and AD-CN-ADNI classification experiments for intra ROI correlated and inter ROI uncorrelated artificial noise with scale factors 150 and 450 for 201 ROI length In the table ST stands for spatial-temporal ROI vectors being used for correlation matrices.....	67
Table 60: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI, and MCI-AD-ADNI classification experiments for intra ROI correlated and inter ROI uncorrelated artificial noise with scale factors	68

Table 61: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-OASIS, AD-CN-ADNI, MCI-CN-ADNI, and MCI-AD-ADNI classification experiments for intra ROI uncorrelated and inter ROI correlated artificial noise with scale factors 450 and 900 In the table ST stands for spatial-temporal ROI vectors being used for correlation matrix..... 70

Table 62: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-ADNI, and AD-CN-OASIS classification experiments for intra ROI uncorrelated and inter ROI correlated artificial noise with scale factors 450 and 900 71

Table 63: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI, and MCI-AD-ADNI classification experiments for intra ROI uncorrelated and inter ROI correlated artificial noise with scale factors 450 and 900 71

LIST OF FIGURES

Figure 1: Summary of previous studies in AD classification	8
Figure 2: Segments of Hilbert curve around AAL-90 center voxels to form ROIs	18
Figure 3: Time-averaged voxel BOLD signals traversed along the Hilbert curve.....	20
Figure 4: Illustration of the formation of spatial-temporal ROI vectors	20
Figure 5: Averaged Time series over all voxels within ROI	21
Figure 6: 4D fMRI BOLD signal acquisition in space and time, fMRI data consist of BOLD signals from M volumes, one for each of M time samples. Each brain volume is constructed from Z “2D” slices of a thickness equal to scan voxel thickness along the z-axis. In both the ADNI and OASIS data sets, the “2D” slices have 64x64 voxels. The number of slices in the ADNI data set is 48 (slice thickness 3.3mm), covering a z-axis range of 158 mm, and in the OASIS data set, it is 36 (slice thickness 4mm), covering a z-axis range of 144m.	24
Figure 7: Slice timing correction for fMRI preprocessing.....	25
Figure 8: Normalization and co-registration of functional and anatomical image for each subject.....	26
Figure 9: Registration of brains to a common space.....	27
Figure 10: Smoothing 2D slices of fMRI images with a Gaussian kernel size 8 .	27
Figure 11: different source of artifact in rs-fMRI and their frequency range [48]	28
Figure 12: Proposed CNN-A, (Top) and CNN-B (below) architecture	30
Figure 13: Proposed Transformer architecture	33
Figure 14: Extracting patches from correlation matrices to feed a sequence of data to the transformer	36
Figure 15: Self-attention for encoder layer of Transformer.....	37
Figure 16: Comparison of results for the five classification experiments based on correlation matrices generated from the three types of ROI vectors: temporal (Solid Line), Spatial (Dash line), and spatial-temporal (Dot Line). The Spatial and spatial-temporal....	52
Figure 17: Accuracy, sensitivity, and specificity change for the AD-CN-ADNI and AD-CN-OASIS, AD-CN-Mixed classification experiments for correlation matrices based on no artificial noise added to voxel intensity values, and on ROI vectors with intra ROI	

correlated and inter ROI uncorrelated artificial noise added to voxel intensity values with noise scale factors of 150 and 450 67

Figure 18: Accuracy, sensitivity, and specificity change for the MCI-CN-ADNI and MCI-AD-ADNI classification experiments for correlation matrices based on no artificial noise added to voxel intensity values and on ROI vectors with intra ROI correlated and inter ROI uncorrelated artificial noise added to voxel intensity values with noise scale factors of 150 and 450 68

Figure 19: Accuracy, sensitivity, and specificity change for the AD-CN-ADNI, AD-CN-OASIS classification experiments for intra ROI uncorrelated and inter ROI correlated artificial noise with scale factors 450 and 900 71

Figure 20: Accuracy, sensitivity, and specificity change for the MCI-CN-ADNI, and MCI-AD-ADNI classification experiments for intra ROI uncorrelated and inter ROI correlated artificial noise with scale factors 450 and 900 72

Figure 21: a) Hilbert curve for a bounding box enclosing the brain, (b) a pruned Hilbert curve covering only the brain 83

Figure 22: Representation of fMRI brain activation traversal by the SFC for the last a) 1000, 3000, 5000 and 7000 voxels (red), overlaid on the total SFC trajectory (gray), representing a sample progression of the SFC trajectory during the last 7000-voxel portion..... 84

Figure 23: Signal intensity along the trajectory for successive voxels 85

1. INTRODUCTION

Alzheimer's disease is progressively degenerative with a character of memory loss, mood and behavior changes, and deepening confusion about time and place. Alzheimer's disease is thought to begin 20 years or more before symptoms arise [1] with minor changes in the brain that are unnoticeable to the person affected. The cause of the disease is that neurons involved in thinking, learning, and memory have been damaged or destroyed [2]. It is estimated that worldwide, about 50 million people are affected by AD, but that only about 25% of those affected have been diagnosed with AD. The lifetime per patient care cost of AD is estimated to be about \$250k [3], and the total cost of care of AD patients could exceed \$1 trillion by 2050. Alzheimer's disease is generally viewed as degraded communication between brain regions termed functional connectivity. "The idea that the human mind/brain is made up of highly specialized components began with the Viennese physician Franz Joseph Gall (1758–1828). Gall proposed that the brain is the seat of the mind, that the mind is composed of distinct mental faculties, and that each mental faculty resides in a specific brain organ. A heated debate on localization of function in the brain raged over the next century. By the early 20th century, a consensus merged that at least basic sensory and motor functions reside in specialized brain regions" [4].

Because of the brain complexity, accurately defining the brain regions and mapping their functions and connections is exceptionally challenging. Brain parcellation methods subdivide the brain into individual regions that can be used to build a network to study its structure and function. Using anatomical or functional connectivity or hierarchal clustering methods aims to offer a meaningful parcellation of the brain at each level of granularity.

Korbinian Brodmann parcellated the human cerebral cortex based on cytoarchitectonic, resulting in regions with similar microarchitecture, distribution, and arrangement of cell bodies in the grey matter [5]. Since then, there have been numerous attempts to identify functionally meaningful regions based on different properties such as anatomy [6], [7], microarchitecture (chemo-architectonic features) [8], and more recently, function [9]. Different brain atlases arise depending on the particular property, the most common listed below [7][10].

1) AAL[7]: Automated Anatomical Labeling of Activations in Statistical Parameter Mapping (SPM) software package [11] using a Macroscopic Anatomical Parcellation of the Montreal Neurological Institute (MNI) MRI Single-Subject Brain.

An anatomical parcellation of the spatially normalized single-subject high-resolution T1 volume provided by the MNI[12] was performed. “The MNI single-subject main sulci were first delineated and further used as landmarks for the 3D definition of 45 anatomical volumes of interest (AVOI) in each hemisphere. This procedure was performed using dedicated software, which allowed a 3D following of the sulci course on the edited brain. Regions of interest were then drawn manually with the same software every 2 mm on the axial slices of the high-resolution MNI single subject. The 90 AVOI were reconstructed and assigned a label “ [7].

2) Brainnetome Atlas [10]: Brainnetome Atlas, with 210 cortical and 36 subcortical subregions, is built upon a connectivity-based parcellation framework containing information on anatomical and functional connections. It is an open resource for researchers for the analysis of whole-brain parcellation, connections, and functions. Brainnetome Atlas provides fine-grained brain sub-regions, which revealed numerous anatomical subdivisions which were

missed in previous brain atlases. Moreover, there are detailed functional connectivity patterns for each area in Brainnetome Atlas.

3) Harvard-Oxford [13]: Probabilistic atlases covering 48 cortical and 21 subcortical structural areas, derived from structural data and segmentations kindly provided by the Harvard Center for Morphometric Analysis. “T1-weighted images of 21 healthy male and 16 healthy female subjects (ages 18-50) were individually segmented by the CMA using semi-automated tools developed in-house. The T1-weighted images were affine-registered to MNI152 space using FLIRT (FSL) [14], and the transforms were then applied to the individual labels. Finally, these were combined across subjects to form population probability maps for each label” [13].

Atlases are derived by statistically summarizing, e.g., averaging, voxel-wise, regional, or global brain MRI measures from several individuals, and they may be used in research as registration targets for functional activation, segmentation, and statistical mapping, for example, in the analysis of population imaging datasets

The International Consortium of Brain Mapping (ICBM) has created a few standard brain templates onto which individual subjects' brains are mapped to compare observations from different individuals. The templates are available from the Montreal Neurological Institute (MNI) [7]. In this work, we use the MNI-152 [15] brain template commonly used in population brain studies.

“Talairach coordinates [16], also known as Talairach space, is a 3-dimensional coordinate system (known as an 'atlas') of the human brain, which is used to map the location of brain

structures independent from individual differences in the size and overall shape of the brain. It is still common to use Talairach coordinates in functional brain imaging studies and to target transcranial stimulation of brain regions. However, alternative methods such as the MNI Coordinate system (originated at the Montreal Neurological Institute and Hospital) have largely replaced Talairach” [17]. “Talairach brain is the brain dissected and photographed for the famous Talairach and Tournoux atlas. The atlas has Brodmann's areas labeled, albeit in a rather approximate way. What the authors did was to look at pictures of the Brodmann map and estimate where the same place was on their brain” [16].

The MNI wanted to define a brain that is more representative of the population. Therefore, the MNI defined a new standard brain by using a large series of MRI scans on normal controls. The MNI305 was the first MNI template. The current standard MNI template is the ICBM152, which is the average of 152 normal MRI scans that have been matched to the MNI305 using a 9-parameter affine transform.

Brain activity for AD classification is captured by functional Magnetic Resonance Imaging (fMRI) that measures Blood Oxygen Level-Dependent (BOLD) magnetization. When an area of the brain is active, the flow of oxygen-rich blood to that region increases, affecting the blood magnetic properties through the binding of oxygen to hemoglobin, thus causing oxygen-rich blood to have a stronger MRI signal. The functional connectivity is inferred from correlations of the observed BOLD signals. The spatial (voxel) resolution in fMRI for AD classification typically is 2 – 4 mm. With female and male brain volumes on average being 1130 cm³ and 1260 cm³, respectively, an fMRI brain scan for AD classification purposes contains in the order of 10⁴ to 10⁵ voxels. The sampling rate typically is 0,5 – 2 Hz [18][1]

for a duration of five to six minutes, creating a 4D data set of ten's of thousands of time series (one per voxel) each of 100 – 200 samples.

Functional connectivity can be assessed using stimuli, task-based observations, or without stimuli, resting-state observation. In recent years, it has become clear that resting-state networks (RSN) provide sufficient information for the good classification of AD and MCI patients [19][20]. In this work, we use resting-state BOLD signals to assess functional connectivity and patient classification.

With the understanding that different regions of the brain are responsible for different functions, it is common to reduce the computational complexity in classification by clustering voxel signals into Regions-Of-Interest (ROIs). The ROIs are commonly taken as anatomical regions [21]. However, ROIs have also been formed based on spatial proximity to selected seed voxels [9], the strength of the BOLD signal correlations with some cut-off criteria [9] which may result in regions of arbitrary shape and size that may not be spatially localized. Methods not based on spatial proximity such as principal component analysis (PCA) [22] and independent component analysis (ICA) [23], or graph theoretical approaches such as clustering coefficients, node degree, betweenness, path lengths, local efficiency, global efficiency, and modularity have been used for reducing the computational effort in classification [24][25][26] [27][28][29].

In this research, we use the 90 regions Automated Anatomic Labeling (AAL) software package and digital atlas [7] to form ROIs. This software package, which is often used in functional neuroimaging, maps voxels as obtained in MRI scans to 90 anatomical regions of the brain.

Most of the studies on AD classification use the Alzheimer's Disease Neuroimaging Initiative (ADNI) data set [30], while some other studies use their in-house collected data. 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).

Machine learning can build a model based on sample data to make decisions without being explicitly programmed to do so. We use Deep Learning algorithms like Convolutional Neural Network and Transformer for AD classification. Unlike machine learning algorithms like SVM, which needs feature engineering, deep learning does not require feature engineering. It can identify multivariate feature combinations that best predict an outcome of interest. Moreover, it can work on high-dimensional connectome data.

2. RELATED WORK

As represented in Figure 1, in previous work, some studies use the brain connection network, and some studies use 2D slices of fMRI/MRI images as input for machine learning approaches. For extracting functional connectivity in group studies, fMRI images from different subjects need to be preprocessed before any group analysis. Before mapping each subject's brain to any anatomical atlas to extract Regions Of Interest for pairwise correlation, slice timing correction, realignment, registration, normalization, resampling, and smoothing is done on each subject's fMRI image to make sure activation in similar locations of the brain from different subjects have overlap and easily comparable. The preprocessing steps for previous studies are reported in Table 2. After preprocessing fMRI data and extracting ROIs by mapping each subject's brain on the AAL90 atlas, a brain connectivity map is defined, which is a temporal correlation among time series BOLD signals of different ROIs. The correlation matrix, which is the connectivity map, is input to the ML algorithms. Different ML approaches like SVM, logistic regression, and random neural networks are trained on correlation matrices and classify and predict test subjects. For 2D slices of fMRI images, they use deep learning approaches like Convolution Neural Network as it is used for RGB images. In this case, CNN is trained on pixel values, and for that, 2D slices of fMRI images need to be normalized and converted to PNG format for processing. After conversion, 2D slices of fMRI images are used for training convolution neural networks. For each subject based on the dataset as an example, we have 150 volumes, and each volume has 48 slices, which results in 7200 2D slices.

For 200 subjects, there would be 1,440,000 2D slices of fMRI image. The data size in our case is approximately equal to the size of ImageNet training data, around 1,200,000. If we use

ResNet [31] for training 1,440,000 2D slices of fMRI image for the AD classification task, we can assume it will take approximately the same time as training ResNet with 1,200,000 ImageNet [32] data. Based on [33], with 8 NVIDIA A100 GPUs, it takes around 36 minutes. This is at least 80 times slower than our reported training time in Table 16. We use only one NVIDIA K40 GPU for training the proposed Transformer mode. Based on different applications and measurements reported in Table 1, the most significant regions are reported below in Tables 3-7.

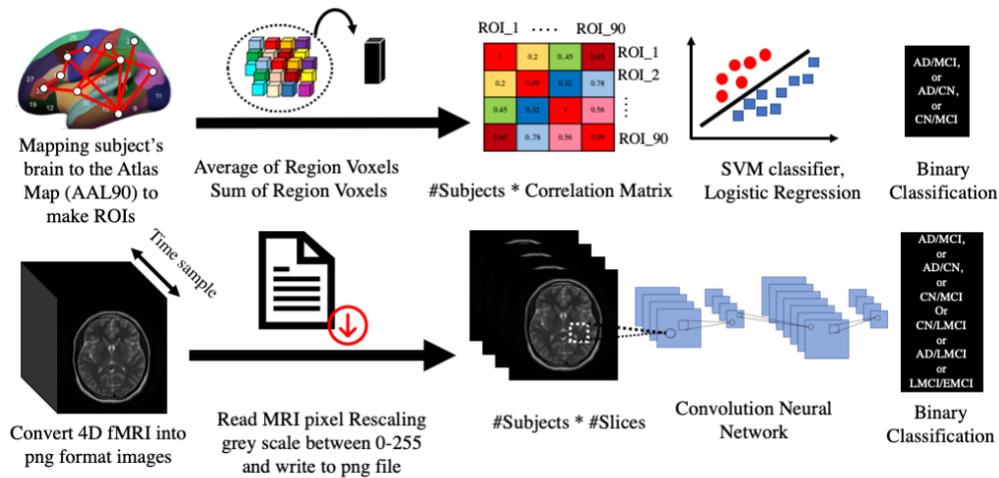


Figure 1: Summary of previous studies in AD classification

Table 1: Comparison of preprocessing steps in our study and related work

Work	Voxel Size	Software	Brain Size	Time Samples	Brain Geom	Resampling (Voxel Size)	Smoothing	Samples Used
(1)[34]	[3 3 4]	SPM8	[64 64 33]	150	MNI	[3 3 3]	Gauss 8	145(first 10 removed)
(2)[35]	[3.13 3.13 5]	None	[64 64 30]	140	MNI	[4 4 4]	Gauss 5	130 (last 10)
(3)[36]	[3.31 3.31 3.31]	None	[64 64 48]	140	MNI	None	Gauss 4	130 (last 10)
(4)[29]	[3 3 2.5]	SPM8	[64 64 32]	220	None	None	Gauss 8	216 (first 4)
(5)[27]	[3.75 3.75 4]	ADNI	[64 64 36]	180	Talairach	None	None	175 (first 5)
(6)[37]	None	DPARSF	[64 64 48]	None	None	None	None	First 10
(7) [38]	[? ? 3.13]	DPARSF, SPM5	[? ? 48]	140	MNI	[3 3 3]	Gauss 4	133 (first 7)
Our Study	[3.3 3.3 3.3], [4 4 4]	SPM12	[64 64 48] [64 64 36]	46-200 164	MNI	[3 3 3]	Gauss 8	46-200 164

Table 2: Related work summary

Paper	ACC	SE	SP	Data Set	Atlas Map	ML Approach	ROI Signal	Modality
(1) [34]	85.8	83.5	83.5	Own	116-AAL	DNN	Average of node voxels	fMRI
(2) [35]	99.99			ADNI	None	CNN	Transform 4D fMRI to 2D PNG format	MRI+fMRI
(3) [36]	92.9	100	86.67	ADNI	Common activity voxels	SVM		fMRI
(4) [29]	MCI vs AD 80%, NC vs MCI 75%	MCI vs AD 90%, NC vs MCI 50%	MCI vs AD 80%, NC vs MCI 100%	Own	82-AAL	Bayesian Gaussian process logistic regression	Sum of region voxels	fMRI
(5) [27]	AD vs CN 87% MCI vs CN 95%	AD vs CN 85% MCI vs CN 93%	AD vs CN 80% MCI vs CN 90%	Own	116-AAL	Fisher linear discriminant analysis	Avg of region voxels	fMRI
(6) [37]	92.31			Own	90-AAL	Random Neural Network Cluster	Avg of region voxels	fMRI
(7) [26]	100	100	100	Own	90-AAL	SVM	Avg of region voxels	fMRI
(8) [38]	83.2	MCI-CN-AD 70,96,70	MCI-CN-AD 100,73,96	Own	90-AAL	SVM	Avg of region voxels	fMRI
(9)[39]	AD vs CN 80% MCI vs AD 99%	AD vs CN 91% MCI vs AD 99%	AD vs CN 83% MCI vs AD 99%	Own	None	ResNet18	Transform 4D fMRI to 2D PNG format	fMRI
Our CNN-B	AD-CN 89%	AD-CN 85%	AD-CN 93%	ADNI	Segments of Hilbert Curve around 90-AAL Center points	CNN	Spatial-Temporal	fMRI
	MCI-CN 90%	MCI-CN 90%	MCI-CN 91%					
	MCI-AD 87%	MCI-AD 82%	MCI-AD 91%					
Our Transformer	AD-CN 87%	AD-CN 85%	AD-CN 89%	ADNI	Segments of Hilbert Curve around 90-AAL Center points	Transformer	Spatial-Temporal	fMRI
	MCI-CN 89%	MCI-CN 89%	MCI-CN 90%					
	MCI-AD 87%	MCI-AD 85%	MCI-AD 89%					

Table 3: Identifying patients with Alzheimer’s disease using resting-state fMRI and graph theory [26]

Brain Regions	Domain
Regions with the best discrimination ability (AAL16)	Dorsolateral of Frontal Gyrus Sup L
	Dorsolateral of Frontal Gyrus Sup R
	Frontal Gyrus Middle L
	Orbital Frontal Gyrus Middle L
	Opercular Frontal Gyrus Inferior R
	Orbital Frontal Gyrus Inferior L
	Olfactory Cortex R
	Frontal Gyrus Sup Medial , R
	Insula R
	Posterior Cingulate Gyrus R
	Calcarine Fissure and Surrounding Cortex L
	Lingual Gyrus R
	Supramarginal Gyrus R
	Caudate Nucleus L
	Temporal Gyrus Middle L
Temporal Gyrus Middle R	

Table 4: The identification of Alzheimer's disease using functional connectivity between activity voxels in resting-state fMRI data [36]

Brain Regions	Domain	% Active Voxels
AAL12	Frontal Sup L	2.04
	Frontal Sup R	1.47
	Frontal Mid L	2.9
	Frontal Mid R	2.12
	Frontal Sup Medial L	1.18
	Frontal Sup Medial R	1.72
	Coneus L	2.9
	Parietal Sup R	1.7
	Preconeus L	1.3
	Preconeus R	2.03
	Temporal Pole Sup L	5.24
	Temporal Pole Sup R	2.75

Table 5: Analysis of Alzheimer’s disease based on random neural network cluster in fMRI [37]

Brain Regions	Domain	Weight
<p>AAL23 The number of features related to a certain region is considered as the weight</p>	Precentral Gyrus L	19
	Middle Frontal Gyrus L	
	Olfactory Cortex L	
	Orbital Sup Frontal Gyrus R	18
	Triangular Inferior Frontal Gyrus L	
	Supplementary Motor Area L	
	Orbital Sup Frontal Gyrus L	17
	Orbital Middle Frontal Gyrus L	
	Precentral Gyrus R	16
	Dorsolateral Sup Frontal Gyrus R	
	Orbital Middle Frontal Gyrus R	
	Orbital Inferior Frontal Gyrus L	
	Orbital Inferior Frontal Gyrus R	
	Olfactory Cortex R	
	Middle Frontal Gyrus R	15
	Rolandic Operculum L	
	Rolandic Operculum R	
	Dorsolateral Sup Frontal Gyrus L	14
	Opercular Inferior Frontal Gyrus L	
	Opercular Inferior Frontal Gyrus R	
Triangular Inferior Frontal Gyrus R		
Supplementary Motor Area R		
Medial Sup Frontal Gyrus L		

Table 6: Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer’s disease [38]

Brain Regions	Domain
AAL (19) + 4 others of 264 regions	Lingual R
	Cerebelum Crus1 R
	Temporal Inf R
	Postcentral L
	Paracentral Lobule L
	Area 50 of 264
	Postcentral R
	Lingual L
	Precuneus R
	Fusiform R
	Temporal Pole Middle R
	Precuneus L
	Area 151 of 264
	Occipital Mid L
	Frontal Mid R
	Putamen L
	Area 230 of 264
	Thalamus L
	Thalamus R
	Area 237 of 264
Putamen R	
Temporal Mid L	
Cerebelum 6 R	

Summary of methods used in Tables 3-7 for AD classification, ACC, SE, SP, and ML approaches are reported in Table 2 and Table 8.

