IEEE Journal of Translational Engineering in Health and Medicine

Received 19 March 2022; revised 17 November 2023 and 10 February 2024; accepted 12 February 2024.
Date of publication 19 February 2024; date of current version 14 March 2024.

Digital Object Identifier 10.1109/JTEHM.2024.3366504

Sparse Deep Neural Network for Encoding and Decoding the Structural Connectome

SATYA P. SINGH[1], SUKRIT GUPTA[2], AND JAGATH C. RAJAPAKSE[3], (Fellow, IEEE), for the Alzheimer's Disease Neuroimaging Initiative

[1]Division of Electronics and Communication Engineering, Netaji Subhas University of Technology, Dwarka, New Delhi 110078, India
[2]Department of Computer Science and Engineering, Indian Institute of Technology Ropar, Rupnagar, Punjab 140001, India
[3]School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798

CORRESPONDING AUTHOR: J. C. RAJAPAKSE (asjagath@ntu.edu.sg)

This work was supported by AcRF Tier-2 of Ministry of Education, Singapore, under Grant MOE T2EP20121-0003.

ABSTRACT Brain state classification by applying deep learning techniques on neuroimaging data has become a recent topic of research. However, unlike domains where the data is low dimensional or there are large number of available training samples, neuroimaging data is high dimensional and has few training samples. To tackle these issues, we present a sparse feedforward deep neural architecture for encoding and decoding the structural connectome of the human brain. We use a sparsely connected element-wise multiplication as the first hidden layer and a fixed transform layer as the output layer. The number of trainable parameters and the training time is significantly reduced compared to feedforward networks. We demonstrate superior performance of this architecture in encoding the structural connectome implicated in Alzheimer's disease (AD) and Parkinson's disease (PD) from DTI brain scans. For decoding, we propose recursive feature elimination (RFE) algorithm based on DeepLIFT, layer-wise relevance propagation (LRP), and Integrated Gradients (IG) algorithms to remove irrelevant features and thereby identify key biomarkers associated with AD and PD. We show that the proposed architecture reduces 45.1% and 47.1% of the trainable parameters compared to a feedforward DNN with an increase in accuracy by 2.6 % and 3.1% for cognitively normal (CN) vs AD and CN vs PD classification, respectively. We also show that the proposed RFE method leads to a further increase in accuracy by 2.1% and 4% for CN vs AD and CN vs PD classification, while removing approximately 90% to 95% irrelevant features. Furthermore, we argue that the biomarkers (i.e., key brain regions and connections) identified are consistent with previous literature. We show that relevancy score-based methods can yield high discriminative power and are suitable for brain decoding. We also show that the proposed approach led to a reduction in the number of trainable network parameters, an increase in classification accuracy, and a detection of brain connections and regions that were consistent with earlier studies.

INDEX TERMS Alzheimer's disease, brain decoding, diffusion tensor imaging, Parkinson's disease, relevancy backpropagation, structural connectome.

Clinical and Translational Impact Statement— The methods proposed in this work are clinically relevant because they can aid in obtaining more generalizable and better performing models for predicting Alzheimer's disease and Parkinson's disease using fewer training samples. Besides the diagnosis, the proposed approach can also be used to provide clinicians the basis for the diagnosis by locating neuroimaging biomarkers that lead to the obtained prediction for the given subject.

I. INTRODUCTION

WITH the advances in neuroimaging techniques and availability of neuroimaging biomarkers for Alzheimer's disease (AD) and Parkinson's disease (PD), the diagnosis of neurological diseases can be streamlined and significantly improved [1], [2]. However, due to paucity and high dimensionality of imaging data, application of deep neural networks (DNN) for analysis leads to increased computational load and overfitting [3]. This situation can be mitigated if the number of trainable parameters and input

parameters are reduced. To this end, we propose a sparse neural network architecture and a recursive feature elimination algorithm which include layer-wise relevance propagation (LRP) [4], DeepLIFT [5], and Integrated Gradients (IG) [6] to obtain a leaner model for encoding and decoding the human structural connectome derived from diffusion tensor imaging (DTI) brain scans.

Recent developments suggest that impairments in white-matter also contribute to AD and PD pathology, leading to investigations on how diffusion tensor imaging (DTI) can aid in early diagnosis of AD/PD and predict cognitive decline [7], [8], [9], [10], [11]. DTI characterizes white-matter microstructures and measures white-matter pathways between brain regions, which are quantified by the structural connectome (SC) [8], [12]. The brain network quantified by the SC is represented as a weighted graph consisting of a set of nodes denoted by brain regions of interest (ROI) and a set of edges with associated weights of white matter connections linking the ROI. White matter connections are represented as a network of nodes and edges, where the nodes represent brain regions and the edges represent the connections between brain regions. The edges are weighted to represent the strength of the connection. Several statistical and machine learning techniques have been explored to detect AD and MCI from DTI brain scans, including support vector machines (SVM) [13], independent component analysis [14], and deep neural networks (DNN) [15], [16] eXtreme gradient boosting (XGBoost) [17]. Recently, deep neural techniques such convolutional neural networks (CNN) [14], [16] and graph neural networks have been explored to diagnose AD and MCI by classifying SC derived features. Several recent approaches have been proposed to assess AD and PD. For example, support vector machine recursive feature elimination (SVM-RFE) [18], CT-GAN [19], SVM-RBF [20], SVM [21], [22] CNN [23], 3D-CNN [23]. These approaches achieve state-of-the-art classification accuracy on similar datasets. However, they all have certain limitations. For example, CNN, 3D-CNN can be computationally expensive to train, especially when using large datasets. Training SVMs can be computationally expensive, especially for large datasets. Deep learning models and SVMs can be difficult to interpret, making it difficult to understand why the model makes certain predictions. This is a major limitation in neuroimaging applications, where interpretation of inferences is crucial. SVMs are good at classifying data that is linearly separable, but they can struggle with data that is not linearly separable. Independent component analysis is good at extracting independent signals from a mixture of signals, but it can be sensitive to noise. DNN are flexible and powerful, but they can be computationally expensive to train. One of the advantages of using DNN to diagnose AD and MCI is that they can learn complex relationships between the features in the data. This can be difficult to do with traditional machine learning techniques. Another advantage of DNN is that they can be trained on large datasets. This holds significance due to the intricate nature of AD and PD, necessitating the training of machine learning models on extensive datasets to attain optimal performance. Nonetheless, utilizing DNN for the diagnosis of AD and PD presents specific challenges. One notable hurdle is the potential for overfitting, wherein the model excessively learns from the training data, impeding its ability to generalize to new data. Moreover, the complexity of DNN makes the interpretation of their predictions challenging, complicating the understanding of the underlying mechanisms. Additionally, these approaches have yet to overcome the limitations of applying DNN to neuroimaging, such as the constraints posed by low sample sizes, the high dimensionality of inputs, and the absence of interpretability during both training and testing phases. To mitigate overfitting caused by limited samples and numerous dimensions, techniques like dropouts and weight regularization are typically employed. However, the black-box nature of DNN in neuroimaging remains largely unaddressed, especially in contexts where the interpretation of inferences holds crucial significance. Recent works on functional connectivity [24] and DTI [12] address this issue and use model interpretability techniques for detecting disease related biomarkers.

Feature selection and dimension reduction strategies can play an important role in diagnosis and finding biomarkers for brain disease [24], [25]. Traditional machine learning techniques such as k-best, principal component analysis (PCA) and recursive feature elimination (RFE) have been successfully utilized for feature selection and dimension reduction in neuroimages but such techniques cannot be used to select features of DNN because of their multilayered nature. For base input space, the relevant features selected are potential disease biomarkers that have neuroscientific relevance. However, techniques like PCA project the input space into a latent space that is hard to interpret in the neuroanatomical sense. Perturbation-based approaches are most intuitive for determining salient features. However, several drawbacks exist in such approaches and the most notable are computational efficiency and susceptibility to saturation effects [12]. To overcome these issues, relevancy (or saliency) backpropagation methods have been proposed in the literature [26]. Since all back-propagation based approaches depend on gradient computation, they face saturation problems and may produce misleading results at the discontinuities of activation functions [27]. Several attempts have been made recently to determine relevancy (importance, contributions, or saliency) of features that are most discriminative for classification using DNN models [26]. These methods differ depending on how gradients are tackled and how relevancy scores are evaluated and include LRP, DeepLIFT, and IG.

