IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 17, 2021, accepted February 5, 2021, date of publication February 16, 2021, date of current version February 25, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3059658

Volumetric Feature-Based Alzheimer's Disease Diagnosis From sMRI Data Using a Convolutional Neural Network and a Deep Neural Network

ABOL BASHER[1], BYEONG C. KIM[2,4], KUN HO LEE[2,3,5], AND HO YUB JUNG[1]

[1]Department of Computer Engineering, Chosun University, Gwangju 61452, South Korea
[2]Gwangju Alzheimer's disease and Related Dementias Cohort Research Center, Chosun University, Gwangju 61452, South Korea
[3]Department of Biomedical Science, Chosun University, Gwangju 61452, South Korea
[4]Department of Neurology, Chonnam National University Medical School, Gwangju 61469, South Korea
[5]Korea Brain Research Institute, Daegu 41062, South Korea

Corresponding author: Ho Yub Jung (hoyub@chosun.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government through MSIT under Grant NRF-2019R1A4A1029769, and in part by the Research Fund of Chosun University 2020.

ABSTRACT Alzheimer's disease (AD) is a progressive neurodegenerative disorder that is mostly prevalent in people older than 65 years. The hippocampus is a widely studied region of interest (ROI) for a number of reasons, such as memory function analysis, stress development observation and neurological disorder investigation. Moreover, hippocampal volume atrophy is known to be linked with Alzheimer's disease. On the other hand, several biomarkers, such as amyloid beta ($a\beta_{42}$) protein, tau, phosphorylated tau and hippocampal volume atrophy, are being used to diagnose AD. In this research work, we have proposed a method to diagnose AD based on slice-wise volumetric features extracted from the left and right hippocampi of structural magnetic resonance imaging (sMRI) data. The proposed method is an aggregation of a convolutional neural network (CNN) model with a deep neural network (DNN) model. The left and right hippocampi have been localized automatically using a two-stage ensemble Hough-CNN. The localized hippocampal positions are used to extract (80 × 80x80 voxels) 3-D patches. The 2-D slices are then separated from the 3-D patches along axial, sagittal, and coronal views. The pre-processed 2-D patches are used to extract volumetric features from each slice by using a discrete volume estimation convolutional neural network (DVE-CNN) model. The extracted volumetric features have been used to train and test the classification network. The proposed approach has achieved average weighted classification accuracies of 94.82% and 94.02% based on the extracted volumetric features attributed to the left and right hippocampi, respectively. In addition, it has achieved area under the curve (AUC) values of 92.54% and 90.62% for the left and right hippocampi, respectively. Our method has outperformed the other methods by a certain margin in the same dataset.

INDEX TERMS Hippocampus, volumetric features, 2-D/3-D patches, hough-CNN, CNN, DNN, MRI, Alzheimer's disease, classification, knowledge transfer.

I. INTRODUCTION

Alzheimer's disease (AD) is aicities. According to [1], 1 out of 85 persons will suffer from AD by the year 2050. It will be an enormous burden in the context of the economy, as well as for families. It is believed that AD can begin developing a decade or more before the appearance of clinical symptoms [2], [3]. Therefore, it is important to diagnose AD patients in the early stages so that the necessary treatments can be provided to delay the progression of AD for a certain period. In addition, an individual's lifestyle practices can deter or slow down AD progression, which is an important aspect that can be monitored if it is possible to know in the early stages that a particular individual is likely to become an AD patient [2]. Therefore, automatic and early diagnosis is an important research endeavor.

From early 2000 to 2010, Alzheimer's disease dementia-related research communities have focused on finding various

The associate editor coordinating the review of this manuscript and approving it for publication was Hiu Yung Wong.

biomarkers for AD [4]. Different biomarkers, such as beta amyloid ($a\beta_{1-42}$) plaque deposition, tau, phosphorylated tau $(ptau)_{181}$, fluoro-deoxy-glucose positron emission tomography (FDG-PET) uptake in PET and hippocampal volume reduction [2], [4], are being investigated to test the effectiveness in detecting early stage AD. Moreover, it remains an open challenge to select the right biomarker from the validated biomarker list. However, these biomarkers can serve as early AD diagnostic indicators. For example, axonal death and neuronal degeneration can cause atrophy in brain regions such as the hippocampus and may increase the levels of intracellular tau protein in the cerebrospinal fluid (CSF) [2], [5], [6]. Hippocampal volume atrophy can be observed from structural magnetic resonance imaging (sMRI), which is an important biomarker. Hippocampal volume measurement is a challenging task. There are few software packages, such as FreeSurfer, that can measure the hippocampal volume automatically/semi-automatically. However, depending on the algorithm used by the software package, the required time may vary significantly.

Computer aided diagnosis is offering a wide variety of abilities to different branches of medical imaging. Several methods [3], [7]–[9] have been developed to diagnose various brain/non-brain diseases based on the different modalities of medical imaging. An sMRI is one of the variants that provides high resolution structural information of the human brain that can be used to diagnose various brain diseases, such as AD, epilepsy and so on. Various semi-automatic and automatic algorithms have been proposed to predict AD. However, machine learning and deep learning based approaches [10]–[13] are offering the most promising performances.

Deep learning algorithms attempt to discover the unknown hidden representation of the input data. The relation between the input data and the outputs is mapped based on the learnable filters [14]. The feature extraction from the input data is an important step for solving computer-aided tasks [15]–[17]. In deep learning, automatic algorithms are being used to extract features without any user intervention. Convolutional neural network (CNN) is widely used in the research communities for image-based problem solving. CNN automatically extracts features which are distilled in nature, i.e., only the relevant important distinguishable information is collected from the input data during training. The extracted features are used for different purposes, such as classification and regression. In addition, distilled knowledge transfer is another important concept [18], [19] that is being utilized to train a network from another trained network's observation. The teacher network guides the student network to learn specific tasks. Moreover, transfer learning [14], [20], [20]–[22] is widely used in the deep learning communities to solve various problems in computer vision. In this research work, a discrete volume estimation convolutional neural network (DVE-CNN) [23] was used to extract volumetric features from the sMRI scans. This network has two models named left hippocampal model (LHM) and right hippocampal model (RHM). DVE-CNN extracts slice-wise number of voxels attributed to the hippocampus from the target sMRI. In this study, this slice-wise voxels are considered as the volumetric features. The proposed deep neural network network (DNN) learns the hidden representation of slice-wise volumetric features and ultimately performs better in classifying the AD and normal control (NC) classes.

In this research paper, we have proposed a deep learning based fully automatic approach for AD diagnosis using sMRI data from Gwangju Alzheimer's and Related Dementia (GARD) dataset. Our proposed approach is the incremental improvement upon our previous works [23], [24], where the first method automatically predicts the position of a region of interest (ROI) inside an sMRI, and the second method estimates the number of voxels (discrete volume) of the corresponding ROI (in this case, the hippocampus) from the 3-channel 2-D patches. The obtained slice-wise hippocampal volumetric features have been used to estimate the probability score for an AD sample class and the NC sample class using DNN model. The graphical illustration of the proposed approach is shown in Fig. 1. The DNN model learns the important distinguishable features generated from the DVE-CNN network, where the learned features are distilled in nature. The trained model is used for final prediction.

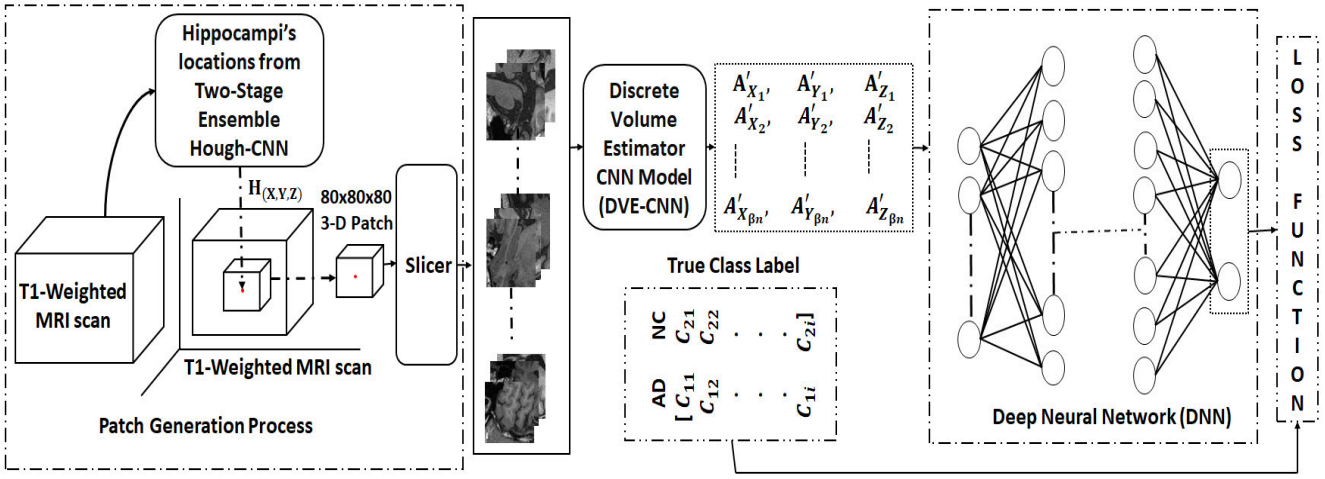
This research work is organized in the following manner. In section II, we have discussed the related literature in this research field. The methodology along with the dataset, data processing, network architecture detail, and loss functions have been illustrated in section III. In the following section, the experimental results for the proposed method are explained with elaborate reasoning, as well as comparative analysis of the proposed method with the other recent methods. In section V, this research work has been summarized.