Table 7: Gaussian process classification of Alzheimer’s disease and mild cognitive impairment from resting state fMRI [29]

Brain Regions	Domain	Inverse Length Scale
<p>AAAL-43</p> <p>For the linear ARD covariance function, inverse length scale of d is a direct measure of how much feature d, contributes to the classification</p> <p>The larger is inverse length scale Parameter is, the greater the contribution feature d makes to the predictive function.</p>	mmse bl	53,392.8
	Frontal inf orb L ↔ temporal Inf R	32,874.0
	Frontal inf orb L ↔ temporal Inf L	12,573.6
	Frontal sup orb R ↔ occipital inf R	9,251.4
	Frontal sup R ↔ fusiform L	6,914.3
	Frontal sup L ↔ occipital inf R	5,426.2
	Occipital inf L ↔ temporal inf R	3,371.7
	Frontal sup R ↔ fusiform R	2,663.3
	Frontal sup orb R ↔ occipital Inf L	339.5
	Frontal sup medial L ↔ parietal sup R	117.1
	Rectus L ↔ temporal inf L	64.4
	Occipital inf R ↔ parietal sup R	47.3
	Frontal mid L ↔ frontal inf orb R	49.7
	Occipital inf R ↔ temporal inf L	49.3
	Frontal sup orb L ↔ temporal inf R	48.6
	Frontal sup orb R ↔ fusiform L	44.7
	Rectus L ↔ temporal inf L	64.4
	Occipital inf R ↔ parietal sup R	47.3
	Frontal inf orb R ↔ temporal inf R	29.8
	Frontal sup medial L ↔ temporal inf R	24.5
	Temporal inf L ↔ temporal inf R	20.1
	Occipital inf L ↔ temporal inf L	15.4
	Frontal mid L ↔ occipital Inf R	13.0
	Frontal sup R ↔ occipital inf L	9.8
	Frontal mid R ↔ occipital inf R	7.5
	Rectus L ↔ fusiform L	6.4
	Rectus L ↔ temporal inf R	5.7
	Frontal sup orb R ↔ temporal inf R	5.0
	Frontal mid R ↔ temporal inf R	3.5
	Frontal sup orb R ↔ fusiform R	3.1
	Rectus R ↔ temporal inf L	2.8
	Rectus L ↔ fusiform R	2.7
	Frontal sup medial R ↔ temporal inf R	2.6
	Frontal sup L ↔ frontal inf orb R	2.5
	Frontal sup L ↔ fusiform R	2.2
	Frontal sup medial R ↔ occipital inf R	1.8
	Fusiform L ↔ paracentral lobule L	1.5
	Frontal sup medial L ↔ occipital inf L	1.5
	Frontal mid L ↔ fusiform R	1.3
	Frontal sup R ↔ temporal inf L	1.3
Occipital inf L ↔ temporal mid R	1.0	
Frontal sup orb L ↔ occipital inf L	1.0	
Frontal sup medial L ↔ occipital inf R	0.7	
Occipital inf R ↔ temporal inf R	0.7	
Frontal mid R ↔ fusiform L	0.7	

Table 8: Summary of methods used in Tables 3-7 for AD classification, ACC, SE, SP, and ML approaches are reported in Table2

Previous Work	AD-MC-CN	Method
Table 7	27-AD 50-MC 30-CN	Each subject's rsfMRI scan is converted to a brain region connectivity feature vector defined as the variance-covariance in BOLD signals between 82 anatomically distinct regions of interest (ROIs). The covariance between each ROI was calculated. Define the most relevant features to the GP-LR model by analyzing the strength of optimized covariance function length-scale parameters λ .
Table 6	34-AD 89-MC 45-CN	1) Brain Local functional network features 2) Brain Global functional network features Total 44 Features Fisher score is used to reduce the number of features and take the most discriminative features
Table 3	20-AD 20-CN	1) Brain Local functional network features 2) Brain Global functional network features Total 21 Features Fisher score To reduce the number of features and take the most discriminative features
Table 5	25-AD 36-CN	The image is divided into 90 regions defined by AAL brain atlas. The Pearson correlation coefficient is defined between two regions as functional connectivity The number of features related to a certain region is considered as the weight
Table 4	67-AD 76-CN	Implement ICA on a single subject (with V voxels and T time points) calculate the location set of activity voxels in the ICs of each subject at a given threshold θ , Obtain the location sets of common activity voxels of subjects in the HC and AD group
ours	298-AD 97-MCI 253-CN	The image is divided into 90 regions by extracting 201,101 length segments of 64x64x64 Hilbert Curve around 90-AAL center points or random points. Taking Pearson correlation between V*T Hilbert Curve ordered voxels of every pair of regions. (V total number of voxels in the region). T is number of time points for each Voxel.

There are multiple ML approaches for different purposes. Language modeling is the task of predicting what word/letter comes next in a sequence model. Unlike CNN, in sequence modeling, the current output is dependent on the previous input, and the length of the input is not fixed. Encoder-Decoder architecture is used for sequence modeling. The final hidden state vector contains all encoded information from the previous hidden representation and previous inputs in the encoder. For example, RNN acts as an encoder. The encoded information is passed to the decoder, which decodes into probability distribution of the next possible word. Encoder-decoder is used for image captioning, in which image is passed to CNN and features are extracted in the form of a feature representation vector. The encoded vector/features are passed through the RNN or LSTM to generate the caption. In machine translation, each character or input is fed into RNN. At the last step of the encoder, the final hidden state representation of all previous inputs will be passed to the decoder. The decoder can be an RNN or LSTM network that decodes the step representation vector and gives the probability distribution of each character.

A Generative Adversarial Network (GANs) is an unsupervised learning method in machine learning that involves automatically discovering and learning the regularities or patterns in input data in such a way that model can be used to generate or output new examples that plausibly could have been drawn from the original dataset. GANs are used in a range of applications such as image, video, and sound generation.

A Convolution Neural network or ConvNet can capture temporal and spatial dependencies in an image through the application of relevant filters. It is most commonly applied to analyze visual imagery. ConvNet reduces the image into a form that is easier to process without losing features which are critical for getting good predictions. CNN is an architecture that is not only

good at learning features but also scalable to the massive datasets. CNNs have applications in image and video recognition, recommender system, image classification, image segmentation, medical image analysis, natural language processing.

In image classification, a CNN is trained on pixel values of images, and kernels learn temporal and spatial dependencies in short-range (3x3 FOV). In our study, our input is 2D correlation matrices with values in the range of -1 to 1 without any sequence dependencies as language models. We use CNNs for AD classification since our input is similar to 2D images used in image classification, and convolution can learn temporal and spatial dependencies among correlation values by using different kernels.

3. BOLD SIGNAL REPRESENTATION FOR MACHINE LEARNING

Our goal is to design computationally efficient Machine Learning (ML) models for accurate AD, MCI, and CN classification. To reach this goal, we base our ML models on ROI-to-ROI correlation of ROI vectors of values derived from the 2D image intensity values generated from the BOLD signals in 2D slice MRI scans of the brain. Each 2D slice is “2.5” D since it corresponds to a 3D brain volume with the “0,5” dimension being the voxel extent in the dimension perpendicular to the 2D image plane. For forming the ROIs, the “2.5D” slices are assembled into a 3D brain. Pearson correlation [40] is used for ROI-to-ROI correlation. Commonly, ML applied to image classification directly uses the 2D intensity images, which generally results in large models for good accuracy, as, for instance, was reported for AD classification in [39].

Most ML approaches to AD classification reported in the literature also use ROI-to-ROI correlation matrices. However, we form ROIs in unique ways, with on average fewer voxels per ROI than approaches using anatomical regions as ROIs. For example, compared to ROIs inferred by the AAL-90 regions, our ROIs with up to 5x fewer voxels per ROI result in classification accuracies comparable to published works using anatomical regions as ROIs. Further, we derive the ROI vector values in unique ways compared to how ROI image intensity values are derived in published works. Below we first describe how we form ROIs, then how ROI vectors are formed.

3.1 ROI formation

Unlike other works, we form spatially localized ROIs based on segments of a Hilbert curve traversing a bounding box of the MNI-152 brain template. The choice of a curve for mapping

sensitivity to the center-voxel selection on the classification accuracy. For 1) ROIs based on Hilbert curve segments of length up to 201 did not overlap. For 2) only non-overlapping Hilbert curve segments were accepted for forming ROIs.

3.2 ROI vector values

Our classification of subjects with respect to having Alzheimer's disease (AD) or Mild Cognitive Impairment (MCI) or being Cognitive Normal (CN) is based on intensity magnitude values in BOLD signal reconstructed images. We use two novel ways to derive ROI vector values from image intensity magnitude values.

- *spatial*: the ROI *vectors* consist of time-averaged voxel intensity values for the ROI Hilbert curve segment voxels.
- *spatial-temporal*: the ROI vectors consist of all time-series intensity values for all ROI Hilbert curve segment voxels are concatenated for all the voxels for a Hilbert curve segment representing an ROI.

Below we define the spatial and spatial-temporal ROI vectors precisely. The voxel intensity value for voxel i , $i = 1, 2, \dots, N$, where N is the Hilbert curve segment length, at time j , $j = 1, 2, \dots, M$ where M is the number of time samples in region r , $r = 1, 2, \dots, R$ is denoted $S(i, j, r)$.

For ROI r , the spatial ROI vector entry i , $V(i, r)$ is

$$V(i, r) = \frac{1}{M} \sum_{j=1}^M S(i, j, r) \text{ for } i \in [1, N] \text{ and } r \in [1, R]$$

The spatial ROI vector for ROI r is of length $L=N$. The formation of the voxel intensity vector for the spatial representation is illustrated in Figure 3.

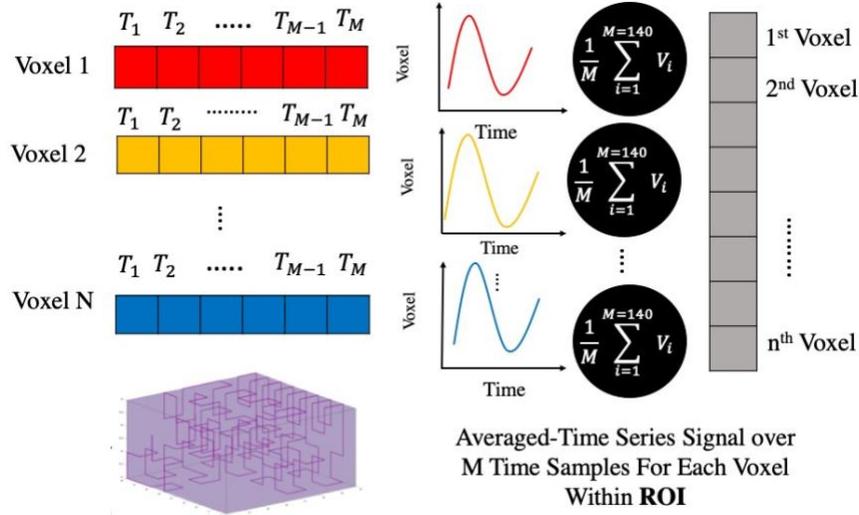


Figure 3: Time-averaged voxel BOLD signals traversed along the Hilbert curve

For ROI r , the spatial-temporal ROI vector is of length $L = N * M$. The voxel intensity values $S(i, j, r)$ are mapped to a vector $V(k, r)$

$$V(k, r) = S(k, r) \quad \text{with} \quad k = i + N * j \quad \text{for} \quad i = 1, \dots, N \quad \text{and} \quad j = 1, \dots, M \quad (k = 1, \dots, N * M) \quad \text{for} \quad r = 1, \dots, R$$

(column major order with voxel time series forming rows).

The formation of the spatial-temporal ROI vectors is illustrated in Figure 4.

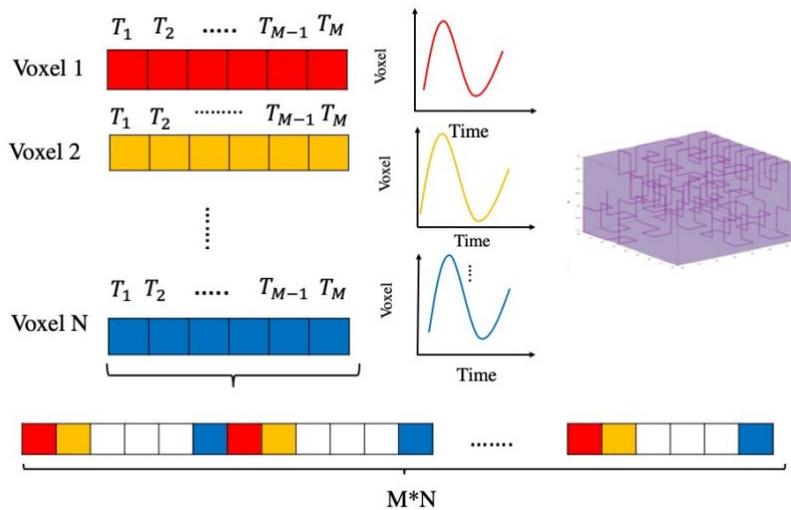


Figure 4: Illustration of the formation of spatial-temporal ROI vectors

Conventionally, ROI vectors are formed by spatially averaging the time series intensity values $S(i,j,r)$, creating temporal ROI vectors of length $L=M$, as shown below

$$V(j, r) = \frac{1}{N} \sum_{i=1}^N S(i, j, r) \text{ for } j \in [1, M] \text{ and } r \in [1, R]$$

The formation of the temporal ROI vectors is illustrated in Figure 5.

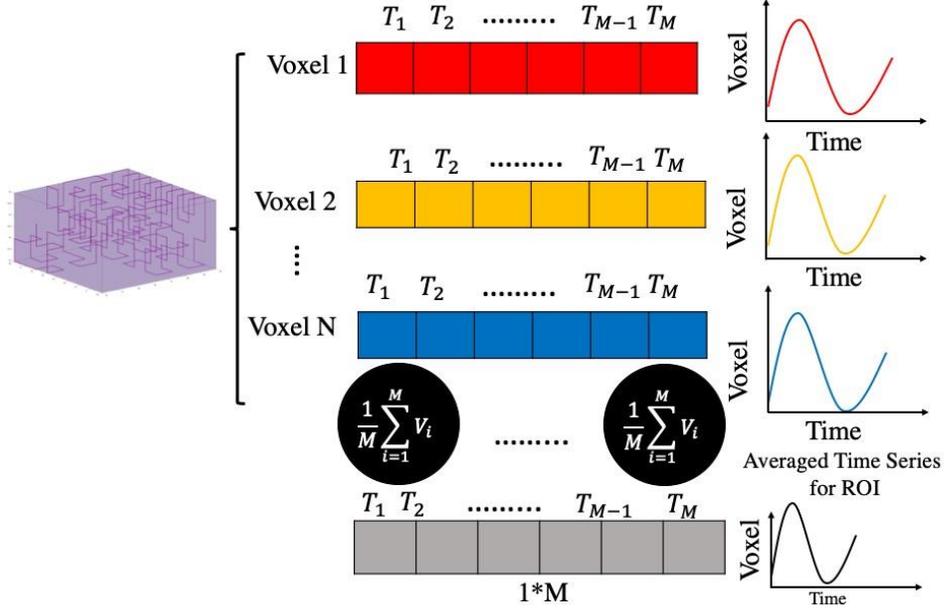


Figure 5: Averaged Time series over all voxels within ROI

3.3 Functional connectivity – Pearson correlation

With the above notation the Pearson correlation for population of ROIs $\rho(n, m)$ [40], $n, m = 1, 2, \dots, R$ can be expressed as

$$\rho(n, m) = \frac{1}{(L-1)} \sum_{l=1}^L \frac{(V(l, n) - \mu(n))(V(l, m) - \mu(m))}{\sigma(n)\sigma(m)} \text{ for } n, m \in [1, R]$$

$$\mu(r) = \frac{1}{L} \sum_{l=1}^L V(l, r) \text{ for } r \in [1, R]$$

$$\sigma(r) = \text{sqrt}\left(\frac{1}{L} \sum_{l=1}^L (V(l, r) - \mu(n))^2\right) \text{ for } r \in [1, R]$$

4. DATA SETS FOR ASSESSMENT OF ROI IMAGE INTENSITY REPRESENTATION AND ML CLASSIFICATION ACCURACY AND COMPUTATIONAL EFFICIENCY

For our assessment of the effectiveness of ROI image intensity representation for classification accuracy using small CNN and Transformer ML models for binary AD, MCI, and CN subject classification we created three data sets from the ADNI (<http://adni.loni.ucla.edu>) and OASIS [41] Resting-State fMRI data sets; one set from each of the ADNI and OASIS data sets and a third by combining the two data sets to create a larger data set to assess the impact of data set size on the accuracy of our ML models. The ADNI data set of 302 subjects contain all three classes of subjects with 101 AD subjects (33%), 97 MCI subject (32%), and 104 CN subjects (34%). The OASIS dataset contains no MCI subjects. In this data set, 149 (43%) of the subjects were classified as AD subjects. The combined data (Mixed) set of 648 subjects has 250 (39%) AD subjects, 97 (15%) MCI subjects, and 301 (46%) CN subjects. Table 9 summarizes the sociodemographic information of the subjects in the two data sets.

Table 9: MRI parameters during scanning for the ADNI and OASIS [41] datasets

Data Set	Imaging	Time Samples	TR(s)	TE (ms)	Flip Angle	Slice Resolution	#Slices	Voxel Size
OASIS	Single-shot gradient echo planar imaging (EPI)	164	2.2	27	90	64 x 64	36	4x4x4
ADNI	Gradient echo (GR) pulse	46-200	3	30	80	64 x 64	48	3.3x3.3x3.3

Table 10: Summary of sociodemographic information of subjects in our three data sets.

OASIS Data Set				
	AD	MCI	CN	Total
Number	149	0	197	346
Male/Female	79/70	0	70/127	149/197
Age	74±10.4		66±10.4	69±9.7
ADNI Data Set				
	AD	MCI	CN	Total
Number	101	97	104	302
Male/Female	35/66	45/52	47/57	127/175
Age	75±7.69	74±8.6	76±7.19	75±7.8
Combined Data Set (OASIS + ADNI)				
	AD	MCI	CN	Total
Number	250	97	301	648
Male/Female	114/136	45/52	117/184	
Age	74.5±9.15	74±8.6	72±8.8	

The ADNI dataset (<http://adni.loni.ucla.edu>) was acquired with a slice resolution of 64 x 64, 3.3 x 3.3 mm voxels of 3.3 mm thickness with 48 slices to cover the brain volume and 46 to 200 time samples using Gradient echo (GR) pulses, with 3s repetition time (TR) (total duration varies from 138s to 600s), a 30s echo time (TE) and 80-degree flip angle (FA). The OASIS dataset [41] was acquired with slice resolution 64 x 64, 4 x 4 mm voxels of 4 mm thickness with 36 slices to cover the brain volume, and 164 whole-brain time samples using a repetition time (TR) of 2.2s (total duration 360 seconds) using single-shot gradient echo-planar imaging (EPI) with 27 ms echo time (TE) and 90-degree flip angle (FA). Both RS-fMRI data sets were acquired using 3 Tesla Siemens scanners. Table 10 summarizes the data acquisition parameters for the ADNI and OASIS data sets.

