NEUROSCIENCE
RESEARCH ARTICLE
A. Mehmood et al. / Neuroscience 460 (2021) 43–52

A Transfer Learning Approach for Early Diagnosis of Alzheimer's Disease on MRI Images ☆

Atif Mehmood,[a] Shuyuan Yang,[a]* Zhixi Feng,[a] Min Wang,[b] AL Smadi Ahmad,[a] Rizwan Khan,[c] Muazzam Maqsood[d] and Muhammad Yaqub[e]

[a] *School of Artificial Intelligence, Xidian University, Xi'an 710071, China*

[b] *Key Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China*

[c] *School of Electronic Information and Communications, HUST University, Wuhan 4370074, China*

[d] *Department of Computer Science, COMSATS University Islamabad, Attock Campus, Attock 43600, Pakistan*

[e] *Faculty of Information Technology, Beijing University of Technology, Beijing 10000, China*

Abstract—Mild cognitive impairment (MCI) detection using magnetic resonance image (MRI), plays a crucial role in the treatment of dementia disease at an early stage. Deep learning architecture produces impressive results in such research. Algorithms require a large number of annotated datasets for training the model. In this study, we overcome this issue by using layer-wise transfer learning as well as tissue segmentation of brain images to diagnose the early stage of Alzheimer's disease (AD). In layer-wise transfer learning, we used the VGG architecture family with pre-trained weights. The proposed model segregates between normal control (NC), the early mild cognitive impairment (EMCI), the late mild cognitive impairment (LMCI), and the AD. In this paper, 85 NC patients, 70 EMCI, 70 LMCI, and 75 AD patients access form the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Tissue segmentation was applied on each subject to extract the gray matter (GM) tissue. In order to check the validity, the proposed method is tested on preprocessing data and achieved the highest rates of the classification accuracy on AD vs NC is 98.73%, also distinguish between EMCI vs LMCI patients testing accuracy 83.72%, whereas remaining classes accuracy is more than 80%. Finally, we provide a comparative analysis with other studies which shows that the proposed model outperformed the state-of-the-art models in terms of testing accuracy.

Key words: Transfer learning, Alzheimer's disease, Image classification, Early diagnosis.

INTRODUCTION

Alzheimer's disease (AD) is a kind of brain disease, causing dementia in the aged population. It is thought to begin 15–20 years before syndromes arise. Syndromes occur due to the destruction of neurons involved in memory, thinking, and learning functions (Wee et al., 2013). Over time, syndromes tend to escalate and become intrusive with performing daily activities such as planning family events, walking, and skill loss. At this stage, cognitive decline is said to have dementia due to Alzheimer's disease. The small changes in the brain that progress normal control (NC) to mild cognitive impairment (MCI) and ultimately reaches the last stage of AD (Zeng et al., 2018). AD is the 6th leading cause of death in the united states, official accounting for 121,404 deaths in 2017. It is predicted that 60 million people will be affected by AD in the next 20 years. According to the World Alzheimer's Report, it will grow to 152 million patients in 2050 (Oh et al., 2019). The total estimated cost for long term health care for dementia patients is about $290 billion. Researchers are ongoing early detection of AD to slow down the abnormal degeneration of neurons of the brain. it also produced the emotional and financial benefit for the patient family (Mehmood et al., 2020). Brain imaging modalities used for AD diagnose, such as functional magnetic resonance imaging (fMRI), magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT), positron emission tomography (PET), and computed tomography (CT). If we compare these

☆ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/howtoapply/ADNIAcknowledgementList.pdf.

*Corresponding author.
E-mail address: syyang@xidian.edu.cn (S. Yang).