The aim of this paper is to propose a computational efficient feedforward DNN architecture for classification of patients from their healthy cognitively normal (CN) counterparts and investigate relevancy backpropagation methods in identifying key structural connectome features that are

IEEE Journal of Translational Engineering in Health and Medicine

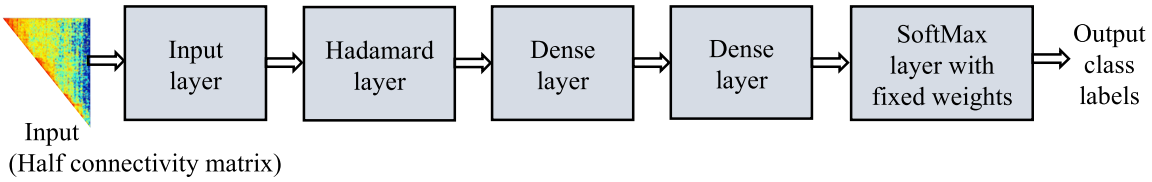

FIGURE 1. Neural network architecture for classification of CN, MCI, and AD.

associated with the respective diseases. We demonstrate the efficacy of the proposed methods for classification of AD and MCI subjects from their healthy CN subjects by using DTI brain scans gathered in Alzheimer's Disease Neuroimaging Initiative[1] (ADNI) [28], and PD subjects from their healthy CN subjects by using DTI brain scans gathered in the Parkinson's Progressive Markers Initiative[2] (PPMI) [29]. We refer to 'decoding' as the process of identifying key brain regions and connections of the connectome that are associated with a disease. We decode structural biomarkers associated with AD and PD. The decoded for both AD and PD are found to be consistent with existing literature and can be used as disease biomarkers.

The proposed architecture uses sparse first hidden layer and an output layer with fixed parameters, which reduces the number of learnable parameters and computational time significantly. We propose a recursive feature elimination (RFE) strategy based on DeepLIFT, IG, and LRP algorithm to remove irrelevant input features and thereby identify biomarkers (i.e., key brain regions and connections) associated with AD and PD. We demonstrate that our approach to removing irrelevant features not only identifies biomarkers associated with brain disease but also increases the accuracies. The proposed approach leads to a compact DNN architecture and an improved classification accuracy.

We make the following contributions in this paper:

1) We present a sparse DNN for encoding the structural connectome of human brain. The key features of the architecture are (i) a sparsely connected Hadamard layer as the first hidden layer, and (ii) a softmax layer with fix connections. We demonstrate that this architecture reduces the number of learnable parameters and increases the classification accuracy.
2) We evaluate three major relevancy backpropagation methods including LRP, DeepLIFT, IG, and demonstrate their efficacies in decoding the structural connectome. We propose a recursive feature elimination approach that iteratively removes irrelevant features from the input and show that an empirically selected subset of features leads to improved accuracy of classification of AD, MCI, and PD. We report state-of-the-art accuracies in binary classification of AD, MCI, and PD, by using selected features from DTI scans from ADNI and PPMI database.
3) By decoding the structural connectome, we identify key brain regions and connections associated with MCI and AD. These biomarkers are consistent with existing literature.

The rest of the paper is organized as follows: Section II discusses the proposed DNN architecture and Section III introduces the relevancy backpropagation methods and the proposed RFE algorithm for brain decoding. Experiments and results are presented in Section IV and finally, Section V discusses the findings and makes conclusions.

II. NEURAL NETWORK ARCHITECTURE

Encoding the connectome involves learning representation of the connectome, leading to classification of the subjects into groups. The structural connectome is represented by a graph $G = (V, E, A)$ where V denotes the set of brain regions of interests (ROI), E denotes the set of edges connecting brain ROI, and A denotes the structural connectivity matrix quantifying edge strengths. The connectivity matrix A is a symmetric matrix and therefore we use the elements in the upper/lower triangular matrices as input features to the DNN.

The proposed DNN architecture is shown in Fig. 1 and consists of (i) an input layer of nodes receiving connectivity features, (ii) a Hadamard layer as the first hidden layer, (iii) two consecutive dense layers, and (iv) a softmax layer with fixed weights. Let the input $x \in \mathbf{R}^n$ where n denotes the number input features, and W^l and b^l denote the weight matrix and bias vector of layers $l = 0, 1, \cdots 4$. The network has 5 layers. $l = 0$ corresponds the input layer and $l = L = 4$ denotes the output layer. Let y^l denote the output of layer l.

Inspired by the pursuit of energy and computation efficient neural networks [30], we use a Hadamard layer as the first hidden layer, which uses element-wise product of weights and inputs to compute the synaptic input. We hypothesize that one to one mapping of feature inputs with trainable weights can scale the strengths of the edges between brain regions and improve the sparsity and the performance of the network. The output y^1 of the Hadamard layer is given by:

$$y^1 = \text{ReLU}\left(W^{1^T} \odot x\right) \tag{1}$$

where $\odot$ denotes the element-wise product and ReLU denotes the ReLU activation function. The Hadamard layer is followed by two consecutive fully connected layers. The outputs y^l of the fully connected layers $l = 2$ and 3 are given by

$$y^l = \text{ReLU}\left(W^{l^T} y^{l-1} + b^l\right) \tag{2}$$

[1] http://adni.loni.usc.edu/
[2] https://www.ppmi-info.org/

Recently, it has been argued that the weights of the output layer can be fixed as non-trainable with no loss in accuracy [31], [32]. The main intuition is that during the training, both input features and weight vectors align simultaneously but fixing output layer weights still is able learn by adapting to input features. Therefore, we fix the weight matrix W^L of the output layer to be an the Hadamard approximation of an orthonormal projection matrix $P \in \mathbf{R}^{n_{L-1} \times K}$ where n_{L-1} is the size of the last hidden layer. The output y of the network is given by [32]:

$$y = \text{softmax}\left(W^L \hat{y}^{L-1}\right) \tag{3}$$

where $W^L = H \approx P$ with $H \in \{-1, 1\}^{n_{L-1} \times K}$ is the truncated Hadamard matrix approximating an arbitrary orthonormal matrix P. $\hat{y}^{L-1} = y^{L-1} / \left\| y^{L-1} \right\|$ is the normalized output of the last hidden layer.

The trainable parameters $\theta = \left\{W^l, b^l\right\}_{l=1,2,3}$ are learned by using mini-batch stochastic gradient descent learning by minimizing the cross-entropy cost J:

$$J = -log\left(E_x\left\{log\left(y = d|x, \theta\right)\right\}\right) \tag{4}$$

where E_x takes the expectation over all the training patterns x and d denotes the corresponding target labels.

III. RELEVANCY BACKPROPAGATION

While neural networks using information from neuroimaging brain scans have been used for classification of patients and healthy subjects, few have attempted to interpret the information flow or identifying salient features contributing to the classification. Recently, there have been increased attention on methods interpreting neural network predictions. We describe three such methods, namely LRP, DeepLIFT, and IG, that assign a relevance score (also referred to as contribution, attribution, or salience score) for each input feature. The relevancy score quantifies the importance of an input feature for the classification task. We will then present a recursive feature elimination (RFE) method that removes irrelevant features recursively for classification and identifies key brain connections associated with AD and MCI.