4.1 Data preprocessing

The BOLD signals for the whole-brain are acquired as a sequence of 2D adjacent slices (36 and 48 in case of the ADNI and OASIS data sets respectively) during the repetition time, as shown in Figure 6. For the classification, we use voxel intensity values derived from the magnitude and phase values of the BOLD signal.

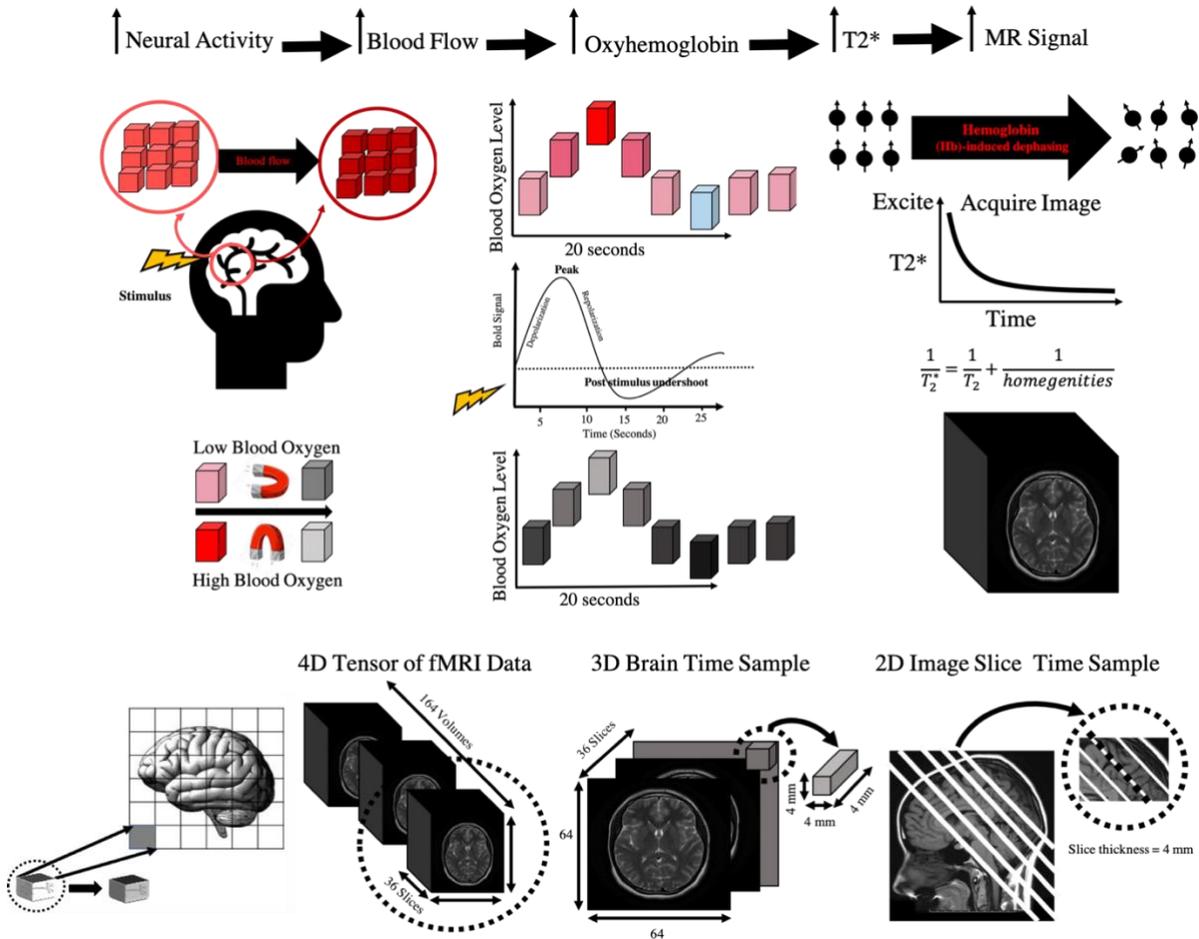


Figure 6: 4D fMRI BOLD signal acquisition in space and time, fMRI data consist of BOLD signals from M volumes, one for each of M time samples. Each brain volume is constructed from Z “2D” slices of a thickness equal to scan voxel thickness along the z-axis. In both the ADNI and OASIS data sets, the “2D” slices have 64x64 voxels. The number of slices in the ADNI data set is 48 (slice thickness 3.3mm), covering a z-axis range of 158 mm, and in the OASIS data set, it is 36 (slice thickness 4mm), covering a z-axis range of 144mm.

In the pre-processing, as represented in Figure 7, the 2D slice acquisition times are adjusted to a common reference time for each whole-brain scan by simply adjusting the actual acquisition time t_k for slice k in the brain scan by shifting the sampling time to that of the middle slice by linear interpolation giving an adjusted time $(Z/2 + 1 - k) \times TR/Z$, with $k = 1, \dots, Z$ where Z is the number of slices and TR is the repetition time for a whole-brain scan.

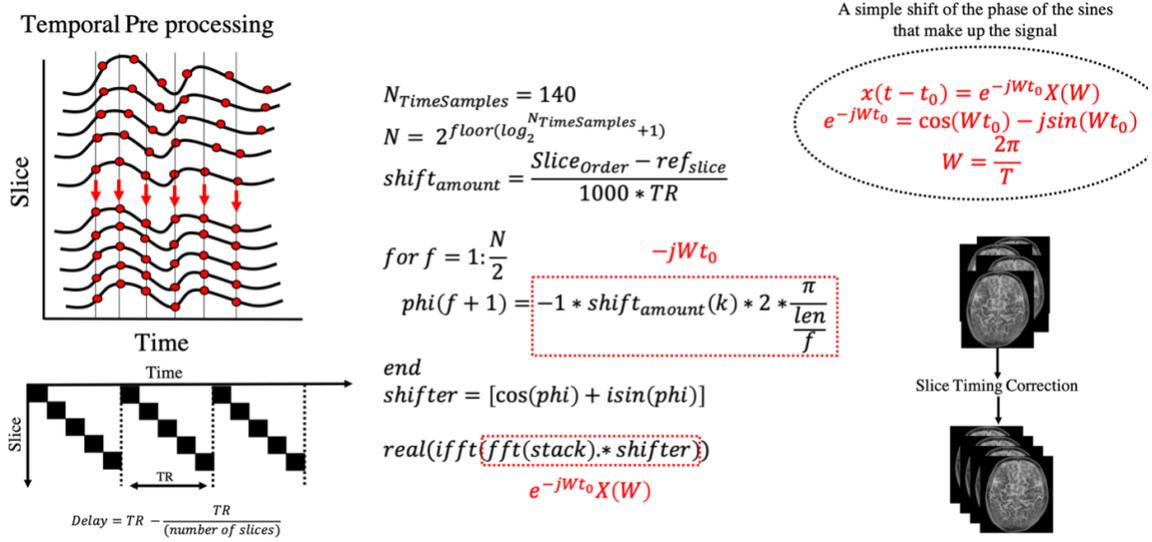


Figure 7: Slice timing correction for fMRI preprocessing

In the preprocessing functional (fMRI) images are first aligned with a structural (MRI) image for each subject, called co-registration, then co-registered fMRI images are mapped to a template brain used for all subjects, the MNI-152 brain in our case. The second step is known as normalization. The co-registration and normalization processes are represented in Figure 8. We use the SPM12-V7771 software package [11] for co-registration and normalization involving affine transformations (rotation, translation, shear, scaling).

The SPM12-V7771 software package uses an entropy and normalized mutual information-based objective function for co-registration, equation (1) [42]. In equation (1) $I(X, Y)$ is the

mutual information between X and Y, and $H(\cdot)$ denotes entropy. For the alignment, misalignment tolerances are given for translation, rotation, scaling, and shear, which in our case are (0.02, 0.001, 0.01, 0.001)

$$\text{NMI} = \frac{I(X, Y)}{\sqrt{H(X)H(Y)}} \quad (1)$$

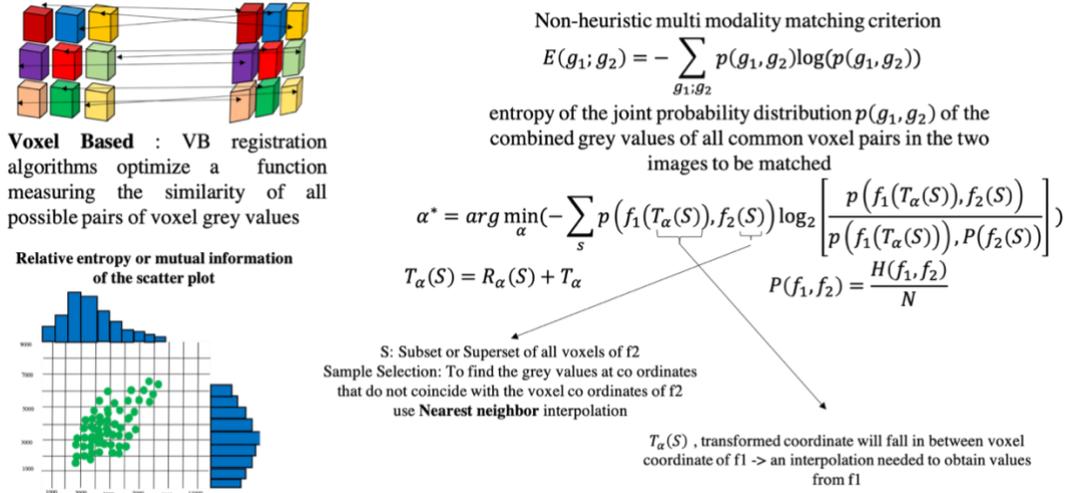


Figure 8: Normalization and co-registration of functional and anatomical image for each subject

As represented in Figure 9, for group studies, the brains of all subjects are normalized to a common space. The normalization uses a tissue probability map containing prior probabilities of all the tissues found in the image. The prior probabilities of different tissue classes at each location in the brain are constructed from a large number of brains mapped to the MNI-152 template. For normalization, we use the log-likelihood criteria that have been proved effective [43]. To compare our findings with previous AD classification studies as reported in Table 2, we use the same 3x3x3 mm voxel size reported in previous studies for comparison requiring a resampling of the ADNI and OASIS voxel sizes in our data sets. The resampling is made as a

3D nearest-neighbor interpolation of the voxel intensity values from the 3.3x3.3x3.3 mm ADNI and the 4x4x4 mm OASIS scanner voxel sizes.

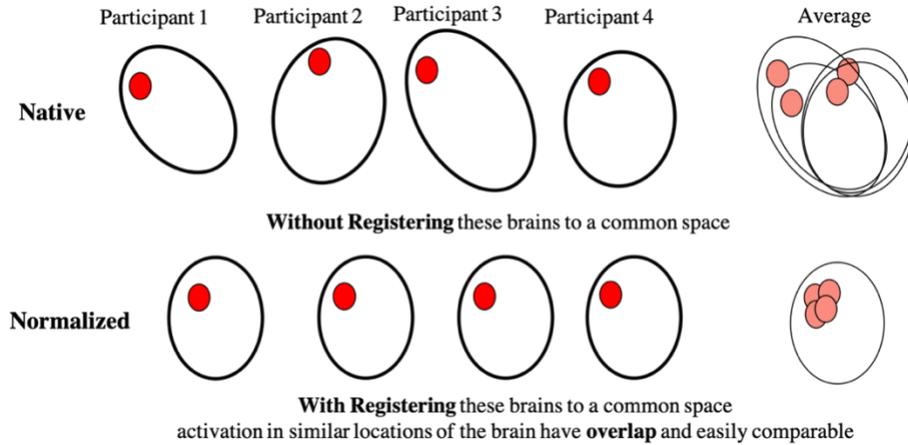


Figure 9: Registration of brains to a common space

After time alignment, co-registration, normalization, and signal interpolation for voxel size, as represented in Figure 10, spatial smoothing of the 2D slices of 3D fMRI image was made using a Gaussian filter with the Full Width at Half Maximum (FWHM) of 8 mm to enhance the signal to noise ratio.

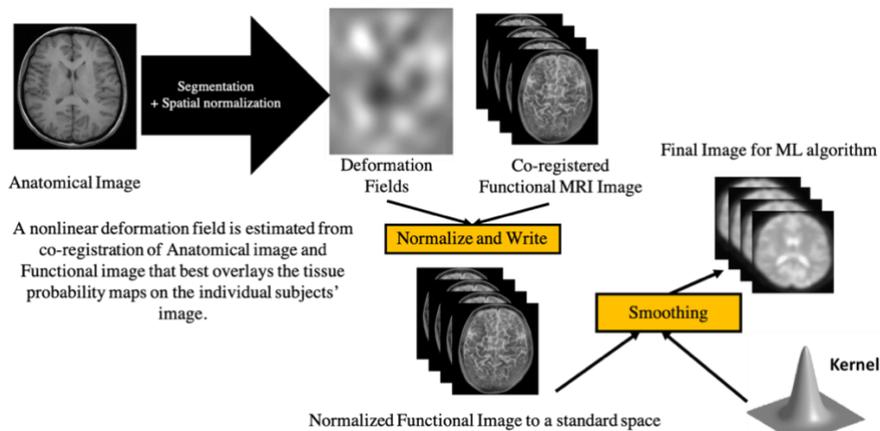


Figure 10: Smoothing 2D slices of fMRI images with a Gaussian kernel size 8

As represented in Figure 11, fMRI data are characterized by small activation fluctuations with superimposed motion, physiological and temporal artifacts [44][45] in addition to scanner

imperfections. Spontaneous neuronal oscillations or low-frequency oscillations have frequencies in the 0.01 to 0.1 Hz range, respiratory signal contributions are in the 0.1-0.5Hz range, and cardiac activities in the 0.6 to 1.2 Hz range. Numerous physiological noise removal techniques, broadly grouped into reference-based models and data-driven approaches [46], are currently widely used. These techniques would selectively remove all artifacts. Low pass filtering cannot entirely remove these artifacts as it would remove neuronal signals as well [47]. For our studies, we report the preprocessing steps in Table 1; we do not apply any temporal or spatial filtering for artifact removal, nor do we use any data-driven or reference-based model to remove artifacts.

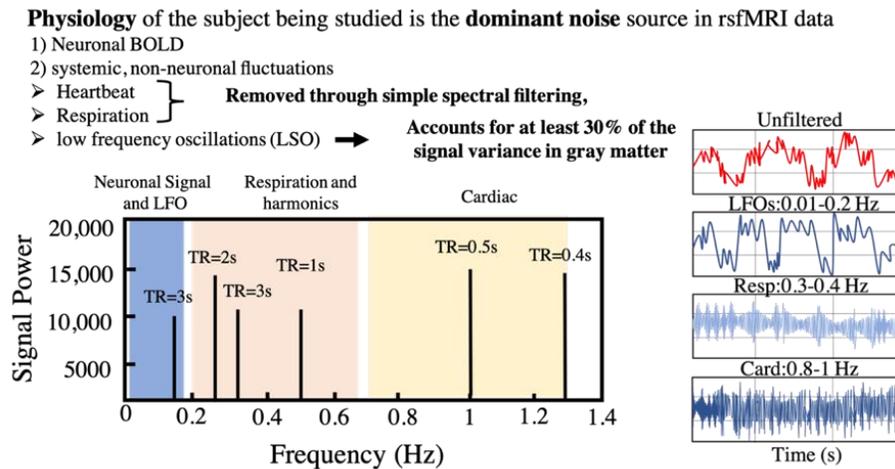


Figure 11: different source of artifact in rs-fMRI and their frequency range [48]

5. MACHINE LEARNING MODELS

A Convolution Neural Network (CNN) is a deep learning algorithm used mostly for computer vision tasks. In a Convolution Neural Network, there are multiple learnable parameters called filters that slide along the input features and output translation equivariant output called feature maps in each convolution layer. In a CNN, they can capture spatial and temporal dependencies in images.

The convolution layer can extract low-level to high-level features from the input image. For example, low-level features such as edges and color are captured in the first layers. As the network gets deeper it captures high-level features to the end. They are also known as shift invariant or space invariant artificial neural networks (SIANN).

A sequence to sequence architecture, transformers are used primarily in natural language tasks and have recently been applied in computer vision tasks. The Seq2seq model consists of an encoder and decoder. The encoder takes the input sequence and maps it to high-dimensional space. That vector is then fed into the decoder, turning into an output sequence.

The Transformer does the same with the help of a self-attention layer. In a transformer, by the attention mechanism, the relationship between every part of the sentence or sequence (token) is measured and provides relevant information about different tokens. For images, the smallest unit of analysis is a pixel. In an image with size 224×224 , there are 50,176 pixels. Suppose we want to input a sequence of 50,176 tokens to the transformer, for the dot product in the self-attention layer. In that case, we need at least 2.5 Billion operations which is not efficient regarding memory and computation. Instead of using pixels for analysis, the image is divided into sections called patches. These patches with positional embedding are used as an input sequence for the transformer. If we use a patch size 14×14 for a 224×224 image, the

sequence length decreases from 50,176 to 256. The sequence with 256 tokens is passed through a positional encoding first to give every token a relative position to other tokens since we need to know the order of tokens in the sequence. At the next step, it passes through a linear layer called the encoding layer to embed into a lower-dimensional space. After embedding the tokens, they are passed through layers of encoders.

5.1 CNN models

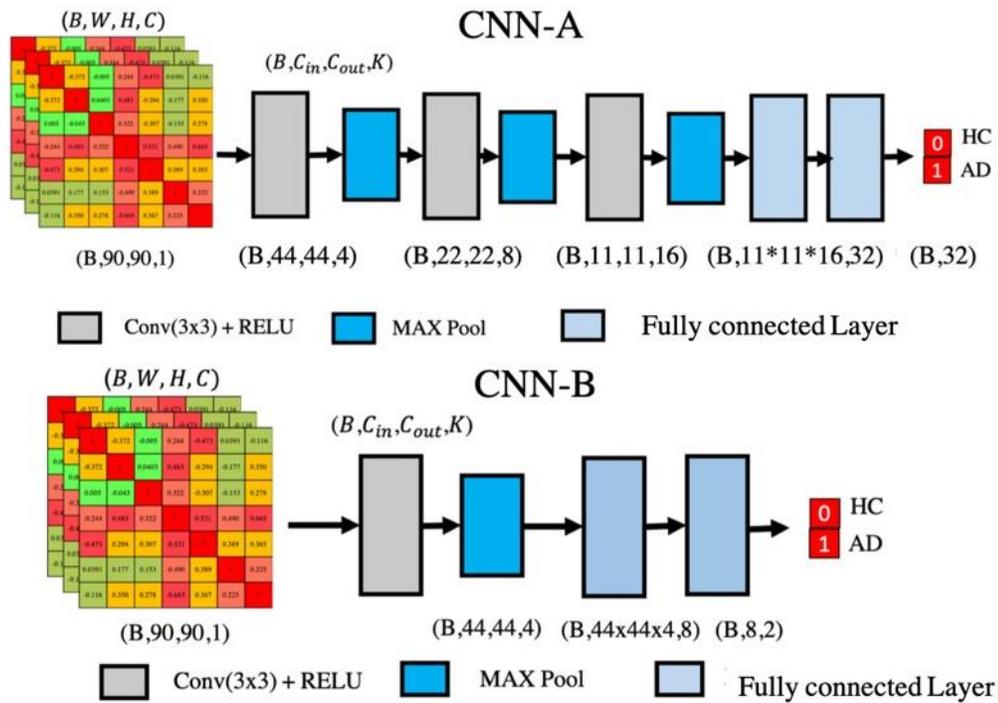


Figure 12: Proposed CNN-A, (Top) and CNN-B (below) architecture

As represented in Figure 12, the input to the CNNs we study is the 90x90 ROI correlation matrices for each subject with one channel, the voxel intensity values. The number of output channels for the convolution layers in CNN-A are 4,8,16, respectively, with each having 3x3 filters. The fourth layer is a fully-connected layer with 32 output channels, and the last layer is a fully-connected layer with two output channels. The last fully-connected layer gets the input

tensor of size (B,32) and maps it to (B,2) tensor, which is the final output. The CNN-B has one convolution layer with 4 output channels and a fully connected layer with 8 output channels, and the last layer is a fully-connected layer with 2 channels. The first fully connected layer gets the input (B,44*44*4,8) which is flattened after the last convolution layer and maps it to (B,8) tensor, which is the input to the last fully connected layer. The last fully-connected layer gets a tensor of size (B,8) and outputs a tensor of size (B,2), the final output. The total number of parameters for a convolution layer using $K \times K$ filters having C_i input channels and C_o output channels are $K \times K \times C_i \times C_o$. Thus, in CNN-A, the first convolution layer with 1 input channel and 4 output channels has $3 \times 3 \times 1 \times 4 = 36$ parameters, the second convolution layer has 288 parameters, and the third convolution layer has 1,152 parameters. The fourth fully connected layer with 32 output channels has $16 \times 9 \times 9 \times 32 = 41,472$ parameters, and the last layer has 64 parameters. The total number of parameters is 43,012 ($36 + 288 + 1,152 + 41,472 + 64$). With single-precision floating-point IEEE data representation (FP32), the CNN-A model size is about 172 kiB. The CNN-B has $3 \times 3 \times 4 + 44 \times 44 \times 4 \times 8 + 8 \times 2 = 62,004$ parameters. Thus, with the FP32 data representation, the CNN-B model size is about 242 kiB. Thus both CNN models are considerably smaller than many other CNN models for image classification [32].