A. CONTRIBUTION

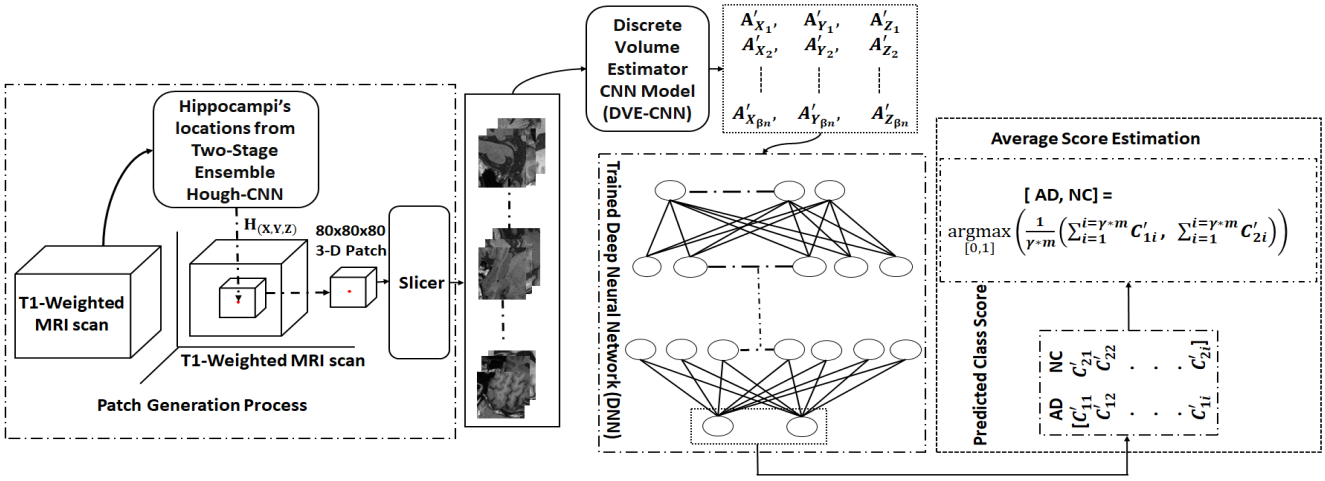
An aggregated system of Hough-CNN, CNN, and DNN has been developed to classify AD and NC from sMRI data. We have used three different deep learning-based models to complete AD diagnosis, where each of the processes is fully automatic and the diagnosis is performed based on the hippocampal volume reduction, an early diagnostic indicator of AD. The DVE-CNN extracted volumetric features are specifically inclined to indicate the hidden cases of Alzheimer's disease attributed with the hippocampal volume reduction. The proposed DNN learns the extracted volumetric information of each slice related to the hippocampus of an sMRI scan and achieves higher accuracy than other compared methods. The proposed approach offers an on-site diagnosis of AD that can be used to make further clinical decisions. In addition, the processing of slice-wise volumetric features is one of the most important contributions to the classification literature, where processed features help to distinguish the AD and NC classes.

II. RELATED WORKS

AD is an irreversible chronic neurodegenerative disease. From the prodromal stage (mild cognitive impairment) to AD



(a) AD versus NC classification: training system



(b) AD versus NC classification: testing system

FIGURE 1. The proposed deep learning model (training and test framework) for AD diagnosis. (a) The two-stage ensemble Hough-CNN model for predicting the position of the hippocampus has been used to extract patches for the DVE-CNN (LHM and RHM) model to predict the number of voxels from each patch attributed to the hippocampus. The observed slice-wise volumetric features along with their corresponding class labels are used to train the DNN model. The trained model is deployed to make inferences for predicting the class label for AD and NC classes. (b) In the testing phase, the trained DNN model predicts the class probability score based on the volumetric features attributed to the hippocampus predicted by the DVE-CNN (LHM and RHM) model from each patch. The average predicted class probability score is used to determine the final prediction for any particular subject.

formation, the progression rate is approximately $10-15\%$ per year [25]. Early diagnosis may attenuate further deterioration and offer an opportunity to provide proper treatment to improve patient quality of life. Greater prevalence of AD is observed among those aged 65 years and older, and therefore, we must manifest proper preparation to give them a healthy environment for life. An efficient and reliable diagnostic method is an important prerequisite to early diagnosis. Several research initiatives are being conducted for the development of methods for automatic and early detection of AD.

A number of methods have been proposed by various research communities from different disciplines to diagnose AD based on different biomarkers [5], [26], [27]. Different research groups [5], [28]–[31] have investigated the pathological beta amyloid ($A\beta_{1-42}$) deposition measurement from CSF to predict AD. On the other hand, metabolic brain alterations were observed using FDG-PET to classify the AD and NC classes [27], [32]. The structural atrophy of brain morphometry (i.e., size and shapes) is another important biomarker that has long been used by many research communities to diagnose AD [9], [33]–[36].

The anatomical size and shape variability/atrophy information of brain ROIs (i.e., left and right hippocampi) extracted from sMRI data [37] is utilized to train the classification/regression model to estimate the future progression of AD. This informative and distinguishable features extraction from sMRI can be performed in the following three possible ways: (a) a voxel-based feature extraction approach [38], [39], which includes the local tissue densities (i.e., white matter, gray matter and CSF) of the brain; (b) an ROI-based/patch-based features extraction technique [13],

[40], [41], which includes the regional cortical thickness estimation, hippocampal volume measurement, gray matter volumes, and so on; (c) the full-MRI-based/multiple feature extraction process [42]–[44], which considers the whole sMRI rather considering the sub-regions of an sMRI. However, complete MRIs with demographic information, such as age, gender and education, can be used jointly for classification and regression.

Next, from the model design and ROI selection perspective, numerous network architectures have been proposed to classify the atrophied state [45] of the human brain, segment the brain regions [46]–[51] and measure the volume of ROIs [23], [52]. A volume-based automatic cortical thickness pipeline was proposed in [52]. This method is known as Advanced Normalization Tools (ANTs). ANTs method estimates the volume-based cortical thickness by conducting multiple operations, such as (a) primary N4 bias correction on input sMRI, (b) brain extraction based on template strategy or hybrid segmentation, (c) performing the pure tissue posterior probability based weighted bias correction using an open source software tools (n-tissue) [53] and N4 bias correction along with altering them with prior-based segmentation, (d) cortical thickness estimatation based on Diffeomorphic registration-based cortical thickness algorithm [54]. Another optional normalization operation is conducted to the specified template/multi-atlas cortical parcellation. A method based on the hippocampal volume atrophy detected from sMRI has been proposed to predict AD in [40]. Yue *et al.* [55] have proposed a hierarchical feature extraction process for early diagnosis of AD. A method based on in vivo mapping of gray matter loss along with voxel-based morphometry for mild cognitive impairment (MCI) patients has been proposed by Baron *et al.* [38]. Afzal *et al.* [56] have proposed a data augmentation framework for imbalanced classes for AD stage detection. Different temporal information has been retrieved from longitudinal MRI scans to diagnose AD in [57].

