Exploiting Discriminative Regions of Brain Slices based on 2D CNNs for Alzheimer’s Disease Classification

FUJIA REN\textsuperscript{1,4}, CHENHUI YANG\textsuperscript{1}, QI QIU\textsuperscript{1}, NIANYIN ZENG\textsuperscript{2}, CHUNTING CAI\textsuperscript{1}, CHAOQUN HOU\textsuperscript{1} AND QUAN ZOU\textsuperscript{3}

\textsuperscript{1}Computer Science Department, Xiamen University, Xiamen, China
\textsuperscript{2}Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, China
\textsuperscript{3}School of Big Data and Computer Science, Guizhou Normal University, Guiyang, China

Corresponding author: Quan Zou (zouquan@nclab.net) and Nianyin Zeng (zny@xmu.edu.cn).

The work was supported by the National Key R&D Program of China (SQ2018YFC090002), the Natural Science Foundation of China (No. 61771331), the International Science and Technology Cooperation Project of Fujian Province of China(No. 2019I0003), and the Fundamental Research Funds for the Central Universities (No.20720190009)

ABSTRACT Convolutional Neural Networks (CNNs)-based classifiers improve the accuracy of diagnosis and prediction for Alzheimer’s Disease (AD). However, exploiting specific brain regions with AD is essential to understand pathological alteration in AD and monitor its progression. This study aims to construct novel AD classification models which have good performances and interpretation on AD diagnosis. We propose three classifiers including a simple broaden plain CNNs (SBPCNNs), a major slice-assemble CNNs (SACNNs) and a multi-slice CNNs (MSCNNs), which record the slice positions but have fewer parameters. Specifically, we integrate the ranking and the random forest methods to find the discriminative region that is consistent with domain knowledge about AD. The results of the visualization explanation of pixel and slice level deliver a clearer understanding of AD to specialists. The experimental results indicate that the proposed models are meaningful for AD classification.

INDEX TERMS Alzheimer’s Disease, CNNs-based Classification, Structural Magnetic Resonance Imaging (sMRI), Visual Explanation

I. INTRODUCTION

When Alzheimer’s disease (AD) presents a stern strategic challenge around the world\cite{1}, computer-aided diagnosis (CAD) has improved AD diagnosis and prediction performance. AD, being incurable, fatal dementia, bring significant pain and cost to patients and their families\cite{2}. Thus, the accurate diagnosis of AD is important to postpone the disease progression and improve the quality of life of people with AD\cite{3}. With the development of computer vision, machine learning, and deep learning techniques, novel models using medical images promote AD diagnostic accuracy\cite{4,5}. However, several difficulties constrain the models’ performance.

Main challenges of data-driven models of AD-related tasks are limited data, multiple dimensions, and sophisticated pre-processing of medical images. Due to the small training dataset, many automatic models, especially for those using pixels or voxels of the images, apply classic augmentation techniques such as rotation and patching from slices. Furthermore, decomposing 3D subjects into 2D slices is another common trick to get more training dataset. To reduce the complexity and large dimensions of brain image, some previous works\cite{6} focus on using local features based on AD domain knowledge such as the hippocampus, frontal pole, and shell core, limit in need for prior knowledge and complex prepossessing. Some researchers extract the region-of-interest (ROI) as features from the available images and effectively reduce the image dimension, while the partial details might be lost. Auto-encoder is widely applied for AD tasks, but it cost more time because it is the two-stage method\cite{7,8}. Summarily, the art of state techniques takes raw features from an extensive, annotated data set, such as a collection of images or genomes\cite{9}–\cite{11}, and use them to create a predictive tool based on patterns buried inside. However, the partial methods might disturb or lost original location information which could interpret the disease.
sMRI Preprocessing
GM, WM, CSF segmentation
Spatial Normalization

FIGURE 1: Pipeline of the proposed approach. Note: (a) Simple broaden plain CNNs; (b) Assembly CNNs; (c) Multi-slice CNNs; AD Alzheimer’s Disease; NC normal cohort; GM - grey matter; CSF cerebrospinal fluid.

