A Multi-Stream Convolutional Neural Network for Classification of Progressive MCI in Alzheimer's Disease Using Structural MRI Images

Mona Ashtari-Majlan , Abbas Seifi , and Mohammad Mahdi Dehshibi

Abstract—Early diagnosis of Alzheimer's disease and its prodromal stage, also known as mild cognitive impairment (MCI), is critical since some patients with progressive MCI will develop the disease. We propose a multi-stream deep convolutional neural network fed with patch-based imaging data to classify stable MCI and progressive MCI. First, we compare MRI images of Alzheimer's disease with cognitively normal subjects to identify distinct anatomical landmarks using a multivariate statistical test. These landmarks are then used to extract patches that are fed into the proposed multi-stream convolutional neural network to classify MRI images. Next, we train the architecture in a separate scenario using samples from Alzheimer's disease images, which are anatomically similar to the progressive MCI ones and cognitively normal images to compensate for the lack of progressive MCI training data. Finally, we transfer the trained model weights to the proposed architecture in order to fine-tune the model using progressive MCI and stable MCI data. Experimental results on the ADNI-1 dataset indicate that our method outperforms existing methods for MCI classification, with an F1-score of 85.96%.

Index Terms—Alzheimer's disease, brain-shaped map, convolutional neural network, multivariate statistical test, transfer learning.

I. INTRODUCTION

ALZHEIMER'S disease (AD) is a progressive neurodegenerative disorder that is one of the leading causes of dementia in the elderly. According to [1], this disorder affects over 30 million people worldwide. Early diagnosis of this disease and its prodromal stage, also known as mild cognitive impairment (MCI), is crucial since 10% to 15% of MCI patients progress to AD, which is classified as progressive MCI (pMCI) [2].

As AD advances, several brain regions develop structural deformation and atrophy [3]. Structural Magnetic Resonance
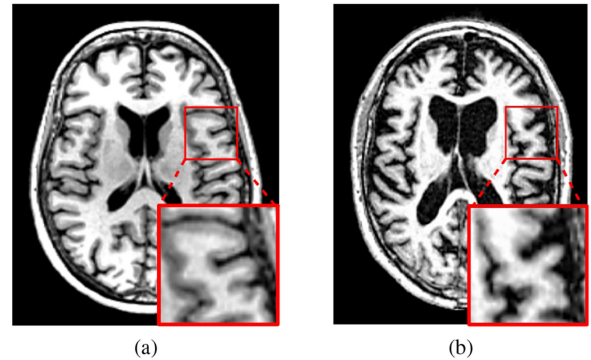
Fig. 1. MRI samples of (a) Cognitively Normal (CN) and (b) AD classes from Alzheimer's Disease Neuroimaging Initiative database (ADNI-1) [10]. In order to demonstrate the subtle brain atrophy, we highlighted the affected regions.

Imaging (sMRI) is one of the most widely employed neuroimaging tools for predicting this disorder through identifying brain atrophy [4] (see Fig. 1). In addition to sMRI (referred to as MRI in this paper), non-invasive biomarkers such as (1) demographic information (*e.g.*, age and education) [5], and (2) cognitive test scores [4] can also be used to provide possible discriminative information for diagnosing AD in the early stages. Several studies [4], [6]–[9] have addressed the MCI-to-AD conversion issue using neuroimaging methods in conjunction with the biomarkers.

Conventional methods in medical image processing typically use prior knowledge to segment brain images into Regions of Interest (ROI) [11] or Voxel-Based Morphometry (VBM) [12] to predict AD. While these methods can classify stable MCI (sMCI) and pMCI, the focus was mainly on Alzheimer's disease. Deep learning algorithms, on the other hand, have made it possible for researchers to combine feature extraction, dimensionality reduction, and classification in an end-to-end way. These algorithms also outperform conventional methods in identifying AD patterns because they can discover hidden representations among multiple regions of neuroimages. Hence, these models have gained prominence in analysing Alzheimer's disease.

In this paper, we propose a multi-stream convolutional neural network (CNN) for classifying sMCI and pMCI, which is fed with patch-based imaging data extracted using a novel data-driven technique. To accomplish so, we first use the multivariate T2 Hotelling test to compare MRI images of AD and cognitively normal (CN) individuals in order to identify distinct

anatomical landmarks. Following that, the statistical test is performed on textural and statistical features extracted from MRI images. These landmarks are then used to generate $19 \times 19 \times 19$ patches, with each landmark serving as the centre of the patch. Finally, the extracted patches are fed into the proposed multi-stream CNN to classify MRI images. To compensate for the lack of pMCI training data, we first train the proposed architecture using AD/CN images that are anatomically similar to the pMCI ones. Then, we transfer the weights of the trained model to the proposed architecture in order to fine-tune the model using pMCI and sMCI data. The contribution of this study is two-fold:

1) Rather than utilising non-rigid registration to identify anatomical landmarks in the brain, we propose employing rigid registration to reduce computational complexity and eliminate morphological structure deformations that could cause inherent errors in the classification step. Therefore, we partition each MRI image during the anatomical landmark detection phase and perform the T2 Hotelling test on these partitions to capture subtle anatomical differences, to account for more information, and to reduce the impact of potential errors caused by inter-subject brain shape variations. Finally, the identified statistically significant landmarks centres from the partitions are used as anchoring forces for selecting the patches fed into the proposed multi-stream CNN.

2) We propose using transfer learning to classify pMCI/sMCI classes in training the proposed architecture to overcome the complexity of learning caused by the subtle structural changes in MCI brains compared to AD and CN. In the Ablation study, we also demonstrate the importance of employing transfer learning. Furthermore, we address intrinsic errors caused by inter-subject brain shape differences by conducting experiments to determine an ideal image patch size in order to feed to the proposed CNN model.

The rest of this paper is organised as follows: Section II surveys the previous studies. Section III describes the proposed method. Experimental results are given in Section IV. Finally, the paper is concluded in Section V.

II. LITERATURE REVIEW

Mild cognitive impairment (MCI) is a stage of cognitive decline that occurs between the predicted cognitive loss associated with normal ageing and the more severe decline associated with dementia. MCI may result in a higher level of developing dementia caused by Alzheimer's disease if the anatomical changes in the brain are proactive. Progressive MCI differs from stable MCI in the progression of functional connectivity values over time. However, classifying pMCI and sMCI patients is challenging due to the subtle anatomical differences in the brain [13]. The four conventional feature extraction approaches usually mentioned in the literature for classifying pMCI and sMCI are voxel-based, slice-based, ROI-based, and patch-based [14], although they are not entirely mutually exclusive. In this section, before surveying recent advances in deep learning-based methods for classifying pMCI and sMCI, we will briefly review these four

approaches by discussing the advantages and disadvantages of each group.