Recently, machine learning- and deep learning-based approaches have shown promising performance in classifying different stages of dementia, especially diagnosing the AD, MCI and NC classes. Typical machine learning-based approaches, such as support vector machine (SVM) and k-nearest neighbor(KNN), are being used to classify AD, NC, stable mild cognitive impairment (sMCI), and progressive mild cognitive impairment (pMCI). Gupta *et al.* [44] have used the machine learning approaches (SVM, KNN, and random forest (RF)) to classify the atrophied states (AD, NC/healthy control (HC), asymptotic Alzheimer's disease(aAD), mild Alzheimer's disease (mAD)) using combined features of voxel-based morphometry (VBM), cortical and subcortical volumetric features (CSC), and hippocampal volumetric (HV) information of T1-weighted sMRI and obtained comparatively high performance with respect to AD diagnosis. 2-D and 3-D deep learning models have been constructed to diagnose AD [10]. A patch-based ensemble classifier has been constructed to predict the AD and NC classes in [13]. Using temporal information of sMRI data, a method has been developed to diagnose AD in [3] based on long short-term memory, a deep learning-based framework. Based on rs-fMRI data analysis, an automatic AD classification network architecture has been proposed using 3-D CNN in [58], [59]. In this research work, we have proposed a deep learning-based classifier for AD versus NC classification through extracting deep features of hippocampal discrete volume by the DVE-CNN (LHM and RHM) model [23].

III. METHODOLOGY

The internal properties of the hippocampus and the cases of its volume reduction are embedded with extracted volumetric features, which are successfully used to diagnose AD in this research work. Hippocampal volume degradation/variability is a well-known biomarker to diagnose AD. A detailed slice-wise volumetric feature extraction procedure is shown in one of our previous works [23]. However, the volumetric feature extraction process has been briefly explained in another section of this research work. We used DVE-CNN to extract the hippocampal volumetric features for NC(171) and ADD(80) subjects of the GARD dataset. The aAD and mAD subjects are not considered in this study. The extracted volumetric features were used to train the proposed DNN model. In the testing phase, the trained DNN model predicted the class probability score for a particular test subject based on the transferred knowledge of volumetric features acquired by DVE-CNN. The detailed procedures of diagnosing AD are illustrated in the following sections.

A. DATASET

The Gwangju Alzheimer's and Related Dementia (GARD) dataset was used in this research work. The dataset was collected from the National Research Center for Dementia (NRCD), South Korea, from January 2014 to March 2018 [44]. The dataset consists of 326 MRI scans with four classes: (1) NC (171 subjects), (2) Alzheimer's disease dementia (ADD) (81 subjects), (3) asymptotic Alzheimer's disease (aAD) (35 subjects), and (4) mild Alzheimer's disease (mAD) (39 subjects). The subjects' average age range was 70.018±6.074 years old. Most of the sMRI scans have the dimensions of 320 × 212x240 with $0.512mm^3$ unit voxel volume.

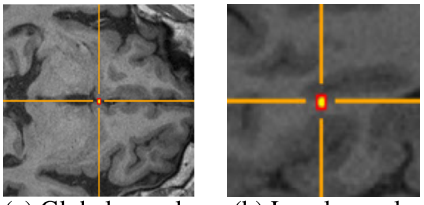
B. DATA PROCESSING

The MRI scans were processed based on the network requirements. The detailed processing of MRI scans and the volumetric feature extraction process are explained in the following sections.

1) PATCH GENERATION: LOCALIZATION

The localization operation is conducted in two phases, as explained in our previous work [24]. The left and right hippocampal positions were previously localized manually. In the first phase, we have extracted 96 × 96(voxels) 2-D patches considering random voxel position from the whole sMRI of axial, coronal and sagittal views and reshaped them

into 32 × 32(voxels) 2-D patches for the global model. After that, we have normalized the extracted patches with mean zero and standard deviation of one. The normalized patches are then reshaped into 32 × 32x1. At the end, we have concatenated the axial, coronal and sagittal patches and reshaped them into 32 × 32x3 3-channel 2-D patches. In the second phase, 32×32 (voxels) 2-D patches have been extracted based on random voxel position in the vicinity of the hippocampal region for the local model. The extracted patches were normalized with mean zero and standard deviation of one. Normalized patches have then been reshaped into 32 × 32x1. The reshaped patches have been concatenated to construct the 32 × 32x3 patches. In both phases, the corresponding displacement vectors have been estimated for the ground truth. 512 patches from each sMRI were extracted for both global and local hippocampal position estimation in current research work. The global and local random samples are shown in Fig 2. The best and worst case automatic position estimation of right hippocampus by two-stage ensemble Hough-CNN is shown in Fig. 3



(a) Global sample (b) Local sample

FIGURE 2. The global samples (96 × 96 voxels) (a) were extracted randomly from whole sMRI scan with their corresponding displacement vectors, whereas, the local samples (32 × 32 voxels) (b) were extracted from the vicinity of the hippocampus along with their corresponding displacement vectors.(@ Both patches has been displayed by resizing them to 1 × 1 inches).

2) PATCH GENERATION: VOLUMETRIC FEATURE ESTIMATION

The 80 × 80x80 (voxels) 3-D patches were extracted from the two-stage ensemble Hough-CNN [24] localized positions of the left and right hippocampi. The extracted 3-D patches were then separated along the axial, coronal and sagittal planes. From each view, 80 patches of size 80 × 80 (voxels) were extracted. The extracted patches were then normalized with zero mean and standard deviation of one. The normalized patches were then augmented by a factor of n (in this case, n = 15). The augmentation operation is conducted with Θ-degree rotation. The degree of freedom for Θ is $-90 < \Theta < 90$. The augmented patches were then reshaped into 32 × 32x1. At the end, the reshaped patches were concatenated along the axis=2 (zero-based axis). The concatenated 32 × 32x3 patches were used to train the DVE-CNN (LHM and RHM) model to predict the slice-wise number of voxels, which were later used as volumetric features for the DNN model training and testing.

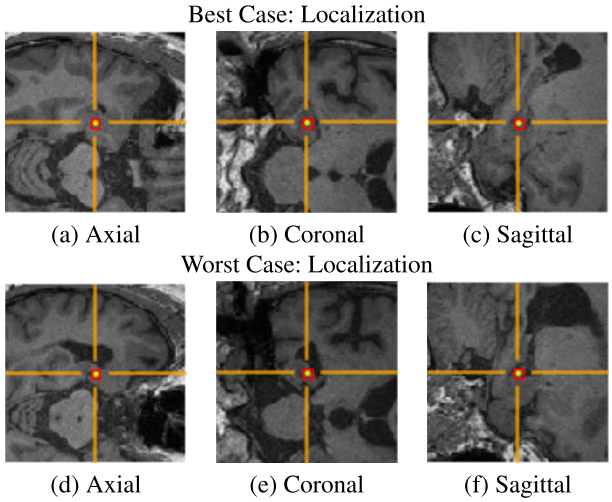


FIGURE 3. The representative automatic localized position of right hippocampus is shown in (a, b, c) and (d, e, f) for best case and worst case (respectively) of the axial, coronal and sagittal views by the two-stage ensemble Hough-CNN of the test MRI scans. Two-stage ensemble Hough-CNN localized position is used to extract patches for DVE-CNN. The DVE-CNN estimated slice-wise volumetric features are used to classify the AD and NC classes.

3) DATA PREPARATION: CLASSIFICATION

We have performed 5-fold cross validation while training the DVE-CNN model. The predicted number of voxels for all 5 folds were combined together with corresponding class labels for training and testing the classification model. The patches extracted from each MRI of the whole GARD dataset, the DVE-CNN (LHM and RHM) model predicted the number of voxels in each slice attributed to the left and right hippocampi. The volumetric features with their corresponding class labels are used to train the proposed DNN models. It has been shown in our previous work [23] that the predicted number of voxels (volumes) exhibit high correlation with the ANTs method [52]. The class labels (AD and NC classes) determined by the clinician for the GARD datast have been used as the ground truth for the current work. The diagnosis criteria and guidelines considered by clinician to analyze the data based on the following research works [60]–[62]. Th Neuro I software package(http://www.infomeditech.com/) is used to analyze the GARD dataset, where the ANTs method is included with this software package. The detail information about subject selection, sMRI acquisition can be found in these research works [26], [58], [59].

C. NETWORK ARCHITECTURE: TWO-STAGE HOUGH-CNN

The original network architecture of two-stage Hough-CNN is modified to simplify the localization model. Instead of using three models for each stage, we have used one model for each phase following previously proposed DVE-CNN model [23]. The global and local models have the same number of layers. There are 6 convolutional layers followed

TABLE 1. GH-CNN and LH-CNN network architectures used to localize the left and right hippocampi.

