Psychiatry Research: Neuroimaging 328 (2023) 111576

Contents lists available at ScienceDirect

Psychiatry Research: Neuroimaging

journal homepage: www.elsevier.com/locate/psychresns

Discriminative patterns of white matter changes in Alzheimer's

Subaramya Srivishagan [a,b], Logiraj Kumaralingam [c], Kokul Thanikasalam [c], U.A.J. Pinidiyaarachchi [d], Nagulan Ratnarajah [a,*], The Alzheimer's Disease Neuroimaging Initiative [1]

[a] Department of Physical Science, Faculty of Applied Science, University of Vavuniya, Vavuniya, Sri Lanka
[b] PGIS, University of Peradeniya, Peradeniya, Sri Lanka
[c] Department of Computer Science, Faculty of Science, University of Jaffna, Jaffna, Sri Lanka
[d] Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka

ARTICLE INFO

Keywords:
Structural brain network
White matter connectivity
Alzheimer's disease
CNN
Class activation map

ABSTRACT

Changes in structural connectivity of the Alzheimer's brain have not been widely studied utilizing cutting-edge methodologies. This study develops an efficient structural connectome-based convolutional neural network (CNN) to classify the AD and uses explanations of CNNs' choices in classification to pinpoint the discriminative changes in white matter connectivity in AD. A CNN architecture has been developed to classify normal control (NC) and AD subjects from the weighted structural connectome. Then, the CNN classification decision is visually analyzed using gradient-based localization techniques to identify the discriminative changes in white matter connectivity in Alzheimer's. The cortical regions involved in the identified discriminative structural connectivity changes in AD are highly covered in the temporal/subcortical regions. A specific pattern is identified in the discriminative changes in structural connectivity of AD, where the white matter changes are revealed within the temporal/subcortical regions and from the temporal/subcortical regions to the frontal and parietal regions in both left and right hemispheres. The proposed approach has the potential to comprehensively analyze the discriminative structural connectivity differences in AD, change the way of detecting biomarkers, and help clinicians better understand the structural changes in AD and provide them with more confidence in automated diagnostic systems.

1. Introduction

Alzheimer's disease (AD) is an irreversible, progressive, neurodegenerative illness that affects primarily older individuals. AD is characterized by a higher impairment in memory or cognitive skills than healthy adults of the same age (Neugroschl and Wang, 2011). The accumulation of abnormal amyloid-β and hyperphosphorylated tau proteins is a pathological feature of AD. Amyloid-β deposition is thought to be the cause of pathological tau formation and subsequent neurodegeneration (Jack et al., 2010; Hardy and Selkoe, 2002). The neurodegeneration leads to molecular neuropathological abnormalities in distinct neuronal brain networks, which results brain network dysfunction (Drzezga, 2018). Understanding the impairments and structural modifications in white matter connectivity, that lead to the formation of these networks, can help to identify the brain network dysfunction and structural biomarkers of AD. The presence and destruction of structural connections in structural brain networks could explain simultaneous molecular, metabolic, and functional alterations (Mito et al., 2018). According to the theory of the AD brain network, the disease attacks susceptible areas in the brain and spreads across intrinsic networks, presumably via specific white matter pathways (Raj et al., 2012; Zhou et al., 2012).

Understanding brain structural connectivity is key to elucidate how neurons and neural networks process information. The disruption of structural and functional connectivity in Alzheimer's disease has always been linked to the structural brain network destruction. Functional MR

* Corresponding author at: Department of Physical Science, Faculty of Applied Science, University of Vavuniya, Vavuniya, Sri Lanka.
E-mail address: nagulanr@vau.ac.lk (N. Ratnarajah).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report.

https://doi.org/10.1016/j.pscychresns.2022.111576
Received 8 September 2022; Received in revised form 12 November 2022; Accepted 22 November 2022
Available online 30 November 2022

imaging has been extensively used in the study of functional connectivity analysis in AD, revealing distinct patterns of functional connectivity impairments (Mondragón et al., 2021; Tang et al., 2021; Sorg et al., 2009). Changes in white matter pathways are likely to mediate this functional network dysfunction; however, due to the difficulties of modelling complex white matter structures, impairments in specific structural connectivity between the cortical regions have not been thoroughly examined in the literature using advanced techniques. In this study, we applied deep neural network classification decisions' explanation to comprehensively investigate the discriminative structural connectivity changes in subjects with AD.

Diffusion MR imaging is now the only approach available to analyze structural changes associated with fiber pathways in vivo and non-invasively. Several diffusion MRI based investigations have shown structural alterations in white matter that have occurred over the development of AD in the last decade, the findings of which have been presented in a number of comprehensive reviews (Chua et al., 2008; Mak et al., 2017). Despite hopeful findings of white matter changes in AD using diffusion models, quantitative analysis of FA and MD has major flaws, making their findings unreliable and anatomically difficult to interpret. Using whole-brain tract-based spatial statistical (TBSS) analysis (Acosta-Cabronero et al., 2010; Bosch, 2012) or directly analyzing individual fiber bundle methods (Daianu et al., 2016; Bendlin et al., 2010) has its own set of challenges, such as automatically segmenting white matter into known fiber bundles, quantifying properties and similarity of a bundle, analyzing a specific bundle for a subject group, and the occurrence of redundant and non-existent fibers or false positives (Campbell et al., 2005) in whole brain tractography.

The structural brain network (Sporns et al., 2005) provides a more appropriate solution, which represents the complete map of the white matter connectivity in the brain. The network not only includes edges as a list of linked regions, but also provides the weight of each connection (Hagmann et al., 2008). Structural brain networks not only have the ability to shed light on the insights of structural connectivity (Srivishagan et al., 2020) but also uncover new information about the principles that govern how distinct functional subunits are organized and interact with one another (Passingham, 2013) and pathological brain conditions (Griffa et al., 2013). Building on our preliminary studies (Subaramya et al., 2021), we used structural brain networks of AD and healthy elderly people to classify and explain the discriminative differences in white matter. Rather than analyzing individual fibers or bundles, we focused on the white matter pathways between pairs of cortical regions, which are naturally provided by the structural brain network.

Deep Neural Networks (DNNs) have produced state-of-the-art outcomes in a variety of medical imaging applications, including the identification of Alzheimer's disease using neuroimaging data (for a review, see Ebrahimighahnavieh et al., 2020; Vieira et al., 2017; Jo et al., 2019). However, the DNNs' decisions are frequently seen as non-transparent (Castelvecchi, 2016), making it challenging to use these algorithms in clinical practice. Recently, several researchers (Zhou et al., 2016, Selvaraju et al., 2017) proposed techniques to visually explain the DNNs' decisions in various tasks. However, only a very few recent studies have explained DNNs' decisions in neuroimaging-based AD classification with different visualization approaches (Böhle et al., 2019; Rieke et al., 2018; Yang et al., 2018). These studies are focused on visually identifying the most influential brain regions in diagnosing Alzheimer's disease based on the decisions of 3D MR image classification. To the best of our knowledge, no research has been done to visually identify the discriminative white matter connectivity changes between cortical regions using structural brain networks.