Recently, numerous studies investigated improving the performance of classifiers, but few studies focused on inspecting pathological changes in computer-aided diagnosis models [12]–[16], especially for Alzheimer’s diagnosis tasks. For example, the discriminable brain region of FDG-PET has been inspected in the machine learning models [17]. In the study, voxels of FDG-PET were firstly part into nonoverlapping groups. Next, they identified essential regions related to AD in group level by important score, which is calculated by the random forest (RF) model. When used CNNs-based models, the 2D-CAM and 2D-Grad-CAM have been revised for 3D CNNs on structural magnetic resonance image (sMRI) [18]. By the approach, the essential region of the brain involving AD was visualized. Recent work uses active maps to determine brain region information, but their solution focuses on specific 60 sagittal slices [13]. As was mentioned in [13], [18], above methods decrease diagnostic accuracy, because they modified and retrained classifiers during the visualization stage. As a result, the ability to interpret trained models would be affected, which might prevent specialists from understanding the brain structure variations in AD progression.

This paper aims to study the informative regions of the brain between AD and normal cohort (NC). We proposed three 2D CNNs-based classifiers, including a simple, broaden plain CNNs (SBPCNNs) using a specific slice, a simple assemble CNNs (SACNNs) and multi-slice CNNs (MSCNNs) using slices combination. Our methods are sketched in Fig 2. Our methods monitor relevant region details in the brain without disturbing brain slice location. Thus, we can systematically explore the corresponding brain regions at the pixel and slice levels by reliable ways. This study makes the following three contributions: (1) We conducted three novel models of monitoring slices location with fewer parameters, which obtained more than 90% performance of AD vs. NC task. (2) We proved that a slice of a specific location or the several slices combination of sMRI could diagnose AD thoroughly. Moreover, we found precise discriminative regions of pixel and slice levels. (3) We use the saliency maps to visualize the slice information of the models of excellent performance. Consequently, we give new clues for understanding abnormal brain region related to AD.
The remainder of this paper is organized as follows. Section II describes the proposed frameworks and analysis methods; Section III reports experiments and presents the results of AD classification and slice importance; Section IV explains the visualization process; Section V concludes this paper and discusses avenues for future research.

II. METHODOLOGY
We focus on CNN-based classifiers and explanation for sMRI changing in grey matter (GM) for AD tasks. GM is chosen, because it is a reliable and classic marker for AD diagnoses, such as its’ local and whole atrophy, and the density changes. In this section, we firstly introduce proposed models, namely simple broaden plain CNNs (SBPCNNs), major slice-assemble CNNs (SACNNs) and multi-slice CNNs (MSCNNs). Next, we present how to determine slices importance. Finally, we describe the visual explanation technique for the models.

A. MAIN NETWORK

1) Model Construction
To promise the diagnostic accuracy, we evaluate prevalent neural networks by AD dataset. The advanced CNNs generally separates into two groups. The first group of CNNs (e.g., AlexNet, LeNet-5, VGG, GoogleNet, etc.) has shallow, sequentially connected layers with bigger receptions. The second group (e.g., Resnet, DenseNet [19], etc.) has deeper layers, and skip connections between blocks with smaller filter kernels. Comparing the two architectures’ advantages, the first group runs faster due to the more straightforward structure, while the second group overcomes the vanishing gradient problem of deeper layers through their particular structure. In practice, the general rule of CNNs is that the framework with deeper layers performs better than those with the broader layer. However, we conduct subnetworks from the two types of CNNs, because the used dataset’s characteristics, which describes in Section III.

Considering the risks of overfitting from the small scale of the dataset, we simplify models and apply regularization techniques. Fig. 2 presents three revised networks. For the first type of CNNs, we experimentally convolutionalize all full connected layer of classic AlexNet. Therefore, the network is flexible to the input size. Additionally, we insert the normal batch layer between convolution layers for regularization reason. Since the used slices on a specific location have similar shapes and less noise, we simultaneously explore a simple broaden plain CNNs (SBPCNNs) which have bigger receptive filed and fewer layers. The second type of CNNs, we experimentally convolutionalize all of the dataset, we simplify models and apply regularization characteristics, which describes in Section III.