Model Name	Network Architecture		Activation Function	Batch Normalization	Optimizer
GH-CNN	GM_{lh}	$I_{32}, C_3^{16}, C_3^{32}, P_2^m, C_3^{64}, P_2^m, C_3^{128}, P_2^m, C_3^{256}, C_3^{512}, F^{512}, F^{128}, F^3$	ReLU	All Layers	Adam
	GM_{rh}	$I_{32}, C_3^{16}, C_3^{32}, P_2^m, C_3^{64}, P_2^m, C_3^{128}, P_2^m, C_3^{256}, C_3^{512}, F^{512}, F^{128}, F^3$			
LH-CNN	LM_{lh}	$I_{32}, C_3^{16}, C_3^{32}, P_2^m, C_3^{64}, P_2^m, C_3^{128}, P_2^m, C_3^{256}, C_3^{512}, F^{512}, F^{128}, F^3$	ReLU	All Layers	Adam
	LM_{rh}	$I_{32}, C_3^{16}, C_3^{32}, P_2^m, C_3^{64}, P_2^m, C_3^{128}, P_2^m, C_3^{256}, C_3^{512}, F^{512}, F^{128}, F^3$			

$I_{\text{sample size}}$ = Network input, $C^{(\#\text{filter})}_{(\text{kernal size})}$ = Convolutional layer, P_2^m = Max Pooling with stride 2, $F^{(\#\text{filter})}$ = Fully connected layer.

The models have dropout layer after first fully connected layer (F^{512})(25 %) and second fully connected layer (F^{128})(35 %).

TABLE 2. LHM and RHM network architectures used with the GARD cohort dataset to measure the discrete volume of the left and right hippocampi.

Model Name	Network Architecture	Activation Function	Batch Normalization	Optimizer
LHM*	$I_{32}, C_3^{32}, C_3^{64}, P_2^m, C_3^{128}, P_2^m, C_3^{256}, P_2^m, C_3^{256}, C_3^{512}, F^{512}, F^{128}, F^3$	ReLU	All Layers	Adam
RHM*	$I_{32}, C_3^{32}, C_3^{64}, P_2^m, C_3^{128}, P_2^m, C_3^{256}, P_2^m, C_3^{256}, C_3^{512}, F^{512}, F^{128}, F^3$	ReLU	All Layers	Adam

$I_{\text{sample size}}$ = Network input, $C^{(\#\text{filter})}_{(\text{kernal size})}$ = Convolutional layer, P_2^m = Max Pooling with stride 2, $F^{(\#\text{filter})}$ = Fully connected layer.
*LHM and RHM stand for left hippocampal model and right hippocampal model, respectively. Both LHM and RHM have dropout layer after first fully connected layer (F^{512})(25 %) and second fully connected layer(F^{128})(35 %).

by a rectified linear unit (ReLU) activation function [63] and batch normalization layer [64]. The max pooling layer is used after the 3^{rd}, 4^{th} and 5^{th} convolutional layers. There are 3 fully connected layers followed by a ReLU activation function and batch normalization layer. After the first and second fully connected layers, a dropout layer (25% and 35%, respectively) has been used. Both networks have been trained with Adam optimizer [65], along with a mean square error loss function. The learning rate used to train the global Hough-CNN (GH-CNN) and local Hough-CNN (LH-CNN) was 1e-4. The network details are shown in Table 1.

D. NETWORK ARCHITECTURE: DVE-CNN (LHM AND RHM)

We have retained the exact same network settings as described in our previous research work [23]. The only modification here is the input size of the patches. We have extracted 80 × 80x80 (voxels) 3-D patches for the current model instead of 64 × 64x64 (voxels) 3-D patches. To construct the discrete volume estimator CNN model (DVE-CNN (LHM and RHM)), 6 convolutional layers and 3 fully connected layers have been used, where each convolutional layer is followed by a batch normalization layer [64] and a ReLU activation function. The fully connected layers are also followed by a batch normalization layer and ReLU activation function [63]. A max pooling layer is used after the 3^{rd}, 4^{th} and 5^{th} convolutional layers. The DVE-CNN (LHM and RHM) detailed architectures are shown in Table 2. DVE-CNN (LHM and RHM) were trained with the Adam optimizer. The considered learning rate was 1e-4.

E. NETWORK ARCHITECTURE: CLASSIFICATION

To diagnose AD, we have constructed a 6-layered DNN model for left and right hippocampal data. The left and right hippocampal DNN models are denoted as LH-DNN and RH-DNN. Each layer of the LH-DNN and RH-DNN models is followed by a ReLU activation function [63] with one exception. The 6^{th} layer is followed by a sigmoid activation function. The number of filters is different in different layers. Adam optimizer [65] is used with its default parameter settings, along with a cross-entropy loss function. The DNN model parameter detail for left and right hippocampi is shown in Table 3.

F. LOSS FUNCTIONS

Three different models were used to perform the AD versus NC classification in completely automatic fashion. First, the localization models have been used to estimate the hippocampal positions in an sMRI scan. After that, discrete volume estimation models have been employed to measure the number of voxels (volumes) contributing to construct the hippocampus. The slice-wise output of the DVE-CNN model is considered as the volumetric features for the left and right hippocampi of each sMRI scan. Finally, using the extracted volumetric features in each patch, the AD versus NC classification operation has been performed.

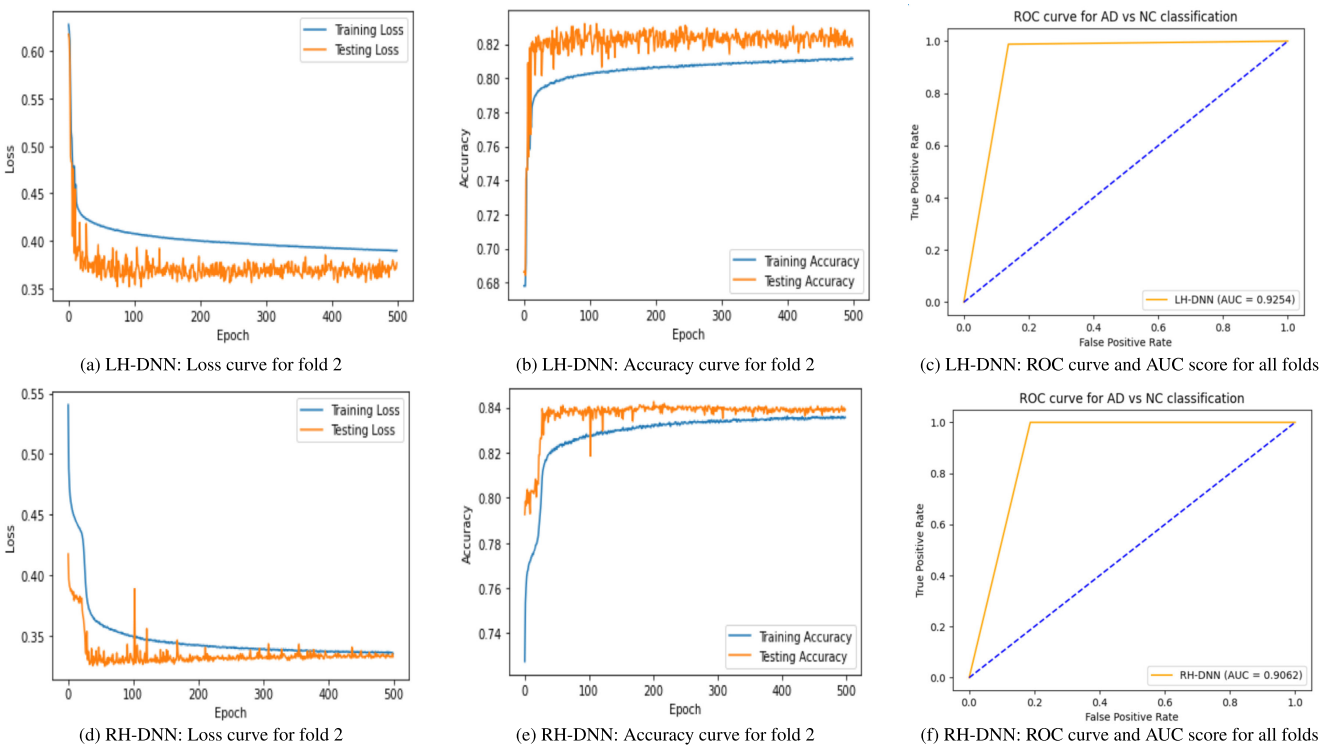
To train the two-stage ensemble localization network model, the mean squared error is considered as the loss function. If q number of sMRI scans are involved in the training process, and the number of patches extracted from each sMRI scan is α, then the mean squared error cost function can be expressed in the following way, where (X_j, Y_j, Z_j) are the target displacement vectors and (X_j', Y_j', Z_j') are the predicted displacement vectors.

$$MSE_{Hippocampus\ localization} = \frac{1}{\alpha * q}\left(\sum_{j=1}^{j=\alpha*q}\right)\left(\frac{1}{3}\left((X_j - X_j')^2 + (Y_j - Y_j')^2 + (Z_j - Z_j')^2\right)\right) \quad (1)$$

TABLE 3. LH-DNN and RH-DNN network architectures used to diagnose Alzheimer's disease.