In this study, we focus on exploring the most influential white matter connectivity changes in AD by visually explaining the convolutional neural network (CNN) decisions in structural brain network based classification. By feeding the features of the structural brain network, a CNN architecture was proposed to distinguish AD from healthy normal subjects. Then the Gradient-weighted Class Activation Mapping (Grad-CAM) technique (Selvaraju et al., 2017) was utilized to visually interpret the classifier's decision. The study investigates inside the black box of classification for AD and explains the CNN decisions regarding which changes in the structural connections will have the most impact on the classification outcome. We show that the overall approach succeeded in illustrating the discriminative pattern of white matter connectivity changes in AD. The discriminative white matter pathways have piqued the interest of researchers, and they are now being used as biomarkers for Alzheimer's disease. The outcome of this study contributes to AD diagnosis and also provides clinicians more faith in automated AD diagnostic systems.

2. Methodology

2.1. MRI Data

Diffusion MR images and the structural T1-weighted images of the cognitively normal controls (NC) and AD subjects were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) project database (http://adni.loni.usc.edu/) (Jack et al., 2008). A summary of the selected NC and AD subjects' demographics and clinical scores (Mini--Mental State Examination (MMSE)) for this study is described in Table 1. There were no age (two-sample t-test, $p = 0.0743$) or gender (Chi-Square test, $p = 0.61$) differences between NC and AD individuals. These findings support the removal of the effects of age and gender from the comparison test models for the other variables in NC and AD participants. The MMSE score showed a significant difference between the groups, as expected ($p < 0.01$).

MRI Acquisition Protocols: (i) T1-weighted images and diffusion MR images of 30 NC and 32 AD subjects were obtained from the ADNI-2 phase. Anatomical T1-weighted images (256×256 matrix; slices = 196; voxel size = 1.2×1.0 ×1.0 mm^3; TI = 400 ms; TE=2.85 ms; TR=6.98 ms; flip angle = 11^0) and diffusion MR images (128×128 matrix reconstructed to 256 × 256 matrix; voxel size: 2.7×2.7×2.7 mm^3; five T2-weighted images without dedicated diffusion sensitization (b0 images) and forty-one diffusion-weighted images (b = 1000 s/mm^2)) were performed on a 3-Tesla GE Medical Systems scanners. (ii) MR images of 30 NC and 30 AD subjects (Siemens 3-Tesla MRI scanners) were obtained from ADNI-3 phase. T1-weighted anatomical images were obtained (1×1×1 mm^3 voxel size; TE=2.98 ms; TR=2300 ms; flip angle = 11^0). For diffusion-weighted MRI, the pulse sequence was acquired in the axial plane (TE=56 ms; TR=7200 ms; 48 diffusion directions (b = 1000 s/mm^2); 7 non–diffusion-weighted images; 2×2×2 mm^3 voxel size). Additional imaging information can be obtained at (http://adni.loni.usc.edu/methods/documents/mri-protocols/).

2.2. Brain network construction

Network Nodes: For all of the participants' T1-weighted and diffusion MRI brain images, the FSL (Smith et al., 2004) was utilized for image preprocessing operations such as noise reduction, correction for subject motion and geometrical distortions, Eddy current correction for diffusion MR imaging, and brain extraction. Each structural T1-weighted MR image was initially linearly registered to the b0 image

Table 1
A summary of the selected Normal Control (NC) and Alzheimer's disease (AD) subjects' demographics and clinical scores (Mini-Mental State Examination (MMSE)) for this study.

Variable	NC	AD
Number of Subjects	60	62
Age (mean ± std)	71.51±5.85	76.89 ± 6.95
Gender (Male: Female)	27:33	32:30
MMSE Score (mean ± std)	29.1 ± 1.0	23.5 ± 1.6

in diffusion tensor imaging (DTI) space, and then the resulting T1-weighted MR images were registered to MNI152 space using a non-linear transformation. Finally, the Desikan-Killiany atlas template (Desikan et al., 2006) in MNI152 space was subjected to inverse transformations, yielding DTI native space parcellations of cortical (and subcortical) regions. Table 2 depicts the details of the constructed 80 cortical (and subcortical) regions (40 for each hemisphere), and each node in the structural brain network is represented by a region.

Network Edges: Diffusion weighted images were used to estimate the diffusion tensor and to calculate fractional anisotropy (FA) maps using CAMINO (http://camino.cs.ucl.ac.uk/) for every subject. Average FA values of white matter in AD (0.32±0.09) and NC (0.36±0.07) indicate the destruction of white matter microstructure integrity in AD. Diffusion tensor-based whole-brain fiber tractography was performed using the FACT (Mori et al., 1999) algorithm, which started from white matter seed points and ended at a voxel with a turning angle greater than 45 degrees or with a value of FA less than 0.1.

The edges of the brain network were considered to be the tractography streamlines linking distinct cortical regions. Adjacency matrices of weighted graphs of brain networks with 80×80 elements were created as follows:

$$w_{ij} = \frac{k}{(S_i + S_j)/2}$$

Table 2
Cortical and subcortical grey matter regions are used as nodes in the structural brain networks, corresponding to the regions defined in the Desikan-Killiany Atlas (Desikan et al., 2006).

Cortical and subcortical regions	Left Hemisphere		Right Hemisphere	
	Short Form	Index	Short Form	Index
ctx- bankssts	B.L	1	B.R	41
ctx- caudalanteriorcingulate	CAC.L	2	CAC.R	42
ctx- caudalmiddlefrontal	CMF.L	3	CMF.R	43
ctx- cuneus	Cu.L	4	Cu.R	44
ctx- entorhinal	Ent.L	5	Ent.R	45
ctx- fusiform	Fu.L	6	Fu.R	46
ctx- inferiorparietal	IP.L	7	IP.R	47
ctx- inferiortemporal	IT.L	8	IT.R	48
ctx- isthmuscingulate	IsC.L	9	IsC.R	49
ctx- lateraloccipital	LO.L	10	LO.R	50
ctx- lateralorbitofrontal	LOF.L	11	LOF.R	51
ctx- lingual	Li.L	12	Li.R	52
ctx- medialorbitofrontal	MOF.L	13	MOF.R	53
ctx- middletemporal	MT.L	14	MT.R	54
ctx- parahippocampal	Phi.L	15	Phi.R	55
ctx- paracentral	PC.L	16	PC.R	56
ctx- parsopercularis	PPe.L	17	PPe.R	57
ctx-parsorbitalis	POr.L	18	POr.R	58
ctx-parstriangularis	Ptr.L	19	Ptr.R	59
ctx-pericalcarine	Per.L	20	Per.R	60
ctx-postcentral	PoC.L	21	PoC.R	61
ctx-posteriorcingulate	Pci.L	22	Pci.R	62
ctx-precentral	PrC.L	23	PrC.R	63
ctx-precuneus	Pcu.L	24	Pcu.R	64
ctx-rostralanteriorcingulate	RAC.L	25	RAC.R	65
ctx-rostralmiddlefrontal	RMF.L	26	RMF.R	66
ctx-superiorfrontal	SF.L	27	SF.R	67
ctx-superiorparietal	SP.L	28	SP.R	68
ctx-superiortemporal	ST.L	29	ST.R	69
ctx-supramarginal	SuM.L	30	SuM.R	70
ctx-frontalpole	Fpo.L	31	Fpo.R	71
ctx-temporalpole	Tpo.L	32	Tpo.R	72
ctx-transversetemporal	TrT.L	33	TrT.R	73
ctx-insula	Ins.L	34	Ins.R	74
Thalamus-Proper	Tha.L	35	Tha.R	75
Caudate	Cau.L	36	Cau.R	76
Putamen	Put.L	37	Put.R	77
Pallidum	Pal.L	38	Pal.R	78
Hippocampus	Hip.L	39	Hip.R	79
Amygdala	Amy.L	40	Amy.R	80

where k is the number of fiber tracts connecting the cortical regions (nodes) i and j, and S_i and S_j are the volumes of the regions i and j respectively. For removing some brain connections that appeared as non-existent fibers or false positives, a non-parametric one-tailed sign test was used (Srivishagan et al., 2020). To retain the particular connectivity at the same point in the matrix, the order of indices (Table 2) is fixed in all brain network matrices.