The voxel-based techniques [15]–[17] use the voxel intensity values from all neuroimaging modalities. Although voxel-based techniques are simple to implement, they typically require spatial co-alignment of the input image to standard 3D space and suffer from high dimension feature space compared to available sample numbers. Ortiz *et al.* [18] used the *t*-test algorithm to partition the brain area into 3D patches in order to address the mentioned drawbacks and eliminate non-significant voxels. The patches were then used to train an ensemble of deep belief networks, and a voting scheme was used to make the final prediction. However, as mentioned by [19], there is an inherent over-fitting challenge with voxel-based techniques.

The sliced-based techniques [20], [21] extract slices from the 3D neuroimaging brain scan by projecting the sagittal, coronal, and axial to the 2D image slices. Indeed, because non-affected regions and normal slices must be chosen as the reference distribution, they cannot account for the disease and may be considered an anomaly [22]. Furthermore, choosing separate 2D slices may neglect the spatial dependencies of voxels in adjacent slices due to inter/intra anatomical variances in the brain images [14]. However, sliced-based techniques allow for the usage of a broader range of conventional and deep learning-based approaches. For instance, different pre-trained deep learning models on ImageNet, such as DenseNet, VGG16, GoogLeNet, and ResNet, can be fine-tuned by 2D slices to classify AD from CN [19]. In [21], researchers extracted features from MRI image slices using a pre-trained 2D CNN and fed the extracted feature sequence to a recurrent neural network (RNN). The RNN was in charge of determining the relationship between the sequence of extracted features corresponding to MRI image slices. However, sliced-based techniques are computationally expensive due to the use of additional learnable parameters which cannot directly benefit from transfer learning.

ROI-based techniques consider brain regions that have been predefined physically or functionally [8], [23], [24]. These methods use spatial information such as automated anatomical labelling [25] and diffusion-weighted imaging in MRI to extract features. The prominent regions that have been considered by almost all ROI-based feature extraction studies on AD prediction are the hippocampus, amygdala, and entorhinal. However, one of the advantages of employing ROI-based approaches is like a double-edged sword which can be a disadvantage because ROI identification requires expert human expertise. Furthermore, these techniques are considered time-consuming due to the need for non-linear registration and brain tissue segmentation. There is also the possibility of information loss because the abnormal region may spread from a single ROI to multiple ROIs.

Patched-based approaches [26], [27] partition the entire brain into multiple patches from which numerous feature vectors are extracted. The extracted patches include shapes, texture, and volume features generated from distinct brain regions or specific patterns. This computational approach eliminates the need for manual ROI identification, and makes it possible to use landmark-based patch extraction or other discriminative patch-based biomarkers [8]. However, selecting useful patches from a complete image is problematic, mainly due to increasing computational complexity when a non-rigid registration approach is

used. Researchers have used rigid registration or embedded the registration step into deep learning-based methods to address this challenge [28]. There is another issue with patched-based approaches related to multiple instance learning. Although this challenge is almost addressed in classifying AD and CN by leveraging patch relationships [29], [30], the difficulty in classifying pMCI in Alzheimer's disease is not still resolved. This issue is linked to the bag's labelling in classifying pMCI from sMCI, where a bag can get a negative label even though it contains informative anatomical landmark(s), but it cannot meet the majority rule.

Convolutional neural networks (CNNs) are a popular deep learning approach for classifying Alzheimer's disease. Islam and Zhang [31] proposed three 2D CNNs to generate three distinct MRI views. Each CNN in their architecture comprised three convolutional layers and four dense blocks, where the final decision was made by majority vote. Oh *et al.* [32] proposed a convolutional autoencoder-based approach for the AD and CN classification task, addressing the pMCI data limitation with transfer learning. Liu *et al.* [8] suggested a coarse-to-fine hierarchical ensemble learning method for simultaneous hippocampus segmentation and Alzheimer's disease classification that employed a multi-task deep CNN architecture and 3D densely connected CNNs. In this method, an MRI image is first divided into multiple slices, and then a pre-trained deep neural network is used to extract features from the slices. The coarse predictions were then used in ensemble learning to obtain refined results for all slices. Ebrahimi *et al.* [21] extracted a sequence of features from 2D MRI slices using a pre-trained ResNet18 [33], which they subsequently trained a temporal convolutional network and several types of RNNs to classify AD/CN. Zhao *et al.* [24] introduced a region ensemble model with three sequential sub-networks to account for a global feature map derived from the entire brain and regional feature maps extracted using a segmentation model. The feature representations were fused in their method, and the classification was performed using an attention-based technique. Researchers employed a data-driven technique to select informative patches in [34], which resulted in specific landmark localisation in brain MRI images. Each landmark patch was then fed into the CNN models, which produced the final classification result using the maximum voting strategy. Pölsterl *et al.* [35] proposed the dynamic affine feature map transform, an auxiliary module for CNNs that dynamically incites or represses each feature map of a convolutional layer based on both image and tabular biomarkers. A more detailed overview of deep learning algorithms for Alzheimer's disease classification can be found in [14], [36], [37].

III. PROPOSED METHOD

Given a dataset of N samples $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$, with $x_i \in \mathbb{R}^{d_x}$ and $y_i \in \mathbb{R}^{d_y}$, our goal is to train a multi-stream deep convolutional neural network $\mathcal{H}(x) = \mathbb{E}[\mathbf{Y}|\mathbf{X} = x]$ to classify sMCI and pMCI by minimising the cross-entropy between the class labels and the softmax output as in 1

$$p(y_i|x; w, b) = \frac{\exp(x^T w_i + b_i)}{\sum_{j \in d_y} \exp(x^T w_j + b_j)}. \qquad (1)$$

where w and b are the network's weights and bias terms, respectively. In this study, we use the baseline 1.5 T T1-weighted MRI images of subjects from the ADNI-1 dataset [10], where the input image $\mathbf{X}$ has a size of $d_x = 185 \times 155 \times 150$ and is labelled by $d_y = \{0, 1\}$. The output label $\mathbf{Y}$ consists of two probability values in the [0,1] range with $\mathcal{H}(x_i) = 0$ if the i-th sample belongs to the sMCI class and $\mathcal{H}(x_i) = 1$ otherwise.

It has been shown [38] that in the early stages of Alzheimer's disease, only certain brain regions are subject to morphological changes caused by the disease. Therefore, we conduct a statistical test to identify these informative landmarks in the MRI images and extract L patches from each MRI image $x_i = \{s_{i,j}\}_{j=1}^L$, with $s_{i,j} \in \mathbb{R}^{19 \times 19 \times 19}$. The proposed data-driven approach for extracting patches from the MRI image, on which a preprocessing step has been performed, is described in the following subsections, followed by the details of the multi-stream CNN. Fig. 2 shows the schematic of the proposed multi-stream architecture.