Model Name	Network Architecture	Activation Function	Optimizer	Cost Function
LH-DNN*	$I_{3x1}, F^{32}, F^{64}, F^{128}F^{256}, F^{512}, F^{2}$	ReLU, Sigmoid(F^2)	Adam	Binary Cross-Entropy
RH-DNN*	$I_{3x1}, F^{32}, F^{64}, F^{128}, F^{256}, F^{512}, F^{2}$	ReLU, Sigmoid(F^2)	Adam	Binary Cross-Entropy

$I_{\text{sample size}}$ = Network input, $F^{(\#\text{filter})}$ = Dense layer.
*LH-DNN and RH-DNN stand for left hippocampal DNN model and right hippocampal DNN model, respectively.
Both LH-DNN and RH-DNN have dropout layer after dense layers (F^{256})(25 %)and (F^{512})(35 %).



(a) LH-DNN: Loss curve for fold 2
(b) LH-DNN: Accuracy curve for fold 2
(c) LH-DNN: ROC curve and AUC score for all folds
(d) RH-DNN: Loss curve for fold 2
(e) RH-DNN: Accuracy curve for fold 2
(f) RH-DNN: ROC curve and AUC score for all folds

FIGURE 4. The representative loss and accuracy curves for Fold 2 are shown above for training and testing sets with ROC curves along with AUC scores for all Folds.

If the numbers of slices/patches contributing to construct the hippocampus from each sMRI scan are β, and each slice/patch was augmented by a factor of n, and if the true and predicted numbers of voxels attributed to the corresponding axial, coronal, and sagittal views are $(A_{X_j}, A_{Y_j}, A_{Z_j})$ and $(A'_{X_j}, A'_{Y_j}, A'_{Z_j})$, respectively, then the cost function for DVE-CNN (LHM and RHM) can be expressed in the following manner.

$$MSE_{Discrete\ volume} = \frac{1}{n*q}\left(\sum_{j=1}^{j=\beta*n*q}\right) * \frac{1}{3}\left((A_{X_j}-A'_{X_j})^2+(A_{Y_j}-A'_{Y_j})^2+(A_{Z_j}-A'_{Z_j})^2\right) \quad (2)$$

If the true distribution of a classification model is y and the estimated distribution by the same model is y', then the cross entropy loss for γ*m number of training samples extracted from each sMRI scan can be expressed in the following way.

$$Cross-Entropy_{AD/NC} = -\left(\frac{1}{\gamma*m*k}\right) * \left(\sum_{j=1}^{j=\gamma*m*k}(y_j log(y'_j)+(1-y_j)log(1-y'_j)\right) \quad (3)$$

Here, γ is the number of observations from the slices/patches used to estimate the volumetric features from each sMRI scan involved in the training process and m is the observation factor of repetition by which the slices/patches have been previously augmented in the discrete volume estimation period. k is the total number of sMRI scans involved in the training of LH-DNN and RH-DNN models.

The training and test operations are performed on an HP Z640 workstation with an Intel(R) Xeon (R) CPU

TABLE 4. The classification accuracy for LH-DNN and RH-DNN models.

Dataset	Model Name	Type	Fold 1		Fold 2		Fold 3		All folds' average result	
			Acc%	WA%	Acc%	WA%	Acc%	WA%	Avg. Acc%	Avg. WA%
GARD	LH-DNN	Training	82.32	86.75	81.15	100	85.99	100	83.15	**94.82**
		Testing	79.73		81.91		67.13		76.26	
	RH-DNN	Training	87.07	84	83.57	100	84.87	98	85.17	**94.02**
		Testing	69.90		83.91		78.21		77.34	

WA% is estimated after averaging the probability scores of the input samples extracted from each sMRI scan of the testing set, whereas, the Acc% is calculated considering each individual input sample of the training and testing sets.

TABLE 5. The precision, recall/sensitivity, F1 score and confusion matrix for the left hippocampus: Fold 1.

Left hippocampus: Fold 1				
Dataset	AD/NC	Precision	Recall/Sensitivity	F1 Score
GARD	AD	100%	58%	0.73%
	NC	84%	100%	91%
Confusion Matrix				
15		11		
0		57		

TABLE 6. The precision, recall/sensitivity, F1 score and confusion matrix for the left hippocampus: Fold 2.

Left hippocampus: Fold 2				
Dataset	AD/NC	Precision	Recall/Sensitivity	F1 Score
GARD	AD	100%	100%	100%
	NC	100%	100%	100%
Confusion Matrix				
26		0		
0		57		

E5-1607 V4 @3.10 GHz (4 CPUs) and 32 GB RAM along with an 8 GB Nvidia Quadro M4000 GPU. The models were implemented using the Keras library (backend: TensorFlow).

IV. EXPERIMENTAL RESULTS

The experimental results for the AD versus NC classification are shown in this section based on the volumetric features of left and right hippocampi extracted by DVE-CNN (LHM and RHM) from the GARD dataset. The comparative analysis has been conducted to validate the proposed method based on the number of matrices explained in the following section.

A. EVALUATION MATRICES

We have used a number of matrices to validate our model and compare it with the state-of-the-art methods. The accuracy, precision, recall/sensitivity, and F1 score are reported for AD versus NC classification. In addition, the value of area under the curve (AUC) has been reported, and the receiver operating curve (ROC) has been plotted for the LH-DNN and RH-DNN models.

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall/Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

$$f1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{7}$$

Here, TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively.

B. WEIGHTED AVERAGE SCORE ESTIMATION

The volumetric features in each patch predicted by the DVE-CNN (LHM and RHM) models consist of multiple augmented patches, where each patch was repeated 15 times by Θ-degree rotation. Therefore, there will be an observation from the proposed model based on each patch. However, the predicted average probability score of the total number of patches generated from each sMRI scan will be considered as the final predicted class for the corresponding sMRI scan. Let us consider the number of patches extracted from each sMRI scan as β, where the corresponding augmentation factor is n. Therefore, the number of observations is γ, where the augmentation factor of the observations is m. Then, the predicted class will be determined in the following manner.

$$[AD, NC] = \underset{[0,1]}{argmax}\left(\frac{1}{\gamma * m}\left(\sum_{i=1}^{i=\gamma*m} C'_{1i}, \sum_{i=1}^{i=\gamma*m} C'_{2i}\right)\right) \tag{8}$$

where C'_1 is the class probability score for the AD subject and C'_2 is the class probability score for the NC subject. The argmax function returns the axis (0/1) of the max value of the weighted average probability score. If the function returns 0, then it is AD, otherwise, it is NC.

C. LEFT HIPPOCAMPAL OBSERVATION

Using the DVE-CNN (LHM) estimated left hippocampal volumetric features attributed to each patch, the trained LH-DNN model has been used to predict the class probability scores. Based on the volumetric features attributed to each patch, the LH-DNN model training and testing accuracy for

TABLE 7. The precision, recall/sensitivity, F1 score and confusion matrix for the left hippocampus: Fold 3.

Left hippocampus: Fold 3				
Dataset	AD/NC	Precision	Recall/Sensitivity	F1 Score
GARD	AD	93%	100%	97%
	NC	100%	96%	98%
Confusion Matrix				
28		0		
2		55		

TABLE 8. The precision, recall/sensitivity, F1 score and confusion matrix for the left hippocampus: all folds.

Left hippocampus: all folds				
Dataset	AD/NC	Precision	Recall/Sensitivity	F1 Score
GARD	AD	97%	86%	91%
	NC	94%	99%	96%
Confusion Matrix				
69		11		
2		169		

TABLE 9. The precision, recall/sensitivity, F1 score and confusion matrix for the right hippocampus: Fold 1.