Let $x = (x_i)$ and $y = (y_k)$ be the input and output of the DNN. The relevance r_i^k of an input feature x_i to the output y_k is defined as $r_i^k = \partial y_k / \partial x_i$. Perturbation based methods, for example [33], compute saliency of an input feature by removing or masking it, running a forward pass on the new input, and then computing the difference with the original output. However, such methods are slow and computationally inefficient. Here, we focus on backpropagation-based methods that compute relevance of input features at the output layer in a single forward and a backward pass through the network to the input layer.

In a DNN, each layer performs a linear transformation of features, followed by a nonlinear mapping f (through the activation function). A path connecting an input and the output unit consists of a sequence of such operations and involves several weights. The backpropagation of relevancy at the output units to input units thus involves the sum of products of all weights and all derivatives of the nonlinearities along individual paths. Following [33], we present below the formulation of relevancy backpropagation methods. We define $z_{ji} = w_{ji}^{l+1,l} u_i^l + b_j^{l+1}$ to be the weighted activation from neuron i in previous layer l to j in layer $l+1$ where $w_{ji}^{l+1,l}$ and b_j^{l+1} denotes the weights and biases connected to layer $l+1$, respectively.

A. LAYER-WISE RELEVANCE PROPAGATION (LRP)

LRP [4] Is computed in a backward pass and starts with assigning the relevance of the targeted neuron k equal to the activation output of the neuron itself $r_i^l = y_k$ and all other neurons to zero. The relevancy of neuron i in layer l in LRP is given by

$$\mathrm{r}_\mathrm{i}^\mathrm{l} = \sum_\mathrm{j} \frac{\mathrm{z_{ji}}}{\sum_\mathrm{i} {}'\mathrm{z_{ji'}} + \epsilon \left(\sum_\mathrm{i} {}'\mathrm{z_{ji'}}\right)} \mathrm{r}_\mathrm{j}^{\mathrm{l}+1} \tag{5}$$

where ϵ is small constant to avoid numerical instability.

Here, i' represents an index iterating over neurons in the same layer as i. The relevancy of the neurons in input layer is computed by recursively applying the above rule from the output layer.

B. DeepLIFT

The basic idea behind DeepLIFT [5] is to explain the "difference from reference value" of the output in terms of the "difference from the reference value" of inputs. The relevancy of the output neuron $r_i^l = y_k - \bar{y}_k$ where $\bar{y}_k$ is the output of k th neuron when the input is the baseline $\bar{x}$

$$\mathrm{r}_\mathrm{i}^\mathrm{l} = \sum_\mathrm{j} \frac{\mathrm{z_{ji}} - \bar{z}_\mathrm{ji}}{\sum_\mathrm{i} {}'\mathrm{z_{ji'}} - \sum_\mathrm{i} {}'\bar{z}_\mathrm{ji'}} \mathrm{r}_\mathrm{j}^{\mathrm{l}+1} \tag{6}$$

The reference value $\bar{z}_{\mathrm{ji}}$ for baseline input $\bar{x}$ is calculated by running a forward pass through the network.

C. INTEGRATED GRADIENT (IG)

IG [6] computes the average partial derivative of each feature as input varies from a baseline to its final value. The relevancy at the output layer is obtained as follows:

$$r_i^L = (x_i - \bar{x}_i) \int_{\alpha=0}^{1} \left.\frac{\partial y_k}{\partial x_i}\right|_{x=\bar{x}+\alpha(x-\bar{x})} d\alpha \tag{7}$$

where $\bar{x}$ is a baseline input and is defined by the user and often taken as zero and $\alpha \in [0, 1]$. The backpropagation formula for IG are similar to those of DeepLIFT except that the relevancy are integrated from the baseline to its value.

D. RECURSIVE FEATURE ELIMINATION (RFE)

We present a RFE algorithm that recursively eliminates irrelevant features so that the most salient features of the classification remain. The algorithm is as follows:

Given a set of input features $\{i\}$

Iterate for each fold:

Train the network, using the train set

IEEE Journal of Translational Engineering in Health and Medicine

Evaluate relevance scores $\{r_i^k\}$
Rank input features based on relevance scores
Remove the most irrelevant features

In our implementation, we removed 10% of features at each step and modified the number of hidden layer neurons, accordingly. We empirically choose 10% removal of features by considering the computational complexity in each elimination step. The optimal set of features for classification was determined by the best accuracy.

We will make the codes used in the manuscript available in public domain upon acceptance of the manuscript.

IV. EXPERIMENTS AND RESULTS

A. DATA DESCRIPTION AND PROCESSING

AD data: Data gathered in ADNI2 and ADNI-GO baseline studies were downloaded from ADNI database [34]. The study included individuals aged between 50 to 90 years, consisting of 67CN, 56 with AD, and 93 with MCI at the baseline assessment. Within the CN group, the average age was 74.5 years. The ADNI Memory (MEM) and ADNI Executive Functioning (EF) scores averaged 0.61 and 0.54 respectively, while the Mini-Mental State Examination (MMSE) score averaged 17.69. For the MCI group, the average age was 73.84 years, with MEM and EF scores averaging 0.30 and 0.26 respectively, and an MMSE score averaging 19.94. As for the AD group, the average age was 76.15 years, with MEM and EF averaging -0.47 and -0.37 respectively, and an MMSE score averaging 9.44. The study involved scans from a total of 221 CN, 155 AD, and 315 MCI subjects across various visits.

PD data: For Parkinson's disease classification, we downloaded 185 PD and 72 CN subjects from PPMI database with both T1 and DTI images available. The mean age for PD and CN subjects were 62.2 and 58.7 years, respectively. In this study we did not include evidence of dopaminergic deficit (SWEDD), genetic cohorts, and prodromal subjects. We use scans of all visits for our experiments while scans from the same subject were kept separate for cross-validation. Therefore, in total, we had scans from 172 CN and 655 PD subjects for experiments.

During data split for test, train, and validation, we have taken care that the scans of the same subject and all visits are not mixed with the others. Only the participants who had both structural MR and DTI images were used in this study. The structural MR images used were T1-weighted images that had been spatially normalized, masked, and nonparametric non-uniform intensity normalization corrected. DTI images had been eddy-current corrected images when downloaded from ADNI. Initial analysis was performed on the selected samples. All preprocessing pipelines were written using Nipype[3] with FSL,[4] and diffusion toolkit[5] tools. Detailed steps for pipeline used for deterministic tractography and cortical parcellation are given in Supplementary Materials in flowcharts S.1 and S.2, respectively.

Fractional anisotropy (FA) images from DTI and AAL transformed MRI images were used to generate connectivity matrices which measure the strength of the connections between different brain regions. In other words, each element of the connectivity matrix represents the strength of the connection between two specific regions. For example, the element at row i, column j of the connectivity matrix represents the strength of the connection between region i and region j. The details about connectivity matrix generation can be found in the supplementary material. The AAL atlas divided the brain into 116 cortical regions and white matter tractography generated connectivity matrices of 116×116 size. Since the connectivity matrices are symmetric, we further eliminated the elements in the lower half triangles of the matrices thereby reducing the number of features to 6786. Removing half of the diagonally symmetric elements in connectivity matrices for FFN training streamlines the dataset by discarding redundant mirrored information, preserving essential connectivity patterns while enhancing computational efficiency and managing data size effectively. We kept diagonal elements in the connectivity matrix to capture internal coherence within brain regions, aiding analysis of brain organization and pathology. This also ensures consistent dimensions for processing and comparison with lower triangular matrices, simplifying visualization of regions of interest.

B. EFFECT OF FEATURE ELIMINATION BASED ON RELEVANCY SCORES

We trained proposed DNN architecture for classification of AD, MCI, and PD, and computed relevancy scores for input features, using each relevancy backpropagation method. We varied the optimizer, the number of neurons in each layer, the number of hidden layers, the learning rate and used nested cross-validation approach for parameter tuning and model selection for each of the relevancy backpropagation method. We divided the dataset into train and test partitions at 4:1 and performed 5-fold cross-validation on the train partition. We obtained the best performance with the Adam optimizer with learning rates 0.01, 0.02, 0.01, and 0.01 for CN vs AD, CN vs MCI, MCI vs AD, and CN vs PD respectively, and with the network configurations given in Table 1 (all features).