The CNN-A and CNN-B model characteristics for inference are summarized in Tables 11 and 12.

Table 11: Proposed CNN-A architecture, and model characteristics

Layer Name	Input	Output	#Params	#MAC
Conv	(B,90,90,1)	(B,90,90,4)	$4 \times 1 \times 3 \times 3 = 36$	$9 \times 2 \times 1 \times 4 \times 90 \times 90 = 583,200$
MAX pooling	(B,90,90,4)	(B,44,44,4)		
Conv	(B,44,44,4)	(B,44,44,8)	$4 \times 8 \times 3 \times 3 = 288$	$9 \times 2 \times 8 \times 4 \times 44 \times 44 = 1,115,136$
MAX pooling	(B,44,44,8)	(B,21,21,4)		
Conv	(B,21,21,8)	(B,21,21,16)	$8 \times 16 \times 3 \times 3 = 1152$	$9 \times 2 \times 8 \times 16 \times 21 \times 21 = 1,016,064$
MAX pooling	(B,21,21,16)	(B,9,9,16)		
Flatten and Reshape	(B,9,9,16)	(B,9x9x16)		0
Linear	(B,1296)	(B,32)	$1296 \times 32 = 41,472$	$2 \times 1296 \times 32 = 82,944$
Linear	(B,32)	(B,2)	$32 \times 2 = 64$	$2 \times 32 \times 2 = 128$
Total			43,012	2,797,472

Table 12: Proposed CNN-B architecture, and model characteristics

Layer Name	Input	Output	#Params	#MAC
Conv	(B,90,90,1)	(B,90,90,4)	$4 \times 1 \times 3 \times 3 = 36$	$9 \times 2 \times 1 \times 4 \times 90 \times 90 = 583,200$
MAX pooling	(B,90,90,4)	(B,44,44,4)		
Flatten and Reshape	(B,44,44,4)	(B,44x44x4)		0
Linear	(B,7744)	(B,8)	$7745 \times 8 = 61,952$	$2 \times 7744 \times 8 = 123,904$
Linear	(B,8)	(B,2)	$8 \times 2 = 16$	$2 \times 8 \times 2 = 32$
Total			62,004	707,136

Table 13 summarizes the computational and memory requirements with parameters represented as FP32 numbers

Table 13: Computational and memory requirements with parameters represented as FP32 numbers

Model	#Params	#MAC	Memory Footprint
CNN-A	43,012	2,797,472	168 KB
CNN-B	62,004	707,136	242 KB

5.2 Transformer model

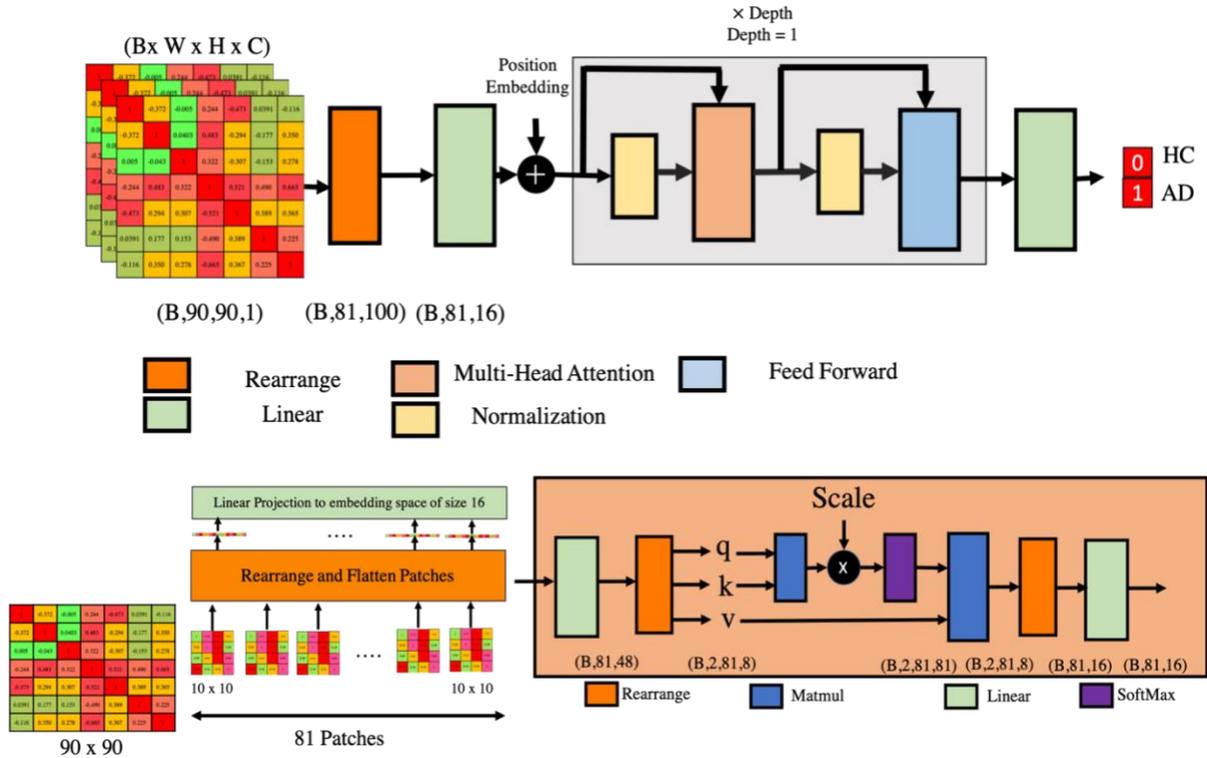


Figure 13: Proposed Transformer architecture

Transformer enables modeling long dependencies between input sequence elements. “Different from convolutional networks, transformers require minimal inductive biases for their design and are naturally suited as set functions “ [49]. As represented in Figure 13, Transformer architectures are based on a self-attention mechanism that learns the relationship between elements of the sequence; thus, they can attend to complete sequences, thereby learning long-range relationships. There is no need for inductive bias in a transformer, unlike a convolution neural network. An inductive bias is a design choice when creating a learning algorithm that relies on an assumption about data being processed. In computer vision, convolution neural networks are translational invariant; thus, they can recognize features regardless of position or angle within an image. However, it results in some undesirable behavior in some cases.

“Transformers has two advantages. First, when applied to images, it has been shown that the attention mechanism is a generalized case of convolution. That means transformers can learn to mimic convolutions. Secondly, transformers can learn the complex spatial relationships between high-level image features. Together, these advantages allow transformers to address the limitations of translation invariance while still being flexible enough to learn operations that can produce a state of the art performance.”[50]

An important feature of these models is their scalability to high complexity models and large-scale datasets. In this section, we explore the proposed transformer's architecture and compare it with a convolution neural network.

Transformers lack the inductive biases of Convolution Neural Networks (CNN), such as translation invariance and a locally restricted receptive field. Invariance means that an object in an image can be recognized even if its appearance or position varies. Translation in computer vision means that each pixel has been moved by a fixed amount in a particular direction. Moreover, convolution is a linear local operator, and only neighbor values indicated by the kernel can be seen. On the other hand, a transformer is by design permutation invariant, so it cannot process grid structure data. So, a sequence is needed, and non-sequential spatial data should be converted to a sequence. The proposed architecture for our study is inspired by this work [51].

To create a sequence of 2d structures, the 90x90 correlation matrices are split into patches. The patches are flattened to a 1D vector, positional embedding is added, and the final sequence is fed to the transformer encoder as input. Positional embedding is used to keep the notion of order. Positioned embeddings are constant numbers added to the embedding vector before the self-attention layer. The encoder block is identical to the original transformer

proposed in [52] for NLP tasks. The number of encoders or the depth of the encoder stack change according to the complexity of the task and dataset size. Transformers are composed of linear and fully connected layers, but the fundamental building block of a transformer is self-attention. Self-attention enables ML to find correlations between different input words that indicate the syntactic and contextual structure of the sentence. In the context of fMRI and brain connectome, we believe that self-attention enables ML to find correlations between different brain regions to capture, map, and understand the organization of neural interactions within the brain.

To feed 2D data to the Transformer encoder, each 2D data set is split into linearized patches. For example, with a 1x1 patch size, the sequence length would be 1x8100, while with a 6x6 patch size, the sequence length would be 225x36, and with a 9x9 patch size, the sequence length would be 81x100. The resulting vector is passed to a linear layer to be embedded in a lower dimension. The Transformer linear layer includes a matrix multiplication operation, in which the sequence from patches is multiplied with a learnable weight. With our focus on computational and memory efficiency, we chose patch sizes that result in square-like matrices. Specifically, we chose 10x10 patches for the correlation matrix, which results in 81 x 100 sequences. With 81 entities (tokens), the proposed Transformer model will have fewer parameters than 90 or 100 entities. The sequence is fed into a linear layer and embedded to size 16.

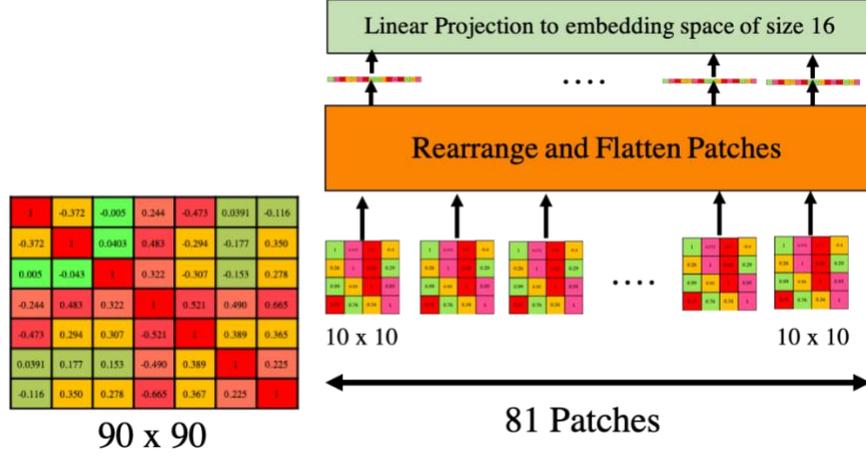


Figure 14: Extracting patches from correlation matrices to feed a sequence of data to the transformer

As represented in Figure14, the correlation matrices are rearranged into 81x100 sequences fed into a Transformer linear layer. The sequence is embedded in $d = 16$ dimensions. After the linear layer, the sequence has 81 entities of size $d = 16$. The sequence of size 81x16 is added to the position embedding vector. The position vector is a vector of constant numbers added to the sequence to keep the order of entities in the sequence. Our Transformer model has one encoder, with the most important components being a self-attention layer and a feed-forward layer. The goal of self-attention is to capture the interaction among all 81 entities (tokens) by encoding each entity in terms of global contextual information. This is done by defining three learnable weight matrices to transform Queries ($W^Q \in \mathbb{R}^{d \times d_q}$), Keys ($W^K \in \mathbb{R}^{d \times d_k}$), and Values ($W^V \in \mathbb{R}^{d \times d_v}$), where $d_q = d_k$. The input sequence $X \in \mathbb{R}^{81 \times 16}$ is first projected onto these weight matrices to get $Q = XW^Q$, $K = XW^K$, $V = XW^V$. The output $Z \in \mathbb{R}^{n \times d_v}$ of the self-attention layer is, $Z = \text{softmax}\left(\frac{QK^T}{\sqrt{d_q}}\right)V$. For a given entity in the sequence, the self-attention computes the dot-product of the query with all keys, which is then normalized using the SoftMax [53] operator to get the attention scores. Each entity then becomes the weighted

sum of all entities in the sequence, where weights are given by the attention scores (Figure 15) [49]. In order to capsule multiple complex relationships among different elements in the sequence, the multi-head attention is comprised of multi self-attention blocks. Each block has its own set of learnable weight matrices $\{W^{Q_i}, W^{K_i}, W^{V_i}\}$ where $i = 0, \dots, h$. In our model $h = 2$ and $d_q = d_k = d_v = 8$. The outputs from two heads are concatenated into a single matrix and projected onto a weight matrix via a linear layer (Figure 14). Transformer filters are dynamically calculated unlike CNN filters which stay the same for any input.

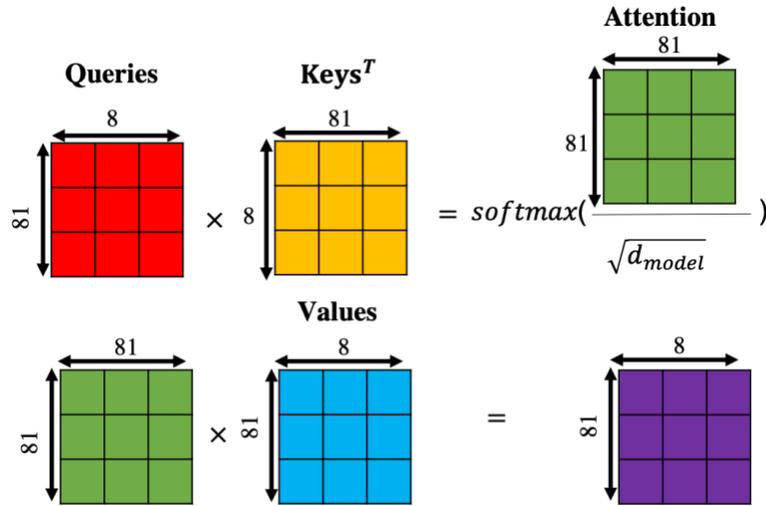


Figure 15: Self-attention for encoder layer of Transformer

The Transformer model characteristics are summarized in Table 14.

Table 14: Proposed Transformer architecture

Layer Name		Input	Output	#Params	#MAC
Rearrange		(B,90,90,1)	(B,81,100)	0	0
Linear		(B,81,100)	(B,81,16)	$101 \times 16 = 1616$	$2 \times 81 \times 16 \times 100 = 259,200$
Pos Embedding		(B,81,16)	(B,81,16)	$82 \times 16 = 1312$	0
Stack 1 (Begin)					
Normalization		(B,81,16)	(B,81,16)	$2 \times 16 = 32$	
Multi-Head Attention	Linear	(B,81,16)	(B,81,3x2x8)	$8 \times 16 \times 3 \times 2 = 768$	$2 \times 81 \times 48 \times 16 = 124,416$
	Dot Product (Q, K)	(B,81,16)	(B,2,81,81)		$2 \text{ (heads)} \times 2 \times 81 \times 81 \times 8 = 209,952$
	Attn= SoftMax	(B,2,81,81)	(B,2,81,81)		
	Dot Product (attn, V)	(B,2,81,81)	(B,2,81,81)		$2 \text{ (heads)} \times 2 \times 81 \times 81 \times 8 = 209,952$
	Rearrange	(B,2,81,81)	(B,81,16)		
	Linear	(B,81,16)	(B,81,16)	$17 \times 16 = 272$	$2 \times 81 \times 16 \times 16 = 41,472$
Normalization		(B,81,16)	(B,81,16)	$2 \times 16 = 32$	
Feedforward	Linear + GELU	(B,81,16)	(B,81,8)	$17 \times 8 = 136$	$2 \times 81 \times 8 \times 16 = 20,736$
	Linear	(B,81,8)	(B,81,16)	$9 \times 16 = 144$	$2 \times 81 \times 8 \times 16 = 20,736$
End of Stack				(Depth=1)x1384	(Depth=1) x886,464
Flatten and reshape		(B,81,16)	(B,1296)		0
Normalization		(B,1296)	(B,1296)	$2 \times 1296 = 2592$	
Linear		(B,1296)	(B,2)	$1297 \times 2 = 2594$	$2 \times 1296 \times 2 = 5184$
Total				9,498	891,648

5.3 Comparison of CNN and Transformer models computational and memory requirements.

The Transformer model has considerably fewer parameters than both CNN models, as seen from Table 15. It requires about 26% more MACs than the CNN-B model but far fewer MACs than the CNN-A model. We will later compare the accuracies achieved with the three models.

Table 15: Comparison of CNNs and Transformer architecture regarding memory footprint, and number of MACs

Model	#Params	#MAC	FP32 Parameter Memory
CNN-A	43,012	2,797,472	168 KB
CNN-B	62,004	707,136	242 KB
Transformer	9498	891,648	37.10 KB
Factor of Reduction Transformer / CNN-A	4.52X	3.137X	
Factor of Reduction Transformer / CNN-B	6.52X	0.79X	

6. DATA SETS FOR STATISTICAL MODEL EVALUATION

To evaluate the classification accuracy achieved by our CNNs, five classification experiments were carried out: AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, MCI-CN-ADNI, MCI-AD-ADNI. In the triplet notation, the first two denote the binary classification carried out, and the last data set used for the classification experiment. The respective dataset is divided into training and testing in all five experiments based on the 80/20 Pareto Principle [54]. The 20% subjects for the test sets were uniformly randomly chosen from the data set, and the remaining subjects in the data set were used for training. To keep the size of the two classes in the training sets approximately balanced, randomly created experimental training/test sets that had a larger than 20% difference between the two classes in the training set were discarded. Thirty training/test sets satisfying these conditions were generated for each of the five classification experiments.

In the first experiment (AD-CN-Mixed), the 97 MCI subjects were excluded from our data set, resulting in a set with 551 subjects with 111 subjects in the test set and 440 subjects in the training set. In the second experiment, AD-CN-OASIS, of the 346 subjects in the OASIS data set, the test set had 70 subjects, and the training sets had 276 subjects. In the third experiment, AD-CN-ADNI, of the subset of ADNI data set in which subjects were classified as either AD or CN subjects (set size 205), the test sets had 41 subjects, and the training sets 164 subjects. In the fourth experiment, MCI-CN-ADNI, the ADNI subset with AD subjects excluded is of size 201 with the test sets having 40 subjects and the training sets 161 Subjects. In the fifth experiment, MCI-AD-ADNI, the ADNI subset with subjects classified as either AD or MCI is of size 198, with the test sets having 39 subjects and 159 Subjects

7 TRAINING

7.1 Training procedure

We use the Adam Optimizer [55] with a fixed learning rate of 0.001 for the CNNs and the Transformer models. Both models are trained with no data augmentation in any of the experiments. We train the models for 15 epochs. On every epoch, one batch of four matrices is picked with batches picked in order to ensure every matrix is used only once. In each epoch, we shuffle the training set, so a different batch of matrices are input to the network in every epoch.

7.2 Training performance

Our two CNNs, CNN-A and CNN-B, and the Transformer network were implemented in Pytorch 1.5.0 with Python 3.7.7 and timings obtained on an NVIDIA K40 GPU.

The training times for the five classification experiments AD-CN-Mixed, AD-CN-ADNI, AD-CN-OASIS, MCI-CN-ADNI, MCI-AD-ADNI for three models, CNN-A, CNN-B, and Transformer, are reported in Table 16. For all five experiments, the Transformer architecture converges faster than the CNN models as expected from the model characteristics in Table 15. The Transformer model requires 14%, 14%, 11%, 7%, and 10% of the training time of the CNN-A model for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, MCI-CN-ADNI, MCI-AD-ADNI classification experiments, respectively.

Compared to the CNN-B model, the Transformer model requires 13%, 14%, 9%, 10%, and 10% of the CNN-B training time for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, MCI-CN-ADNI, MCI-AD-ADNI classification experiments, respectively.

Table 16: Training time for the CNN-A, CNN-B, and Transformer models for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, MCI-CN-ADNI, and MCI-AD-ADNI classification experiments

Data Set	Training Time (Sec)			Factor of reduction Transformer to CNN-B	Std		
	CNN-B	CNN-A	Transformer		CNN-B	CNN-A	Transformer
AD-CN-Mixed	201.72	188.88	27.41	6.71X	12.49	21.47	8.40
AD-CN-OASIS	121	119.81	17.11	7.07X	8.7	9.30	7.7
AD-CN-ANDI	65.10	53.29	5.96	10.92X	10.05	22.26	0.78
MCI-CN-ADNI	66.80	93.48	6.75	9.89X	5.06	6.33	1.68
MCI-AD-ADNI	63.70	68.96	6.9	9.23X	10.04	3.67	1.65

7.3 Parameter sparsity

By sparsity, we refer to the property that a subset of the model parameters has sufficiently small values approximating zero. With zero parameter values, multiplication and addition (which dominate neural network computation) can be optimized. Further, models can be stored and transmitted in a sparse format. It has been shown that neural networks can tolerate high levels of sparsity [56]. The smallest positive subnormal number represented in the IEEE FP32 standard is 2^{-149} . Subnormal numbers are the subset of denormalized numbers that fill the underflow gap around zero. In evaluating sparsity, we approximate any value smaller than 2^{-149} with zero. We measure the sparsity level of each convolution layer in our CNNs and Transformer models. The results are shown in Tables 17 and 18 for CNN-A, CNN-B. Based on Figure 12 in the proposed CNN-B, there are one convolution and two fully connected layers, and in the proposed CNN-A, there are three convolution and two fully connected layers. For each layer, we measure the sparsity of the input, input gradient (delta), weight, and weight gradient. In the CNN-B model, for the last fully connected layer, the sparsity level is all zeros; however, for the convolution and first fully connected layer, we observe some sparsity levels. For the convolution layer, the sparsity level of the input delta changes around 83% to 79%. The sparsity level of weight gradient and input change between 16% to 0.02%

and 34% to 29% for the first fully connected layer. For the proposed Transformer model, the sparsity level of different layers is zero.