A. Preprocessing

We pre-process the MRI images to use them in the proposed method. There are four steps in the preprocessing phase: (1) anterior commissure-posterior commissure correction using the 3D Slicer software;[1] (2) intensity inhomogeneity correction using N4ITK [39], an enhanced version of nonparametric nonuniform normalisation; (3) skull stripping using a pre-trained U-Net[2] to remove both the skull and the dura; and (4) rigid registration, which involves linearly aligning MRI images to the Colin27 template and resampling them to a size of $155 \times 185 \times 150$ with a resolution of $1 \times 1 \times 1$ mm³. Fig. 3 shows a sample of MRI image from ADNI-1 dataset on which the preprocessing is performed.

B. Anatomical Landmark Detection

We must first identify the anatomical locations in the brain that are most influenced by the disease before we can classify sMCI and pMCI patients. As a result, we randomly divide samples from AD and CN individuals into the train, validation, and test sets with sizes of $0.7 \times N$, $0.1 \times N$, and $0.2 \times N$, respectively, where N is the total number of samples. Then, we select $M = 0.7 \times N$ (*i.e.*, the training set) and propose a novel data-driven landmark detection method in which MRI images are partitioned into $5 \times 5 \times 5$ patches. As mentioned in [40], [41], when a patient is diagnosed with Alzheimer's disease, several regions in the brain are subject to anatomical degeneration. At the pMCI stage, the same regions undergo anatomical changes, but the degeneration is not as severe as those seen at the onset of Alzheimer's disease. With respect to this fact, we use identical anatomical locations for classifying sMCI and pMCI patients.

Each partition is then represented by a 29-dimensional feature vector. This feature vector includes the Gray-Level Co-Occurrence Matrix (GLCM) [42], Structural Similarity Index Measure (SSIM) [43], Mean Square Error (MSE), entropy, and the mean and standard deviation of the partition voxels. To

---

[1][Online]. Available: http://www.slicer.org/
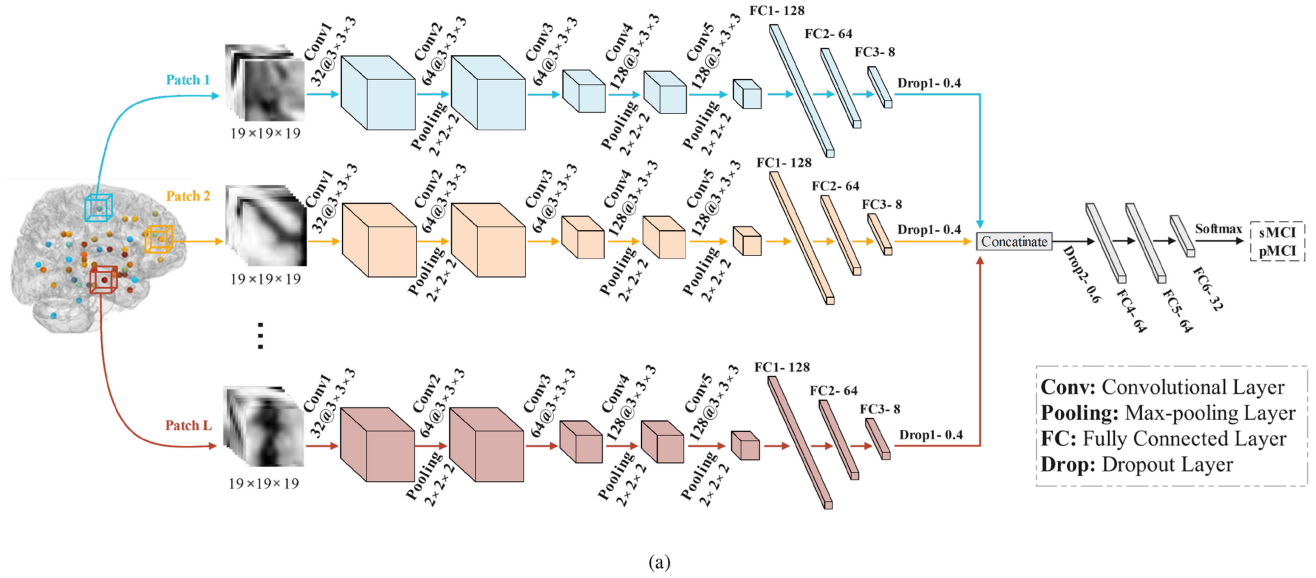[2][Online]. Available: https://github.com/iitzco/deepbrain

(a)

Fig. 2. The schematic of the proposed multi-stream convolutional neural network.
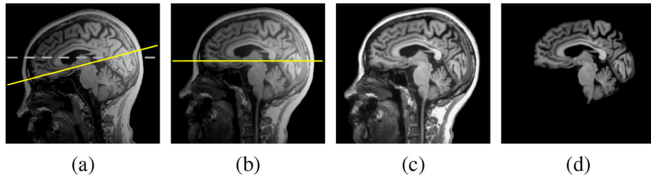

(a)   (b)   (c)   (d)

Fig. 3. The visual representation of the preprocessing steps for an MRI sample. (a) A raw MRI image, (b) the MRI image with anterior commissure-posterior commissure correction (c) the MRI image with intensity inhomogeneity correction, and (4) skull stripped MRI image. The yellow line in (a) and (b) depicts the anterior commissure-posterior commissure line.

Fig. 4. A brain-shaped *p-value* map in which top 50 landmark locations are represented by spheres in a gradient colour from red to blue, with *p-values* ranging from 0 to 0.001. Each *p-value* is paired with a landmark (l_x, l_y, l_z).

extract the GLCM elements of the feature vector, we generate six GLCM matrices with three adjacency directions, namely horizontal, vertical, and in-depth, each of which is associated with two distance values. Then, we extract contrast, correlation, homogeneity, and entropy, from each GLCM matrix resulting in a total of 24 elements. The Colin27 template [44] is used as the reference image for measuring SSIM and MSE at each patch location. Finally, we apply the multivariate T2 Hotelling statistical test [45] to generate a brain-shaped *p-value* map (see Fig. 4). Algorithm 1 details the steps of generating the brain-shaped *p-value* map.

After obtaining $\mathcal{P}$ set, we exclude landmarks with a spatial Euclidean distance of less than 15 to reduce the redundancy of overlapped adjacent patches to identify the most discriminative anatomical landmarks in the brain. The top 50 landmarks with the lowest *p-values* are then chosen (see Fig. 4).