In the forward stage of the adapted CNNs extracts features from slices, then generate the disease probabilities through the final classification layers. In the backward stage, we use a cross-entropy loss function to measure the distance between the ground-truth labels and the prediction results. Thus, the network will learn the parameter values through an optimization process. The loss function is defined as $l = -\sum c_i \log (f(x_i, w))$, where $c_i$ denotes the real label.

2) Architecture of SACNNs
After getting the results of the pre-trained models on every slice, we combine top K classifiers outputs for the final result. Fig. 3 decipher MSCNNs architecture. According to the ranking in terms of accuracy, we get the results of the final diagnosis of the test stage:

$$y_{pre} = \sum_{i=0}^{k} y_i \times p_i$$

where $p_i$ is the probability of prediction for a given slice $i$, and $k$ denotes the number of selected pre-trained classifier. $y_{pre}$ value denotes final results.

3) Architecture of MSCNNs.

The first method explores the contribution of non-correlated slices to AD diagnosis, while the second method examines the contribution of the slices combination. MSCNNs is inspired by multi-view CNN algorithm [20], [21]. By fusing several views and compacting representation, the models can apply views’ correlation and pixel details simultaneously. Besides, the MSCNNs of the end-end model is more efficient than the SACNNs of the two-stage model.

Medical images have 3D and 2.5D characteristics. For this reason, the multi-view algorithm can be applied for the disease tasks [22]. For example, an adapted multi-view model is used to detect pulmonary nodules [23], and a recent model used views from the axial, sagittal, and coronal directions to enable better disease diagnoses. Liu2018Classification. In this paper, we develop MSCNNs based on multi-view
The slices importance refer to how effect those areas make practically the prediction accuracy of the \( k \) slices, then analyze the pixel-level relevance to AD.

The second method decide slice importance by using the RF classifier, which is a common classification tool in Biomedical fields [25]–[30]. SACNNs focuses on the local slices with the best performance, while RF identifies all slices correlated with disease. The importance score from the re-trained RF classifiers identify the slice importance including the Mean Decrease in Impurity (MDI)/Gini and the Mean Decrease in Accuracy (MDA). Our method uses MDI. It is calculated as follow: [31]:

\[
Imp(x_m) = \frac{1}{NT} \sum_{t \in T} \sum_{p(t)} (p(t) \Delta i(s_t, t))
\]  

(2)

,where \( p(t) \) denotes the proportion \( \frac{N_t}{N} \) of samples \( t \), and \( v(s_t) \) is a function of split \( s_t \). \( \Delta i(s_t, t) \) is the impurity reduction at node \( N \), which can be described as:

\[
\Delta i(s_t, t) = i(t) - p_{Left}(t_{Left}) - p_{Right}(t_{Right})
\]  

(3)

,where \( i(t) \) is the impurity measure, \( p_{Left} = \frac{N_{Left}}{N_t} \), and \( p_{Right} = \frac{N_{Right}}{N_t} \). Note that, when using one of the correlated features, the importance of the rest features will drop. The reason is that the first feature has removed the impurity, then the feature importance score would be biased towards other rest related features.

For approach 2, the slice combination with the optimal performance contains important and supplemental information for the final diagnosis. We examined two sets of slice combination experiments: 1) select several slices in a fixed interval from three directions, which start from the center of the brain to cover abroad region; 2) choose the slices from the essential areas of the standard brain image, which rely on specialist’s domain knowledge about AD. As a consequence, the diagnostic accuracy will reveal the most critical slice combination.

C. VISUAL EXPLANATION USING SALIENCY MAPS

Saliency maps interpret the trained models and describe the discriminative regions of the brain at the pixel level for AD classification. According to the method described by [32], we follow four steps: (1) fetch an image through the trained model; (2) calculate the score of every class; (3) set the output of the selected class to 1 and the output of all other classes to 0; and (4) backpropagate this derivative to obtain the value of the gradients.