For the final model architectures, we observed the performance indices and evaluated relevance scores, using relevancy backpropagation methods. The relevance scores were sorted from high to low values and a subset of 10% features with the least relevance scores were eliminated in each step of RFE. The RFE was performed for each fold, such that we removed 10% features for the same fold till we were left with just 1% of the input features. The average performance scores across the folds were reported. After each feature removal step, training was performed. The relevancy backpropagation methods (LRP, DeepLIFT, and IG) were evaluated in terms of mean accuracy after feature selection

[3]https://nipype.readthedocs.io/en/latest/
[4]https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET
[5]http://trackvis.org/dtk/

TABLE 1. Classification performance with feature selection using DeepLift, IG, and LRP approaches (RS = Relevance score, Acc. = Accuracy, Sens. = Sensitivity, Spec. = Specificity).

RS Method	DNN architecture	Acc. %	Sens. %	Spec. %
ADNI dataset CN vs AD				
DeepLift	487-104-53	79.6± 1.3	72.4± 1.3	84.2± 1.5
IG	438-93-48	79.2± 2.4	71.7± 1.9	84.6± 1.7
LRP	394-84-43	78.4± 1.7	70.7± 2.4	84.4± 2.1
All features	6786-1450-750	77.5± 2.1	70.8± 3.9	81.9± 3.7
ADNI dataset CN vs MCI				
DeepLift	668-147-78	72.0± 2.1	82.8± 1.9	56.4± 2.2
IG	438-96-51	71.8± 1.4	81.5± 1.7	56.2± 1.8
LRP	916-202-108	70.8± 1.8	82.1± 2.5	55.6± 2.7
All features	6786-1500-800	68.6± 2.1	86.4± 3.3	44.3± 2.5
ADNI dataset MCI vs AD				
DeepLift	742-136-71	75.4± 1.9	53.8± 1.3	87.1± 1.9
IG	668-123-64	74.8± 1.4	53.3±1.2	85.0± 1.6
LRP	668-123-64	74.7± 1.3	52.8± 2.7	85.3± 2.3
All features	6786-1250-650	72.4± 2.9	45.4± 3.5	83.3± 3.5
PPMI dataset (CN vs PD)				
DeepLift	438-96-51	82.7± 2.3	87.8± 2.2	76.6± 2.5
IG	438-96-51	82.4± 3.4	87.3±1.2	76.0± 3.6
LRP	394-87-46	82.4± 2.3	86.8± 2.7	76.3± 2.3
All features	6786-1500-800	78.7± 2.8	85.5± 3.1	70.3± 3.4

and the number of selected features. We repeated the experiments with 10 random seeds.

Fig. 2 shows the behavior of mean test accuracy during the recursive elimination of features against feature elimination steps. In each recursive step, we drop 10% less irrelevant features and the test accuracy is noted. The results are also summarized in Table 1. In Table 1, we report peak mean test accuracy during recursive feature selection. We also show the test accuracy without any feature selection (all features). Column 2 in Table 1 showing the model architecture for peak test accuracy. For CN vs AD classification, all DeepLIFT (p-value $< 10^{-7}$), IG (p-value $< 10^{-5}$), and LRP (p-value $< 10^{-4}$) show a statistically significant improvement in accuracy. Using DeepLIFT, 79.6% (acc. improved by 2.1%) accuracy was achieved with only 487 (7.1% of initial features) features. We were also able to maintain initial accuracy (test accuracy with all 6786 features) using only 233 most discriminative feature for DeepLIFT and IG, and 258 for LRP. For CN vs MCI, DeepLIFT (p-value $< 10^{-11}$) was the most consistent performer in a certain range, but all methods show a statistically significant improvement in test accuracy ranging from 3.2 % to 2.2% (IG p-value $< 10^{-11}$, and LRP p-value $< 10^{-9}$). DeepLIFT shows a 3.4% (acc. 72.0%) improvement in the acc. with 668 (9.8% of initial features) features. We were able to maintain initial test accuracy using most discriminative features as low as 170 (for DeepLIFT and IG), and 259 (for LRP). In case of MCI vs AD, we witnessed a statistically significant increase of 2.3 to 3% in test accuracy with LRP (p-value $< 10^{-8}$), DeepLIFT (p-value $< 10^{-9}$) and IG (p-value $< 10^{-9}$). We were able to maintain initial test accuracy with 259 features for LRP and IG, and 233 for DeepLIFT. We also noticed a significant improvement in the accuracy for CN vs PD. Proposed method shows 82.7% (4% improvement), 82.4% (3.7% improvement), and 86.8% (3.7% improvement) accuracy using DeepLIFT, IG, and LRP respectively. We were, therefore, able to remove approximately 90 to 95% irrelevant features while getting higher classification accuracy. Since, DeepLIFT showed the best performance among all methods, the rest of the results in the manuscript are presented using only the DeepLIFT. However, results with other relevancy-based methods are available in the Supplementary material.

C. EFFECT OF HADAMARD LAYER AND FIXED TRANSFORMED LAYER

We ran several experiments for evaluation and assessment of the impact of elementwise multiplication layer at input layer and fixed transform at output softmax layer. We prepared three models: (i) the proposed architecture in Fig. 1; (ii) the fixed transform at the softmax layer is replaced by a dense trainable layer; and (iii) the element-wise multiplication Hadamard layer is replaced by a fully connected dense layer. For training different network models, we used Adam optimizer and a dropout of 0.02 before the softmax layer. The results with different configurations of network architecture are shown in Table 2. Results support our argument made in neural network architecture section that the weights of the output layer can be fixed as non-trainable with no loss in accuracy ((a) and (b) in Table 2). The small decrease in accuracy and other indices outweighs the reduction in trainable parameters, computational complexity, and reduction in training time. The inclusion of element-wise multiplication ((a) and (c) in Table 2) in the input layer shows an increase in accuracy of 2.6 %, 2.3%, 2.4%, and 3.1% for CN vs AD, CN vs MCI, and MCI vs AD, and CN vs PD respectively. A detailed comparison of different network architectures in terms of parameter savings is given in supplementary Table S.1. The use of a Hadamard layer at the input reduces trainable parameters by 45.1%, 46.1%, 47.1%, and 46.1% for CN vs AD, CN vs MCI, MCI vs AD, and CN vs PD respectively.

Although the fixed transform layer also contributes in the reduction of trainable parameters, it will lead to a higher reduction in the number of trainable parameters in case of multiclass classification. In summary, the proposed architecture has several computational benefits with comparatively less memory requirement and trainable parameters and hence requires comparatively less training time. In conclusion, our proposed architecture contributes an increase of ∼2.5-3.1 % in accuracy with ∼45% savings in trainable parameters. Also, RFE using relevancy backpropagation further contributes to ∼2.1 to 4.0 % improvement in accuracy while removing ∼93% irrelevant features. Thereby, our proposed approach improves accuracy of DNN by ∼4.5-7% with significant reduction in number of trainable parameters and irrelevant features.