Right hippocampus: Fold 1				
Dataset	AD/NC	Precision	Recall/Sensitivity	F1 Score
GARD	AD	100%	50%	67%
	NC	81%	100%	90%
Confusion Matrix				
13		13		
0		57		

TABLE 10. The precision, recall/sensitivity, F1 score and confusion matrix for the right hippocampus: Fold 2.

Right hippocampus: Fold 2				
Dataset	AD/NC	Precision	Recall/Sensitivity	F1 Score
GARD	AD	100%	100%	100%
	NC	100%	100%	100%
Confusion Matrix				
26		0		
0		57		

TABLE 11. The precision, recall/sensitivity, F1 score and confusion matrix for the right hippocampus: Fold 3.

Right hippocampus: Fold 3				
Dataset	AD/NC	Precision	Recall/Sensitivity	F1 Score
GARD	AD	100%	93%	96%
	NC	97%	100%	98%
Confusion Matrix				
26		2		
0		57		

TABLE 12. The precision, recall/sensitivity, F1 score and confusion matrix for the right hippocampus: all folds.

Right hippocampus: all folds				
Dataset	AD/NC	Precision	Recall/Sensitivity	F1 Score
GARD	AD	100%	81%	90%
	NC	92%	100%	96%
Confusion Matrix				
65		15		
0		171		

Fold 1, Fold 2, and Fold 3 are shown in Table 4. The average predicted class score for all patches of an sMRI scan has been used to determine the final class label (Weighted Accuracy) for AD versus NC classification. The weighted accuracy values for the testing set of the LH-DNN model for Fold 1, Fold 2, and Fold 3 are 86.75%, 100%, and 100%, respectively. The average weighted accuracy for the testing set of all three folds is 94.82%. The precision, recall/sensitivity and F1 score are reported for Fold 1, Fold 2, and Fold 3 in Tables 5 – 7, respectively. The weighted average values of precision, recall/sensitivity, and F1 score are reported in Table 8.

D. RIGHT HIPPOCAMPAL OBSERVATION

Similarly, for the right hippocampal volumetric features attributed to each patch, the RH-DNN model is used to infer the class probability scores for AD and NC classes. By considering the attributed volumetric features in each patch, the training and testing accuracy are reported for Fold 1, Fold 2, and Fold 3 in Table 4. The weighted average accuracy values for the test sets of Fold 1, Fold 2, and Fold 3 are 84%, 100%, and 98%, respectively. The average weighted accuracy for the test set of Fold 1, Fold 2, and Fold 3 is 94.02%. The precision, recall/sensitivity, and F1 score of the RH-DNN model are reported for Fold 1, Fold 2 and Fold 3 in Tables 9 – 11, respectively. The estimated average weighted values of precision, recall/sensitivity, and F1 score of the RH-DNN model for the whole dataset are shown in Table 12.

E. COMPARISON AND DISCUSSION

The GARD dataset is not available for public usage, therefore, only a few research works have been conducted using

TABLE 13. Comparison between existing methods and the proposed method.

Method	Dataset	Considered Regions/Feature Type	Accuracy%	Precision%	Recall/Sensitivity%	F1 Score %
Gupta et al. [44]	GARD	(HV(LH+RH)+VBM+CSC)*	93.06	90.62	87.87	89.23
S. Ahmed et al. [13]	GARD	LH(Patch)	83.27	76.88	90.50	84.34
		RH(Patch)	81.56	76.88	87.56	81.87
		LH(Patch) + RH(Patch)	86.28	82.95	91.19	86.88
		Ensemble Classifier(Patch)	90.06	91.11	90.22	90.66
Proposed Method(LH-DNN)	GARD	LH(Voxels/Volume)	**94.82**	**97**	86	**91**
Proposed Method(RH-DNN)	GARD	RH(Voxels/Volume)	**94.02**	**100**	81	90

HV= Hippocampal Volume, LH= Left hippocampus, RH= Right hippocampus,
VBM= Voxel-based morphometry, CSC= Cortical and subcortical volumetric features

this dataset. In addition, all scanned subjects (male and female) are of Korean origin. The inter-subject racial variability does not exist, therefore, the dataset is homogeneous in nature. We have performed 3-fold cross validation and reported the accuracy results for left and right hippocampi in Table 4 and the precision, recall/sensitivity, and F1 score in Tables 5 - 8 and Tables 9 - 12. Moreover, we have compared our proposed approach with two other methods [13], [44] which were published in the literature in 2019 to differentiate the AD class from NC for the GARD dataset. We have reported the comparative results in Table 13.

The proposed method obtained average weighted accuracy values of 94.85% and 94.02% for the extracted volumetric features from each corresponding slice attributed to the left and right hippocampi, respectively. Gupta *et al.* [44] have reported accuracy of 93.06% by using both the left and right hippocampi's features along with voxel-based morphometry and cortical and sub-cortical volumetric features. On the other hand, Ahamed *et al.* [13] have proposed an ensemble-based classifier to distinguish the AD and NC classes from the GARD dataset and obtained accuracy of 90.06%. Our proposed approach has outperformed both methods in the context of accuracy. In addition, Gupta *et al.* have used multiple features to improve the accuracy, where multiple software packages such as FreeSurfer and SPM have been used to extract features, which may require a large amount of time. Moreover, Ahmed *et al.* have used manual localization, which is a complicated task. The proposed approach is completely automatic and requires comparatively less time than the other two methods.

The reported values of precision, recall, and F1 score of the proposed approach are either superior or comparable to those of the other methods. Furthermore, the confusion matrices have been reported for each individual fold, as well as for the whole dataset. It can be determined that out of 80 ADD subjects, 69 were correctly classified using left hippocampal volumetric information, whereas 65 were identified correctly using right hippocampal volumetric information. The proposed method achieved 86.25% and 81.25% accuracy on diagnosing the AD class using left and right hippocampal volumetric information in the GARD dataset, respectively. However, it is also seen that the proposed method obtained low accuracy in the first fold for both the left and right hippocampal volumetric data in comparison with the other two folds. This may be due to the low correlation with the manually measured volumetric data estimated by the automatic DVE-CNN(LHM and RHM) method. However, the obtained results for both hippocampi are comparable to the state-of-the-art literature on this dataset.

The proposed DNN model with Hough-CNN and DVE-CNN models offers a completely automatic system to classify the AD and NC classes. In addition, hippocampal volume atrophy is an important biomarker used in this study that proved the viability of the biomarker. Furthermore, using single features, we have achieved better results than the multi-featured model proposed by Gupta *et al.* [44], as well as the ensemble-based classifier proposed by Ahmed *et al.* [13].

F. LIMITATION

Although the proposed approach offers high performance in classifying AD versus NC classes, there are still a few limitations that must be addressed. The proposed method highly depends on the accuracy of the previously proposed automatic localization (Hough-CNN) and discrete volume estimation (DVE-CNN (LHM and RHM)) methods, where the performances of those two methods can be problematic for the current method. In addition, the proposed method has been validated on a relatively small private dataset. However, the embedded hidden correlation of volume reduction with Alzheimer's disease is useful to diagnose AD, which is shown by the proposed approach.

V. CONCLUSION

In this research paper, we have proposed an aggregated approach of Hough-CNN, CNN and DNN models to diagnose Alzheimer's disease based on volumetric features from sMRI data. The proposed DNN model used the volumetric features extracted by the DVE-CNN model to classify the AD and NC classes. We reported the obtained average weighted accuracy values of 94.82% and 94.02% based on the volumetric

features attributed to left and right hippocampi of the GARD dataset, respectively. Three fold cross validation has been performed, demonstrating the obtained results. The proposed method achieved AUC values of 92.54% and 90.62% for the left and right hippocampal data, respectively. Our method successfully diagnosed 69 subjects and 65 subjects from 80 AD subjects using left and right hippocampal volumetric features, respectively. The advantage of the proposed method is that it is fully automatic and obtains comparatively higher accuracy than the other methods proposed in the literature on the same dataset. Moreover, it is also shown that the slice-wise volumetric features of the hippocampus are important biomarkers, and that it is possible to diagnose Alzheimer's disease using the volumetric features. However, in the future, we will apply our method on large open source datasets along with data of other modalities, such as positron emission tomography (PET), and when combining the volumetric features of left and right hippocampi with other regional features, such as VBM and CSC.