Given inputs of grayscale brain images, \( I_{m \times n} \) (\( m \), \( n \)) denote rows and columns, \( c \) is specific class, the saliency map can be calculated as:

\[
M_{ij} = |w_{h(i,j)}|
\]  

(4)

where \( w \) represents the derivative and \( h(i, j) \) denotes the index of \( w \) on the image pixel \( (i, j) \). Because our models use softmax layers to generate the probability of all classes, the saliency map displays the most discriminative pixels represent class \( c \) under the extreme value of class \( c \).
Because the feature maps of the first convolutional layer are usually interpretable and describe simple visual properties, we highlight the magnitude of their gradient on the same image. Then we can know the apparent regions on which the models focus.

### III. EXPERIMENTAL RESULTS

#### A. SUBJECT AND IMAGE PROCESSING

Labeled sMRI obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI)\(^1\). The three datasets, including AD images, NC images, and Mild Cognitive Impairment (MCI) images, are listed in Table 1. The ADNI was built by the National Institute on Aging (NIA), a non-profit organization, which provide data related to brain diseases such as serial MRI images, serial PET images, clinical and neuropsychological assessments, and additional biological markers. In this work, used sMRI have been preprocessed with skull-stripping and alignment, and are "spatially normalized, masked, and N3-corrected."\(^{[33]}\)

In experiments, we choose the first image, in case of images from the same person at different stages might induce "noise" to the models and affect their precision. Next, we renormalized the voxel value of each image to the range \([0,1]\)\(^{[34]}\). Finally, all profiles were then segmented into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) using the Computational Anatomy Toolbox (CAT12 Box). After removing irrelevant details, the GM density was aligned to the International Consortium for Brain Mapping (MNI-ICBM-152) atlas\(^{[4]}\). A total of 387 2D slices then were extracted from the axial, coronal, and sagittal planes of the original 121 × 145 × 121 medical images.

#### B. IMPLEMENTATION DETAILS

Our models implemented on an AMD A8 CPU with 16 GB RAM and an NVIDIA GTX 1080 under the Linux operating system with the PyTorch framework. We set a mini the batch size of 32, a learning rate of 0.0001, weight decay of 0.005, and a momentum factor of 0.9 with Adam optimization. To reduce the risk of overfitting, we ran an early stopping criterion. Owing to retain the original brain image structure and increase training dataset, we utilized Mixup technology\(^{[35]}\) to extend train-set instead of classic data augmentation techniques such as patching, rotation, and cropping.

We split all 2D image slices into training and validation sets in the ratio of 80:20. On the first stage of ASCCNNS, we pretrained and validated the model on the training sets; on the voting stage, we test it on validation sets. For avoiding prediction bias, the training and validation sets do not have the images from the same subject simultaneously. The results were averaged over ten evaluation cycles.

Experimentally, we use confusion matrix, a common evaluation standard for binary and ternary classification, to analysis classifiers performance. Table 2 explains confusion matrix. By making use of value of confusion matrix during test stage, sensitive (Sen) and specific (Spe) are computed:

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (5)
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \quad (6)
\]

In practice, we compared accuracy, sensitivity, specificity, and area under the Receiver Operating Characteristic Curve (ROC-AUC) of classifiers for different tasks on 10 cross folds. ROC-AUC was calculated using library of sklearn package.

#### C. MODEL INFERENCE

![Figure 5](image-url)  
**FIGURE 5:** Comparison of classification performance for AD vs. NC on the 81th slices in the coronal plane. Note: (a) DenseNet; (b) classic AlexNet; (c) full convolutional AlexNet; (d) simple broaden plain CNNs locate the left to the right respectively; Y-axis denotes the value of accuracy, sensitivity, specificity and area under curve.

We experimented with AlexNet, full convolutional AlexNet, DenseNet of 30 layers and proposed simple broaden plain

---

\(^1\)http://adni.loni.usc.edu/
CNNs (SBPCNNs). On the first layer of SBPCNNs, the kernel size was set as huge as 39, and the number of feature maps is 18. Fig. 3 displays differences in models of time cost and the number of parameters. SBPCNNs is the simplest model having significant conceived fields and fewer layers. DenseNet runs slower associated with the complex network structure which produces numerous feature maps during the training stage. Fig. 5 shows the comparative performance of those models via a specific slice of a fold. The accuracy of the diagnosis of SBPCNNs surpassed other models significantly. By input several slices from different folds, results also showed that SBPCNNs achieved the best performance. For SACNNs, we examined the effects of the number of classifiers required. The prediction accuracy was stable in using 9 classifiers. More than 9, the practice shows the number of classifiers increased while the performance decreased slightly. For MSCNNs, in consideration of the computational cost and slices’ diversity, 30 slices were selected from the original 387 images.