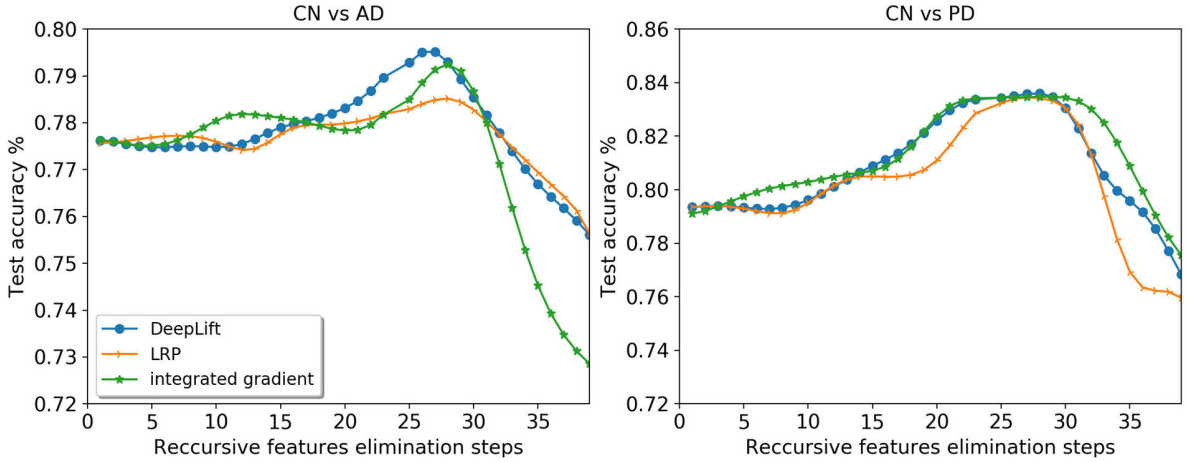

FIGURE 2. Variation in mean test accuracy during recursively eliminating of irrelevant features.

TABLE 2. Effect of Hadamard layer and fixed transform layer in the proposed model on test accuracy with feature selection using DeepLift.

Classification	Acc %	Sens %	Spec %
(a) Proposed approach			
CN vs. AD	79.6± 1.3	72.4± 1.3	84.2± 1.5
CN vs. MCI	72.0± 2.1	82.8± 1.9	56.4± 2.2
MCI vs. AD	75.4± 1.9	53.8± 1.3	87.1± 1.9
CN vs. PD	82.7± 2.3	87.8± 2.2	76.6± 2.5
(b) Fixed transform softmax layer is replaced by dense trainable softmax layer			
CN vs. AD	79.6± 1.8	72.2± 3.2	85.2± 2.7
CN vs. MCI	72.1± 1.7	74.8± 2.2	67.0± 2.5
MCI vs. AD	75.7± 2.2	48.2± 2.9	89.8± 1.2
CN vs. PD	82.8± 2.2	87.2± 2.3	76.3± 2.3
(c) The Hadamard layer is replaced by the dense layer			
CN vs. AD	77.0± 2.2	68.3± 2.1	82.1± 1.7
CN vs. MCI	69.7± 3.2	66.5± 2.4	74.0± 1.3
MCI vs. AD	73.0± 1.2	46.6± 2.3	89.0± 2.3
CN vs. PD	79.6± 3.3	84.2± 3.5	72.4± 4.3

TABLE 3. Comparison of performances with FFN and SVM methods including feature selection using DeepLift and RFE, respectively.

Method	features	Acc. %	Sens. %	Spec. %
CN vs AD				
FFN	DeepLift	77.1± 1.4	71.5± 1.5	82.5± 1.8
SVM	RFE	78.4± 1.2	63.4± 0.8	87.9± 0.5
XGBoost	RFE	76.21±	58.2±	86.4±
ours	DeepLift	79.6± 1.3	72.4± 1.3	84.2± 1.5
CN vs MCI				
FFN	DeepLift	70.2± 1.7	87.7± 2.2	45.1± 2.1
SVM	RFE	70.8± 0.9	80.0± 1.2	61.3± 1.1
XGBoost	RFE	68.5±1.1	72.5±2.1	66.8±1.7
ours	DeepLift	72.0± 2.1	82.8± 1.9	56.4± 2.2
MCI vs AD				
FFN	DeepLift	73.3± 1.9	46.4± 2.5	85.0± 2.1
SVM	RFE	73.9± 0.9	58.8± 1.0	89.2± 1.1
XGBoost	RFE	72.9±1.9	68.7±2.2	82.7±1.7
ours	DeepLift	75.4± 1.9	53.8± 1.3	87.1± 1.9
CN vs PD				
FFN	DeepLift	79.5± 2.7	83.7± 3.5	73.6± 2.4
SVM	RFE	79.6± 1.3	72.4± 1.3	84.2± 1.5
XGBoost	RFE	76.3±2.4	78.2±1.8	80.4±1.7
ours	DeepLift	82.7± 2.3	87.8± 2.2	76.6± 2.5

D. COMPARISON WITH SVM, XGBoost, AND FEEDFORWARD NETWORKS

There is a large variation of classification accuracies due to the number of samples, different steps in preprocessing, cross validation approaches, and different DTI features reported in the literature [35]. Mostly, SVM, XGBoost, and FFN) are the most widely used machine learning approaches for classification of AD [35], [36] and PD [37], [38]. We therefore compare our results with SVMXGBoost, and FFN. We also implemented our RFE approach with FFN architecture using features selected with different relevancy backpropagation methods. Results using RFE with DeepLIFT are presented in Table 3 and detailed analysis is presented in supplementary Table S.2. The results using SVM and XGBoost are shown in Table 3 and detailed results and experimental settings are presented in supplementary Table S.3). We employ RFE for feature selection in both SVM and XGBoost [17] experiments. The optimal parameters for SVM and XGBoost were determined through a Grid search approach. Additionally, a 5-fold cross-validation was executed, and the average results across these folds are presented in Table-3. We observed a very similar trend in terms of test accuracy and the performance of relevancy-based methods as with our approach. Results clearly show that our method offers a better performance than the FFN not only in terms of test accuracy, but also in terms of the number of trainable parameters as discussed in the previous section. In our experiments, XGBoost performance was not good as SVM and FFN. The performance of SVM was better than simple FFN without feature selection and equally well with our proposed approach without any feature selection. However, when we compare the performance of SVM with our approach after feature selection, our approach performs much better than SVM (approx. 1.2%, 1.8%, 1.9% and 3.2% high accuracy for

CN vs AD, CN vs MCI, MCI vs AD, and CN vs PD) for all classification tasks.

E. DECODING BRAIN REGIONS AND CONNECTIONS ASSOCIATED WITH CN, MCI, AD, AND PD

The features used in the classifications were white-matter connections. To investigate the key regions identified by our method using DeepLIFT, we plotted relevance scores for the connections between different regions from the AAL atlas. We observed that the connections between regions in the frontal, precentral, and supplementary areas are crucial for distinguishing AD and MCI from CN brains and from each other. We found that the connections from the right (R) supplementary motor area and left (L) and R precentral gyrus, and middle frontal gyrus are the most distinguishing for CN vs AD, CN vs MCI and MCI vs AD. This was also confirmed by the feature relevance scores from other relevancy backpropagation methods. In order to understand which regions are the most crucial for classification, the anatomical locations of the regions with top 10% relevance scores for (a) DeepLift, (b) Integrated Gradient, and (c) LRP are plotted and are given in supplementary Fig. S.1). We observed that besides the regions from the precentral and supplementary areas, the regions from the frontal lobe (middle, inferior and superior) are crucial for distinguishing MCI and AD patients from the CN subjects. Although, all relevancy backpropagation methods pointed to identical regions as being crucial, the regions identified by DeepLIFT and IG were especially similar.

For more clarity and better comparative analysis, Fig. 3 shows the anatomical labels of the top 10% most relevant brain regions and corresponding relevance score for classification of CN vs AD and CN vs PD with DeepLIFT. For the highest relevance scores, all methods show common distinctive regions (Supplementary Fig. S.2). For CN vs AD, the relevance scores obtained by all methods are high for regions mostly in the frontal lobe (precentral gyrus, superior and middle frontal gyrus, and middle temporal gyrus). For PD, we found that the connections between the regions para hippocampal [39] left and right are crucial for distinguishing AD followed by praecuneus gyrus, frontal superior medial gyrus [40], right lingual gyrus [41], and heschl's gyrus [42].