The registration and landmark detection steps in Algorithm 1 are affected by the MRI image partitioning size, *i.e.*, $5 \times 5 \times 5$. In the case of selecting a smaller partition size, the lack of adequate morphological variations could lead to discarding informative landmarks. In contrast, larger partition sizes cause intrinsic physiological differences to eclipse subtle disease-related changes.
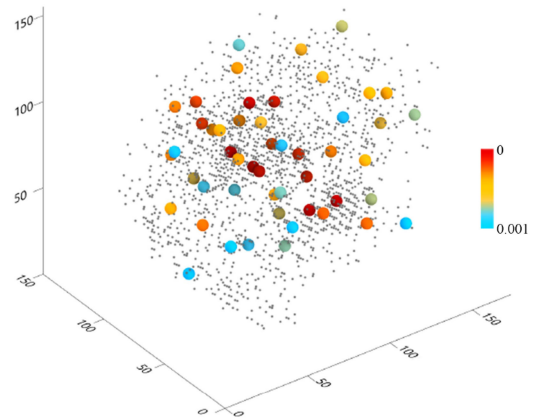
Therefore, after obtaining the top 50 landmarks, we sample 27 3D image patches with a $3 \times 3 \times 3$ displacement around the centre of each landmark to increase the size of patches to $19 \times 19 \times 19$ with two intentions: (i) compensating for regions that may unintentionally be discarded in anatomical landmark detection, and (ii) providing sufficient morphological structures for each stream of the proposed CNN to construct a discriminative latent space.

C. Multi-Stream Classifier Architecture

We propose a multi-stream CNN architecture with L streams, each fed with the patch $s_{i,j}$ extracted from the input image x_i and centred on the identified landmark location $(l_{x,j}, l_{y,j}, l_{z,j})$. We construct the patch $s_{i,j}$ with a size of $19 \times 19 \times 19$, surrounding the corresponding landmark location, in order to better represent

Algorithm 1: Generating Brain-Shaped *p-value* Map.

Input: $AD = \{x_1^{AD}, x_2^{AD}, \cdots, x_M^{AD}\}$
$CN = \{x_1^{CN}, x_2^{CN}, \cdots, x_M^{CN}\}$
$K = 34,410$, which is the total number of
patches with a size of $5 \times 5 \times 5$.
Output: $\mathcal{P}$: A set of *p-value*, forming the brain-shaped
p-value map.

Step 1: Partitioning
1　$\mathcal{V}^{AD} \leftarrow \texttt{Partition} AD$;
　// $\mathcal{V}^{AD} = \{p_{1,1}, \cdots, p_{1,K}, \cdots, p_{M,1}, \cdots, p_{M,K}\}$.
2　$\mathcal{V}^{CN} \leftarrow \texttt{Partition} CN$;
　// $\mathcal{V}^{CN} = \{q_{1,1}, \cdots, q_{1,K}, \cdots, q_{M,1}, \cdots, q_{M,K}\}$.
Step 2: Feature extraction & T2 Hotelling test;
3　**for** $j \leftarrow 1$ **to** K **do**
4　　**for** $i \leftarrow 1$ **to** M **do**
5　　　$f_{i,j}^{AD} \leftarrow \texttt{Feature-Extraction}\ p_{i,j}$;
6　　　$f_{i,j}^{CN} \leftarrow \texttt{Feature-Extraction}\ q_{i,j}$;
7　　**end**
8　　$p-value_j \leftarrow \texttt{Hotelling-Test}\ f_{M,j}^{AD}, f_{M,j}^{CN}$;
　　// $f_{M,j}$ is a $M \times 29$ matrix,
　　representing extracted features
　　form the jth patch.
9　**end**
10　$\mathcal{P} \leftarrow \texttt{Sort}\ p-value$, asc;
　// Each $p-value$ is paired with a
　landmark (l_x, l_y, l_z).
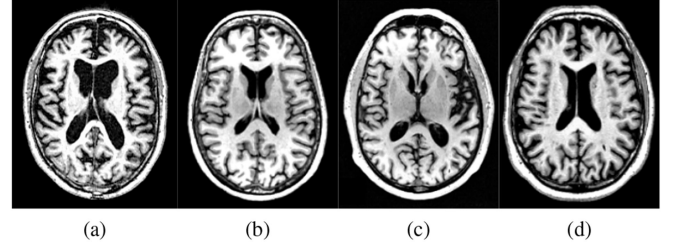11　**return** $\mathcal{P}$

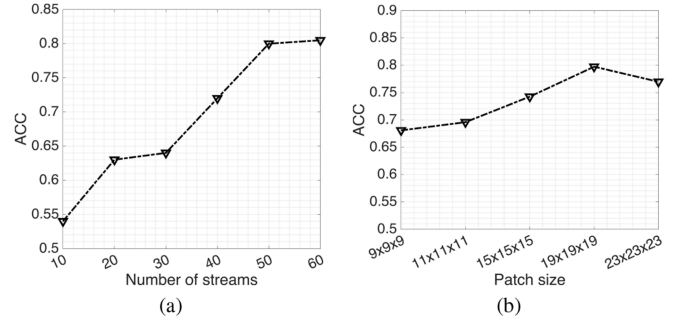Fig. 5.　Four samples from ADNI-1 dataset [10] (a) AD, (b) CN, (c) pMCI, and (d) sMCI.

Fig. 6.　The accuracy of the proposed multi-stream CNN classifier as a function of (a) the number of streams, and (b) the patch size.

morphological variations in the MRI images. The local spectral-spatial feature is extracted from each 3D image patch by each stream of the proposed CNN architecture.

As depicted in Fig. 2, the proposed multi-stream CNN has $L = 50$ streams, with an identical structure. Each stream has five convolutional layers (Conv), followed by a rectified linear unit (ReLU) activation function. The convolutional layers have 32, 64, 64, 128, and 128 $3 \times 3 \times 3$ convolution filters, respectively. After Conv2, Conv4, and Conv5, we consider batch normalisation and $2 \times 2 \times 2$ max-pooling layers. There are three fully connected layers with 128, 64, and 8 units at the end of each stream, which are followed by a dropout layer with a ratio of 0.4 to prevent overfitting. Although the architecture of all 50 streams is the same, their weights are tuned and updated separately, where the input patches for each stream are randomly selected to avoid the unintentional bias towards ordering streams. We concatenate the outputs of 50 streams and add a dropout layer with a ratio of 0.6 to fuse the locally extracted spectral-spatial features. Before passing the feature vector into the softmax function for the final classification, we add three fully connected layers with 64, 64, and 32 units, respectively.