D. COMPARISON RESULTS

Table 4 compares our approaches with several advance methods. Those researchers used different datasets which are various in the numbers of subjects and distribution on gender, age, education, and degree of impairment of the brain. For this reason, this table only shows rough contrast results. In this table, GoogleNet with ROI input have the best performance [4], however it needs extra image process. When using sMRI, 3D CCNs achieved the best performance but it used bigger dataset. Notely, our models also have closely accuracy to them.

TABLE 4: Performance comparison among state-of-the-art methods.

<table>
<thead>
<tr>
<th>Study</th>
<th>NC/AD</th>
<th>Modality</th>
<th>Classification algorithm</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[4]</td>
<td>226/186</td>
<td>ROI</td>
<td>GoogleNet</td>
<td>97.77%</td>
</tr>
<tr>
<td>[36]</td>
<td>226/186</td>
<td>ROI</td>
<td>DeepESRNet</td>
<td>91.02%</td>
</tr>
<tr>
<td>[6]</td>
<td>188/188</td>
<td>ROI</td>
<td>2D CNN fusion</td>
<td>85.94%</td>
</tr>
<tr>
<td>[37]</td>
<td>94/92</td>
<td>sMRI</td>
<td>SVM CNN</td>
<td>93.01%</td>
</tr>
<tr>
<td>[7]</td>
<td>755/755</td>
<td>sMRI</td>
<td>3D CNN</td>
<td>95.39%</td>
</tr>
<tr>
<td>[38]</td>
<td>50/61</td>
<td>sMRI</td>
<td>3DResNet</td>
<td>88.00%</td>
</tr>
<tr>
<td>Proposed model 1</td>
<td>200/229</td>
<td>sMRI</td>
<td>SBPCNNs</td>
<td>88.55%</td>
</tr>
<tr>
<td>Proposed model 2</td>
<td>200/229</td>
<td>sMRI</td>
<td>SACNNs</td>
<td>91.68%</td>
</tr>
<tr>
<td>Proposed model 3</td>
<td>200/229</td>
<td>sMRI</td>
<td>MSCNNs</td>
<td>93.75%</td>
</tr>
</tbody>
</table>

Note: NC normal cohort; AD people with Alzheimer’s disease; 3DResNet 3D Residual and plain convolutional neural networks; DeepESRNet Deep ensemble sparse regression network.

TABLE 5: Comparison of diagnosis performance on the same datasets.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Dataset</th>
<th>Acc</th>
<th>Sen</th>
<th>Spe</th>
<th>ROC-AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>AD vs. NC</td>
<td>89.33%</td>
<td>90.63%</td>
<td>88.76%</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>MCI vs. NC</td>
<td>77.54%</td>
<td>84.18%</td>
<td>75.72%</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>MCI vs. AD</td>
<td>73.46%</td>
<td>72.78%</td>
<td>69.69%</td>
<td>0.72</td>
</tr>
<tr>
<td>SBPC</td>
<td>AD vs. NC</td>
<td>85.55%</td>
<td>89.93%</td>
<td>86.24%</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>MCI vs. NC</td>
<td>74.57%</td>
<td>77.89%</td>
<td>70.17%</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>MCI vs. AD</td>
<td>71.23%</td>
<td>76.00%</td>
<td>64.20%</td>
<td>0.69</td>
</tr>
<tr>
<td>3D ResNet</td>
<td>AD vs. NC</td>
<td>95.21%</td>
<td>96.11%</td>
<td>93.13%</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>MCI vs. NC</td>
<td>90.20%</td>
<td>89.36%</td>
<td>89.61%</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>MCI vs. AD</td>
<td>86.64%</td>
<td>77.80%</td>
<td>90.87%</td>
<td>0.84</td>
</tr>
<tr>
<td>SACNN (MV)</td>
<td>AD vs. NC</td>
<td>91.68%</td>
<td>92.24%</td>
<td>91.50%</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>MCI vs. NC</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>MCI vs. AD</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MSCNN</td>
<td>AD vs. NC</td>
<td>93.75%</td>
<td>94.23%</td>
<td>93.40%</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>MCI vs. NC</td>
<td>88.50%</td>
<td>82.16%</td>
<td>85.24%</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>MCI vs. AD</td>
<td>85.32%</td>
<td>78.79%</td>
<td>87.30%</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Note: Acc: Accuracy; Sen: Sensitivity; Spe: Specificity; MV: Majority voting; RF: Random forest; ROC-AUC: area under the Receiver Operating Characteristic Curve.