Table 17: Sparsity level of different layers of the proposed CNN-A model

Layer		Sparsity
Convolution layer1	Weight	0%
	Weight grad	0%
	Input	0%
	Input delta	90%-84%
Convolution layer 2	Weight	0%
	Weight grad	0.1%-0%
	Input	0%
	Input delta	88%-80%
Convolution layer 3	Weight	0%
	Weight grad	51%-0.06%
	Input	50%-12%
	Input delta	90%-83%
FC1	Weight	0%
	Weight grad	54% - 13%
	Input	58%-29%
	Input delta	0%
FC2	Weight	0%
	Weight grad	0%
	Input activation	0%
	Input delta	0%

Table 18: Sparsity level of different layers of the proposed CNN-B model

Layer		Sparsity
Convolution layer	Weight	0%
	Weight grad	0%
	Input	0%
	Input delta	83%-79%
FC1	Weight	0%
	Weight grad	16%- 0.02%
	Input	34%-29%
	Input delta	0%
FC2	Weight	0%
	Weight grad	0%
	Input activation	0%
	Input delta	0%

8 EXPERIMENTAL RESULTS

We studied the effectiveness of the ROI formation on the classification accuracy for each of our ML models on each of the five classification experiments. For the ROI formation, we report the outcome for AAL-90 region center voxel used as center-voxel for the Hilbert curve segments of length 101 and 201 voxels. Similarly, we report the outcome for randomly chosen Hilbert curve segment center-voxels for 101 and 201 segment lengths. We also studied segments of lengths 51 and 301. Segment length 51 generally resulted in significantly lower accuracy, and we only include some of the results for this segment length. Segment lengths of 301 generally did not produce measurably better results than length 201 and are not included in reported results. For each of these four cases, we report the results for the CNN-A, CNN-B, and Transformer models for the AD-CN-Mixed, AD-CN-ADNI, AD-CN-OASIS, MCI-CN-ADNI, and MCI-AD-ADNI classification experiments. In addition, we report the average and standard deviation of True Positives (TP), True Negative (TN), False Positive (FP), False Negative (FN), Specificity (SP) $TN/(TN+FP)$, Sensitivity (SE) $TP/(TP+FN)$ and Accuracy (ACC) $(TP+TN)/(TP+TN+FP+FN)$.

8.1 AAL-90 region center-voxel based ROIs

8.1.1 CNN models

8.1.1.1 ROIs represented by vectors of time-averaged voxel intensity values (spatial)

As is seen in Tables 19 and 20, the accuracy for the ROIs of size 101 and 201 is comparable for all five classification experiments and both CNN models. However, generally, the accuracy for ROIs of size 101 and the CNN-B model is somewhat higher. The results are similar for sensitivity and specificity, though the results are somewhat more favorable for the

ROIs of size 101 than size 102 and CNN-B than for CNN-A. For example, an accuracy of 92% (the highest) was achieved for the MCI-CN-ADNI classification experiment using ROI size 101 and the CNN-B model. The ROI of size 51 combined with the CNN-A model gave the lowest accuracy of 78%.

The True Positives (TP), True Negative (TN), False Positive (FP), and False Negative (FN) results are reported in Tables 21 and 22.

Table 19: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for AD-CN-OASIS, AD-CN-ADNI, and AD-CN-Mixed subjects for spatial correlation matrices

Network	Path	AD-CN-OASIS (276 Train 70Test)			AD-CN -ADNI (164 Train 41 Test)			AD-CN-Mixed (440 Train 111 Test)		
		ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
CNN-A	201	85±4	82±7	88±5	78±5	79±9	78±9	85±3	84±5	86±4
	101	86±5	82±9	90±7	86±5	82±10	92±6	86±3	83±6	89±4
CNN-B	201	86±4	82±8	89±6	86±5	86±7	87±8	88±3	86±5	89±5
	101	87±4	82±9	90 ±5	89±5	87±9	91±7	88±2	85±4	90±4

Table 20: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for MCI-CN-ADNI and MCI-AD-ADNI subjects for spatial correlation matrices

Network	Path	MCI-CN-ADNI (161 Train 40Test)			MCI-AD-ADNI (159 Train 39Test)		
		ACC	SE	SP	ACC	SE	SP
CNN-A	201	80±6	81±9	79±11	80±4	81±8	80±8
	101	92±4	93±6	91±6	82±5	80±8	85±10
CNN-B	201	84±5	83±9	85±9	83±4	84±8	84±6
	101	92±4	92±7	92±6	84±6	80±8	90±9

Table 21: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI classification experiment for spatial correlation matrices

Network	Path	MCI-CN-ADNI (47% vs 52%) (161 Train 40 Test)			
		TN	TP	FP	FN
CNN-A	201	79±11	81±9	21±11	19±9
	101	91±6	93±6	9±6	7±6
CNN-B	201	85±9	83±9	15±9	17±9
	101	92±6	92±7	8±6	8±7

Table 22: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, and MCI-AD-ADNI classification experiments for spatial correlation matrices

		AD-CN-Mixed (54% vs 46%) (440 Train 111 Test)				AD-CN-OASIS (56% vs 44%) (276 Train 70 Test)			
Network	Path	TN	TP	FP	FN	TN	TP	FP	FN
CNN-A	201	86±4	84±5	14±4	16±5	88±5	82±7	12±5	18±7
	101	89±4	83±6	11±4	17±6	90±7	82±9	10±7	18±9
CNN-B	201	89±5	86±5	11±5	14±5	89±6	82±8	11±6	18±8
	101	90±4	85±5	10±4	15±5	90±5	82±9	10±5	18±9
		AD-CN-ADNI (52% vs 47%) (164 Train 41 Test)				MCI-AD-ADNI (49% vs 50%) (156 Train 39 Test)			
Network	Path	TN	TP	FP	FN	TN	TP	FP	FN
CNN-A	201	78±9	79±9	22±9	21±9	80±8	81±8	20±8	19±8
	101	92±6	82±10	8±6	18±10	85±10	80±8	15±10	20±8
CNN-B	201	87±8	86±7	13±8	14±7	84±6	84±8	16±6	16±8
	101	91±7	87±10	9±7	13±10	90±9	80±8	10±9	20±8

8.1.1.2 ROIs represented by vectors of concatenated voxel intensity values (spatial-temporal)

The results for ROIs of size 101 and 201 and the CNN-A and CNN-B models for the five classification experiments for correlation matrices based on the spatial-temporal ROI vectors are reported in Tables 23, 24, 25, and 26.

As is shown in Tables 23 and 24, the achieved accuracies are comparable for ROIs of sizes 101 and 201. Though the difference is slight, correlation matrices based on ROI spatial-temporal vectors of size 201 in a few cases result in higher accuracy than 101 size vectors for the CNN-B model. On the other hand, the 101-size vector-based correlation matrices for the CNN-A model generally result in slightly higher accuracy. For sensitivity and specificity, the results are mixed as to which ROI vector size yields better results for the two CNN models.

Table 23: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, and AD-CN-ADNI classification experiments for correlation matrices based on spatial-temporal ROI vectors

Net work	Path	AD-CN-OASIS (276 Train 70Test)			AD-CN-ADNI (164 Train 41 Test)			AD-CN-Mixed (440 Train 111 Test)		
		ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
CNN-A	201	84±3	81±8	87±5	86±6	83±9	90±7	85±3	83±5	87±4
	101	84±4	80±9	88±5	87±5	84±8	90±6	86±3	83±6	89±5
CNN-B	201	85±4	82±7	88±5	88±5	87±7	89±8	87±3	87±6	89±4
	101	86±4	83±8	90±4	89±5	85±9	93±6	86±2	85±4	89±4

Table 24: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the MCI-CN-ADNI and MCI-AD-ADNI classification experiments for correlation matrices based on spatial-temporal ROI vectors

Network	Path	MCI-CN-ADNI (161 Train 40Test)			MCI-AD-ADNI (159Train 39Test)		
		ACC	SE	SP	ACC	SE	SP
CNN-A	201	88±4	90±5	87±7	85±5	80±9	90±7
	101	89±4	89±7	90±5	86±4	82±9	90±8
CNN-B	201	90±3	90±7	91±5	88±4	84±7	92±5
	101	89±4	88±8	89±5	87±5	82±8	91±7

Table 25: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, and MCI-AD-ADNI classification experiments for correlation matrices based on spatial-temporal ROI vectors

Network	Path	AD-CN-Mixed (54% vs 46%) (440 Train 111 Test)				AD-CN-OASIS (56% vs 44%) (276 Train 70 Test)			
		TN	TP	FP	FN	TN	TP	FP	FN
CNN-A	201	87±4.4	83±5.5	13±4.4	17±5.5	87±5.9	81±8.2	13±5.9	19±8.2
	101	89±5	83±6.2	11±5	17±6	88±5.4	80±9	12±5	20±9
CNN-B	201	89±4.3	87±6	11±4.3	13±6	88±5.6	82±7.8	12±5.6	18±7.8
	101	89±4	85±4	11±4	15±4.9	90±4.6	83±8.8	10±4.4	17±6.6
Network	Path	AD-CN-ADNI (52% vs 47%) (164 Train 41 Test)				MCI-AD-ADNI (49% vs 50%) (159 Train 39 Test)			
CNN-A	201	90±7	84±9.7	10±7.2	16±9.7	90±7	80±9	10±7	20±9
	101	91±6.6	84±8.2	9±6.6	16±8.2	90±8	82±9	10±8	18±9
CNN-B	201	89±8.1	87±7.7	11±8	13±7	92±5	85±7.8	8±5.4	15±7.8
	101	93±6.2	85±8.8	7±6.2	15±8.8	92±7	83±8	8±7.6	17±8.3

Table 26: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI classification experiment for correlation matrices based on spatial-temporal ROI vectors

		MCI-CN-ADNI (47% vs 52%) (161 Train 40 Test)			
Network	Path	TN	TP	FP	FN
CNN-A	201	87±7	91±5	13±7	9±5
	101	90±5	90±7	10±5	10±7
CNN-B	201	91±5	90±7	9±5	10±7
	101	90±5	89±8	10±5	11±8

8.1.1.3 ROIs represented by vectors of space averaged voxel intensity values (temporal)

As is shown in Tables 27 and 28, accuracies are generally comparable for the two ROI sizes and the two CNN models. However, in most cases, the CNN-B model yields somewhat higher accuracy than the CNN-A model. In some cases, the 101 ROI size yields higher accuracy; in others, the 201 sizes, though generally, the difference is within the error tolerance. Sensitivity and specificity are also generally higher for the CNN-B model.

The True Positives (TP), True Negative (TN), False Positive (FP), and False Negative (FN) results are reported in Tables 29 and 30.

Table 27: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI classification experiments for correlation matrices based on ROI temporal vectors

		AD-CN -OASIS (276 Train 70Test)			AD-CN -ADNI (164 Train 41 Test)			AD-CN -combined (440 Train 111 Test)		
Network	Path	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
CNN-A	201	69±4	60±10	76±7	64±8	66±12	65±15	67±4	62±9	72±7
	101	69±5	63±11	73±6	69±6	69±13	70±10	69±4	65±8	71±7
CNN-B	201	74±5	68±9	80±7	73±7	73±11	76±14	73±4	72±8	74±7
	101	75±5	69±8	79±7	71±6	71±12	73±9	73±4	69±6	75±5

Table 28: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the MCI-CN-ADNI and MCI-AD-ADNI classification experiments for correlation matrices based on ROI temporal vectors

		MCI-CN-ADNI (161 Train 40Test)			MCI-AD-ADNI (159 Train 39Test)		
Network	Path	ACC	SE	SP	ACC	SE	SP
CNN-A	201	70±6	73±11	69±9	73±5	72±11	75±9
	101	68±5	69±10	67±13	76±6	75±10	78±8
CNN-B	201	75±6	79±9	72±9	74±7	73±12	77±8
	101	71±4	73±8	69±10	74±8	73±11	76±11

Table 29: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, and MCI-AD-ADNI classification experiments for correlation matrices based on ROI temporal vectors

		AD-CN-Mixed (54% vs 46%) (440 Train 111 Test)				AD-CN-OASIS (56% vs 44%) (276 Train 70 Test)			
Network	Path	TN	TP	FP	FN	TN	TP	FP	FN
CNN-A	201	72±7	62±9	28±7	38±9	76±7	60±10	24±7	40±10
	101	71±7	65±8	29±7	35±8	73±6	63±11	27±06	37±11
CNN-B	201	74±7	72±8	26±7	28±8	80±7	68±9	20±7	32±9
	101	75±5	69±6	25±5	31±6	79±7	69±8	21±7	31±8
		AD-CN-ADNI (52% vs 47%) (164 Train 41 Test)				MCI-AD-ADNI (49% vs 50%) (159 Train 39 Test)			
Network	Path	TN	TP	FP	FN	TN	TP	FP	FN
CNN-A	201	65±15	66±12	35±15	34±12	75±9	72±11	25±9	28±11
	101	70±10	69±13	30±10	31±13	78±8	75±10	22±8	25±10
CNN-B	201	76±14	73±11	24±14	27±11	77±8	73±12	23±8	27±12
	101	73±9	71±12	27±09	29±12	76±11	73±11	24±11	27±11

Table 30: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI classification experiment for correlation matrices based on ROI temporal vectors.

		MCI-CN-ADNI (47% vs 52%) (161 Train 40 Test)			
Network	Path	TN	TP	FP	FN
CNN-A	201	69±9	73±11	31±10	27±12
	101	67±13	69±10	33±13	31±10
CNN-B	201	72±9	79±9	28±9	21±10
	101	69±10	73±8	31±10	27±8

8.1.1.4 Discussion of ROI vector representation of voxel intensity values and CNN models

Table 31: smROI, Summary of CNN model and ROI size for best accuracy, sensitivity, and specificity for the three types of ROI vectors for AD-CN-OASIS, AD-CN-ADNI, AD-CN-Mixed

	AD-CN-OASIS			AD-CN-ADNI			AD-CN-Mixed		
	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
spatial-temporal	B/101	B/101	B/101	B/101	B/101	B/201	B/201	B/201	B/101/201
spatial	B/101	A/B/101/201	B/101	B/101	B/101	B/101	B/101	B/201	B/101
temporal	B/101	A/B/101/201	B/201	B/201	B/201	B/201	B/101/201	B/201	B/101

Table 32: smROI, Summary of CNN model and ROI size for best accuracy, sensitivity, and specificity for the three types of ROI vectors for MCI-CN-ADNI, MCI-AD-ADNI

	MCI-CN-ADNI			MCI-AD ADNI		
	ACC	SE	SP	ACC	SE	SP
spatial-temporal	B/201	A/201	B/201	B/101	B/101	B/101
spatial	B/101, A/201	A/101	B/101	B/101	B/201	B/101
temporal	B/201, A/101	B/201	B/201	A/101	A/101	A/101

Tables 31 and 32, for smROI, represent the summary of CNN models and ROI sizes for best accuracy, sensitivity, and specificity for the three types of ROI vectors.

As is seen from Tables 31 and 32, smROI for most classification experiments, the CNN-B model, and ROI vector size of 101 yielded the best results with respect to the accuracy, sensitivity, and specificity. The corresponding percentages are summarized in Tables 33 and 34, for smROI-2.

Table 33: smROI-2, Summary of best accuracy, sensitivity, and specificity for the three types of ROI vectors and the two CNN models for AD-CN-OASIS, AD-CN-ADNI, AD-CN-Mixed

	AD-CN-OASIS			AD-CN-ADNI			AD-CN-Mixed		
	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
spatial-temporal	86/4	83/8	90/4	89/5	87/7	93/6	87/3	87/6	89/4
spatial	87/4	82/7	90/5	89/5	87/9	92/6	88/2	86/5	90/4
temporal	75/5	69/8	80/7	73/7	73/11	76/14	73/4	72/8	75/5

Table 34: smROI-2, Summary of best accuracy, sensitivity and specificity for the three types of ROI vectors and the two CNN models for MCI-CN-ADNI, MCI-AD-ADNI

	MCI-CN-ADNI			MCI-AD-ADNI		
	ACC	SE	SP	ACC	SE	SP
spatial-temporal	90/3	90/5	91/5	88/4	84/7	92/5
spatial	92/4	93/6	92/6	84/6	84/8	90/9
temporal	75/6	79/9	72/9	76/6	75/10	78/8

As is seen from Tables 33 and 34, for smROI-2, both the spatial and spatial-temporal types of ROI vectors are superior with respect to the accuracy, sensitivity, and specificity compared to the conventional temporal type ROI vectors. Further, the differences between the spatial and spatial-temporal ROI vector results are slight. One or the other gives better results for the five classification experiments with no apparent pattern.

Comparison of results for the five classification experiments based on correlation matrices generated from the three types of ROI vectors: temporal (Solid Line), Spatial (Dash line), and spatial-temporal (Dot Line). The Spatial and spatial-temporal are represented in Figure 16.

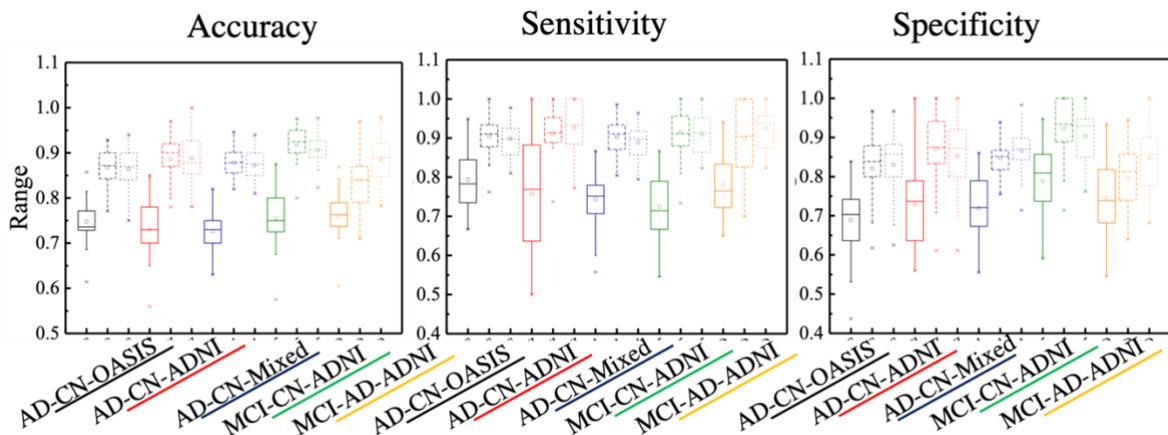


Figure 16: Comparison of results for the five classification experiments based on correlation matrices generated from the three types of ROI vectors: temporal (Solid Line), Spatial (Dash line), and spatial-temporal (Dot Line). The Spatial and spatial-temporal are represented in Figure 16.

8.1.2 The Transformer model

8.1.2.1 ROIs represented by vectors of time-averaged voxel intensity values (spatial)

As is seen in Tables 35 and 36, the accuracy for the ROIs of size 101 and 201 is comparable for all five classification experiments for the Transformer model. For MCI-CN-ADNI, with ROI size 201, we can achieve 100% accuracy, and with ROI size 101, the accuracy is 90%. However, generally, the accuracy for ROIs of size 101 is somewhat higher. The results are similar for sensitivity and specificity, though the results are somewhat more favorable for the ROIs of size 101 than size 102. An accuracy of 100% (the highest) was achieved for the MCI-CN-ADNI classification experiment using ROI size 201.

The True Positives (TP), True Negative (TN), False Positive (FP), and False Negative (FN) results are reported in Tables 37 and 38.