IV. EXPERIMENTS

A. ADNI-1 Dataset

In this study, we use the baseline 1.5 T T1-weighted MRI images of subjects from the ADNI-1 dataset [10]. The volumetric 3D MPRAGE protocol is used to acquire sagittal T1-weighted MRI images with an in-plane spatial resolution of

1.25×1.25 mm^2 and 1.2 mm thick sagittal slices. The imaging dataset contains baseline images from 695 participants including 200 Alzheimer's disease, 231 cognitively normal, 164 progressive MCI, and 100 stable MCI. Fig. 5 shows four samples from this dataset, and Table I presents the demographic and clinical information of subjects in ADNI-1.

B. Architecture Details and Evaluation Metrics

The proposed architecture is implemented using Python based on the Keras package,[3] on a computer with Intel(R) Core(TM) i7-4790 K @4.00 GHz CPU and 16 G RAM. We trained the network using Adam optimiser [46] with the first momentum of 0.9 and the second momentum of 0.999. The initial learning rate and the constant for numerical stability are set to 10^{-3} and 10^{-6}, respectively. We set the maximum number of training epochs to 40 and used a mini-batch-size of 5 at each iteration, where the training data was shuffled before each training epoch. There are two other hyper-parameters which are the number of streams L and the patch size.

We evaluate the performance of the proposed architecture with the number of streams in the range $\{10, 20, 30, 40, 50, 60\}$ and the size of patches in the range $\{9, 11, 15, 19, 23\}$ using the validation set from the AD and CN classes. It is worth noting that because there are two hyperparameters, one must be fixed in order to generate the other. Fig. 6 shows the accuracy of the proposed multi-stream CNN as a function of these two hyper-parameters.

While increasing the number of streams improves classification accuracy (see Fig. 6(a)), it also increases computational complexity. We can also observe in Fig. 6(b) that a larger patch size (*i.e.*, $23 \times 23 \times 23$) shows a performance drop when

---

[3][Online]. Available: https://github.com/fchollet/keras

Class	Male/Female	Age	ADAS	CDR-sb	FAQ	MMSE	NPI-Q
AD	103/97	75.64 ± 7.71	13.02 ± 5.23	4.39 ± 1.6	13.16 ± 6.71	23.31 ± 2.03	3.44 ± 3.27
CN	119/112	76.155 ± 4.99	28.51 ± 4.89	0.03 ± 0.12	0.13 ± 0.59	29.09 ± 0.98	0.36 ± 0.95
sMCI	66/34	75.44 ± 7.27	22.93 ± 5.78	1.24 ± 0.62	1.65 ± 3.00	27.65 ± 1.70	1.45 ± 2.40
pMCI	97/67	74.54 ± 7.05	17.67 ± 5.14	1.87 ± 0.96	5.64 ± 5.15	26.62 ± 1.71	2.30 ± 3.11

Values are Reported as Mean $\pm$ Standard Deviation.
ADAS: Alzheimer's Disease Assessment Scale.
CDR-sb: Clinical Dementia Rating 'sum of boxes'.
FAQ: Functional Activities Questionnaire.
MMSE: Mini-Mental State Examination.
NPI-Q: Neuropsychiatric Inventory Questionnaire.

compared with the $19 \times 19 \times 19$ patch size. This is owing to the fact that enlarging the patch size includes tissues in that patch of the brain whose intrinsic morphological similarity is dominant to subtle changes that aid in the early diagnosis of the disease. Therefore, instead of discriminating the subtle changes induced by the disease's development in the early stages, the network discriminates irrelevant tissues. As a result, to establish a trade-off between computational complexity and accuracy, we set the number of streams L to 50 and the patch size to $19 \times 19 \times 19$ and preserve these values in the rest of the experiments.

To evaluate the performance of the proposed architecture, we use the metrics as in 2.

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{SEN} = \frac{TP}{TP + FN},$$

$$\text{SPE} = \frac{TN}{TN + FP},$$

$$\text{F1} - \text{score} = \frac{2 \times \text{SEN} \times \text{SPE}}{\text{SEN} + \text{SPE}}. \tag{2}$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively. We also report area under receiver operating characteristic curve (AUC).

C. Experimental Results

To assess the performance of the proposed architecture, we performed three steps: (1) training the proposed multi-stream CNN with MRI images of AD and CN subjects, (2) transferring weights from the trained model in Step 1 to the identical architecture and fine-tuning it with the data related to sMCI and pMCI patients, and (3) adding biomarkers as an auxiliary modality to examine the reciprocal influence of spectral-spatial features and biomarkers.

1) Classification of AD and CN subjects: To train the proposed multi-stream CNN, we randomly select 70% of the MRI patches from the two classes of AD and CN as the training set, 20% as the test set, and 10% as the validation set. We also compare the trained model in this step with 10 additional approaches, including regions of interest (ROI) [47], Voxel-Based Morphometry (VBM) [12], and five deep learning-based methods. We used the FAST algorithm [48] to segment brain MRI images into three different tissues for the ROI and VBM comparison: White Matter

(WM), Gray Matter (GM), and Cerebrospinal Fluid (CSF). We followed the implementations described by the researchers for the deep learning-based approaches.

For the RIO-based method, we align the anatomical automated labelling template [25] with the native space of each MRI image. This template contains 116 predefined regions in which the extracted GM tissue volumes are normalised using the summation of GM, WM, and CSF volumes. This normalised tissue volume is used as the representative feature for the MRI images. For the VBM method, we use affine registration with 12 degrees of freedom to align MRI images with the Colin27 template [44] in order to extract the GM density as the representative feature. To reduce the dimensionality of this feature vector, we perform the *t*-student statistical test on the extracted features from AD and CN subjects and chose GM densities with *p-values* less than 0.001.

Finally, we classify AD and CN using these two feature representations using a linear support vector machine (SVM) [50] with soft-margin and a multilayer perceptron (MLP). The MLP comprises two hidden layers, with 13 and 15 neurons, and a binary output layer with a logistic activation function. We train these classifiers using the 5-fold cross-validation strategy. To ensure a fair comparison for MCI conversion prediction, we train the VBM and ROI models on AD/CN and test them on an independent test set of pMCI/sMCI.

2) Classifying sMCI and pMCI using transfer learning: The availability of MRI images for sMCI and pMCI patients is less than that of AD because the symptoms of Alzheimer's disease are not as severe at these stages. As a result, many patients may not be asked to get an MRI test. Furthermore, structural changes in MCI brains caused by dementia may be very subtle compared to CN and AD, making the convergence of the proposed multi-stream CNN challenging. To overcome these limitations, we first trained our model using data from CN and AD MRI images and fine-tune the model with data gathered from sMCI and pMCI patients.

Since the pre-trained streams of the proposed architecture are designed to extract structural features that correspond to AD, we freeze the initial layers and only re-train the last three fully connected layers. To fine-tune the model, we used 70% of the pMCI and sMCI MRI patches and tested the fine-tuned model with the remaining data.