E. SLICE IMPORTANCE

As the previous section argued, our models are the benefit of exploit informative region with AD in slice and pixel level. By ranking of the pertained 387 classifiers accuracy, the slices with greater than 80% accuracy are found to located in the sagittal (38 – 50, 73 – 84), coronal (59 – 90), and axial (24 – 47) regions. Referred to atlas, Table 6 presents the distribution of slices with good diagnosis performance.

TABLE 5: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS

MSCNNs and SACNNs, SACNNs spends considerable time on pertaining for whole slices including useless slices. All together, MSCNNs could be a good alternative to the voxel model for 3D CNN-based classification.

Renset or Denset based 3D CNNs faced serious problem of overfitting and high computational cost. Compared to them, our models have closely accuracy to them.
TABLE 6: Discriminative pixel overlap values for cross-registration on MNI ICMB 152.

<table>
<thead>
<tr>
<th>Region</th>
<th>Pixel Freq</th>
<th>Overlap</th>
<th>Region</th>
<th>Pixel Freq</th>
<th>Overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>L Background</td>
<td>2899</td>
<td>15.10%</td>
<td>R Background</td>
<td>2389</td>
<td>13.51%</td>
</tr>
<tr>
<td>L Frontal.InfOrb</td>
<td>71</td>
<td>0.37%</td>
<td>R Frontal.InfOrb</td>
<td>4</td>
<td>0.02%</td>
</tr>
<tr>
<td>L Olfactory</td>
<td>116</td>
<td>0.60%</td>
<td>R Olfactory</td>
<td>63</td>
<td>0.33%</td>
</tr>
<tr>
<td>L Insula</td>
<td>68</td>
<td>0.35%</td>
<td>R Insula</td>
<td>36</td>
<td>0.19%</td>
</tr>
<tr>
<td>L Hippocampus</td>
<td>1636</td>
<td>8.52%</td>
<td>R Hippocampus</td>
<td>1481</td>
<td>7.71%</td>
</tr>
<tr>
<td>L Parahippocampal</td>
<td>2182</td>
<td>11.36%</td>
<td>R Parahippocampal</td>
<td>2121</td>
<td>11.05%</td>
</tr>
<tr>
<td>L Amygdala</td>
<td>463</td>
<td>2.41%</td>
<td>R Amygdala</td>
<td>465</td>
<td>2.42%</td>
</tr>
<tr>
<td>L Lingual gyrus</td>
<td>13</td>
<td>0.07%</td>
<td>R Lingual gyrus</td>
<td>15</td>
<td>0.08%</td>
</tr>
<tr>
<td>L Fusiform</td>
<td>757</td>
<td>3.94%</td>
<td>R Fusiform</td>
<td>829</td>
<td>4.32%</td>
</tr>
<tr>
<td>L Putamen</td>
<td>396</td>
<td>2.06%</td>
<td>R Putamen</td>
<td>465</td>
<td>2.42%</td>
</tr>
<tr>
<td>L Pallidum</td>
<td>233</td>
<td>1.21%</td>
<td>R Pallidum</td>
<td>251</td>
<td>1.31%</td>
</tr>
<tr>
<td>L Temporal_Pole_Sup</td>
<td>438</td>
<td>2.28%</td>
<td>R Temporal_Pole_Sup</td>
<td>226</td>
<td>1.18%</td>
</tr>
<tr>
<td>L Temporal_Pole_Mid</td>
<td>116</td>
<td>0.06%</td>
<td>R Temporal_Pole_Mid</td>
<td>137</td>
<td>0.71%</td>
</tr>
<tr>
<td>L Cerebellum_Crus2</td>
<td>0</td>
<td>0.00%</td>
<td>R Cerebellum_Crus2</td>
<td>41</td>
<td>0.21%</td>
</tr>
<tr>
<td>L Cerebellum_3</td>
<td>510</td>
<td>2.66%</td>
<td>R Cerebellum_3</td>
<td>376</td>
<td>1.96%</td>
</tr>
<tr>
<td>L Cerebellum_4_5</td>
<td>80</td>
<td>0.42%</td>
<td>L Cerebellum_4_5</td>
<td>113</td>
<td>0.59%</td>
</tr>
</tbody>
</table>