The normalized characteristic path length λ and the normalized clustering coefficient γ (Srivishagan et al., 2020) for NC and AD structural brain networks were found to be nearly one ($p<0.01$) and larger than one ($p<0.01$), respectively, when compared to random networks with the same nodes and degree distribution, indicating that the NC and AD brain networks exhibited strong small-world properties.

2.3. Convolution neural network architecture for AD-NC classification

A CNN architecture was developed from scratch for classifying AD patients and NC individuals. The details of network architecture, training, and an evaluation metric are described below.

Architecture: The proposed CNN architecture is depicted in Fig. 1. It contains three convolutional layers and three fully-connected layers. A dropout layer and an average pooling layer are located in between fully-connected and convolutional layers, respectively. In each convolutional layer, a 3×3 kernel was employed, and the number of neurons in convolutional layers was set in increasing order: 32, 64, and 128 as shown in Fig. 1. The fully connected layers' neurons were set to 256, 64, and 2 respectively. In fully connected and convolutional layers, the Rectified Linear Activation Function (ReLU) was used. In the last fully-connected layer, we employed the softmax activation function to obtain the classification scores. For each AD patient and NC subject, their structural brain network was generated and then fed to the proposed CNN architecture with the dimension of an 80 × 80 matrix.

Training: The proposed CNN architecture has been trained from scratch by feeding the structural brain network features. The training process is more challenging as the structural brain network matrices consist of a considerable number of zeros. Several optimization functions have been attempted to train the proposed network architecture, and the Adam optimizer has been selected since it showed excellent validation accuracy. During the training, the learning rate was set at 0.001 and the dropout parameter was set at 0.5. The entire architecture was trained for 50 epochs with a batch size of 32. Furthermore, all the hyperparameters of the proposed CNN architecture were tuned based on the validation results. We have followed the 5-fold cross-validation technique to evaluate the model to avoid selection bias. As the first step of cross-validation, the entire dataset was split into five random groups. Then four groups of data were used for training and validation (to tune the hyperparameters) by selecting 80% and 20%, respectively. The remaining group of data was used to test the model. We have repeated this process until each group serves as a test set. At the end, the average of the test accuracies of each fold was considered the final accuracy of the model.

Evaluation Metric: The classification accuracy is used to measure the classification performance as follows:

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

where TP, FP, TN and FN are the true positive, false positive, true negative and false negative, respectively.

2.4. Visualizing the CNN classifier's decision using gradient-based localization

Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2017) is a well-known technique used to identify the class-specific discriminative input features for any CNN architecture.

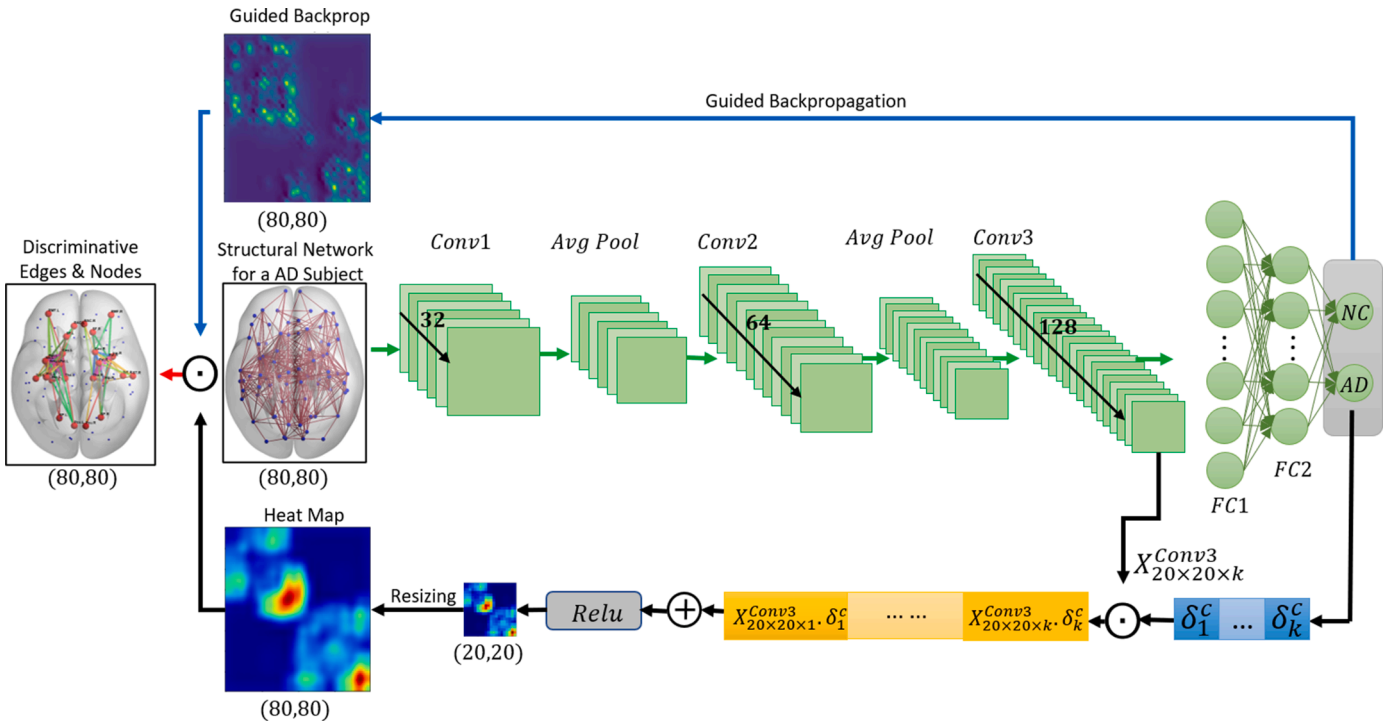

Fig. 1. The proposed CNN architecture for AD and NC classification and the procedure for visualizing the class activation maps. The brain structural adjacency matrices of AD and NC subjects are fed to the proposed CNN architecture to train the classifier. To visually identify the discriminative edges and nodes of an AD subject in classification, the adjacency matrix of a test subject is input to the trained CNN classifier, and then the guided backpropagation and heatmap are obtained and then multiplied through an element-wise multiplication operation to obtain the Grad-CAM activation map.