3) Unleashing the impact of biomarkers: In addition to MRI images, non-invasive biomarkers such as demographic information and cognitive test scores [4] can also be used to provide potentially discriminatory information for diagnosing

| Method	AD vs. CN (%)					pMCI vs. sMCI (%)				
	ACC	SEN	SPE	F1-score	AUC	ACC	SEN	SPE	F1-score	AUC
ROI + SVM	70.30	68.51	67.50	67.83	79.39	59.47	66.86	68.90	67.87	60.99
VBM + SVM	82.84	82.54	80.50	81.34	90.39	70.08	81.48	67.07	73.58	74.38
ROI + MLP	73.08	70.00	71.19	70.59	77.52	58.75	70.00	66.04	67.96	53.93
VBM + MLP	83.85	78.33	85.45	81.74	89.92	70.00	72.00	78.26	75.00	73.93
Shmulev et al. [49]	-	-	-	-	-	76.00	70.00	88.00	77.97	86.00
DM^{2}L [5]	91.09	93.50	88.05	90.69	95.86	76.90	82.43	42.11	55.74	77.64
H-FCN [6]	90.30	**96.50**	82.40	85.08	95.10	80.90	**85.40**	52.60	65.10	78.10
HybNet [28]	91.90	82.40	94.50	91.50	96.50	82.70	57.90	86.60	69.39	79.30
Zhao et al. [24]	-	-	-	-	-	**85.90**	50.00	91.60	64.68	85.40
Proposed architecture $\oplus$ biomarkers	97.54	95.54	99.40	97.43	99.38	69.21	68.56	93.15	78.98	77.35
Proposed architecture	**97.78**	95.59	**99.82**	**97.66**	**99.97**	79.90	75.55	**99.70**	**85.96**	**94.39**
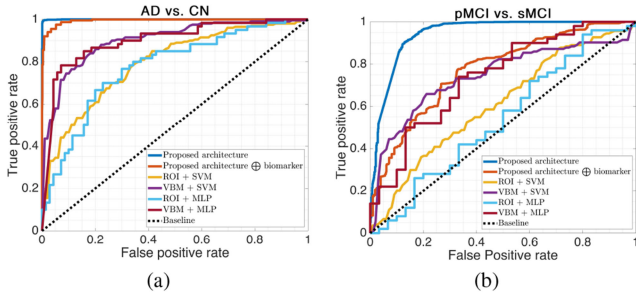
Fig. 7. ROC curves of the proposed multi-stream architecture, the proposed architecture integrated with biomarkers, ROI, and VBM-based methods for (a) AD/CN classification, and (b) pMCI/sMCI classification.

Alzheimer's disease in its early stages. We propose adding biomarkers as an auxiliary modality to the architecture in order to evaluate the reciprocal influence of spectral-spatial features and biomarkers. Therefore, we introduce an additional input stream to the proposed architecture to incorporate numerical values of {Age, ADAS, CDR-sb, FAQ, MMSE, NPI-Q} (see Table I). However, a subset of the MRI images in the ADNI-1 dataset does not include biomarkers, resulting in missing values when incorporating them into the proposed CNN architecture. We use k-Nearest Neighbour (kNN) as a preprocessing step to handle missing values. First, we execute the kNN with $k = 6$ (analogous to 6 biomarkers) on the available data. Then, for those entries that do not have a biomarker, we fill in the value with the average values of the k nearest neighbours. Finally, we incorporate these biomarkers into the architecture at the 'Concatenate' layer while preserving the learning parameters as described in Section IV-B.

Table II shows the performance of the proposed architectures compared with 10 other approaches. As shown in Table II, the proposed multi-stream CNNs achieve considerably better F1-scores than both conventional and deep learning-based approaches. While the biomarkers can improve the classification performance, the lack of large-scale data is a barrier to incorporating them into the proposed method. Furthermore, as previously stated, the cognitive test scores of individuals suffering from various types of MCI differ slightly. This subtle difference could explain why the multi-stream architecture combined with biomarkers performs poorly in distinguishing pMCI and sMCI when compared with the AD and CN classification. We also plot the ROC curves in Fig. 7 for classifying both AD vs. CN and pMCI vs. sMCI.

In the case of classifying AD and CN, Fig. 7(a) shows that the proposed architecture achieves an AUC of 99.97%, outperforming ROI, VBM, and multi-stream architecture integrated with biomarkers, which achieve AUCs of 79.39%, 90.39%, and 99.38%, respectively. Furthermore, as demonstrated in Fig. 7(b), when we use transfer learning, the fine-tuned model can provide appropriate discriminative information to classify pMCI and sMCI subjects. This evidence can also convey that AD and pMCI have structural deformations similar to the extracted features from various landmarks in different streams of the proposed architecture.

The results in Table II show that the proposed multi-stream architecture combined with biomarkers is slightly less efficient than the multi-stream CNN in distinguishing pMCI from sMCI when compared with the AD and CN classification. To highlight these subtle differences and better comprehend the reciprocal influence of spectral-spatial features and biomarkers, we have visualised the class-discriminative localisation map and the latent space in Fig. 8 using Grad-CAM [51] and t-SNE [52], respectively. Fig. 8(a) shows an example of a class-discriminative localisation map for the last convolutional layers in the proposed multi-stream CNN in which the salient regions of each brain patch have been identified so that the classifier could distinguish between AD/CN and pMCI/sMCI with high potential. Furthermore, we have observed in Fig. 8(b) that introducing biomarkers as an auxiliary modality causes the clusters (i.e., AD/CN and pMCI/sMCI) to become tighter with a degree of overlap than without the biomarkers. This evidence can explain why the proposed multi-stream CNN without biomarkers performs marginally better, where the latent space provides for clean cluster separation.