Table 35: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI classification experiments for correlation matrices based on ROI spatial vectors with Transformer model

Network	Path	AD-CN -OASIS (276 Train 70Test)			AD-CN -ADNI (164 Train 41 Test)			AD-CN -combined (440 Train 111 Test)		
		ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
Transformer	201	86±4	84±7	87±7	85±5	87±8	84±7	87±3	85±8	89±8
	101	87±4	85±7	90±5	87±4	87±7	87±7	87±3	85±5	89±5

Table 36: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the MCI-CN-ADNI, MCI-AD-ADNI classification experiments for correlation matrices based on ROI spatial vectors with Transformer model

Network	Path	MCI-CN-ADNI (161 Train 40Test)			MCI-AD-ADNI (159 Train 39Test)		
		ACC	SE	SP	ACC	SE	SP
Transformer	201	100±0	100±0	100±0	85±5	80±10	91±7
	101	90±4	93±6	88±7	84±5	80±9	89±8

Table 37: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, and MCI-AD-ADNI classification experiments for correlation matrices based on ROI spatial vectors with Transformer model

		AD-CN-Mixed (54% vs 46%) (440 Train 111 Test)				AD-CN-OASIS (56% vs 44%) (276 Train 70 Test)			
Network	Path	TN	TP	FP	FN	TN	TP	FP	FN
Transformer	201	89±8	85±6	11±5	15±6	87±7	84±7	13±7	16±7
	101	89±5	85±5	11±5	15±5	90±5	85±7	10±5	15±7
		AD-CN-ADNI (52% vs 47%) (164 Train 41 Test)				MCI-AD-ADNI (49% vs 50%) (159 Train 39 Test)			
Network	Path	TN	TP	FP	FN	TN	TP	FP	FN
Transformer	201	84±7	87±7	16±7	13±8	91±7	80±10	9±7	20±10
	101	88±7	87±7	12±7	13±7	89±8	80±9	11±8	20±9

Table 38: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI classification experiments for correlation matrices based on ROI spatial vectors with Transformer model

		MCI-CN-ADNI (47% vs 52%) (161 Train 40 Test)			
Network	Path	TN	TP	FP	FN
Transformer	201	100±0	100±6	0±0	0±0
	101	88±7	93±6	12±7	7±6

8.1.2.2 ROIs represented by vectors of concatenated voxel intensity values (spatial-temporal)

The results for ROIs of size 101 and 201 and the Transformer model for the five classification experiments for correlation matrices based on the spatial-temporal ROI vectors are reported in Tables 39, 40, 41, and 42.

As is shown in Tables 39 and 40, the achieved accuracies are comparable for ROIs of sizes 101 and 201. Though the difference is negligible, the difference is slight in order of 1%-2%. However, for MCI-AD-ANDI classification, ROI size 201 results in 87% accuracy, and ROI size 101 results in 76% accuracy, which is at least 10% degradation in the accuracy. For

sensitivity and specificity, the results are mixed as to which ROI vector size yields better results.

The True Positives (TP), True Negative (TN), False Positive (FP), and False Negative (FN) results are reported in Tables 41 and 42.

Table 39: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI classification experiments for correlation matrices based on ROI spatial-temporal vectors with Transformer model

Network	Path	AD-CN -OASIS (276 Train 70Test)			AD-CN -ADNI (164 Train 41 Test)			AD-CN -combined (440 Train 111 Test)		
		ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
Transformer	201	86±4	83±7	88±6	86±5	87±9	85±8	88±3	87±5	89±4
	101	86±4	83±8	89±5	87±4	85±8	89±6	87±3	84±6	89±5

Table 40: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-MCI-CN-ADNI, MCI-AD-ADNI classification experiments for correlation matrices based on ROI spatial-temporal vectors with Transformer model

Network	Path	MCI-CN-ADNI (161 Train 40Test)			MCI-AD-ADNI (159 Train 39Test)		
		ACC	SE	SP	ACC	SE	SP
Transformer	201	89±4	89±7	90±6	87±5	85±9	89±8
	101	88±5	88±14	89±11	76±5	73±11	79±12

Table 41: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, and MCI-AD-ADNI classification experiments for correlation matrices based on ROI spatial-temporal vectors with Transformer model

Network	Path	AD-CN-Mixed (54% vs 46%) (440 Train 111 Test)				AD-CN-OASIS (56% vs 44%) (276 Train 70 Test)			
		TN	TP	FP	FN	TN	TP	FP	FN
Transformer	201	89±4	87±5	11±4	13±5	88±6	83±7	12±6	17±7
	101	89±5	84±6	11±5	16±6	89±5	83±8	12±6	17±7
Network	Path	AD-CN-ADNI (52% vs 47%) (164 Train 41 Test)				MCI-AD-ADNI (49% vs 50%) (159 Train 39 Test)			
		TN	TP	FP	FN	TN	TP	FP	FN
Transformer	201	85±8	87±9	15±8	13±9	89±8	85±9	11±8	15±9
	101	89±6	85±8	11±6	15±8	79±12	73±11	21±12	27±11

Table 42: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI classification experiments for correlation matrices based on ROI spatial-temporal vectors with Transformer model

		MCI-CN-ADNI (47% vs 52%) (161 Train 40 Test)			
Network	Path	TN	TP	FP	FN
Transformer	201	89±11	88±14	11±11	12±4
	101	88±7	93±6	12±7	7±6

8.1.2.3 ROIs represented by vectors of space averaged voxel intensity values (temporal) Add results for Transformer on temporal ROI vectors

As is shown in Tables 43 and 44, accuracies are generally comparable for the two ROI sizes and the Transformer model. In some cases, the 201 ROI size yields higher accuracy; in others, the 101 sizes, though generally, the difference is within the error tolerance.

Table 43: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI classification experiments for correlation matrices based on ROI temporal vectors with Transformer model

		AD-CN -OASIS (276 Train 70Test)			AD-CN -ADNI (164 Train 41 Test)			AD-CN -combined (440 Train 111 Test)		
Network	Path	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
Transformer	201	75±6	68±10	80±8	74±5	73±11	77±14	73±4	68±7	77±7
	101	77±5	73±10	79±6	74±8	74±18	74±8	73±4	71±7	75±6

Table 44: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the MCI-CN-ADNI, MCI-AD-ADNI classification experiments for correlation matrices based on ROI temporal vectors with Transformer model

		MCI-CN-ADNI (161 Train 40Test)			MCI-AD-ADNI (159 Train 39Test)		
Network	Path	ACC	SE	SP	ACC	SE	SP
Transformer	201	76±5	77±10	77±9	75±7	78±7	72±12
	101	74±5	77±9	73±13	73±8	74±11	72±12

The True Positives (TP), True Negative (TN), False Positive (FP), and False Negative (FN) results are reported in Tables 45 and 46.

Table 45: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI, and MCI-AD-ADNI classification experiments for correlation matrices based on ROI temporal vectors with Transformer model

		AD-CN-Mixed (54% vs 46%) (440 Train 111 Test)				AD-CN-OASIS (56% vs 44%) (276 Train 70 Test)			
Network	Path	TN	TP	FP	FN	TN	TP	FP	FN
Transformer	201	77±7	68±7	23±7	32±7	80±8	68±10	20±8	32±10
	101	75±6	71±7	25±6	29±7	79±6	73±10	21±6	27±10
		AD-CN-ADNI (52% vs 47%) (164 Train 41 Test)				MCI-AD-ADNI (49% vs 50%) (159 Train 39 Test)			
Network	Path	TN	TP	FP	FN	TN	TP	FP	FN
Transformer	201	77±14	73±10	23±14	27±11	72±12	78±7	28±12	22±7
	101	75±9	71±10	25±9	29±18	72±12	74±11	28±12	26±11

Table 46: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for MCI-CN-ADNI classification experiments for correlation matrices based on ROI temporal vectors

		MCI-CN-ADNI (47% vs 52%) (161 Train 40 Test)			
Network	Path	TN	TP	FP	FN
Transformer	201	77±9	77±10	23±9	23±10
	101	73±13	77±9	27±13	23±9

8.1.2.4 Comparison of Transformer results for different ROI vector types and sizes

As seen from Table 47 and 48 smROI, ROI vector size of 101 yielded the best results with respect to the accuracy, sensitivity, and specificity for most classification experiments. The corresponding percentages are summarized in Table 49 and 50, smROI-2

Table 47: smROI, Summary ROI size for best accuracy, sensitivity, and specificity for the two types of ROI vectors with Transformer model for AD-CN-OASIS, AD-CN-ADNI, AD-CN-Mixed

		AD-CN-OASIS			AD-CN-ADNI			AD-CN-Mixed		
		ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
spatial-temporal		101/201	101/201	101	101	201	101	201	201	101/201
spatial		101	101	101	101	101	101	101/201	101/201	101/201
temporal		101	101	201	101/201	101	201	101/201	101	201

Table 48: smROI, Summary ROI size for best accuracy, sensitivity, and specificity for the two types of ROI vectors with Transformer model for MCI-CN-ADNI, MCI-AD-ADNI

	MCI-CN-ADNI			MCI-AD ADNI		
	ACC	SE	SP	ACC	SE	SP
spatial-temporal	201	201	201	201	201	201
spatial	201	201	201	201	101/201	201
temporal	101	101	201	101/201	101	201

As seen from Tables 47-50, smROI-2, both the spatial and spatial-temporal type of ROI vectors are superior with respect to the accuracy, sensitivity, and specificity compared to the conventional temporal type ROI vectors. Further, the differences between the spatial and spatial-temporal ROI vector results are slight. Therefore, one or the other gives better results for the five classification experiments with no apparent pattern.

Table 49: smROI-2, Summary of best accuracy, sensitivity, and specificity for the two types of ROI vectors with the transformer model for AD-CN-OASIS, AD-CN-ADNI, AD-CN-Mixed

	AD-CN-OASIS			AD-CN-ADNI			AD-CN-Mixed		
	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
spatial-temporal	86/4	83/8	89/5	87/4	87/9	89/6	88/3	87/5	89/4
spatial	87/4	85/7	90/5	87/4	87/7	87/7	87/3	85/5	89/5
temporal	77/5	73/10	79/6	74/5	74/18	77/14	73/4	71/7	77/7

Table 50: smROI-2, Summary of best accuracy, sensitivity, and specificity for the two types of ROI vectors and the transformer model for MCI-CN-ADNI, MCI-AD-ADNI

	MCI-CN-ADNI			MCI-AD-ADNI		
	ACC	SE	SP	ACC	SE	SP
spatial-temporal	89/4	89/7	90/6	87/5	85/9	89/8
spatial	100/0	100/0	100/0	85/5	80/9	91/7
temporal	77/5	73/10	80/8	74/5	74/18	77/14

8.2 Uniformly randomly selected center-voxels for ROI Hilbert curve segments

For this experiment, we select 90 random voxels as center points, and we make ROIs around these voxels by using segments of the Hilbert curve with 100 voxels before and after the center voxel. The classification accuracies for 5 five different groups of subjects AD-CN-Mixed, AD-CN-ADNI, AD-CN-OASIS, MCI-CN-ADNI, MCI-AD-ADNI with proposed Transformer and CNN-B model, ROI size 201, spatial-temporal correlations with random seed voxels, and AAL90 seed voxels are reported in Tables 51 and 52. The accuracy in classification of AD-CN-Mixed with 440 subjects in the training set and 111 subjects in the test set is 86%, 86% for AAL-90, and random center points, respectively. The accuracy in classification of AD-CN-OASIS with 276 subjects in the training set and 70 subjects in the test set is 88%, 85% for AAL-90, and random center points, respectively. The accuracy in classification of AD-CN-ADNI with 164 subjects in the training set and 41 subjects in the test set is 86%, 87% for AAL-90, and random center points, respectively.

The accuracy in the classification of MCI-CN-ADNI with 161 subjects in the training set and 40 subjects in the test set is 91%, 87% for AAL-90, and random center points, respectively.

The accuracy in the classification of MCI-AD-ADNI with 159 subjects in the training set and 39 subjects in the test set is 88%, 87% for AAL-90, and random center points, respectively.

Random seed voxels' ROI-ROI spatial-temporal correlation can result in the same accuracy as AAL-90 seed voxels' ROI-ROI spatial-temporal correlation, with only 2% to 4% difference.

Table 51: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-Mixed, AD-CN-OASIS, AD-CN-ADNI classification experiments for groups of subjects for two different seed voxels

Network	Seed voxel	AD-CN -OASIS (276 Train 70Test)			AD-CN -ADNI (164 Train 41 Test)			AD-CN-Mixed (440 Train 111 Test)		
		ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
Transformer	Random	85±4	82±8	88±5	87±4	85±10	88±8	86±4	83±6	88±5
	AAL90	88±4	85±7	90±4	86±4	86±7	86±8	86±4	83±7	89±5
CNN-B	Random	87±4	83±10	90±6	88±4	87±8	88±8	87±3	85±8	89±8
	AAL90	85±4	82±9	88±5	86±5	86±9	86±5	87±3	85±6	89±5

Table 52: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the MCI-CN-ADNI and MCI-AD-ADNI classification experiments

Network	Seed voxel	MCI-CN-ADNI (161 Train 40Test)			MCI-AD-ADNI (159 Train 39Test)		
		ACC	SE	SP	ACC	SE	SP
Transformer	Random	87±4	88±7	86±9	87±4	84±8	92±6
	AAL90	91±4	92±8	91±6	88±5	84±8	92±8
CNN-B	Random	88±4	89±6	86±7	86±6	85±8	87±12
	AAL90	91±4	92±8	91±6	87±6	84±8	90±14

9 ARTIFACT INFLUENCE ON CLASSIFICATION ACCURACY

9.1 Artifacts in fMRI

In Resting-State fMRI, establishing brain functional connectivity from observed BOLD signals may be measurably affected by various artifacts contributing significantly to the signal. Considerable efforts [57][58] have been devoted to reducing the impact of physiological signal contributions such as cardiac and pulmonary activities, head motion, temporal noise, and scanner imperfections on the observed BOLD signal. Head motions tend to increase short-range BOLD signal correlation and weaken long-range correlations [45]. Physiological contributions [44] to the BOLD signal contribute to BOLD signal correlation in space and time. Temporal noise [9], [44], [59], [60] is generally viewed as being uncorrelated with a Gaussian distribution [61]. Further, preprocessing such as voxel resampling from observed voxel sizes to the voxel size used for the reference brain, MNI-152 in our case, affect BOLD signal correlation, as does spatial smoothing [58]. Parallel imaging also impacts BOLD signal correlations [62][63]. Thus, there are many contributing factors to the BOLD signal correlation. Even though great care is taken to remove the impact of artifacts in signal acquisition and preprocessing, some influence will remain. It has been shown that a better signal to noise is achieved by using small voxel sizes in combination with spatial smoothing rather than large voxels since the physiological signal contribution is small relative to temporal signal contribution for small voxels [64][65]. The smoothing reduces the temporal noise since it is uncorrelated. Larger voxels increase the physiological contributions to the BOLD signal relative to the temporal contribution since physiological signal contributions are spatially correlated.

In this dissertation, functional connectivity and classification are based on correlation matrices of image intensity magnitude values in brain reconstructed images obtained through inverse Fourier transforms. The image intensity magnitude values generally have a non-central Chi distribution [62]. They are computed as the square root of the sum of the image's squared real and imaginary components resulting from the inverse Fourier transform. In the case of BOLD signals resulting from zero-mean Gaussian noise from multiple coils, the intensity magnitude has a central Chi distribution [55]. For a single coil, the intensity magnitude distribution is Rician, which becomes a Rayleigh distribution for a Gaussian zero-mean BOLD signal [63]. The deviation from the Gaussian distribution of the image intensity magnitude values increases with the number of coils [64] but decreases with the signal strength [66].

9.2 ROI intensity magnitude representation impact on Pearson correlation matrices

Artifact contributions to the BOLD signals and the corresponding image intensity magnitude values influence the ROI vector types differently. The classification results regarding the accuracy, sensitivity, and specificity were better for correlation matrices based on ROI vectors of type spatial and spatial-temporal than for type temporal. To relate this observation to properties of the Pearson correlation matrices, we computed the condition numbers of every temporal and spatial-temporal correlation matrix corresponding to each subject.

The results are shown in Tables 53 and 54 for correlation matrices generated for four classification experiments: MCI-CN-ADNI, MCI-AD-ADNI, AD-CN-OASIS, and AD-CN-ADNI. The condition number of the spatial correlation matrix corresponding to each subject was very close to the condition number of the spatial-temporal correlation matrix of the

subject; thus, we do not include the condition number of spatial correlation matrices in Table 53.

Table 53: Average, standard deviation, min, max of condition numbers for correlation matrices based on ROI vectors of type temporal and spatial-temporal for all subjects in the MCI-CN-ADNI and MCI-AD-ADNI classification experiments

	MCI-CN-ADNI		MCI-AD-ADNI	
	spatial-temporal	temporal	spatial-temporal	temporal
Mean	2,460.84	1,583,423,949.13	2,390.24	1,208,911,902.08
STD	601.52	14,143,448,524.59	660.86	10,609,540,371.99
MAX	4,327.08	178,419,136,240.55	4,431.32	144,267,680,104.60
MIN	885.52	5,252.91	382.71	8,330.48

Table 54: Average, standard deviation, min, max of condition numbers for correlation matrices based on ROI vectors of type temporal and spatial-temporal for all subjects in the AD-CN-OASIS and AD-CN-ADNI classification experiments

	AD-CN-OASIS		AD-CN-ADNI	
	spatial-temporal	temporal	spatial-temporal	temporal
Mean	3,892.32	39,230,928.99	2,445.08	128,748,047.03
STD	35,775.26	728,786,129.86	655.03	1,320,772,115.64
MAX	667,361.49	13,556,256,743.47	4,431.32	17,296,434,470.60
MIN	848.25	7,480.35	382.71	5,252.91

The condition numbers are smaller for the correlation matrices based on spatial-temporal ROI vectors. Given this fact, it is not surprising that using correlation matrices based on spatial-temporal ROI vectors for classification yields higher accuracy.

9.3 Noise impact on Pearson correlation matrices

To assess the impact of noise on the classification accuracy, we evaluated the impact of artificial Gaussian noise added to the image intensity magnitude voxel values on the accuracy of the CNN-B model. The noise was generated using the *randn* MATLAB function, which

returns normally distributed random values with zero mean and a standard deviation of one. Two noise scenarios were investigated: a) noise spatially correlated within each ROI but uncorrelated between ROIs, b) noise uncorrelated within each ROI but correlated between ROIs. For case a) the same random value was added to the voxel intensity magnitude value for every voxel along the Hilbert curve segment representing an ROI with a different random number used for different ROIs. For b) the same random number was added to corresponding Hilbert curve segment voxels in every ROI with different random numbers added to different voxels in the ROI segments. Thus, for a) 90 random numbers were used for each time instant, and for b) the number of random numbers for each time instant equaled the Hilbert curve segment length. For each time instant, a new set of random numbers were generated in both cases.

In addition to studying the impact of noise correlation, we also studied the impact of the magnitude of the added artificial noise by multiplying the *randn* generated numbers by 150, 450, and 900.

The voxel intensity magnitude characteristics for the four cases are summarized in Table 55 for case a), artificial noise correlated within ROIs but uncorrelated between ROIs, and in Table 56 for case b), artificial noise uncorrelated within ROIs but correlated between ROIs.

Table 55: Impact on voxel intensity statistics of artificial noise correlated within ROIs but uncorrelated between ROIs

	Measured voxel intensity			Measured voxel intensity + noise, scale factor 150			Measured voxel intensity + noise, scale factor 450			Measured voxel intensity + noise, scale factor 900		
	Mean	Max	Std dev	Mean	Max	Std dev	Mean	Max	Std dev	Mean	Max	Std dev
ADNI	1266	2794	1152	1274	3188	1133	2308	4329	1497	1274	6343	1130
OASIS	592	952	98	590	1495	103	590	2934	103	590	5183	85

Table 56: Impact on voxel intensity statistics of artificial noise uncorrelated within ROIs but correlated between ROIs

	Measured voxel intensity			Measured voxel intensity + noise, scale factor 150			Measured voxel intensity + noise, scale factor 450			Measured voxel intensity + noise, scale factor 900		
	Mean	Max	Std dev	Mean	Max	Std dev	Mean	Max	Std dev	Mean	Max	Std dev
ADNI	1266	2794	1266	1813	2801	1648	1272	2810	1128	1224	2721	968
OASIS	592	952	98	591	951	93	590	952	98	589	952	98

The relative magnitude of the peak artificial noise to the mean and max voxel intensity for the ADNI and OASIS data sets are summarized in Table 57. The voxel intensities for the ADNI data set are more than twice that of the OASIS data set, causing the artificial noise to impact the voxel intensities significantly. For artificial noise with a 900-scale factor, the peak artificial noise is larger than the max measured voxel intensity.

Table 57: Peak artificial noise magnitude relative to the mean and max voxel intensities for the ADNI and OASIS data sets

	Peak artificial noise 150	Peak artificial noise 450	Peak artificial noise 900
ADNI			
Mean ADNI	11%	34%	69%
Max ADNI	5%	16%	32%
OASIS			
Mean OASIS	25%	76%	106%
Max OASIS	15%	47%	94%

The noise magnitude relative to the voxel intensities of ADNI was varied between approximately 5%-32% percent of the max and 11%-69% percent of the mean voxel intensities of ADNI by scaling the *randn* numbers with 150, 450, and 900.

The noise magnitude relative to the voxel intensities of OASIS was varied between approximately 15%-94% percent of the max and 25%-106% percent of the mean for OASIS dataset by scaling the *randn* numbers with 150, 450 and 900.