Without any additional CNN architecture or layers, it can be used to visually explain the decisions of a CNN model. Based on the Grad-CAM technique, we have developed an approach to determine the most influential structural white matter brain connectivity in classifying AD patients from NC subjects.

In the proposed visualization technique, initially the gradient score of the AD class, y^{AD}, was computed with respect to the feature map $X^{Conv3}_{20\times20\times128}$ of the third convolutional layer of the proposed CNN architecture, *i.e.* $\frac{\partial y^{AD}}{\partial X^{Conv3}}$. Then the gradients of the AD class were global average pooled to obtain a vector δ^{AD}_k, where k = [1,2, …128], which gives the information about how important channel k is with regard to classifying AD. In the next step, the feature map $X^{Conv3}_{20\times20\times128}$ and δ^{AD}_k were combined through a channel-wise multiplication operation. Then the ReLU function was applied to obtain the heatmap $h_{20\times20}$ since we need only to consider the features that positively influence the AD classification. For the purpose of visualization, the obtained heatmap $h_{20\times20}$ was normalized between 0 and 1 and then resized to 80 × 80 and denoted as $h_{80\times80}$. Finally, to obtain the high-resolution class activation map $Map^{AD}_{80\times80}$ (also called as Guided Grad-CAM), guided backpropagation ($G_{80\times80}$) of AD with respect to the input was obtained, and then it was multiplied with the resized heatmap $h^{Re}_{80\times80}$ through an element-wise multiplication operation. As illustrated in Fig. 2, the high-resolution class activation map $Map^{AD}_{80\times80}$ is used to identify the feature map regions that are influenced to detect AD in classification.

Generating Class Activation Maps and Heat Maps: Since we followed the 5-fold cross validation, all of the available AD samples serve as the test data for the corresponding CNN trained model. In each fold of cross validation, heat maps of AD samples and their corresponding high-resolution class activation maps were obtained. In the next step, an average of these heatmaps $h^{AD\_avg}_{80\times80}$ and high-resolution class activation maps $Map^{AD_avg}_{80\times80}$were generated. Since the locations of nodes are the same in all input adjacency matrices of AD and NC subjects, the average of the output class activation map will provide accurate results. Finally, we used a thresholding function to remove the weakly contributed features in $h^{AD\_avg}_{80\times80}$, and $Map^{AD\_avg}_{80\times80}$ in order to identify the features of brain connections that were primarily influencing the identification of the AD in the classification. We considered white matter connections that contributed at or above 75% of the AD classification to identify these most discriminative brain connections in the input. Because there are no connection changes across the same region, we make diagonal elements of the activation map zero in some places for the analysis (Fig. 3 (a)).

3. Results

Classification: The Keras-TensorFlow library was utilized to implement the CNN architecture in Python. The NVIDIA Tesla K80 GPU was used to run the algorithms on the Google Colab cloud platform. The proposed CNN architecture showed 95.68% average classification accuracy. The high accuracy rate demonstrates that we have effectively designed and enhanced the CNN architecture and tuned the parameters to obtain a satisfactory result from the structural brain network matrices. CNN classifier decisions, therefore, can be used to identify the discriminative differences in white matter connectivity in AD.

Visualizing the CNN classifier's decision: Five randomly selected AD subjects' structural brain networks, heatmaps, guided backpropagation maps, and class activation maps are illustrated in Fig. 2 (a). Discriminant connectivity found in the activation maps of five Alzheimer's patients is consistent. The small variations in the activation maps could indicate distinct stages of the disease. Fig. 2 (b) shows the average result of the class activation map $Map^{AD\_avg}_{80\times80}$ and the maps after removing weakly contributed connections (65%, 70%, and 75%). The color map range of these images is set from blue to red, and they represent less to more contributing features in AD prediction in the classification. Based on this figure, we can observe that the most discriminant regions show a similar pattern within the left and right hemispheres. Even though there are some differences among the

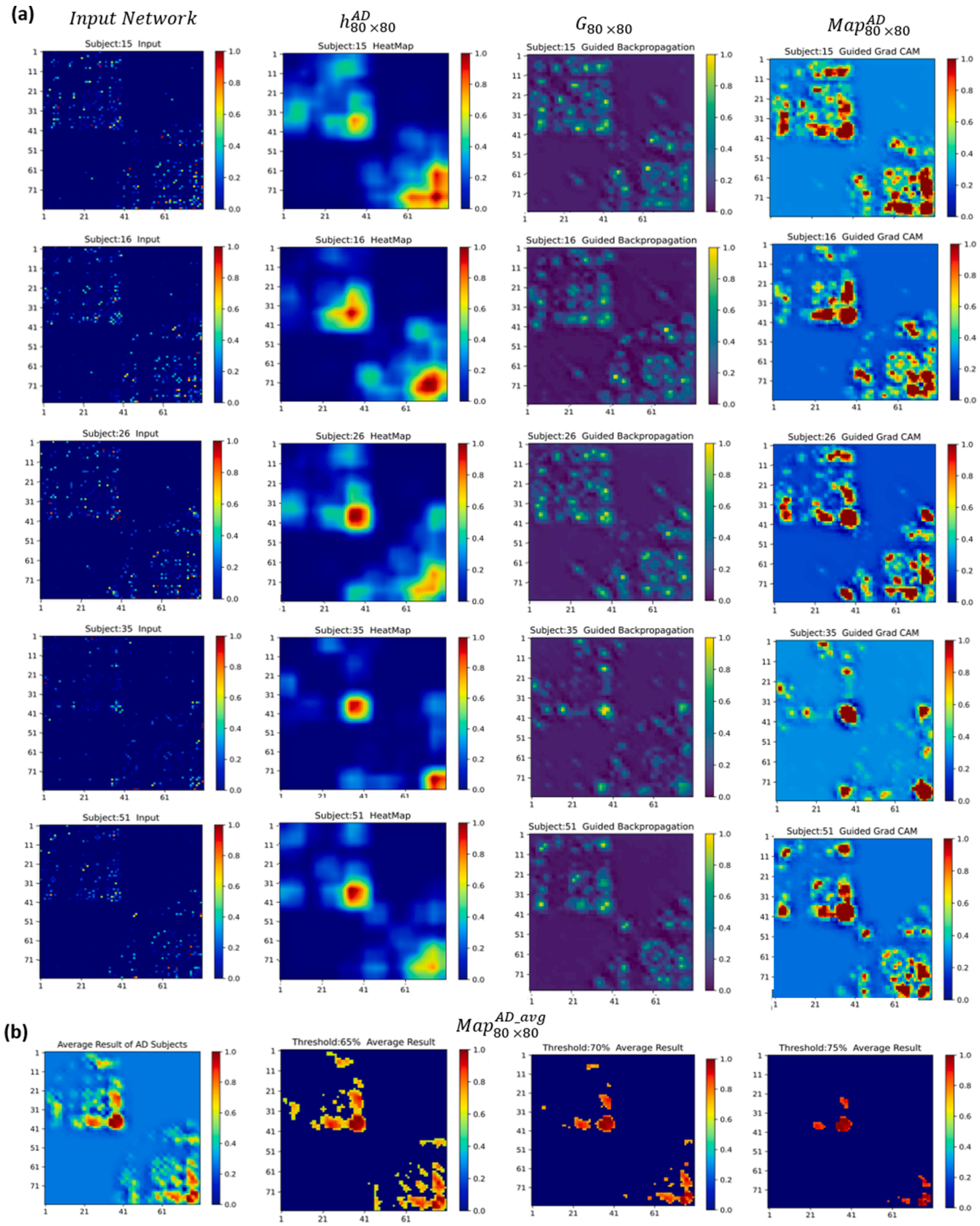

Fig. 2. (a) Visualization results of five randomly selected AD subjects. First Column: input adjacency matrices, Second Column: corresponding resized heatmaps, Third Column: corresponding guided backpropagation, Fourth Column: corresponding class activation maps. (b) Average of Class activation maps of all AD subjects, and resulting maps of thresholding at 65%, 70%, and 75%.