D. Ablation Study

In the ablation study, we conduct two sets of experiments to better understand the efficacy of the proposed multi-stream CNN for the classification of pMCI in Alzheimer's disease. We therefore examine:

— *Contribution of transfer learning:* In this experiment, instead of using transfer learning, we test the baseline performance when the model is directly trained on pMCI/sMCI data. We have 100 and 164 sMCI and pMCI samples, respectively, and use the same data split as in the transfer learning (70/30) for a fair comparison. The other learning parameters are preserved as described in Section IV-B. The evaluation metrics (i.e., ACC:
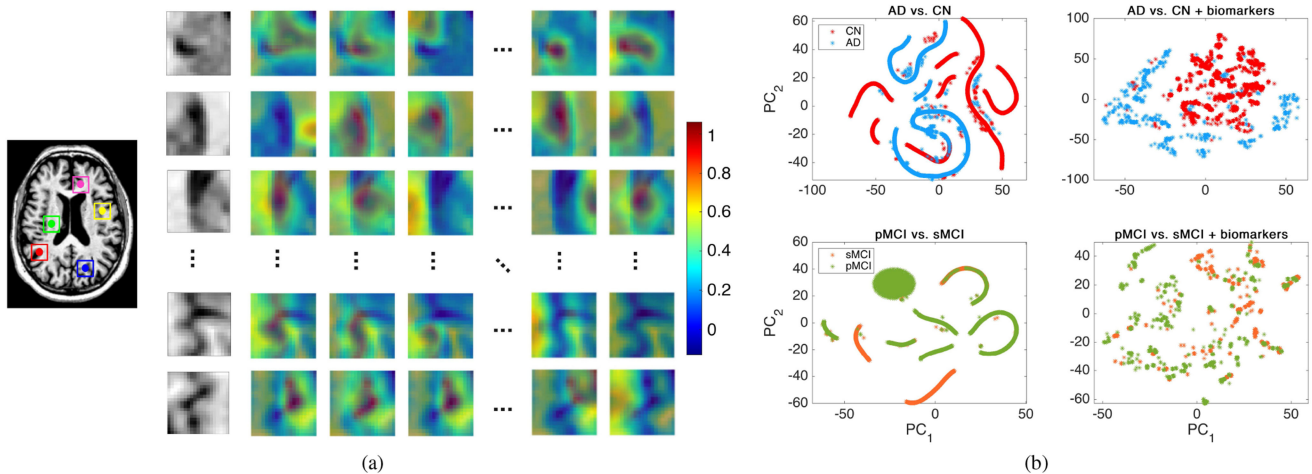
Fig. 8. (a) An example of a class-discriminative localisation map for the last convolutional layers in the proposed multi-stream CNN using Grad-CAM [51]. Each stream has 128 filters. However, in order to keep the publication page limits, we only illustrate five out of 50 patches, each representing five filters. The colourbar represents the relevance of each pixel in the Grad-CAM, where the red colour has the highest importance in the trained network's final decision and the blue colour has the least importance. (b) Visualisation of the feature space using t-SNE [52] for the classification of AD/CN and pMCI/sMCI with and without biomarkers.

63.53%, SEN: 100%, SPE: 0.0%, and F1-score: 77.70%) reveal that the model cannot converge to classify pMCI/sMCI classes without using transfer learning. In fact, the trained model predicts the same optimal output regardless of the input, with the average answer minimising loss.

— *Contribution of multi-stream architecture:* In this experiment, instead of the proposed multi-stream CNN, we use a single stream CNN to which 3D patches are fed. The model is trained on the AD/CN dataset and fine-tuned using pMCI/sMCI data with the same data split used in the previous experiments. The evaluation metrics (*i.e.*, ACC: 56.96%, SEN: 67.35%, SPE: 40.00%, and F1-score: 66.00%) show that the model is unable to converge to classify pMCI/sMCI classes, and the trained model is predicting random class for all the data points regardless of the input. These metrics highlight two main facts: (1) with the multi-stream architecture, the model can build a more discriminative latent space as a result of focusing on the landmarks identified in the anatomical landmark detection phase, which are more likely to be signs of Alzheimer's disease. However, with a single-stream architecture, the model cannot distinguish between anatomical abnormalities caused by the disease and those caused by the morphology of the brain. Indeed, the model's sensitivity and specificity are inadvertently changed due to the morphological similarity of the inputs; (2) input patches to the single-stream architecture predominantly capture local information of the image, and the global relationship between different landmark locations is no longer taken into account.

V. CONCLUSION

Alzheimer's disease is the most common type of dementia, resulting in memory impairment and cognitive decline. Mild cognitive impairment is a prodromal stage of AD, also known as the transition stage. MCI patients either progress to AD or remain at the same stage over time. Therefore, it is critical to distinguish between progressive MCI and stable MCI in early stages to prevent rapid progression of the disease.

In this study, we have proposed a method for classifying pMCI and sMCI patients using MRI images. The proposed method consists of two main steps. (1) We have developed a novel data-driven approach based on the multivariate T2 Hotelling statistical test to identify anatomical landmarks in MRI images and generate a brain-shaped *p-value* map. Each landmark is paired with a 3D coordinate, allowing us to extract patches of $19 \times 19 \times 19$. (2) We have proposed a multi-stream deep convolutional neural network in which each stream is fed by one of the patches. This multi-stream CNN employed transfer learning to classify pMCI and sMCI patients using the ADNI-1 dataset. We assessed the proposed method in three experimental steps and demonstrated the significance of our contributions to transfer learning and multi-stream architecture in the Ablation study. We have performed several experiments to evaluate the efficiency of the proposed architecture based on the best practices. Experimental results have shown that our method outperformed existing approaches, particularly in the classification of MCI patients. Thus, this method can assist practitioners to expand investigating on various diseases associated with structural atrophy.

REFERENCES

[1] D. E. Barnes and K. Yaffe, "The projected effect of risk factor reduction on Alzheimer's disease prevalence," *Lancet Neurol.*, vol. 10, no. 9, pp. 819–828, 2011.

[2] R. C. Petersen *et al.*, "Mild cognitive impairment: Ten years later," *Arch. Neurol.*, vol. 66, no. 12, pp. 1447–1455, 2009.

[3] M. A. DeTure and D. W. Dickson, "The neuropathological diagnosis of Alzheimer's disease," *Mol. Neurodegener.*, vol. 14, no. 1, pp. 1–18, 2019.

[4] K. Hett, V.-T. Ta, I. Oguz, J. V. Manjón, and P. Coupé, "Multi-scale graph-based grading for Alzheimer's disease prediction," *Med. Image Anal.*, vol. 67, pp. 101850:1–101850:13, 2021.

[5] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1195–1206, May 2019.

[6] 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.

[7] Q. Li *et al.*, "Novel iterative attention focusing strategy for joint pathology localization and prediction of MCI progression," in *Med. Image Comput. Comput. Assist. Interv. - MICCAI 2019*. Berlin, Germany: Springer, 2019, pp. 307–315.

[8] M. Liu *et al.*, "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease," *NeuroImage*, vol. 208, pp. 116459:1–116459:15, 2020.

[9] S. Qiu *et al.*, "Development and validation of an interpretable deep learning framework for Alzheimer's disease classification," *Brain*, vol. 143, no. 6, pp. 1920–1933, 2020.

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

[11] C. Cabral, P. M. Morgado, D. C. Costa, and M. Silveira, "Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages," *Comput. Biol. Med.*, vol. 58, pp. 101–109, 2015.

[12] J. Ashburner and K. J. Friston, "Voxel-based morphometry-the methods," *Neuroimage*, vol. 11, no. 6, pp. 805–821, 2000.