9.4 CNN-B classification accuracy with intra ROI (spatially local) correlated but inter ROI and time uncorrelated noise added to the measured Voxel intensity

The data indicate that the accuracy of CNN-B classification based on ROI temporal vector correlation matrices degrades more significantly with intra ROI correlated and inter ROI uncorrelated noise than the accuracy of classification based on ROI spatial-temporal vector correlation matrices.

As represented in Figure 17 and Tables 58 and 59, for AD-CN-ADNI subjects, with spatial-temporal correlation, accuracy drops from 88% to 87%, sensitivity drops from 87% to 85%, and specificity changes between 89% and 91%. However, with temporal correlation, accuracy, sensitivity, and specificity drops from 73% to 54%, 73% to 48%, and 76% to 63%.

For AD-CN-OASIS subjects with spatial-temporal correlation, accuracy drops from 85% to 69%, sensitivity drops from 82% to 46%, and specificity changes between 88% to 89%. With temporal correlation, accuracy drops from 74% to 51%, sensitivity drops from 68% to 26%, and specificity drops from 80% to 72%.

Figure 18 represents the accuracy, sensitivity, and specificity of two different groups of subjects, MCI-CN-ADNI and MCI-AD-ADNI. The experiments are spatial-temporal/temporal correlations between noise-free signals, spatial-temporal/temporal correlations between signals and artificially added noise inside ROIs in which noise peak is 150, and spatial-temporal/temporal correlations between signals and artificially added noise inside ROIs in which noise peak is 450.

As represented in Figure 18 and Tables 58 and 60, for MCI-CN-ADNI subjects, with spatial-temporal correlation, accuracy drops from 90% to 82%, sensitivity drops from 90% to 84%, and specificity drops from 91% to 81%. With temporal correlation, accuracy drops from 75% to 57%, sensitivity drops from 79% to 45%, and specificity drops from 72% to 68%.

For MC-AD-ADNI subjects with spatial-temporal correlation, accuracy drops from 88% to 86%, sensitivity changes between 84% to 78%, and specificity drops from 92% to 95%. With temporal correlation, accuracy drops from 74% to 48%, sensitivity drops from 73% to 53%, and specificity drops from 77% to 45%.

Table 58: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-OASIS, AD-CN-ADNI, MCI-CN-ADNI, and MCI-AD-ADNI classification experiments for intra ROI correlated and inter ROI uncorrelated artificial noise with scale factors 150 and 450 for 201 ROI length In the table ST stands for spatial-temporal ROI vectors being used for correlation matrices

Correlation	AD-CN-OASIS (276 Train 70Test)			AD-CN-ADNI (164 Train 41 Test)			MCI-CN-ADNI (161 Train 40Test)			MCI-AD-ADNI (159 Train 39 Test)		
	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
Spatial-Temporal	85±4	82±7	88±5	88±5	87±7	89±8	90±3	90±7	91±5	88±4	84±7	92±5
ST Noisy-150	84±4	77±9	89±5	87±5	86±7	89±8	90±4	90±7	90±6	87±5	83±9	93±6
ST Noisy-450	69±5	46±16	88±10	87±6	85±8	91±8	82±6	84±8	81±12	86±5	78±8	95±5
Temporal	74±5	68±9	80±7	73±7	73±11	76±14	75±6	79±9	72±9	74±7	73±12	77±8
Temporal Noisy-150	56±5	26±19	80±14	64±6	55±13	75±10	68±7	60±18	77±15	71±6	64±15	80±11
Temporal Noisy-450	51±5	27±17	72±17	54±5	48±15	63±14	57±8	45±11	68±18	48±6	53±18	45±21

Table 59: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-OASIS, and AD-CN-ADNI classification experiments for intra ROI correlated and inter ROI uncorrelated artificial noise with scale factors 150 and 450 for 201 ROI length In the table ST stands for spatial-temporal ROI vectors being used for correlation matrices

Correlation	AD-CN-OASIS (56% vs 44%) (276 Train 70 Test)				AD-CN-ADNI (52% vs 47%) (164 Train 41 Test)			
	TN	TP	FP	FN	TN	TP	FP	FN
ST	88±5	82±7	11±5	17±5	87±7	89±8	12±4	11±7
ST Noisy-150	89±5	77±9	11±5	23±9	89±8	86±7	11±8	14±7
ST Noisy-450	88±10	46±16	12±10	54±17	91±8	85±9	9±8	15±9
Correlation	AD-CN-OASIS (56% vs 44%) (276 Train 70 Test)				AD-CN-ADNI (52% vs 47%) (164 Train 41 Test)			
	TN	TP	FP	FN	TN	TP	FP	FN
Temporal	80±7	68±9	20±7	32±9	76±14	73±11	24±14	27±11
Temporal Noisy-150	80±14	26±19	20±15	74±19	75±10	55±13	25±10	45±13
Temporal Noisy-450	72±18	27±18	28±18	73±18	63±15	48±15	37±15	52±15

Table 60: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI, and MCI-AD-ADNI classification experiments for intra ROI correlated and inter ROI uncorrelated artificial noise with scale factors

	MCI-CN-ADNI (47% vs 52%) (161 Train 40 Test)				MCI-AD-ADNI (49% vs 50%) (159 Train 39 Test)			
Correlation	TN	TP	FP	FN	TN	TP	FP	FN
ST	91±5	90±7	9±5	10±7	92±5	85±7.8	8±5.4	15±7.8
ST Noisy-150	90±7	90±8	10±7	10±8	93±6	83±9	7±6	17±9
ST Noisy-450	81±12	84±8	19±12	15±8	95±5	78±8	5±5	22±8
	MCI-CN-ADNI (47% vs 52%) (161 Train 40 Test)				MCI-AD-ADNI (49% vs 50%) (159 Train 39 Test)			
Correlation	TN	TP	FP	FN	TN	TP	FP	FN
Temporal	72±9	79±9	28±9	21±10	77±8	73±12	23±8	27±12
Temporal Noisy-150	77±15	60±18	23±15	40±18	80±11	64±15	20±12	36±15
Temporal Noisy-450	68±18	45±11	32±18	55±11	45±21	53±18	55±21	47±19

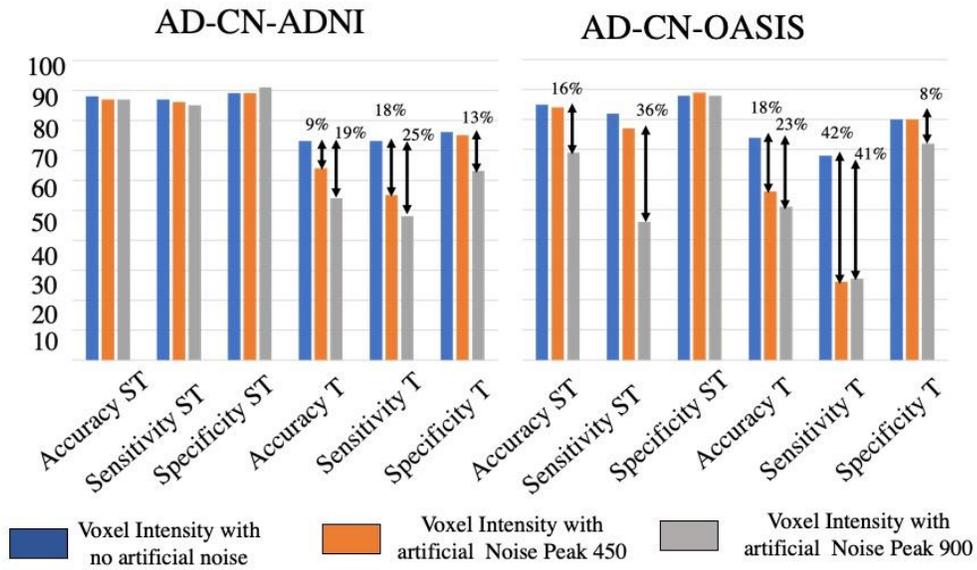


Figure 17: Accuracy, sensitivity, and specificity change for the AD-CN-ADNI and AD-CN-OASIS, AD-CN-Mixed classification experiments for correlation matrices based on no artificial noise added to voxel intensity values, and on ROI vectors with intra ROI correlated and inter ROI uncorrelated artificial noise added to voxel intensity values with noise scale factors of 150 and 450

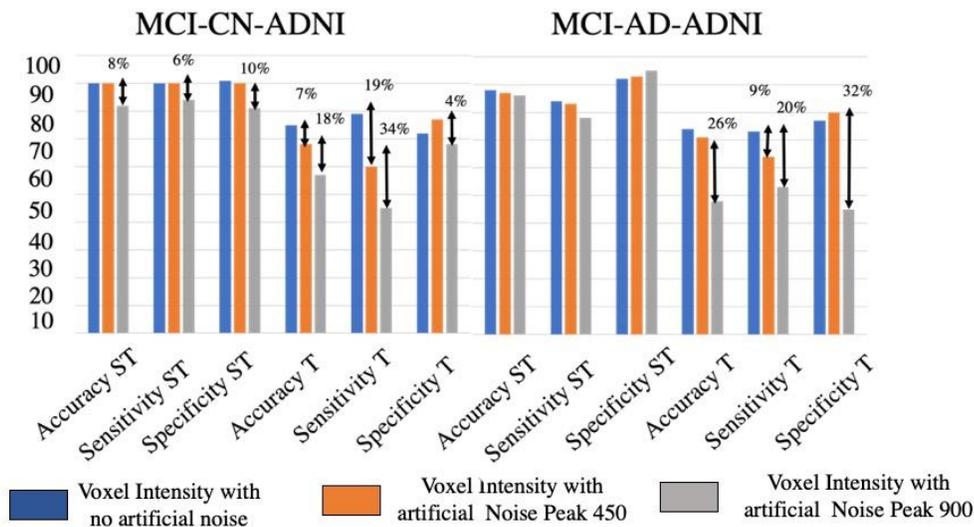


Figure 18: Accuracy, sensitivity, and specificity change for the MCI-CN-ADNI and MCI-AD-ADNI classification experiments for correlation matrices based on no artificial noise added to voxel intensity values and on ROI vectors with intra ROI correlated and inter ROI uncorrelated artificial noise added to voxel intensity values with noise scale factors of 150 and 450

9.5 CNN-B classification accuracy with intra ROI and temporally uncorrelated but inter ROI correlated noise added to the measured voxel intensity

With intra ROI uncorrelated and inter ROI correlated noise added to the voxel intensity values, the impact on classification accuracy is mixed with respect to basing correlation matrices on ROI spatial-temporal or temporal vectors. For some classification experiments, the relative impact is greater for classification based on correlation matrices based on ROI spatial-temporal vectors; in other cases, the noise impact is the opposite with respect to the basis for the correlation matrices.

As represented in Figure 19 and Tables 61 and 62, for AD-CN-ADNI subjects, with spatial-temporal correlation, accuracy changes between 89% to 88%, sensitivity changes between 88% and 87%, and specificity changes between 89% and 91%. However, with temporal

correlation, accuracy, sensitivity, and specificity drops from 73% to 68%, 73% to 68%, and 76% to 69%.

For AD-CN-OASIS subjects with spatial-temporal correlation, accuracy drops from 85% to 75%, sensitivity drops from 82% to 64%, and specificity drops from 88% and 84%. With temporal correlation, accuracy drops from 74% to 69%, sensitivity drops from 68% to 53%, and specificity changes between 80% to 81%.

Based on Figure 20 and Tables 61 and 63, for MCI-CN-ADNI subjects with spatial-temporal correlation, accuracy drops from 90% to 86%, sensitivity drops from 90% to 87%, and specificity drops from 91% to 86%. For temporal correlation, accuracy drops from 75% to 74%, sensitivity changes between 79% and 78%, and specificity drops from 72% to 71%.

For MCI-AD-ADNI subjects with spatial-temporal correlation, accuracy drops from 88% to 86%, sensitivity changes between 84% and 82%, and specificity changes between 92% to 91%. With temporal correlation, accuracy changes between 74% to 76%, sensitivity changes between 73% and 75%, and specificity changes between 78% to 77%.

Table 61: Average (%) and standard deviation (%) of accuracy, sensitivity, and specificity on thirty different test sets for the AD-CN-OASIS, AD-CN-ADNI, MCI-CN-ADNI, and MCI-AD-ADNI classification experiments for intra ROI uncorrelated and inter ROI correlated artificial noise with scale factors 450 and 900 In the table ST stands for spatial-temporal ROI vectors being used for correlation matrix

Correlation	AD-CN-OASIS (276 Train 70Test)			AD-CN-ADNI (164 Train 41 Test)			MCI-CN-ADNI (161 Train 40Test)			MCI-AD-ADNI (159 Train 39 Test)		
	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
Spatial- Temporal	85±4	82±7	88±5	88±5	87±7	89±8	90±3	90±7	91±5	88±4	84±7	92±5
ST Noisy-450	79±5	72±11	85±07	89±5	88±8	91±7	90±5	91±6	90±08	86±5	80±8	93±6
ST Noisy-900	75±5	64±8	84±7	89±6	88±7	90±10	86±5	87±7	86±9	86±5	82±9	92±6
Temporal	74±5	68±9	80±7	73±7	73±11	76±14	75±6	79±9	72±9	74±7	73±12	77±8
Temporal Noisy-450	71±6	60±10	79±07	73±6	71±13	77±8	77±8	78±15	76±09	75±8	75±09	76±12
Temporal Noisy-900	69±5	53±10	81±8	68±7	68±13	69±10	74±7	78±12	71±11	76±8	75±10	78±11

Table 62: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the AD-CN-ADNI, and AD-CN-OASIS classification experiments for intra ROI uncorrelated and inter ROI correlated artificial noise with scale factors 450 and 900

		AD-CN-OASIS (56% vs 44%) (276 Train 70 Test)				AD-CN-ADNI (52% vs 47%) (164 Train 41 Test)			
Correlation		TN	TP	FP	FN	TN	TP	FP	FN
ST		85±4	82±7	15±5	17±5	85±4	82±7	15±5	17±5
ST	Noisy-450	85±07	72±11	15±7	28±11	91±7	88±8	09±7	12±8
ST	Noisy-900	84±7	64±8	16±7	36±8	90±10	88±7	10±10	12±07
		AD-CN-OASIS (56% vs 44%) (276 Train 70 Test)				AD-CN-ADNI (52% vs 47%) (164 Train 41 Test)			
Correlation		TN	TP	FP	FN	TN	TP	FP	FN
Temporal		80±7	68±9	20±7	32±9	76±14	73±11	24±14	27±11
Temporal	Noisy-450	79±07	60±10	21±07	40±10	77±8	71±13	23±08	29±13
Temporal	Noisy-900	81±8	53±10	19±8	47±10	69±10	68±13	31±10	32±13

Table 63: Average (%) and standard deviation (%) of TN, TP, FP, FN on thirty different test sets for the MCI-CN-ADNI, and MCI-AD-ADNI classification experiments for intra ROI uncorrelated and inter ROI correlated artificial noise with scale factors 450 and 900

		MCI-CN-ADNI (47% vs 52%) (161 Train 40 Test)				MCI-AD-ADNI (49% vs 50%) (159 Train 39 Test)			
Correlation		TN	TP	FP	FN	TN	TP	FP	FN
ST		91±5	90±7	9±5	10±7	92±5	85±7.8	8±5.4	15±7.8
ST	Noisy-450	90±08	91±6	10±8	9±6	93±6	80±8	7±6	20±8
ST	Noisy-900	86±9	87±7	14±9	13±7	92±6	82±9	8±6	18±9
		MCI-CN-ADNI (47% vs 52%) (161 Train 40 Test)				MCI-AD-ADNI (49% vs 50%) (159 Train 39 Test)			
Correlation		TN	TP	FP	FN	TN	TP	FP	FN
Temporal		72±9	79±9	28±9	21±10	77±8	73±12	23±8	27±12
Temporal	Noisy-450	76±09	78±15	24±09	22±15	76±12	75±09	24±12	25±09
Temporal	Noisy-900	71±11	78±12	29±11	22±12	78±11	75±10	22±11	25±10

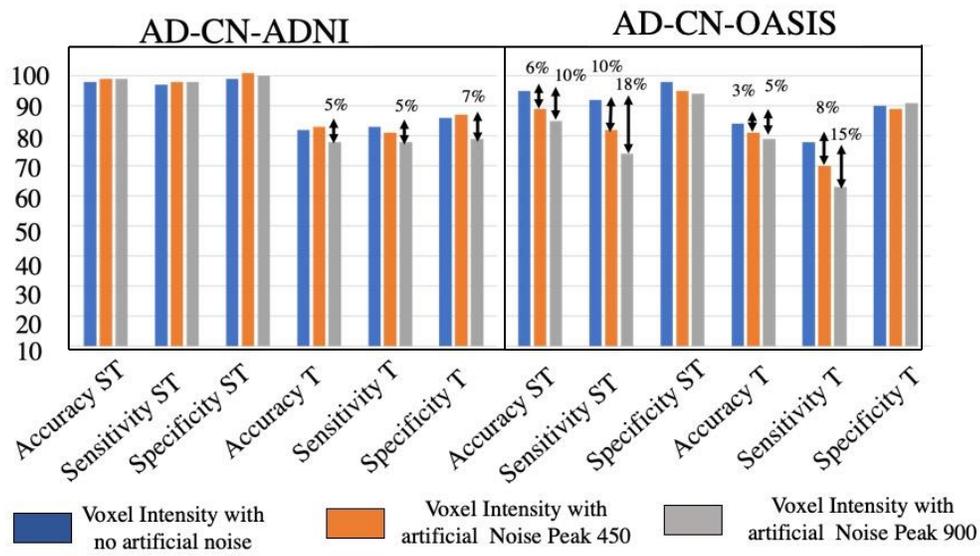


Figure 19: Accuracy, sensitivity, and specificity change for the AD-CN-ADNI, AD-CN-OASIS classification experiments for intra ROI uncorrelated and inter ROI correlated artificial noise with scale factors 450 and 900

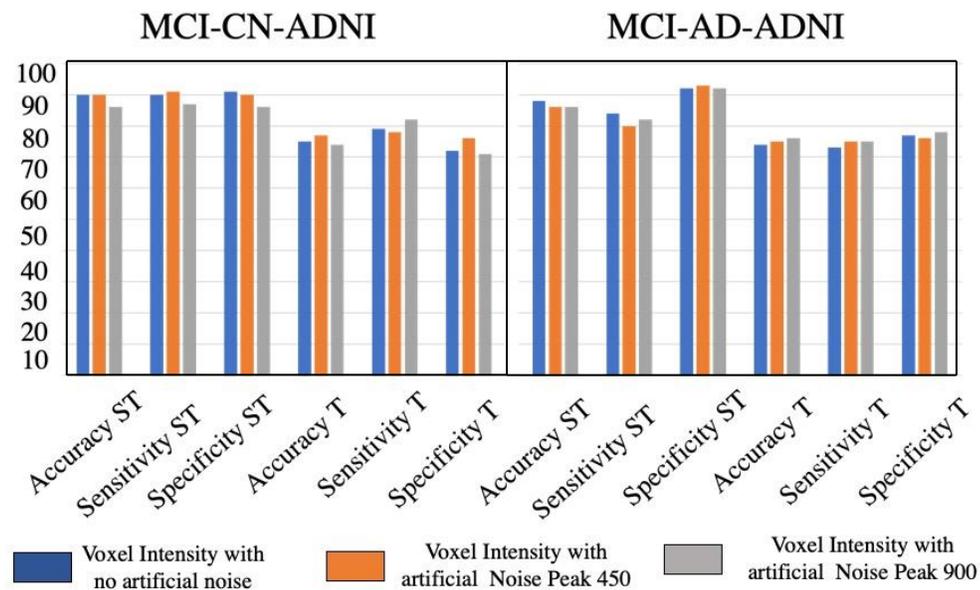


Figure 20: Accuracy, sensitivity, and specificity change for the MCI-CN-ADNI, and MCI-AD-ADNI classification experiments for intra ROI uncorrelated and inter ROI correlated artificial noise with scale factors 450 and 900

9.6 Summary of the artificial noise study

Generally, as expected, noise hurts classification accuracy. But it is interesting to note that in our artificial noise study intra ROI correlated and inter ROI uncorrelated noise appears to have a greater negative impact on classification accuracy than intra ROI uncorrelated but inter ROI correlated noise. For intra ROI correlated and inter ROI uncorrelated noise, the largest negative impact on classification accuracy is for correlation matrices based on temporal ROI vectors. For intra ROI uncorrelated and inter ROI correlated noise, the impact on classification accuracy is mixed with respect to which ROI vector type is used for the ROI-to-ROI correlation matrices.