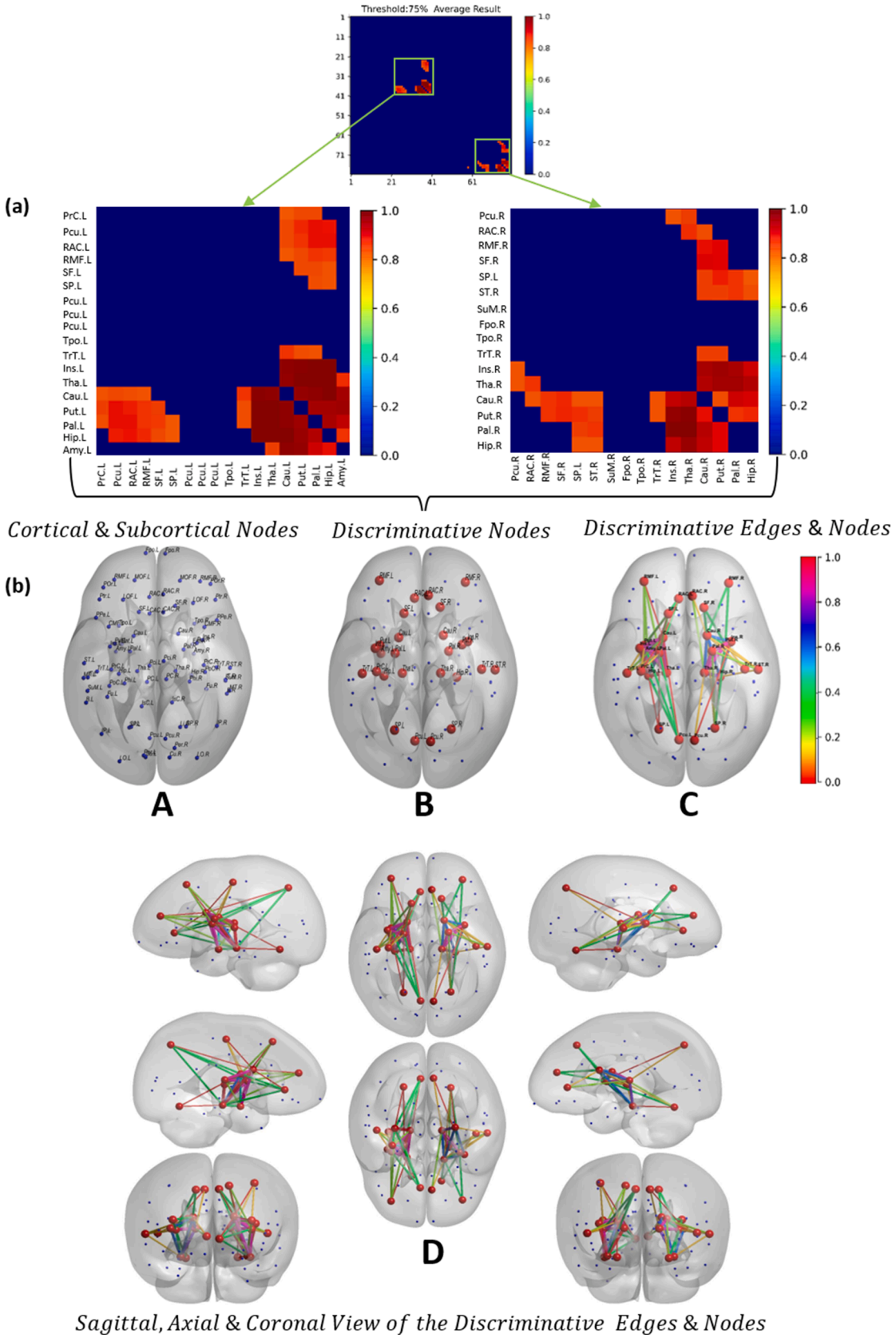

Fig. 3. (a) Illustration of the resultant average class activation map of all AD subjects and the predicted discriminative connectivity with corresponding cortical and subcortical nodes after thresholding at 75% for the left and right hemispheres. (b) (A) Axial representation of all 80 cortical and subcortical nodes (B) Discriminant nodes in axial view (C) Discriminant structural connections and corresponding discriminant edge weights in axial view (D) Sagittal, axial, and coronal views of discriminant structural connections and discriminant nodes in the left and right hemispheres.

hemispheres, a clear and prominent pattern has resulted. The resultant class activation maps appeared symmetric in general, confirming the correctness of the result since the connectivity between two cortical regions is undirected. The most discriminant regions are exclusively found in the intra-hemispheric areas in both the left and right hemispheres, while the interhemispheric regions do not have a high value prediction zone.

In the average class activation map of AD, most discriminant regions decrease very smoothly starting from the 65% threshold (Fig. 2 (b)). Furthermore, when compared to the entire number of connections (80×80), our choice for the analysis is a 75% threshold, indicating very strong connectivity differences, which is consistent with the fact that the number of biomarkers should be small and significant.

Discriminative structural white matter connectivity: Fig. 3 depicts the discriminant structural connections as well as the related cortical and subcortical regions in great detail. Fig. 3 (a) shows the resultant average class activation map at a 75% threshold, which illustrates discriminative connectivity with corresponding cortical and subcortical nodes for the left and right hemispheres. Fig. 3 (b) illustrates the discriminative nodes (the cortical and subcortical regions involved in the discriminative structural connectivity) and the discriminative connectivity in different views of the brain. The discriminative connectivity that resulted do not appear in random positions; instead, they followed a definite pattern. The cortical and subcortical regions involved in these white matter connections are mainly from the temporal/subcortical regions (Tha, Cau, Put, Pal, Hip, Amy.L, TrT, Ins, ST.R), frontal lobe (RAC, RMF, SF, PrC.L) and parietal lobe (SP, Pcu). Despite the fact that the level of discriminant connectivity varies, the discriminant intra-hemispheric white matter pathways connecting within the temporal/subcortical regions, from temporal/subcortical to frontal, and temporal/subcortical to parietal lobes, display a coherent pattern (Fig. 3 (b)). Table 3 depicts these connections in further detail, with each connection name and the group with the percentage of the contribution in the classification decision. The number of discriminant connections and the number of cortical and subcortical regions involved in theses connections in the left hemisphere is slightly higher than the right hemisphere connections and regions.