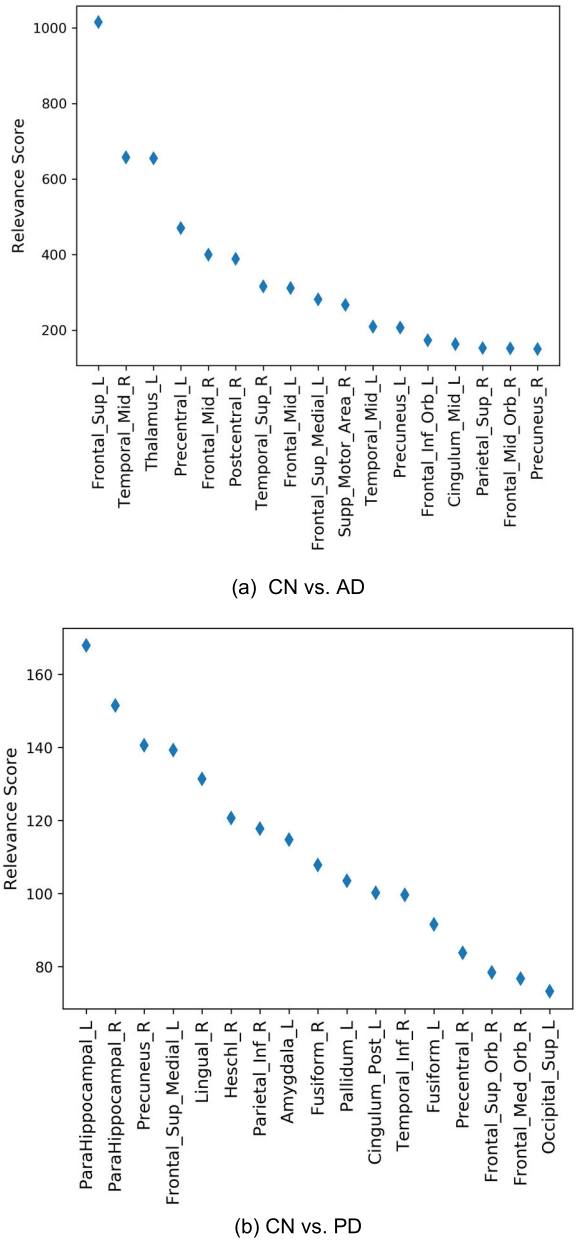

FIGURE 3. Anatomical labels of top 10% most relevant brain regions and corresponding relevance score during classification of CN vs. AD and CN vs. PD using DeepLift.

V. DISCUSSION AND CONCLUSION

The rudimentary symptoms of dementia include confusion and memory loss, followed by a decreased ability to perform daily tasks. However, these symptoms are not unique to dementia, thus making it hard for clinicians to make an accurate diagnosis at the pre-clinical stage. With advances in neuroimaging techniques and information about biological markers for MCI, the diagnosis of neurological diseases can be streamlined and significantly improved. However, shortage of imaging data and high number of features can lead to poor classification accuracy and overfitting especially with DNNs which involve a large number of trainable parameters. This situation can be tackled by reducing the number of trainable parameters in the network and reducing the number of input features. To this affect, we proposed a novel neural network architecture and used relevancy backpropagation to obtain a leaner model for classification of MCI and AD patients. We demonstrated multiple relevancy backpropagation methods, namely DeepLIFT, IG, and LRP, to generate the feature relevance scores and used the scores to recursively eliminate unimportant features. Using the proposed approach, not only were we able to reduce the number of trainable parameters in comparison to the initial model, but also achieved state-of-the-art results in terms of classification accuracy.

Due to different data processing approaches, direct comparison with existing studies is not possible. However, there are similar approaches reported in the literature, which we can compare to gain further insights into our proposed approach. Zhang and Zhan [18], utilized DTI data from 37 AD patients and 36 healthy controls to identify white matter fiber tracts that are damaged in AD employing SVM-RFE. 78.8% accuracy was shown using FA+DA with 75 features selected by RFE. In [19], Zuo et al. proposed CT-GAN by enhancing the fusion of fMRI and DTI brain images for AD analysis. Our proposed approach relatively simple to train, but CT-GANs are more complex due to the presence of the cross-modal transformer and are more prone to overfitting. FFNs are less prone to overfitting than CT-GANs. This is because CT-GANs have more parameters and are more complex.

In [20], Xang et al. employed SVM-RBF to classify AD, MCI, and NC achieving high accuracy and specificity. In another research [21], Tsai et al. discusses machine learning to distinguish PD and Parkinson-plus syndromes, revealing the potential of DTI for the differential diagnosis of Parkinsonism with 87.4% acc. The dataset was different; however, our results are also inline. Yang et al. [22] investigates white matter changes in depressed PD patients, utilizing a DTI-based SVM model for individualized diagnosis with 73% acc. Zhao et al. [23] assesses a CNN's diagnostic efficacy on PD using DTI, demonstrating its potential with a satisfactory performance. In [43], employs a 3D CNN on DTI data for neuroimaging-based brain age prediction, highlighting robust performance and providing insights into age-related changes in axonal structure. 3D-CNN are more complex and computationally expensive to train; more prone to overfitting; more difficult to interpret compared to our proposed approach. However, one potential drawback of our approach compared to 3D-CC that FFN cannot learn as informative features from the 3D DTI data as 3D CNNs; less robust to noise and artifacts in the DTI data. In the comparison between XGBoost [17] our approach across various classification tasks (CN vs AD, CN vs MCI, MCI vs AD, CN vs PD), our method consistently demonstrated higher accuracy and specificity.

On investigating features that distinguish the diseased patients (AD and MCI) from CN subjects, we found that all the three relevancy methods pointed to the same regions and connections. The most salient regions were in the frontal lobe, precentral and supplementary motor area. The different regions in the frontal lobe, known for their role in cognition and decision-making, have been reported to experience significant disruption in white matter connectivity (superior frontal gyrus [44], [45], [46], dorsolateral middle frontal gyrus [47], [48] and orbital inferior frontal gyrus [49], [50]. Further, connections to the precentral, postcentral and supplementary motor areas, known for their role in sensory and motor functions were also found to be crucial in differentiating MCI patients from AD patients, which is in line with previous results suggesting that patients in more advanced stages of AD are more likely to show more motor deficits than in early disease stages [51]. Also, amyloid depositions and neurofibrillary tangles have been found in the primary motor cortex in case of AD [52] and MCI [53], [54] which may cause white matter changes in input and output projections.

The identification of significant brain regions in our classification task, particularly in the frontal lobe, precentral gyrus, and supplementary motor area, sheds light on the neuroanatomical underpinnings of cognitive and motor dysfunction in AD, MCI and related disorders. Disruptions in white matter connectivity within these regions highlight their pivotal role in disease progression and severity, emphasizing the complex interplay between cognitive and motor impairments. From a clinical perspective, these findings provide valuable insights into potential biomarkers for disease progression and treatment response, informing diagnostic protocols and therapeutic interventions aimed at preserving cognitive and motor function in affected individuals. Overall, the identification and clinical interpretation of these significant brain regions contribute to a deeper understanding of disease mechanisms and may facilitate the development of targeted interventions for these debilitating conditions.

Relevancy backpropagation methods play an important role in feature selection and handling high dimensionality of data. An important point is that relevancy-based scoring methods can be used to select an optimal subset of most discriminative features for classification while giving a comparable/higher accuracy with all features. We conclude that relevancy score-based methods can yield high discriminative power and are suitable for brain decoding. We also note that our approach led to a reduction in the number of trainable network parameters, an increase in classification accuracy, and a detection of brain connections and regions that were consistent with earlier studies. The proposed methods can be applied to other neuroimaging modalities and neurological disorders.

The need for precision in articulating the diagnostic implications of this work is evident. While the proposed method shows promise in enhancing image-analytic assessment accuracy for AD and PD, it is crucial to recognize that conventional behavioral tests, like the MMSE for AD or hand-writing and speech tests for PD, maintain significantly higher sensitivity and specificity. Despite the potential, the superior diagnostic performance of simple behavioral tests should not be overlooked. Future studies should validate the proposed method on larger and diverse datasets, exploring the diagnostic relevance of image-analytic assessment for AD and PD.