Abbreviations: EMCI, Early Mild Cognitive Impairment; LMCI, Late Mild Cognitive Impairment; GM, Gray Matter; WM, White Matter; LMCI, Mild cognitive impairment; MRI, Magnetic resonance image; CNN, Convolutional Neural Network.

https://doi.org/10.1016/j.neuroscience.2021.01.002
0306-4522/

modalities MRI images generally available in a standardized form for clinical practice (Bi et al., 2019). The researcher developed functional connectivity modeling for AD diagnose, such as sparse representation method, graphical methods, and partial correlation-based technique (Yue et al., 2019). The cortical thickness, as well as gray matter density, ventricles enlargements, and brain atrophy, are used by researchers. On the other hand, three main tissue in brain images such as white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) is of fundamental importance. In contrast, researchers found GM atrophy correlates more with cognitive decline in MCI (Khedher et al., 2015).

Mild cognitive impairment (MCI) is an intermediate point to damage the memory neurons, more likely to progress dementia due to AD. The investigated six-year conversion rate between MCI to AD is 80%, respectively (Wang et al., 2019). It is an ongoing topic for AD-related researchers to identify MCI patients that are further divided into two stages, such as early mild cognitive impairment (EMCI) and late mild cognitive impairment (LMCI) (Wang et al., 2010). The diagnosis at an early stage of NC and MCI provides the information to clinicians for treatment and take decisions on time. It was also helpful to reduce costs and offer longtime care (Ahmed et al., 2017).

Previous researcher studies have shown that the machine learning algorithm predicts better results for the classification of AD as compared to clinicians. The early achievement of AD classification has been demonstrated by the support vector machine (SVM) (López et al., 2011). Recently, deep learning-based methods such as sparse autoencoder and convolutional neural network (CNN) provide optimal solutions for classification in many domains such as computer vision, speech recognition, and natural language processing (Xu et al., 2019). However, deep learning methods have some limitations during training the model on scratch data because the model required a massive amount of annotated medical images. Due to privacy and cost issues, a vast amount of annotated data availability complicated, alternative solution to overcome this issue by using transfer learning techniques for classification on medical scans (Khan et al., 2019). The concept behind transfer learning is to use the pre-trained model on different problems with a smaller dataset (Liu et al., 2019).

In this paper, we investigate the transfer learning framework, which is based on the most profound CNN architecture for classification of Alzheimer's images into four classes: NC, EMCI, LMCI, and AD. The fundamental motivation behind transfer learning is to transfer features from nature images to Alzheimer's images and introduce the new technique for the classification of AD, which can assist the fresh physicians in creating objective opinion and correct diagnosis. Our primary purpose of getting state-of-the-art results by using a smaller quantity of a dataset without overfitting. To fulfill this requirement, we used the data augmentation technique (Mehmood et al., 2020), which helps us to avoid the overfitting problem, and we achieve the desired results (Hernández-García and König, 2018). We apply layer-wise transfer learning on a deep CNN architecture, where we redesign the last fully connected layer and classifier layer. The proposed model is divided into two groups, gradually trained on some layers, whereas the rest of the others are frozen. Applying transfer learning in this way, we predict the best results on binary classification such as NC, EMCI, LMCI, and AD. Another prominent problem faced in previous studies is to overcome the less training data issue and check the robustness of transfer learning and avoid overfitting. This study is based on GM scans obtain from MRI which correlate more with cognitive performance to help out the early diagnosis of AD.

RELATED WORK

In the last decades, many types of modalities are used for disease prediction in medical fields. Positron Emission Tomography (PET), MRI, and Diffusion Tensor Image (DTI) are used by the researcher in Alzheimer's neuroimaging tools for classification of AD stages (McGeer, 1986). Recently many development ongoings in the field of computer vision to extract useful features by using a machine learning algorithm and developing models for the detection and classification of Alzheimer's disease. These models are working on manually designed features, for this purpose required professional expertise and the need to allocate maximum resources. These approaches are divided into three main categories such as support vector machine (SVM), regression-based, and Bayesian methods (Chaddad et al., 2018). The SVM approach is generally used for classification purposes. Many researchers used SVM to find out the MCI conversion rate. Young et al. (2013) have been developed gaussian processes for predicting stable mild cognitive impairment (sMCI). Experimental results have been shown 74% accuracy for the prediction of AD conversion between three years of sMCI and converted MCI(cMCI). In Badakhshannoory and Saeedi (2011) this study random forest classification algorithm used for MCI classification and achieved 82.3% accuracy. Wang et al. (2010) described during the training of these models many shortcomings occur, several machine learning algorithms perform better results on binary classification, but accuracy declined when applies on multi-classification images.