Table 3
The discriminative white matter connections in left and right hemispheres with each connection name and the group (pattern) with the percentage (%) of the contribution in the classification decision.

Left Hemisphere		Right Hemisphere	
Parietal ↔ Temporal/ Subcortical	**%**	**Parietal ↔ Temporal/ Subcoritcal**	**%**
Pcu.L ↔ Cau.L	77	*Pcu.R ↔ Ins.R*	75
Pcu.L ↔ Put.L, Pal.L	82	*Pcu.R ↔ Tha.R*	78
Pcu.L ↔ Hip.L	81	*SP.R ↔ Cau.R*	78
SP.L ↔ Pal.L, Hip.L	75	*SP.R ↔ Put.R*	81
Frontal ↔ Temporal/ Subcortical		*SP.R ↔ Pal.R*	79
PrC.L ↔ Cau.L	76	*SP.R ↔ Hip.R*	75
PrC.L ↔ Put.L	78	**Frontal ↔ Temporal/ Subcortical**	
PrC.L ↔ Pal.L	77	*RAC.R ↔ Tha.R*	80
RAC.L ↔ Cau.L	76	*RAC.R ↔ Cau.R*	76
RAC.L ↔ Put.L	81	*RMF.R ↔ Cau.R*	82
RAC.L ↔ Pal.L, Hip.L	82	*RMF.R ↔ Put.R*	82
RMF.L ↔ Cau.L	76	*SF.R ↔ Cau.R*	83
RMF.L ↔ Put.L, Hip.L	80	*SF.R ↔ Put.R*	83
RMF.L ↔ Pal.L	79	**Temporal/ Subcortical ↔ Temporal/ Subcortical**	
SF.L ↔ Put.L, Pal.L	78	*ST.R ↔ Cau.R*	78
SF.L ↔ Hip.L	77	*ST.R ↔ Put.R*	78
Temporal/ Subcortical ↔ Temporal/Subcortical		*ST.R ↔ Pal.R*	79
TrT.L ↔ Cau.L	79	*ST.R ↔ Hip.R*	76
TrT.L ↔ Put.L, Pal.L	76	*TrT.R ↔ Cau.R*	78
Ins.L ↔ Cau.L, Put.L	92	*TrT.R ↔ Put.R*	78
Ins.L ↔ Pal.L, Hip.L	93	*Ins.R ↔ Cau.R*	88
Tha.L ↔ Cau.L	91	*Ins.R ↔ Put.R*	91
Tha.L ↔ Put.L	92	*Ins.R ↔ Pal.R*	92
Tha.L ↔ Pal.L, Hip.L	93	*Ins.R ↔ Hip.R*	86
Tha.L ↔ Amy.L	80	*Tha.R ↔ Cau.R*	87
Cau.L ↔ Put.L, Pal.L	91	*Tha.R ↔ Put.R*	90
Cau.L ↔ Hip.L	93	*Tha.R ↔ Pal.R*	90
Cau.L ↔ Amy.L	92	*Tha.R ↔ Hip.R*	90
Put.L ↔ Pal.L	89	*Cau.R ↔ Put.R*	82
Put.L ↔ Hip.L, Amy.L	90	*Cau.R ↔ Pal.R*	87
Pal.L ↔ Hip.L	88	*Cau.R ↔ Hip.R*	86
Pal.L ↔ Amy.L	86	*Put.L ↔ Pal.R*	82
Hip.L ↔ Amy.L	82	*Put.L ↔ Hip.R*	82

4. Discussion

In this study, we introduced a method of explaining CNN decisions in AD classification to explore the discriminative white matter connectivity differences in AD subjects compared to cognitively normal control subjects. The class activation map localizes the elements of the structural brain network's adjacency matrices that contribute when the predictions are made. An element of the structural brain network's adjacency matrix represents the strength of the structural connectivity between a pair of cortical regions. Each individual's class activation maps indicate the importance of each element of the adjacency matrix for the respective classification decision. We analyzed the class activation maps at higher levels of detail with respect to the discriminative brain structural connectivity between pairs of cortical regions and the cortical regions involved in the connectivity.

4.1. Discriminant cortical and sub cortical regions

The cortical and subcortical regions involved in the identified discriminative structural connectivity changes are highly covered in the temporal/subcortical lobe, including the Insula, thalamus, hippocampus, and amygdala. The primary feature of the memory impairment is thought to be caused by temporal lobe and hippocampus damage. Previous studies have shown that AD-related volume loss in the temporal lobe and hippocampal (Killiany et al., 1993), temporal lobe atrophy (Erkinjuntti et al., 1993; Ramos Bernardes et al., 2017), and recently, MR imaging-based activation maps revealed the most discriminant regions have been found in areas of the temporal lobe and subcortical regions (Böhle et al., 2019; Yang et al., 2018). Some parietal cortical regions, particularly the precuneus area and superiorparietal (Jacobs et al., 2012; Prawiroharjo et al., 2020), as well as the frontal lobe (Cajanus et al., 2019), were identified as discriminative connectivity involved cortical regions.

All these identified regions' morphometric changes have been linked to disease progression and cognitive deterioration (Ledig et al., 2018). These changes appear to be correctly used by our CNN framework for making predictions and the Grad-CAM based class activation maps for visualizing the CNN decisions. Connectivity related to the Amygdala in the right hemisphere is not given any significant discriminant, which may be the reason it is on the border of the brain structural matrices. Although the input data contains important boundary features, they vanish throughout the convolution operation, even we employed zero padding in convolutional layers.

4.2. Discriminative structural connectivity changes

In addition to grey matter morphological changes in AD, cerebral white matter neuropathological alterations such as axonal loss, demyelination, and cell death have been observed (Ihara et al., 2010; Englund E. 1998). Here we report how Alzheimer's disease affects structural brain connections based on the resultant discriminant connectivity, which gives a specific pattern of alterations (within temporal/subcortical regions and from temporal/subcortical to frontal and parietal regions) as temporal/subcortical regions are involved in all the discriminant connectivity.

The intra-hemispheric structural connections within the temporal/subcortical regions are strongly varied between NC and AD subjects in both left and right hemispheres. In the white matter connections of the subcortical regions, especially in the limbic system networks, impaired white matter connectivity is reported in Li et al. (2016), in which the limbic system networks are associated with behavioral and emotional responses from memory. The fornix is a projection fiber bundle that belongs to the limbic system, one of the most important anatomical structures related to memory. The fornix is vulnerable in AD and the impairment of the fornix is an early sign of AD (Hopper and Vogel, 1976) and is linked to the cholinergic dysfunction characteristic of Alzheimer's (Bunce et al., 2003; Colom et al., 2010). The structural and functional features of the fornix have naturally took researchers' attention, seeking diagnostic markers of Alzheimer's. The limbic system not only connects the fornix but is also connected by the major fiber bundles such as the cingulum, uncinate fasciculus, and anterior thalamic peduncle.