[13] U. R. Acharya *et al.*, "Automated detection of Alzheimer's disease using brain MRI images- A study with various feature extraction techniques," *J. Med. Syst.*, vol. 43, no. 9, pp. 1–14, 2019.

[14] X. Zhao, C. K. E. Ang, U. R. Acharya, and K. H. Cheong, "Application of artificial intelligence techniques for the detection of Alzheimer's disease using structural MRI images," *Biocybernetics Biomed. Eng.*, vol. 41, no. 2, pp. 456–473, 2021.

[15] S. Basheera and M. S. S. Ram, "A novel CNN based Alzheimer's disease classification using hybrid enhanced ICA segmented gray matter of MRI," *Computerized Med. Imag. Graph.*, vol. 81, pp. 101713:1–101713:15, 2020.

[16] J. Ruiz, M. Mahmud, M. Modassir, and M. S. Kaiser, "3D DenseNet ensemble in 4-way classification of Alzheimer's disease," in *Proc. 13th Int. Conf. Brain Informat.*, Springer, 2020, pp. 85–96.

[17] W. Li, X. Lin, and X. Chen, "Detecting Alzheimer's disease based on 4D FMRI: An exploration under deep learning framework," *Neurocomputing*, vol. 388, pp. 280–287, 2020.

[18] A. Ortiz, J. Munilla, J. M. Górriz, and J. Ramírez, "Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease," *Int. J. Neural Syst.*, vol. 26, no. 7, 2016, Art. no. 1650025.

[19] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review," *Comput. Methods Program. Biomed.*, vol. 187, 2020, Art. no. 105242.

[20] Y. Zhang, Q. Teng, Y. Liu, Y. Liu, and X. He, "Diagnosis of Alzheimer's disease based on regional attention with SMRI gray matter slices," *J. Neurosci. Methods*, vol. 365, pp. 109376:1–109376:8, 2022.

[21] C. R. A. D. N. I. Ebrahimi A and S. Luo, "Deep sequence modelling for Alzheimer's disease detection using MRI," *Comput. Biol. Med.*, vol. 134, 2021, Art. no. 104537.

[22] R. Dey and Y. Hong, "ASC-Net: Adversarial-based selective network for unsupervised anomaly segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2021, pp. 480–489.

[23] V. S. Rallabandi *et al.*, "Automatic classification of cognitively normal, mild cognitive impairment and Alzheimer's disease using structural MRI analysis," *Informat. Med. Unlocked*, vol. 18, pp. 100305:1–100305:7, 2020.

[24] Y.-X. Zhao, Y.-M. Zhang, M. Song, and C.-L. Liu, "Region ensemble network for MCI conversion prediction with a relation regularized loss," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2021, pp. 185–194.

[25] N. Tzourio-Mazoyer *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.

[26] M. Liu, J. Zhang, D. Nie, P.-T. Yap, and D. Shen, "Anatomical landmark based deep feature representation for MR images in brain disease diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1476–1485, Sep. 2018.

[27] J. Wen *et al.*, "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Med. Image Anal.*, vol. 63, pp. 101694:1–101694:19, 2020.

[28] C. Lian, M. Liu, Y. Pan, and D. Shen, "Attention-guided hybrid network for dementia diagnosis with structural MR images," *IEEE Trans. Cybern.*, pp. 1–12, 2020, doi: 10.1109/TCYB.2020.3005859.

[29] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, 2018.

[30] M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 2127–2136.

[31] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Informat.*, vol. 5, no. 2, pp. 1–14, 2018.

[32] K. Oh, Y.-C. Chung, K. W. Kim, W.-S. Kim, and I.-S. Oh, "Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning," *Sci. Rep.*, vol. 9, no. 1, pp. 1–16, 2019.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[34] M. Liu, C. Lian, and D. Shen, "Anatomical-landmark-based deep learning for Alzheimer's disease diagnosis with structural magnetic resonance imaging," in *Deep Learning in Healthcare: Paradigms and Applications*. Berlin, Germany: Springer, 2020, pp. 127–147.

[35] S. Pölsterl, T. N. Wolf, and C. Wachinger, "Combining 3D image and tabular data via the dynamic affine feature map transform," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2021, pp. 688–698.

[36] N. Goenka and S. Tiwari, "Deep learning for Alzheimer prediction using brain biomarkers," *Artif. Intell. Rev.*, vol. 54, no. 7, pp. 1–45, 2021.

[37] G. Chamarajan *et al.*, "Alzheimer's disease: A survey," *Int. J. Artif. Intell.*, vol. 8, no. 1, pp. 33–39, 2021.

[38] H. Braak, E. Braak, D. Yilmazer, R. A. I. De Vos, E. N. H. Jansen, and J. Bohl, "Pattern of brain destruction in Parkinson's and Alzheimer's diseases," *J. Neural Transmiss.*, vol. 103, no. 4, pp. 455–490, 1996.

[39] N. J. Tustison *et al.*, "N4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.

[40] Y. Fan, N. Batmanghelich, C. M. Clark, and C. Davatzikos, "Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline," *NeuroImage*, vol. 39, no. 4, pp. 1731–1743, 2008.

[41] M. L. Schroeter, T. Stein, N. Maslowski, and J. Neumann, "Neural correlates of Alzheimer's disease and mild cognitive impairment: A systematic and quantitative meta-analysis involving 1351 patients," *NeuroImage*, vol. 47, no. 4, pp. 1196–1206, 2009.

[42] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[43] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[44] C. J. Holmes, R. Hoge, L. Collins, R. Woods, A. W. Toga, and A. C. Evans, "Enhancement of MR images using registration for signal averaging," *J. Comput. Assist. Tomography*, vol. 22, no. 2, pp. 324–333, 1998.

[45] H. Hotelling, *The Generalization of Student's Ratio*. New York, NY, USA: Springer, 1992, pp. 54–65.

[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, 2015, pp. 1–15.

[47] D. Zhang *et al.*, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.

[48] S. M. Smith *et al.*, "Advances in functional and structural MR image analysis and implementation as FSL," *NeuroImage*, vol. 23, pp. S208–S219, 2004.

[49] Y. Shmulev *et al.*, "Predicting conversion of mild cognitive impairments to Alzheimer's disease and exploring impact of neuroimaging," in *Graphs in Biomedical Image Analysis and Integrating Medical Imaging and Non-Imaging Modalities*. Berlin, Germany: Springer, 2018, pp. 83–91.

[50] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[51] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

[52] G. Hinton and S. Roweis, "Stochastic neighbor embedding," in *Proc. 15th Int. Conf. Neural Inf. Process. Syst.*, Cambridge, MA, USA, MIT Press, 2002, pp. 857–864.