The purple donate the first important area in brain with AD, and the yellow is the second important area respectively.

FIGURE 6: Comparison of slice importance score of whole slices from RF on three planes. Note: (a) axial; (b) sagittal; (c) coronal. y axis is Gini purity, and X axis is slice location.

FIGURE 7: Distribution of classification accuracy. Horizontal axis is number of classifiers.

by the orange line (axial plane), the frontal slices have the best performance for AD diagnosis.

Fig. 8 displays the performance of five slice combinations via the MSCNNs. The image on the left in Fig. 8, shows the comparison results of slices combination at fixed intervals in three planes, and results of coronal having the highest accuracy. 8 is in line with the results in Fig. 7; the image on the right in Fig. 8 is in line with the results in Fig. 7; the image on the right in Fig. 8 shows the comparison results of slices combination at fixed intervals along the coronal plane perform better than slices based on domain knowledge. We think the reason is slices combination at fixed intervals contain diversity information of brain.

IV. EVALUATION VISUALIZATION

This section describes the use of saliency maps, which interpret and display the related brain regions, by visualizing models with higher gradient values of a specific label. We set a threshold as the 99th percentile on the gradient image. Fig. 9 and Fig. 10 show where the pixels correspond strongly to specific labels and what is seen on the slices by the selected models.

The images in Fig. 9 was generated on the first layers of two models via MSCNNs of different folds. Every picture includes saliency maps. The similarity among them indi-
FIGURE 9: Visualization of the saliency maps generated via the two selected MSCNNs. Note: Slice location indexes range from 13 to 100 in 3-slice intervals along the coronal direction.

FIGURE 10: The saliency maps on the slices. X-axis index of the brain slices. Note: (a) DenseNet, (b) SBPCNNs, (c) MSCNNs.

cates that these models are stable, with the area of highlighted pixels denoting essential regions in the AD or NC groups.

Fig. 10 sequentially presents visualization results from several saliency maps using SPBCNNs, MSCNNs, and DenseNet via the NC samples. In the visualization results (a) and (b), the highlighted regions from DenseNet covers those from SBPCNNs. Thus the saliency maps of DenseNet might be less useful in recognizing the relative locations of AD. In the images on (b) and (c), there are significant highlighted areas that affect the AD diagnosis given by SBPCNNs and MSCNNs. Since details of their locations and shapes are similar, these models produce stable location information, and these brain regions might vary during the progress of AD. The visualization results provide some explicit visual
expressions, which expose the discriminative areas of sMRI in the view of the models.

**V. CONCLUSIONS AND FUTURE WORK**

In this paper, we have developed three 2D CNNs-based classifiers for enhancing the diagnostic accuracy and determining precise location effects for AD. Our models simultaneously record the slice positions and achieve good performance in AD classification tasks. We studied the contributions of different slices combinations to AD classification and derived visual explanations of the proposed models. We identified which brain regions have the largest differentiation between the AD and NC groups in slice-level 2D views. Additionally, visualizing what is “seen” by the different models will provide insights to experts and help with the design of “understanding” models.

The current work was limited in using a single modality of sMRI, which provide restricted clues for MCI conversion. At this point, the future work is to utilize the systematic methods in AD, MCI, and other related disease classification tasks [41], [42]. Moreover, combining multiple medical images with genetic data for AD and other diseases [43]–[46] classification would be an excellent idea and future works [47]–[49].

**REFERENCES**


***