The other set of intra-hemispheric structural connections altered significantly is the connectivity from temporal/subcortical regions to some frontal (superiorfrontal, rostralmiddlefrontal, rostralanteriorcingulate) and parietal (superiorparietal, precuneus) regions. As the clinical condition advances, the white-matter changes are first limited to the medial temporal limbic association tracts, but they tend to expand to the temporal and parietal white matter (Demirhan et al., 2015; Konukoglu et al., 2016). The cingulum bundle is a major white matter bundle that connects the frontal, parietal, and temporal lobes and connects subcortical nuclei to the cingulate gyrus, which is also affected by AD (Dalboni et al., 2020; Kantarci et al., 2017). Other major fiber bundles that connect or pass through temporal/subcortical, frontal, and parietal regions, such as the superior longitudinal fiber (between frontal and temporal language regions), the internal capsule and corona radiata (fibers from the thalamus to the cerebral cortex), and the uncinate fasciculus (connects the anterior temporal lobe with the medial and lateral orbitofrontal cortex), also showed white matter abnormalities in AD (Madhavan et al., 2016).

Because the left hemisphere has a higher number of resultant discriminative intra-hemispheric connectivity than the right, there is a bigger loss of neuronal connectivity in the left hemisphere than in the right. According to the literature, grey matter atrophy happens earlier and advances quicker in the left hemisphere than in the right, resulting in poor verbal memory and language impairment (Yang et al., 2017, Noah Lubben et al., 2021; Rogers et al., 2007). The intra-hemispheric regions showed higher discriminative changes in connectivity, confirming that memory impairment is strongly linked to intra-hemispheric connectivity.

The study does have certain limitations. First, it may be argued that the technique used for obtaining "real" fibers is not the most accurate, even though the method produced outstanding classification performance. By manually checking the MR images, parcellation results, and tractography outputs for quality assurance, we ensure that errors are minimized. Metrics of small-worldness values also suggest that the NC and AD brain networks showed prominent small-world properties (the real network). Another important issue is that the boundary characteristics of a brain network may disappear during the convolution process, even when there is zero padding in the convolutional layers.

5. Conclusion

Understanding how neurons and neural networks process information requires a thorough understanding of brain structural connectivity. Investigating changes in specific white matter fiber pathways in Alzheimer's disease is a great method to evaluate the disease and target therapeutic interventions, but it's also important to address the disease's global, network-based dysfunctions. This research visually explains the Grad-CAM based CNN decision in structural brain network based classification to show discriminative white matter connectivity variations in AD. White matter changes are shown within the temporal/subcortical areas and from the temporal/subcortical regions to frontal and parietal regions in both the left and right hemispheres, revealing a distinct pattern in the discriminative changes in structural connectivity of AD. The temporal/subcortical grey matter regions are densely covered in the observed discriminative structural connectivity alterations in AD. This approach has the potential to comprehensively analyze the discriminative structural connectivity differences, provide a better understanding of network-based changes over the course of Alzheimer's disease, and change the way biomarkers are detected, allowing clinicians to better understand the structural changes that occur in AD and giving them more confidence in automated AD diagnostic systems.

Ethical Approval and Consent to Participate

Details can be found at adni.loni.usc.edu.

Consent to Publish

Not Applicable.

Funding

PGIS research grant programme 2020 Grant Number: PGIS/2020/06.

Availability of data and materials

Data generated or analyzed during this study are included in this published article and remaining are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Subaramya Srivishagan: Conceptualization, Software, Methodology, Formal analysis, Writing – review & editing. **Logiraj Kumaralingam:** Methodology, Writing – review & editing. **Kokul Thanikasalam:** Writing – review & editing, Supervision. **U.A.J. Pinidiyaarachchi:** Writing – review & editing, Supervision. **Nagulan Ratnarajah:** Data curation, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Acosta-Cabronero, J., Williams, G.B., Pengas, G., et al., 2010. Absolute diffusivities define the landscape of white matter degeneration in Alzheimer's disease. Brain 133, 529–539.

Bendlin, B.B., Ries, M.L., Canu, E., et al., 2010. White matter is altered with parental family history of Alzheimer's disease. Alzheimers Dement. 6 (5), 394–403.

Bernardes da Silva Filho, R., Oliveira Barbosa, S., Rondinoni, J.H., et al., 2017. Neurodegeneration profile of Alzheimer's patients: a brain morphometry study. Neuroimage Clin. 15 (3), 15–24.

Böhle, M., Eitel, F., Weygandt, M., et al., 2019. Layer-wise relevance propagation for explaining deep neural network decisions in MRI-based Alzheimer's disease classification. Front. Aging Neurosci. 11, 194.

Bosch, B., Arenaza-Urquijo, E.M., Rami, L., et al., 2012. Multiple DTI index analysis in normal aging, amnestic MCI and AD. Relationship with neuropsychological performance. Neurobiol. Aging 33, 61–74, 21.

Bunce, J.G., Sabolek, H.R., Chrobak, J.J., 2003. Intraseptal infusion of oxotremorine impairs memory in a delayed-non-match-to-sample radial maze task. Neuroscience 121, 259–267.

Cajanus, A., Solje, E., Koikkalainen, J., et al., 2019. The association between distinct frontal brain volumes and behavioral symptoms in mild cognitive impairment, Alzheimer's disease, and frontotemporal dementia. Front. Neurol. 3 (10), 1059.

Campbell, J.S., Siddiqi, K., Rymar, V.V., et al., 2005. Flow-based fiber tracking with diffusion tensor and q-ball data: validation and comparison to principal diffusion direction techniques. Neuroimage 27, 725–736.