Recently, deep learning (DL) techniques overcome the limitation for many medical computer-aided diagnosis (CAD) systems, to extract the discriminative features automatically on the raw image data. In end to end learning four major steps involved to make an accurate prediction of diseases such as feature extraction, segmentation, skull stripping, normalization, and smoothing (Hosseini-Asl et al., 2016). Many architectures have been demonstrating classification results on 1000 categories in the ImageNet dataset (Deng et al., 2009). The initial won the ImageNet challenge with a seven-layer convolutional neural network and developed efficient GPU implementation. They produced a 10% improvement as compared to the previous winner. He et al. (2016) developed the Deep Residual Network

(DRN) to solved the degradation of training accuracy. CNN requires a large number of training data, which is difficult to apply directly on medical imaging due to the shortage of annotated datasets (Kingma and Ba, 2019). Suk et al. (n.d.) produced promising results on binary classification such as MCI vs NC, MCI converter, and stable MCI by using a deep Boltzmann machine(DBM). They obtained 95.35% accuracy using MRI and PET modalities.

(Gupta et al., 2013) has been developed as a key technique for AD classification. They used a sparse autoencoder on natural image for learning the set of bases and convolution applied for feature extraction on MRI scans. The diagnostic classification in three categories: i) AD versus NC, ii) MCI versus NC, and iii) AD versus MCI. In each task produced superior classification results such as 94.74%, 86.35%, and 88.10%. Payan and Montana (2015) introduced the combination of a sparse autoencoder and convolutional neural networks (CNNs) to improved AD classification results. They investigated 2D and 3D convolution and obtained an accuracy of the system was 92.11% on NC vs MCI classification. In other recent work (Li et al., 2015), the same classification problem has been investigated to identify different AD stages on two modalities. Anthimopoulos et al. (2016) used CNN architecture and shown an average success rate of 85.61% and consumed significant efforts for labeling the training data. However, when only a small training dataset available of medical scans, to create the overfitting problems (Lyndon et al., 2015). Shin et al. (2016) shown the impact of transfer learning when applied to medical image classification. During an investigation on different modalities, they shown the fine-tuning processes produced outperforms results. Chen et al. (2015) introduced the transfer learning strategy applied to localize plans in ultrasound scans, which can transfer the knowledge on fewer layers. In researcher (Maqsood et al., 2019) developed a transfer learning technique by utilizing a pre-trained model for multi-class classification of AD. They achieved a 92.80% success rate on un-segmented scans. Aderghal et al. (2018) proposed a cross model transfer learning technique to reduce the overfitting issue during less number of training data. They trained the model on structural MRI and transferred on the diffusion tensor imaging (DTI) dataset. They have been investigated the model on two modalities and attained a 92% performance rate on NC vs AD, 85% AD vs MCI, and 80% on MCI vs NC. In (Phong et al., 2017) researcher proposed three models based on LeNet, Inception ResNet and GoogLeNet. During the training phase, they train only fully connected layers of two models instead of scratch but LeNet train all layers on medical images. They achieved very promising results in terms of accuracy of 99.70%, 98.20%, and 99.20%.

EXPERIMENTAL PROCEDURES

Image dataset

Individuals data used in this study were collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) publically available database (http://adni.loni.usc.edu). ANDI began in 2004 with the help of a public–private partnership under control of Dr. Michael W. Weiner. The primary aims of ADNI to analyze more authentic and sensitive techniques on different biomarkers such as MRI, PET, structural magnetic resonance imaging (sMRI), and clinical assessment to measure the progression of MCI and early stages of AD (Jack et al., 2019). Secondly, introduced the new innovative data-access policy without restraint to all researchers in the world. In this research work, we used 300 T1-weighted MRI subjects, and all demographic information related to four groups such as normal control (NC), EMCI, LMCI, and AD are shown in Table 1.