REFERENCES

[1] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, "Forecasting the global burden of Alzheimer's disease," *Alzheimer's Dementia*, vol. 3, no. 3, pp. 186–191, Jul. 2007.

[2] K. Y. Liang, M. A. Mintun, A. M. Fagan, A. M. Goate, J. M. Bugg, D. M. Holtzman, J. C. Morris, and D. Head, "Exercise and Alzheimer's disease biomarkers in cognitively normal older adults," *Ann. Neurol.*, vol. 68, no. 3, pp. 311–318, Sep. 2010.

[3] X. Hong, R. Lin, C. Yang, N. Zeng, C. Cai, J. Gou, and J. Yang, "Predicting Alzheimer's disease using LSTM," *IEEE Access*, vol. 7, pp. 80893–80901, 2019.

[4] A. Caroli and G. B. Frisoni, "The dynamics of Alzheimer's disease biomarkers in the Alzheimer's disease neuroimaging initiative cohort," *Neurobiol. Aging*, vol. 31, no. 8, pp. 1263–1274, Aug. 2010.

[5] A. Olsson, H. Vanderstichele, N. Andreasen, G. De Meyer, A. Wallin, B. Holmberg, L. Rosengren, E. Vanmechelen, and K. Blennow, "Simultaneous measurement of β-Amyloid(1–42), total tau, and phosphorylated tau (Thr181) in cerebrospinal fluid by the xMAP technology," *Clin. Chem.*, vol. 51, no. 2, pp. 336–345, Feb. 2005.

[6] H. Hampel, K. Bürger, S. J. Teipel, A. L. W. Bokde, H. Zetterberg, and K. Blennow, "Core candidate neurochemical and imaging biomarkers of Alzheimer's disease," *Alzheimer's Dementia*, vol. 4, no. 1, pp. 38–48, Jan. 2008.

[7] I. Krashenyi, A. Popov, J. Ramirez, and J. M. Gorriz, "Application of fuzzy logic for Alzheimer's disease diagnosis," in *Proc. Signal Process. Symp. (SPSympo)*, 2015, pp. 1–4.

[8] D. Brambilla, B. Le Droumaguet, J. Nicolas, S. H. Hashemi, L.-P. Wu, S. M. Moghimi, P. Couvreur, and K. Andrieux, "Nanotechnologies for Alzheimer's disease: Diagnosis, therapy, and safety issues," *Nanomedicine, Nanotechnol., Biol. Med.*, vol. 7, no. 5, pp. 521–540, Oct. 2011.

[9] L. Xu, Z. Yao, J. Li, C. Lv, H. Zhang, and B. Hu, "Sparse feature learning with label information for Alzheimer's disease classification based on magnetic resonance imaging," *IEEE Access*, vol. 7, pp. 26157–26167, 2019.

[10] C. Feng, A. Elazab, P. Yang, T. Wang, F. Zhou, H. Hu, X. Xiao, and B. Lei, "Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM," *IEEE Access*, vol. 7, pp. 63605–63618, 2019.

[11] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, and M. Filippi, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage: Clin.*, vol. 21, 2019, Art. no. 101645.

[12] 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, Dec. 2019.

[13] S. Ahmed, K. Y. Choi, J. J. Lee, B. C. Kim, G.-R. Kwon, K. H. Lee, and H. Y. Jung, "Ensembles of patch-based classifiers for diagnosis of alzheimer diseases," *IEEE Access*, vol. 7, pp. 73373–73383, 2019.

[14] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proc. Workshop Unsupervised Transf. Learn. (ICML)*, 2012, pp. 17–36.

[15] Y. Xian, T. Lorenz, B. Schiele, and Z. Akata, "Feature generating networks for zero-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5542–5551.

[16] L. Zhang and S. Rusinkiewicz, "Learning to detect features in texture images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6325–6333.

[17] J. Spencer, R. Bowden, and S. Hadfield, "Same features, different day: Weakly supervised feature learning for seasonal invariance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6459–6468.

[18] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*. [Online]. Available: http://arxiv.org/abs/1503.02531

[19] J. Yim, D. Joo, J. Bae, and J. Kim, "A gift from knowledge distillation: Fast optimization, network minimization and transfer learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4133–4141.

[20] S. Chen, C. Zhang, and M. Dong, "Coupled end-to-end transfer learning with generalized Fisher information," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4329–4338.

[21] Z. Cao, M. Long, J. Wang, and M. I. Jordan, "Partial transfer learning with selective adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2724–2732.

[22] K.-H. Lee, X. He, L. Zhang, and L. Yang, "CleanNet: Transfer learning for scalable image classifier training with label noise," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5447–5456.

[23] A. Basher, B. C. Kim, K. H. Lee, and H. Y. Jung, "Automatic localization and discrete volume measurements of hippocampi from MRI data using a convolutional neural network," *IEEE Access*, vol. 8, pp. 91725–91739, 2020.

[24] A. Basher, K. Y. Choi, J. J. Lee, B. Lee, B. C. Kim, K. H. Lee, and H. Y. Jung, "Hippocampus localization using a two-stage ensemble Hough convolutional neural network," *IEEE Access*, vol. 7, pp. 73436–73447, 2019.

[25] B. Jie, M. Liu, J. Liu, D. Zhang, and D. Shen, "Temporally constrained group sparse learning for longitudinal data analysis in Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 238–249, Jan. 2017.

[26] K. Y. Choi, J. J. Lee, T. I. Gunasekaran, S. Kang, W. Lee, J. Jeong, H. J. Lim, X. Zhang, C. Zhu, S. Y. Won, and Y. Y. Choi, "Apoe promoter polymorphism-219t/g is an effect modifier of the influence of apoe ε4 on alzheimer's disease risk in a multiracial sample," *J. Clin. Med.*, vol. 8, no. 8, p. 1236, 2019.

[27] S. De Santi, M. J. de Leon, H. Rusinek, A. Convit, C. Y. Tarshish, A. Roche, W. H. Tsui, E. Kandil, M. Boppana, K. Daisley, G. J. Wang, D. Schlyer, and J. Fowler, "Hippocampal formation glucose metabolism and volume losses in MCI and AD," *Neurobiol. Aging*, vol. 22, no. 4, pp. 529–539, Jul. 2001.

[28] A. M. Fjell, K. B. Walhovd, C. Fennema-Notestine, L. K. McEvoy, D. J. Hagler, D. Holland, J. B. Brewer, and A. M. Dale, "CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer's disease," *J. Neurosci.*, vol. 30, no. 6, pp. 2088–2101, Feb. 2010.

[29] L. M. Shaw, H. Vanderstichele, M. Knapik-Czajka, C. M. Clark, P. S. Aisen, R. C. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, R. Dean, E. Siemers, W. Potter, V. M.-Y. Lee, and J. Q. Trojanowski, "Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects," *Ann. Neurol.*, vol. 65, no. 4, pp. 403–413, Apr. 2009.

[30] N. Mattsson, H. Zetterberg, O. Hansson, N. Andreasen, L. Parnetti, M. Jonsson, S. K. Herukka, W. M. van der Flier, M. A. Blankenstein, M. Ewers, and K. Rich, "CSF biomarkers and incipient alzheimer disease in patients with mild cognitive impairment," *Jama*, vol. 302, no. 4, pp. 385–393, 2009.

[31] F. H. Bouwman, W. M. van der Flier, N. S. M. Schoonenboom, E. J. van Elk, A. Kok, F. Rijmen, M. A. Blankenstein, and P. Scheltens, "Longitudinal changes of CSF biomarkers in memory clinic patients," *Neurology*, vol. 69, no. 10, pp. 1006–1011, Sep. 2007.

[32] J. C. Morris, M. Storandt, J. P. Miller, D. W. McKeel, J. L. Price, E. H. Rubin, and L. Berg, "Mild cognitive impairment represents early-stage alzheimer disease," *Arch. Neurol.*, vol. 58, no. 3, pp. 397–405, Mar. 2001.

[33] L. K. McEvoy, C. Fennema-Notestine, J. C. Roddey, D. J. Hagler, D. Holland, D. S. Karow, C. J. Pung, J. B. Brewer, and A. M. Dale, "Alzheimer disease: Quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment," *Radiology*, vol. 251, no. 1, pp. 195–205, Apr. 2009.

[34] F. Ren, C. Yang, Q. Qiu, N. Zeng, C. Cai, C. Hou, and Q. Zou, "Exploiting discriminative regions of brain slices based on 2D CNNs for Alzheimer's disease classification," *IEEE Access*, vol. 7, pp. 181423–181433, 2019.