Castelvecchi, D., 2016. Can we open the black box of AI? Nature 538, 20–23. https://doi.org/10.1038/538020a.
Chua, T.C., Wen, W., Slavin, M.J., et al., 2008. Diffusion tensor imaging in mild cognitive impairment and Alzheimer's disease: a review. Curr. Opin. Neurol. 21, 83–92.
Colom, L.V., Castañeda, M.T., Bañuelos, C., et al., 2010. Medial septal beta-amyloid 1-40 injections alter septo-hippocampal anatomy and function. Neurobiol. Aging 31, 46–57.
Daianu, M., Mendez, M.F., Baboyan, V.G., et al., 2016. An advanced white matter tract analysis in frontotemporal dementia and early-onset Alzheimer's disease. Brain Imaging Behav 10 (4), 1038–1053.
Dalboni da Rocha, J.L., Bramati, I., Coutinho, G., et al., 2020. Fractional anisotropy changes in parahippocampal cingulum due to Alzheimer's disease. Sci Rep 10 (1), 2660.
Demirhan, A., Nir, T.M., Zavaliangos-Petropulu, A., et al., 2015. Alzheimer's Disease Neuroimaging Initiative, 2015. Feature selection improves the accuracy of classifying Alzheimer disease using diffusion tensor images. Proc. IEEE Int. Symp. Biomed. Imaging. 126e130.
Desikan, R.S., Ségonne, F., Fischl, B., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31 (3), 968–980.
Drzezga, A., 2018. The network degeneration hypothesis: spread of neurodegenerative patterns along neuronal brain networks. J. Nucl. Med. 59, 1645–1648.
Ebrahimighahnavieh, M.A., Luo, S., Chiong, R., 2020. Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review. Comput. Methods Progr. Biomed. 187, 105242.
Englund, E., 1998. Neuropathology of white matter changes in Alzheimer's disease and vascular dementia. Dement. Geriatr. Cogn. Disord. 9 (Suppl 1), 6–12.
Erkinjuntti, T., Lee, D.H., Gao, F., et al., 1993. Temporal lobe atrophy on magnetic resonance imaging in the diagnosis of early Alzheimer's disease. Arch. Neurol. 50 (3), 305–310.
Griffa, A., Baumann, P.S., Thiran, J.P., et al., 2013. Structural connectomics in brain diseases. Neuroimage 80, 515–526.
Hagmann, P., Cammoun, L., Gigandet, X., et al., 2008. Mapping the structural core of human cerebral cortex. PLoS Biol 6 (7), e159.
Hardy, J., Selkoe, D.J., 2002. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science 297, 353–356.
Hopper, M.W., Vogel, F.S., 1976. The limbic system in Alzheimer's disease. A neuropathologic investigation. Am. J. Pathol. 85, 1–20.
Ihara, M., Polvikoski, T.M., Hall, R., et al., 2010. Quantification of myelin loss in frontal lobe white matter in vascular dementia, Alzheimer's disease, and dementia with Lewy bodies. Acta Neuropathol 119, 579–589, 8.
Jack, C.R.J., Bernstein, M.A., Fox, N.C., et al., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27 (4), 685–691.
Jack, C.R., Knopman, D.S., Jagust, W.J., et al., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 9, 119–128.
Jacobs, H.I., Van Boxtel, M.P., Jolles, J., et al., 2012. Parietal cortex matters in Alzheimer's disease: an overview of structural, functional and metabolic findings. Neurosci. Biobehav. Rev. 36 (1), 297–309.
Jo, T., Nho, K., Saykin, A.J., 2019. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Front. Aging Neurosci. 11, 220.
Kantarci, K., Murray, M.E., Schwarz, C.G., et al., 2017. White-matter integrity on DTI and the pathologic staging of Alzheimer's disease. Neurobiol. Aging 56, 172–179.
Killiany, R.J., Moss, M.B., Albert, M.S., et al., 1993. Temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. Arch. Neurol. 50 (9), 949–954.
Konukoglu, E., Coutu, J.P., Salat, D.H., et al., 2016. Multivariate statistical analysis of diffusion imaging parameters using partial least squares: application to white matter variations in Alzheimer's disease. Neuroimage 134, 573e586.
Ledig, C., Schuh, A., Guerrero, R., et al., 2018. Structural brain imaging in Alzheimer's disease and mild cognitive impairment: biomarker analysis and shared morphometry database. Sci. Rep. 8, 11258.
Li, X., Wang, H., Tian, Y., et al., 2016. Impaired white matter connections of the limbic system networks associated with impaired emotional memory in Alzheimer's disease. Front. Aging Neurosci. 8, 250.
Madhavan, A., Schwarz, C.G., Duffy, J.R., et al., 2016. Characterizing white matter tract degeneration in syndromic variants of Alzheimer's disease: a diffusion tensor imaging study. J. Alzheimers Dis. 49 (3), 633–643.
Mak, E., Gabel, S., Mirette, H., et al., 2017. Structural neuroimaging in preclinical dementia: from microstructural deficits and grey matter atrophy to macroscale connectomic changes. Ageing Res. Rev. 35, 250–264.
Mito, R., Raffelt, D., Dhollander, T., et al., 2018. Fibre-specific white matter reductions in Alzheimer's disease and mild cognitive impairment. Brain 141, 888–902.
Mondragón, J.D., Marapin, R., De Deyn, P.P., et al., 2021. Short- and long-term functional connectivity differences associated with Alzheimer's disease progression. Dement. Geriatr. Cogn. Dis. Extra 11 (3), 235–249.
Mori, S., Crain, B.J., Chacko, V.P., et al., 1999. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Ann. Neurol 45, 265–269.
Neugroschl, J., Wang, S., 2011. Alzheimer's disease: diagnosis and treatment across the spectrum of disease severity. Mt. Sinai J. Med. 78, 596–612.
Lubben, N., Ensink, E., Coetzee, G.A., et al., 2021. The enigma and implications of brain hemispheric asymmetry in neurodegenerative diseases. Brain Commun. 3 (3), fcab211.
Passingham, R.E., 2013. What we can and cannot tell about the wiring of the human brain. Neuroimage 80, 14–17.
Prawiroharjo, P., Yamashita, K.I., Yamashita, K., et al., 2020. Disconnection of the right superior parietal lobule from the precuneus is associated with memory impairment in oldest-old Alzheimer's disease patients. Heliyon 6 (7), e04516.
Raj, A., Kuceyeski, A., Weiner, M.A., 2012. Network diffusion model of disease progression in dementia. Neuron 73, 1204–1215.
Rieke, J., Eitel, F., Weygandt, M., et al., 2018. Visualizing convolutional networks for MRI-based diagnosis of Alzheimer's disease. Understanding and Interpreting Machine Learning in Medical Image Computing Applications. Springer, Granada, pp. 24–31.
Selvaraju, R.R., Cogswell, M., Das, A., et al., 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: IEEE International Conference on Computer Vision (ICCV), pp. 618–626.
Smith, S.M., Jenkinson, M., Woolrich, M.W., et al., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 (Suppl 1), S208–S219.
Sorg, C., Riedl, V., Perneczky, R., et al., 2009. Impact of Alzheimer's disease on the functional connectivity of spontaneous brain activity. Curr. Alzheimer Res. 6 (6), 541–553.
Subaramya, S., Kokul, T., Nagulan, R., et al., 2021. Detection of Alzheimer's Disease using Structural Brain Network and Convolutional Neural Network. ICIAfS, pp. 173–178. https://doi.org/10.1109/ICIAfS52090.2021.9606008.
Srivishagan, S., Perera, A.A.I., Hojjat, A., et al., 2020. Brain network measures for groups of nodes: application to normal aging and Alzheimer's disease. Brain Connect. 10 (6), 316–327.
Tang, F., Zhu, D., Ma, W., et al., 2021. Differences changes in cerebellar functional connectivity between mild cognitive impairment and Alzheimer's disease: a seed-based approach. Front. Neurol. 12, 645171.
Vieira, S., Pinaya, W.H., Mechelli, A., 2017. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. Neurosci. Biobehav. Rev. 74, 58–75.
Yang, C., Zhong, S., Zhou, X., et al., 2017. The abnormality of topological asymmetry between hemispheric brain white matter networks in Alzheimer's disease and mild cognitive impairment. Front. Aging Neurosci. 9, 261.
Yang, C., Rangarajan, A., Ranka, S., 2018. Visual explanations from deep 3D convolutional neural networks for Alzheimer's disease classification. arXiv arXiv: 1803.02544.
Zhou, J., Gennatas, E.D., Kramer, J.H., et al., 2012. Predicting regional neurodegeneration from the healthy brain functional connectome. Neuron 73, 1216–1227.
Zhou, B., Khosla, A., Lapedriza, A., et al., 2016. Learning deep features for discriminative localization. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, pp. 2921–2929.