BIBLIOGRAPHY

- [1] L. L. Beason-Held *et al.*, “Changes in brain function occur years before the onset of cognitive impairment,” *J. Neurosci.*, vol. 33, no. 46, pp. 18008–18014, 2013.
- [2] R. C. Mohs and K. L. Davis, “Reliability of the family history method in genetic studies of Alzheimer’s disease and related dementias,” *Am J Psychiatry*, vol. 143, no. 1, pp. 1279–1282, 1986.
- [3] A. Association, “2019 Alzheimer’s disease facts and figures,” *Alzheimer’s Dement.*, vol. 15, no. 3, pp. 321–387, 2019.
- [4] N. Kanwisher, “Functional specificity in the human brain: a window into the functional architecture of the mind,” *Proc. Natl. Acad. Sci.*, vol. 107, no. 25, pp. 11163–11170, 2010.
- [5] K. Brodmann, *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth, 1909.
- [6] R. S. Desikan *et al.*, “An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest,” *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.
- [7] N. Tzourio-Mazoyer *et al.*, “Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain,” *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.
- [8] A. Henssen *et al.*, “Cytoarchitecture and probability maps of the human medial orbitofrontal cortex,” *Cortex*, vol. 75, pp. 87–112, 2016.
- [9] R. C. Craddock, G. A. James, P. E. Holtzheimer III, X. P. Hu, and H. S. Mayberg, “A whole brain fMRI atlas generated via spatially constrained spectral clustering,” *Hum.*

- Brain Mapp.*, vol. 33, no. 8, pp. 1914–1928, 2012.
- [10] L. Fan *et al.*, “The human brainnetome atlas: a new brain atlas based on connectional architecture,” *Cereb. cortex*, vol. 26, no. 8, pp. 3508–3526, 2016.
- [11] J. Ashburner *et al.*, “SPM12 manual,” *Wellcome Trust Cent. Neuroimaging, London, UK*, vol. 2464, 2014.
- [12] D. L. Collins *et al.*, “Design and construction of a realistic digital brain phantom,” *IEEE Trans. Med. Imaging*, vol. 17, no. 3, pp. 463–468, 1998.
- [13] “[https://neurovault.org/collections/262/.](https://neurovault.org/collections/262/)”
- [14] “[https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT,](https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT)” .
- [15] G. Grabner, A. L. Janke, M. M. Budge, D. Smith, J. Pruessner, and D. L. Collins, “Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2006, pp. 58–66.
- [16] “[https://brainmap.org/training/BrettTransform.html.](https://brainmap.org/training/BrettTransform.html)”
- [17] “[https://en.wikipedia.org/wiki/Talairach_coordinates.](https://en.wikipedia.org/wiki/Talairach_coordinates)”
- [18] S. Ogawa, T.-M. Lee, A. R. Kay, and D. W. Tank, “Brain magnetic resonance imaging with contrast dependent on blood oxygenation,” *Proc. Natl. Acad. Sci.*, vol. 87, no. 24, pp. 9868–9872, 1990.
- [19] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, “Functional connectivity in the motor cortex of resting human brain using echo-planar MRI,” *Magn. Reson. Med.*, vol. 34, no. 4, pp. 537–541, 1995.
- [20] C. F. Beckmann, M. DeLuca, J. T. Devlin, and S. M. Smith, “Investigations into resting-state connectivity using independent component analysis,” *Philos. Trans. R.*

- Soc. B Biol. Sci.*, vol. 360, no. 1457, pp. 1001–1013, 2005.
- [21] J. A. Etzel, V. Gazzola, and C. Keysers, “An introduction to anatomical ROI-based fMRI classification analysis,” *Brain Res.*, vol. 1282, pp. 114–125, 2009.
- [22] K. Pearson, “On lines of closes fit to system of points in space, London, E dinb,” *Dublin Philos. Mag. J. Sci*, vol. 2, pp. 559–572, 1901.
- [23] T.-W. Lee, “Independent component analysis,” in *Independent component analysis*, Springer, 1998, pp. 27–66.
- [24] Q. Yu, E. A Allen, J. Sui, M. R Arbabshirani, G. Pearlson, and V. D Calhoun, “Brain connectivity networks in schizophrenia underlying resting state functional magnetic resonance imaging,” *Curr. Top. Med. Chem.*, vol. 12, no. 21, pp. 2415–2425, 2012.
- [25] Y. Xiang, J. Wang, G. Tan, F.-X. Wu, and J. Liu, “Schizophrenia identification using multi-view graph measures of functional brain networks,” *Front. Bioeng. Biotechnol.*, vol. 7, p. 479, 2020.
- [26] A. Khazaei, A. Ebrahimzadeh, and A. Babajani-Feremi, “Identifying patients with Alzheimer’s disease using resting-state fMRI and graph theory,” *Clin. Neurophysiol.*, vol. 126, no. 11, pp. 2132–2141, 2015.
- [27] G. Chen *et al.*, “Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging,” *Radiology*, vol. 259, no. 1, pp. 213–221, 2011.
- [28] B. Jie, D. Zhang, W. Gao, Q. Wang, C.-Y. Wee, and D. Shen, “Integration of network topological and connectivity properties for neuroimaging classification,” *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 576–589, 2013.
- [29] E. Challis, P. Hurley, L. Serra, M. Bozzali, S. Oliver, and M. Cercignani, “Gaussian

process classification of Alzheimer’s disease and mild cognitive impairment from resting-state fMRI,” *Neuroimage*, vol. 112, pp. 232–243, 2015.

- [30] “<http://adni.loni.usc.edu>,” .
- [31] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012.
- [33] “<https://mlcommons.org/en/training-normal-10/>.”
- [34] J. Kim, V. D. Calhoun, E. Shim, and J.-H. Lee, “Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia,” *Neuroimage*, vol. 124, pp. 127–146, 2016.
- [35] S. Sarraf, G. Tofghi, A. D. N. Initiative, and others, “DeepAD: Alzheimer’s disease classification via deep convolutional neural networks using MRI and fMRI,” *BioRxiv*, p. 70441, 2016.
- [36] Y. Shi, W. Zeng, J. Deng, W. Nie, and Y. Zhang, “The Identification of Alzheimer’s Disease Using Functional Connectivity Between Activity Voxels in Resting-State fMRI Data,” *IEEE J. Transl. Eng. Heal. Med.*, vol. 8, pp. 1–11, 2020.
- [37] X. Bi, Q. Jiang, Q. Sun, Q. Shu, and Y. Liu, “Analysis of Alzheimer’s disease based on the random neural network cluster in fMRI,” *Front. Neuroinform.*, vol. 12, p. 60, 2018.
- [38] A. Khazae, A. Ebrahimzadeh, and A. Babajani-Feremi, “Application of advanced

- machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease," *Brain Imaging Behav.*, vol. 10, no. 3, pp. 799–817, 2016.
- [39] M. Odusami, R. Maskeliūnas, R. Damaševičius, and T. Krilavičius, "Analysis of Features of Alzheimer's Disease: Detection of Early Stage from Functional Brain Changes in Magnetic Resonance Images Using a Finetuned ResNet18 Network," *Diagnostics*, vol. 11, no. 6, p. 1071, 2021.
- [40] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*, Springer, 2009, pp. 1–4.
- [41] P. J. LaMontagne *et al.*, "OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease," *medRxiv*, 2019.
- [42] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, 1999.
- [43] J. Ashburner and K. J. Friston, "Unified segmentation," *Neuroimage*, vol. 26, no. 3, pp. 839–851, 2005.
- [44] G. Krüger and G. H. Glover, "Physiological noise in oxygenation-sensitive magnetic resonance imaging," *Magn. Reson. Med. An Off. J. Int. Soc. Magn. Reson. Med.*, vol. 46, no. 4, pp. 631–637, 2001.
- [45] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion," *Neuroimage*, vol. 59, no. 3, pp. 2142–2154, 2012.
- [46] L. Kasper *et al.*, "The PhysIO toolbox for modeling physiological noise in fMRI data," *J. Neurosci. Methods*, vol. 276, pp. 56–72, 2017.

- [47] U. Agrawal, E. N. Brown, and L. D. Lewis, “Model-based physiological noise removal in fast fMRI,” *Neuroimage*, vol. 205, p. 116231, 2020.
- [48] Y. Tong, L. M. Hocke, and B. B. Frederick, “Low frequency systemic hemodynamic ‘noise’ in resting state BOLD fMRI: characteristics, causes, implications, mitigation strategies, and applications,” *Front. Neurosci.*, vol. 13, p. 787, 2019.
- [49] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, “Transformers in vision: A survey,” *arXiv Prepr. arXiv2101.01169*, 2021.
- [50] “<https://techblog.ezra.com/transformers-in-medical-computer-vision-643b0af8fc41>.”
- [51] A. Dosovitskiy *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv Prepr. arXiv2010.11929*, 2020.
- [52] A. Vaswani *et al.*, “Attention Is All You Need.” *Advances in Neural Information Processing Systems (pp. 5998-6008)*, 2017.
- [53] E. Jang, S. Gu, and B. Poole, “Categorical reparameterization with gumbel-softmax,” *arXiv Prepr. arXiv1611.01144*, 2016.
- [54] R. Sanders, “The Pareto principle: its use and abuse,” *J. Serv. Mark.*, 1987.
- [55] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv Prepr. arXiv1412.6980*, 2014.
- [56] T. Gale, E. Elsen, and S. Hooker, “The state of sparsity in deep neural networks,” *arXiv Prepr. arXiv1902.09574*, 2019.
- [57] T. T. Liu, “Noise contributions to the fMRI signal: An overview,” *Neuroimage*, vol. 143, pp. 141–151, 2016.
- [58] L. L. Wald and J. R. Polimeni, “Impacting the effect of fMRI noise through hardware and acquisition choices--Implications for controlling false positive rates,” *Neuroimage*,

- vol. 154, pp. 15–22, 2017.
- [59] D. I. Hoult and P. C. Lauterbur, “The sensitivity of the zeugmatographic experiment involving human samples,” *J. Magn. Reson.*, vol. 34, no. 2, pp. 425–433, 1979.
- [60] A. Kumar, W. A. Edelstein, and P. A. Bottomley, “Noise figure limits for circular loop MR coils,” *Magn. Reson. Med. An. Off. J. Int. Soc. Magn. Reson. Med.*, vol. 61, no. 5, pp. 1201–1209, 2009.
- [61] N. R. Goodman, “Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction),” *Ann. Math. Stat.*, vol. 34, no. 1, pp. 152–177, 1963.
- [62] C. D. Constantinides, E. Atalar, and E. R. McVeigh, “Signal-to-noise measurements in magnitude images from NMR phased arrays,” *Magn. Reson. Med.*, vol. 38, no. 5, pp. 852–857, 1997.
- [63] O. Dietrich, J. G. Raya, S. B. Reeder, M. F. Reiser, and S. O. Schoenberg, “Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters,” *J. Magn. Reson. Imaging An. Off. J. Int. Soc. Magn. Reson. Med.*, vol. 26, no. 2, pp. 375–385, 2007.
- [64] C. Triantafyllou, J. R. Polimeni, and L. L. Wald, “Physiological noise and signal-to-noise ratio in fMRI with multi-channel array coils,” *Neuroimage*, vol. 55, no. 2, pp. 597–606, 2011.
- [65] C. Triantafyllou *et al.*, “Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters,” *Neuroimage*, vol. 26, no. 1, pp. 243–250, 2005.
- [66] R. M. Henkelman, “Measurement of signal intensities in the presence of noise in MR images,” *Med. Phys.*, vol. 12, no. 2, pp. 232–233, 1985.

- [67] U. Sakoglu, L. Bhupati, N. Beheshti, N. Tsekos, and L. Johnsson, “An adaptive space-filling curve trajectory for ordering 3D datasets to 1D: application to brain magnetic resonance imaging data for classification,” in *International Conference on Computational Science*, Springer, Cham, 2020, pp. 635–646.

APPENDIX

An adaptive space-filling curve for MRI based schizophrenia classification

In [67], we successfully performed binary classification of subjects diagnosed with Schizophrenia vs. healthy subjects. The classification was based on voxel signals, with voxels ordered with spatial locality as a starting point in mapping brain voxels to a 1D vector. Using locality-preserving space-filling curves to map higher-dimensional spaces to 1D has been used for many problems, such as e.g., ordering data array accesses to minimize DRAM page faults, partitioning data structures for minimizing communication network loads distributed computing environments, and data compression. In our work on MRI-based Schizophrenia classification, we developed an adaptive space-filling curve using a 3D Hilbert curve as the starting point for developing an adaptive space-filling curve (SFC) that minimizes the incremental squared signal differences in space traversal.

First, we remove voxels outside the brain in the Hilbert curve traversing the bounding box enclosing the brain. Figure 21 illustrates the pruning of the Hilbert curve with respect to voxels outside the brain volume.

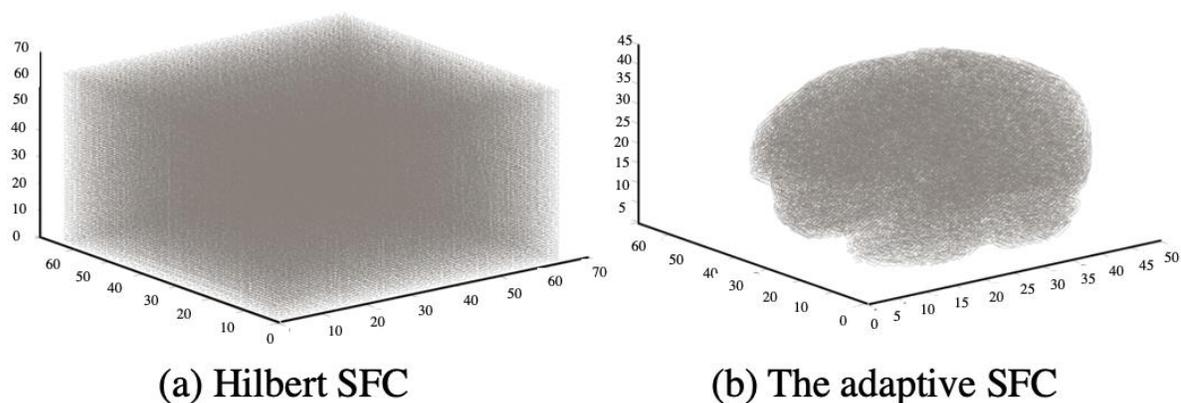


Figure 21: a) Hilbert curve for a bounding box enclosing the brain, (b) a pruned Hilbert curve covering only the brain

In a second step, we generate an SFC by a greedy approach. The neighbor voxel in a voxel's 3x3x3 neighborhood with the minimum absolute signal difference to the voxel is selected as the next voxel along the SFC unless it is already part of the SFC. For example, suppose all 26 neighbors are already included in the SFC. In that case, the algorithm traverses back in the list of voxels selected for the adaptive SFC until one voxel is found for which a neighbor voxel can be included in the SFC. For the backtracking, we use a hash table with key value 1 if a voxel is included in the SFC; otherwise, the value is zero. A snapshot of an adaptive SFC obtained from an MRI image is presented in Figure 21. Figure 22 shows SFC segments of lengths 1000, 3000, 5000, and 7000 (in red color) overlaid on the full-length adaptive SFC (in gray color).

The signal intensities and absolute signal intensity differences along with Hilbert and adaptive SFCs are shown in Figure 23. The adaptive SFC has a 50.5 times less total squared signal intensity difference than the Hilbert SFC, and it is 34% shorter. Using the adaptive SFC for classification resulted in higher accuracy than classification based on the Hilbert SFC [67]

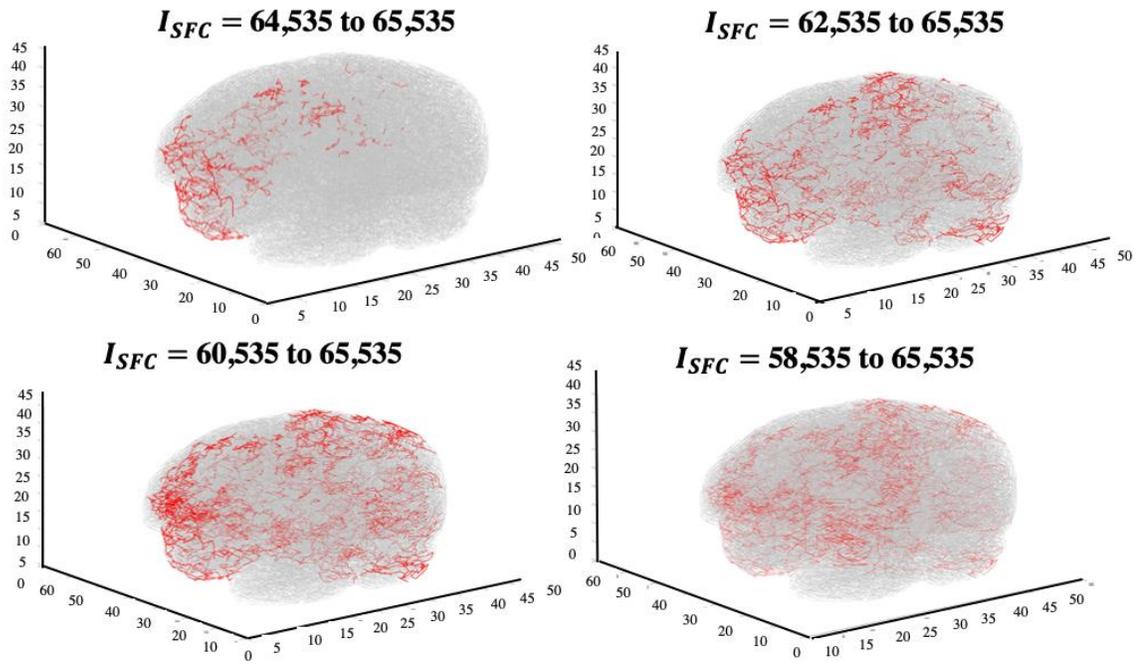


Figure 22: Representation of fMRI brain activation traversal by the SFC for the last a) 1000, 3000, 5000 and 7000 voxels (red), overlaid on the total SFC trajectory (gray), representing a sample progression of the SFC trajectory during the last 7000-voxel portion

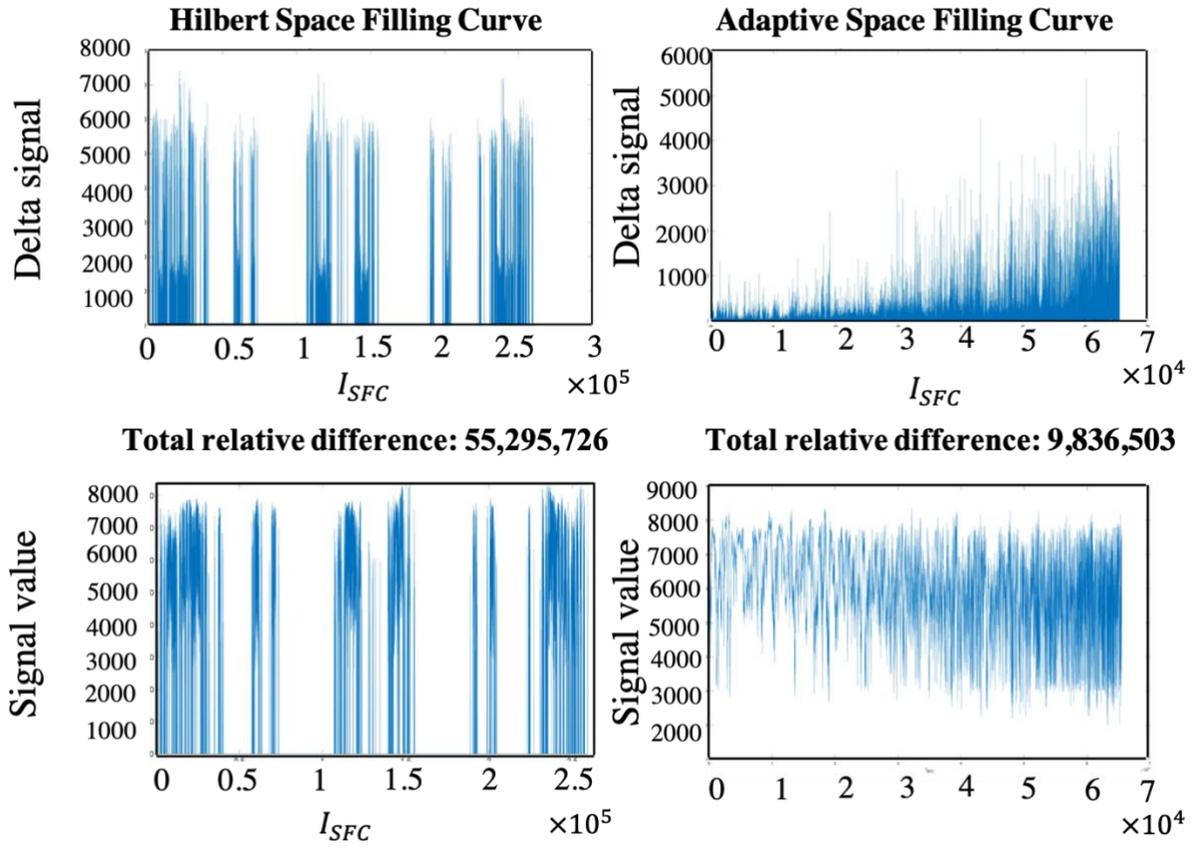


Figure 23: Signal intensity along the trajectory for successive voxels