The proposed sparse DNN architecture tailored for encoding and decoding the human brain's structural connectome resulted in a substantial reduction in trainable parameters and computational time. Our model allows for more efficient processing and analysis of brain connectivity data. The evaluation of prominent relevancy backpropagation techniques—LRP, DeepLIFT, and IG—showcased their effectiveness in

decoding the structural connectome, presenting a comprehensive comparative analysis of these methods for brain decoding purposes. Furthermore, the proposal of a recursive feature elimination methodology iteratively identifies relevant biomarkers linked to AD and PD, offering potential for developing more concise and interpretable models for brain decoding tasks. Achieving state-of-the-art accuracies in binary classification of AD, MCI, and PD using selected features from DTI scans from ADNI and PPMI database underscores the effectiveness of the suggested approaches in accurately classifying Alzheimer's disease and related disorders.

The computational expense associated with the relevancy backpropagation methods utilized here could impede their practical application, particularly with larger-scale datasets. Moreover, the absence of evaluation for the proposed methods on alternative neuroimaging modalities, such as fMRI and PET, leaves uncertainties regarding the methods' generalizability, urging further validation and exploration on diverse imaging modalities to affirm their effectiveness across broader contexts.

DECLARATIONS OF INTEREST

All authors declare there are no conflicts of interest.

DATA AND CODE AVAILABILITY STATEMENT

The data for the manuscript was downloaded from Alzheimer's Disease Neuroimaging Initiative (ADNI). The codes used for the experiments will be made publicly available upon acceptance of the manuscript.

ACKNOWLEDGMENT

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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 the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

The data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (a detailed acknowledgment is found in the Supplementary Material).

REFERENCES

[1] R. P. Guimarães et al., "Is diffusion tensor imaging a good biomarker for early Parkinson's disease?" *Frontiers Neurol.*, vol. 9, p. 626, Aug. 2018, doi: 10.3389/fneur.2018.00626.

[2] S. E. Rose, A. L. Janke PhD, and J. B. Chalk, "Gray and white matter changes in Alzheimer's disease: A diffusion tensor imaging study," *J. Magn. Reson. Imag.*, vol. 27, no. 1, pp. 20–26, Jan. 2008, doi: 10.1002/jmri.21231.

[3] S. Gupta, Y. H. Chan, and J. Rajapakse, "Decoding brain functional connectivity implicated in AD and MCI," in *Proc. Int. Conf. Med. Image Comput. Comput. Interv.*, Jul. 2019, pp. 781–789, doi: 10.1007/978-3-030-32248-9_87.

[4] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PLoS ONE*, vol. 10, no. 7, Jul. 2015, Art. no. e0130140, doi: 10.1371/journal.pone.0130140.

[5] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 4844–4866. [Online]. Available: http://goo.gl/RM8jvH

[6] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proc. 34th Int. Conf. Mach. Learn.*, Mar. 2017, pp. 5109–5118.

[7] C. M. Torgerson, A. Irimia, S. Y. M. Goh, and J. D. Van Horn, "The DTI connectivity of the human claustrum," *Hum. Brain Mapping*, vol. 36, no. 3, pp. 827–838, Mar. 2015, doi: 10.1002/hbm.22667.

[8] X. Sui and J. C. Rajapakse, "Profiling heterogeneity of Alzheimer's disease using white-matter impairment factors," *NeuroImage, Clin.*, vol. 20, pp. 1222–1232, Jan. 2018, doi: 10.1016/j.nicl.2018.10.026.

[9] J. Escudero, E. Ifeachor, J. P. Zajicek, C. Green, J. Shearer, and S. Pearson, "Machine learning-based method for personalized and cost-effective detection of Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 164–168, Jan. 2013, doi: 10.1109/TBME.2012.2212278.

[10] Y. Jin et al., "3D tract-specific local and global analysis of white matter integrity in Alzheimer's disease," *Hum. Brain Mapping*, vol. 38, no. 3, pp. 1191–1207, Mar. 2017, doi: 10.1002/hbm.23448.

[11] J. M. Mateos-Pérez, M. Dadar, M. Lacalle-Aurioles, Y. Iturria-Medina, Y. Zeighami, and A. C. Evans, "Structural neuroimaging as clinical predictor: A review of machine learning applications," *NeuroImage, Clin.*, vol. 20, pp. 506–522, Jan. 2018, doi: 10.1016/j.nicl.2018.08.019.

[12] S.-H. Chu, K. K. Parhi, and C. Lenglet, "Function-specific and enhanced brain structural connectivity mapping via joint modeling of diffusion and functional MRI," *Sci. Rep.*, vol. 8, no. 1, p. 4741, Mar. 2018, doi: 10.1038/s41598-018-23051-9.

[13] L. O'Dwyer et al., "Using support vector machines with multiple indices of diffusion for automated classification of mild cognitive impairment," *PLoS ONE*, vol. 7, no. 2, Feb. 2012, Art. no. e32441, doi: 10.1371/journal.pone.0032441.

[14] T. M. Schouten et al., "Individual classification of Alzheimer's disease with diffusion magnetic resonance imaging," *NeuroImage*, vol. 152, pp. 476–481, May 2017, doi: 10.1016/j.neuroimage.2017.03.025.

[15] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020, doi: 10.1109/TPAMI.2018.2889096.

[16] J. Kawahara et al., "BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, Feb. 2017, doi: 10.1016/j.neuroimage.2016.09.046.

[17] B. Chen et al., "Detection of mild cognitive impairment in Parkinson's disease using gradient boosting decision tree models based on multilevel DTI indices," *J. Transl. Med.*, vol. 21, no. 1, p. 310, May 2023, doi: 10.1186/s12967-023-04158-8.

[18] Y. Zhang and F. Zhan, "Diffusion tensor imaging (DTI) analysis based on tract-based spatial statistics (TBSS) and classification using multimetric in Alzheimer's disease," *J. Integr. Neurosci.*, vol. 22, no. 4, p. 101, Jul. 2023, doi: 10.31083/j.jin2204101.

[19] Q. Zuo, S.-Q. Wang, J. Pan, and S. Wang, "Alzheimer's disease prediction via brain structural-functional deep fusing network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 4601–4612, 2023. [Online]. Available: https://www.researchgate.net/profile/Qiankun-Zuo-2/publication/374452330_Alzheimer's_Disease_Prediction_via_Brain_Structural-Functional_Deep_Fusing_Network/links/651ec2d33ab6cb4ec6bdc2e3/Alzheimers-Disease-Prediction-via-Brain-Structural-Functional-Deep-Fusing-Network.pdf

[20] X. Yang, Y. Xia, Z. Li, L. Liu, Z. Fan, and J. Zhou, "Classification of Alzheimer's disease based on white matter connectivity network," *Appl. Sci.*, vol. 13, no. 21, p. 12030, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/21/12030

[21] C. C. Tsai, Y. L. Chen, C. S. Lu, J. S. Cheng, and Y. H. Weng, "Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning," *Biomed. J.*, vol. 46, no. 3, 2023, Art. no. 100541. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2319417022000919

[22] Y. Yang et al., "Identifying depression in Parkinson's disease by using combined diffusion tensor imaging and support vector machine," *Frontiers Neurol.*, vol. 13, Jun. 2022, Art. no. 878691, doi: 10.3389/fneur.2022.878691.

[23] H. Zhao et al., "Deep learning based diagnosis of Parkinson's disease using diffusion magnetic resonance imaging," *Brain Imag. Behav.*, vol. 16, no. 4, pp. 1749–1760, 2022, doi: 10.1007/s11682-022-00631-y.

[24] S. Gupta, Y. H. Chan, J. C. Rajapakse, and Alzheimer's Disease Neuroimaging Initiative, "Decoding brain functional connectivity implicated in AD and MCI," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Oct. 2019, pp. 781–789.