[35] S. Sarraf, D. D. Desouza, J. A. E. Anderson, and C. Saverino, "MCADNNet: Recognizing stages of cognitive impairment through efficient convolutional fMRI and MRI neural network topology models," *IEEE Access*, vol. 7, pp. 155584–155600, 2019.

[36] X. Fang, Z. Liu, and M. Xu, "Ensemble of deep convolutional neural networks based multi-modality images for Alzheimer's disease diagnosis," *IET Image Process.*, vol. 14, no. 2, pp. 318–326, Feb. 2020.

[37] R. Bansal, A. J. Gerber, and B. S. Peterson, "Brain morphometry using anatomical magnetic resonance imaging," *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 47, no. 6, p. 619, 2008.

[38] J. C. Baron, G. Chételat, B. Desgranges, G. Perchey, B. Landeau, V. de la Sayette, and F. Eustache, "*in vivo* mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease," *NeuroImage*, vol. 14, no. 2, pp. 298–309, Aug. 2001.

[39] S. Kloppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. J. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681–689, Feb. 2008.

[40] C. R. Jack, R. C. Petersen, P. C. O'Brien, and E. G. Tangalos, "Mr-based hippocampal volumetry in the diagnosis of Alzheimer's disease," *Neurology*, vol. 42, no. 1, p. 183, 1992.

[41] M. Atiya, B. T. Hyman, M. S. Albert, and R. Killiany, "Structural magnetic resonance imaging in established and prodromal Alzheimer disease: A review," *Alzheimer Disease Associated Disorders*, vol. 17, no. 3, pp. 177–195, Jul. 2003.

[42] R. Wolz, P. Aljabar, J. V. Hajnal, J. Lötjönen, and D. Rueckert, "Nonlinear dimensionality reduction combining MR imaging with non-imaging information," *Med. Image Anal.*, vol. 16, no. 4, pp. 819–830, May 2012.

[43] 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.

[44] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, and G. R. Kwon, "Early diagnosis of Alzheimer's disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of MRI t1 brain images," *PLoS ONE*, vol. 14, no. 10, Oct. 2019, Art. no. e0222446.

[45] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, "BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, Feb. 2017.

[46] E. R. Mulder, R. A. de Jong, D. L. Knol, R. A. van Schijndel, K. S. Cover, P. J. Visser, F. Barkhof, and H. Vrenken, "Hippocampal volume change measurement: Quantitative assessment of the reproducibility of expert manual outlining and the automated methods FreeSurfer and FIRST," *NeuroImage*, vol. 92, pp. 169–181, May 2014.

[47] J. L. Winterburn, J. C. Pruessner, S. Chavez, M. M. Schira, N. J. Lobaugh, A. N. Voineskos, and M. M. Chakravarty, "A novel *in vivo* atlas of human hippocampal subfields using high-resolution 3T magnetic resonance imaging," *NeuroImage*, vol. 74, pp. 254–265, Jul. 2013.

[48] J. Dolz, C. Desrosiers, and I. Ben Ayed, "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study," *NeuroImage*, vol. 170, pp. 456–470, Apr. 2018.

[49] C. Wachinger, M. Reuter, and T. Klein, "DeepNAT: Deep convolutional neural network for segmenting neuroanatomy," *NeuroImage*, vol. 170, pp. 434–445, Apr. 2018.

[50] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel, and N. Navab, "Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound," *Comput. Vis. Image Understand.*, vol. 164, pp. 92–102, Nov. 2017.

[51] Y. Shao, J. Kim, Y. Gao, Q. Wang, W. Lin, and D. Shen, "Hippocampal segmentation from longitudinal infant brain MR images via classification-guided boundary regression," *IEEE Access*, vol. 7, pp. 33728–33740, 2019.

[52] N. J. Tustison, P. A. Cook, A. Klein, G. Song, S. R. Das, J. T. Duda, B. M. Kandel, N. van Strien, J. R. Stone, J. C. Gee, and B. B. Avants, "Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements," *NeuroImage*, vol. 99, pp. 166–179, Oct. 2014.

[53] B. B. Avants, N. J. Tustison, J. Wu, P. A. Cook, and J. C. Gee, "An open source multivariate framework for n-Tissue segmentation with evaluation on public data," *Neuroinformatics*, vol. 9, no. 4, pp. 381–400, Dec. 2011.

[54] S. R. Das, B. B. Avants, M. Grossman, and J. C. Gee, "Registration based cortical thickness measurement," *NeuroImage*, vol. 45, no. 3, pp. 867–879, Apr. 2009.

[55] L. Yue, X. Gong, J. Li, H. Ji, M. Li, and A. K. Nandi, "Hierarchical feature extraction for early Alzheimer's disease diagnosis," *IEEE Access*, vol. 7, pp. 93752–93760, 2019.

[56] S. Afzal, M. Maqsood, F. Nazir, U. Khan, F. Aadil, K. M. Awan, I. Mehmood, and O.-Y. Song, "A data augmentation-based framework to handle class imbalance problem for Alzheimer's stage detection," *IEEE Access*, vol. 7, pp. 115528–115539, 2019.

[57] K. Trojachanec, I. Kitanovski, I. Dimitrovski, and S. Loshkovska, "Longitudinal brain MRI retrieval for Alzheimer's disease using different temporal information," *IEEE Access*, vol. 6, pp. 9703–9712, 2018.

[58] M. N. I. Qureshi, S. Ryu, J. Song, K. H. Lee, and B. Lee, "Evaluation of functional decline in Alzheimer's dementia using 3D deep learning and group ICA for RS-fMRI measurements," *Frontiers Aging Neurosci.*, vol. 11, pp. 1–9, Feb. 2019, Art. no. 8.

[59] N. T. Duc, S. Ryu, M. N. I. Qureshi, M. Choi, K. H. Lee, and B. Lee, "3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI," *Neuroinformatics*, vol. 18, no. 1, pp. 71–86, Jan. 2020.

[60] G. M. McKhann, D. S. Knopman, H. Chertkow, C. R. Jack Jr., C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub, C. H. Phelps "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 263–269, 2011.

[61] M. S. Albert, S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, N. C. Fox, A. Gamst, D. M. Holtzman, W. J. Jagust, R. C. Petersen, P. J. Snyder, M. C. Carrillo, B. Thies, and C. H. Phelps, "The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 270–279, May 2011.

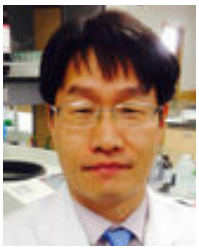
[62] R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack, Jr., J. Kaye, T. J. Montine, and D. C. Park, "Toward defining the preclinical stages of alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 280–292, 2011.

[63] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 1–8.

[64] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[65] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

ABOL BASHER received the B.S. degree in electrical and electronics engineering from the Mymensingh Engineering College, University of Dhaka, Bangladesh, in 2015, and the M.S. degree in computer engineering from Chosun University, Gwangju, South Korea. He is currently working as a Research Assistant with the Computer Vision Laboratory, Chosun University. His research interests include medical image processing, computer vision, and deep learning.



BYEONG C. KIM received the M.D., M.S., and Ph.D. degrees from the Department of Neurology, Chonnam National University Medical School, Gwangju, Republic of Korea, in 1991, 1994, and 2000, respectively. He is currently a Professor with the Department of Neurology, Chonnam National University Medical School. His current research interests include structural and molecular neuroimaging and CSF biomarkers in patients with neurodegenerative diseases.

KUN HO LEE received the B.S. degree from the Department of Genetic Engineering, Korea University, Seoul, Republic of Korea, in 1989, and the M.S. and Ph.D. degrees from the Department of Molecular Biology, Seoul National University, Seoul, in 1994 and 1998, respectively. He is currently an Associate Professor with the Department of Biomedical Science, Chosun University, Gwangju, Republic of Korea. He also works with the National Research Center for Dementia, Chosun University. His current research interests include brain image analysis and the development of prediction model for neurodegenerative diseases based on MRI and genetic variants.

HO YUB JUNG received the B.S. degree in electrical engineering from The University of Texas at Austin, in 2002, and the M.S. and Ph.D. degrees in electrical engineering and computer science from Seoul National University, in 2006 and 2012, respectively. He was with Samsung Electronics as a Senior Engineer, for two years. Since 2017, he has been an Assistant Professor with the Department of Computer Engineering, Chosun University. His research interests include computer vision, machine learning, and medical imaging.