[25] M. Böhle, F. Eitel, M. Weygandt, and K. Ritter, "Layer-wise relevance propagation for explaining deep neural network decisions in MRI-based Alzheimer's disease classification," *Frontiers Aging Neurosci.*, vol. 11, p. 194, Jul. 2019, doi: 10.3389/fnagi.2019.00194.

[26] M. Ancona, E. Ceolini, C. Oztireli, and M. Gross, "Towards better understanding of gradient-based attribution methods for deep neural networks," in *Proc. NIPS*, Nov. 2017, pp. 1–16, doi: 10.3929/ETHZ-B-000237705.

[27] A. Gotsopoulos et al., "Reproducibility of importance extraction methods in neural network based fMRI classification," *NeuroImage*, vol. 181, pp. 44–54, Nov. 2018, doi: 10.1016/j.neuroimage.2018.06.076.

[28] M. W. Weiner et al., "The Alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement," *Alzheimer's Dementia*, vol. 13, no. 5, pp. 561–571, 2017, doi: 10.1016/j.jalz.2016.10.006.

[29] K. Marek et al., "The Parkinson progression marker initiative (PPMI)," *Prog. Neurobiol.*, vol. 95, no. 4, pp. 629–635, Dec. 2011, doi: 10.1016/j.pneurobio.2011.09.005.

[30] T. C. Deveci, S. Cakir, and A. Enis Cetin, "Energy efficient Hadamard neural networks," 2018, *arXiv:1805.05421*.

[31] E. Hoffer, I. Hubara, and D. Soudry, "Fix your classifier: The marginal value of training the last weight layer," 2018, *arXiv:1801.04540*.

[32] F. Pernici, M. Bruni, C. Baecchi, and A. Del Bimbo, "Fix your features: Stationary and maximally discriminative embeddings using regular polytope (fixed classifier) networks," 2019, *arXiv:1902.10441*.

[33] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision—ECCV* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2014, pp. 818–833, doi: 10.1007/978-3-319-10590-1_53.

[34] C. R. Jack et al., "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, 2008, doi: 10.1002/jmri.21049.

[35] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, Jul. 2017, doi: 10.1016/j.neuroimage.2017.03.057.

[36] B. Mwangi, T. S. Tian, and J. C. Soares, "A review of feature reduction techniques in neuroimaging," *Neuroinformatics*, vol. 12, no. 2, pp. 229–244, 2014, doi: 10.1007/s12021-013-9204-3.

[37] R. Das, "A comparison of multiple classification methods for diagnosis of Parkinson disease," *Exp. Syst. Appl.*, vol. 37, no. 2, pp. 1568–1572, Mar. 2010, doi: 10.1016/j.eswa.2009.06.040.

[38] A. Ozcift, "SVM feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease," *J. Med. Syst.*, vol. 36, no. 4, pp. 2141–2147, Aug. 2012, doi: 10.1007/s10916-011-9678-1.

[39] L. Christopher et al., "Salience network and parahippocampal dopamine dysfunction in memory-impaired Parkinson disease," *Ann. Neurol.*, vol. 77, no. 2, pp. 269–280, Feb. 2015, doi: 10.1002/ana.24323.

[40] R. P. Guimarães et al., "Pattern of reduced functional connectivity and structural abnormalities in Parkinson's disease: An exploratory study," *Frontiers Neurol.*, vol. 7, p. 243, Jan. 2017, doi: 10.3389/fneur.2016.00243.

[41] K. Kawabata et al., "Visuoperceptual disturbances in Parkinson's disease," *Clin. Parkinsonism Rel. Disorders*, vol. 3, 2020, Art. no. 100036, doi: 10.1016/j.prdoa.2020.100036.

[42] D. Zhang, J. Wang, X. Liu, J. Chen, and B. Liu, "Aberrant brain network efficiency in Parkinson's disease patients with tremor: A multimodality study," *Frontiers Aging Neurosci.*, vol. 7, p. 31, Aug. 2015, doi: 10.3389/fnagi.2015.00169.

[43] Y. Wang, J. Wen, J. Xin, Y. Zhang, H. Xie, and Y. Tang, "3DCNN predicting brain age using diffusion tensor imaging," *Med. Biol. Eng. Comput.*, vol. 61, no. 12, pp. 3335–3344, 2023, doi: 10.1007/s11517-023-02915-x.

[44] G.-H. Jahng, S. Xu, M. W. Weiner, D. J. Meyerhoff, S. Park, and N. Schuff, "DTI studies in patients with Alzheimer's disease, mild cognitive impairment, or normal cognition with evaluation of the intrinsic background gradients," *Neuroradiology*, vol. 53, no. 10, pp. 749–762, Oct. 2011, doi: 10.1007/s00234-011-0845-3.

[45] S. J. Teipel et al., "Anatomical MRI and DTI in the diagnosis of Alzheimer's disease: A European multicenter study," *J. Alzheimer's Disease*, vol. 31, no. 3, pp. 33–47, Sep. 2012, doi: 10.3233/jad-2012-112118.

[46] Y. Xie, K. Cai, J. Dai, and G. Wei, "Enhanced integrity of white matter microstructure in mind-body practitioners: A whole-brain diffusion tensor imaging study," *Brain Sci.*, vol. 13, no. 4, p. 691, 2023. [Online]. Available: https://www.mdpi.com/2076-3425/13/4/691

[47] D. L. Sultzer et al., "Delusional thoughts and regional frontal/temporal cortex metabolism in Alzheimer's disease," *Amer. J. Psychiatry*, vol. 160, no. 2, pp. 341–349, Feb. 2003, doi: 10.1176/appi.ajp.160.2.341.

[48] A. Bakkour, J. C. Morris, D. A. Wolk, and B. C. Dickerson, "The effects of aging and Alzheimer's disease on cerebral cortical anatomy: Specificity and differential relationships with cognition," *NeuroImage*, vol. 76, pp. 332–344, Aug. 2013, doi: 10.1016/j.neuroimage.2013.02.059.

[49] Y. Sun et al., "Prediction of conversion from amnestic mild cognitive impairment to Alzheimer's disease based on the brain structural connectome," *Frontiers Neurol.*, vol. 9, p. 1178, Jan. 2019, doi: 10.3389/fneur.2018.01178.

[50] A. R. Hoy et al., "Microstructural white matter alterations in preclinical Alzheimer's disease detected using free water elimination diffusion tensor imaging," *PLoS ONE*, vol. 12, no. 3, Mar. 2017, Art. no. e0173982, doi: 10.1371/journal.pone.0173982.

[51] G. B. Frisoni, A. Prestia, P. E. Rasser, M. Bonetti, and P. M. Thompson, "In vivo mapping of incremental cortical atrophy from incipient to overt Alzheimer's disease," *J. Neurol.*, vol. 256, no. 6, pp. 916–924, Jun. 2009, doi: 10.1007/s00415-009-5040-7.

[52] E. Orta-Salazar, A. I. Feria-Velasco, and S. Díaz-Cintra, "Alteraciones en la corteza motora primaria en la enfermedad de Alzheimer: Estudio en el modelo 3xTg-AD," *Neurología*, vol. 34, no. 7, pp. 429–436, Sep. 2019, doi: 10.1016/j.nrl.2017.02.016.

[53] J. Golaz, C. Bouras, and P. R. Hof, "Motor cortex involvement in presenile dementia: Report of a case," *J. Geriatric Psychiatry Neurol.*, vol. 5, no. 2, pp. 85–92, Apr. 1992, doi: 10.1177/002383099200500205.

[54] N. T. Aggarwal, R. S. Wilson, T. L. Beck, J. L. Bienias, and D. A. Bennett, "Motor dysfunction in mild cognitive impairment and the risk of incident Alzheimer disease," *Arch. Neurol.*, vol. 63, no. 12, pp. 1763–1769, Dec. 2006, doi: 10.1001/archneur.63.12.1763.