Dataset preprocessing operations

In this research work, we applied a complete pipeline for preprocessing on the T1-weighted images taken from the ADNI database. We used the statistical parameter mapping (SPM12; https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) for preprocessing and all data in neuroimaging informatics technology initiative (NIFTI) format. Our work focuses on gray matter (GM) because GM segmentation of the brain would be useful to demonstrate early changes in sporadic AD. During preprocessing segmentation, applied on brain data and dividing them into three major parts such as WM, GM, and CSF (Young et al., 2013). During processing the bias regularization set on very light regularization (0.0001), bias full width at half maximum (FWHM) is 60 mm cutoff, and the ICBM space template is used for affine regularization on all datasets. We used the Montreal Neurological Institute (MNI) space for spatial normalization. In this study image, voxel size is (2 2 2), and finally, Gaussian kernel used for smoothing the images. The shape of the data samples after segmentation is 256 $\times$ 240. We resized all images and get the final images in the form of 224 $\times$ 224 that is used for training and testing in our proposed model.

Convolutional neural networks and transfer learning

Convolutional neural networks (CNNs) are multilayered structures working in a group form. These multilayered include convolution layer, pooling layer, number of consecutive fully connected, and lastly softmax layer. The main mechanism of CNNs to extract the local features with convolution layers from input data. These low-level features are extracted through intermediate layers and used in pattern recognition problems to build high-level features (Sezer and Sezer, 2019). In artificial neurons, each neuron is connected to the next layer of over-weighted connections. The CNNs mechanism can increase the depth and breadth size of those images which have a complex structure (Ieracitano et al., 2019). For the reduction of computational complexity, another important CNN parameter is pooling layers, which is mostly used with nonlinear function in the form of max and min pooling. The pooling layer provides another benefit in term of prevention of overfitting in the model because the amount of computation and parameters are reduced (Feng et al., 2019). In many studies, the max-

Table 1. Demographic and clinical information from the ADNI dataset. Total of 300 patients data used for this study based on four classes. N shows the number of subjects in each class. M and F represent the male and female subjects, ± standard deviations, and mini-mental state examination (MMSE) score

	NC	EMCI	LMCI	AD
N	85	70	70	75
Age	72.13 ± 8.4	73 ± 7.60	72.15 ± 8.20	74 ± 9.25
Gender [M/F]	50/35	40/30	42/28	45/25
MMSE	28.4 ± 1.24	28 ± 1.5	27.5 ± 1.74	23.5 ± 2.15

pooling layer is commonly used with an activation function. In this study rectified linear unit (RELU) activation function used because it converts the negative values of the feature into zero and improved the speed of convergence of CNN.

The modern CNNs based model is manually designed by researchers with several different layers and optimization approaches. During training the model with varying parameters and learning rate, batch size, and weight decay over ImageNet dataset (Krizhevsky and Sutskever, n.d). Generally, in CNN lower layers can produce the general feature extraction ability, and higher layers capable to carry more relative information related to the specific classification task (Chougrad et al., 2019). Transfer learning has been produced promising results on medical images such as classification of precancerous disease, cardiac images, and lung disease classification. In researcher introduced the technique for classification of medical imaging by using CNN and transfer learning. All these results have been shown by the transfer learning produced the high accuracy for classification in medical domains and also achieved maximum results on AD classification with less number of the dataset (Liu et al., 2018).

Proposed transfer learning model

GroupA: block 1–3 are freezing.

GroupB: block 1–4 are freezing.

Due to promising results by CNN, many well-established models have been developed by researchers to solve the binary and multi-class classification problems. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) benchmark provides a big breakthrough in object recognition. The major challenge is to classify the 1000 different objects. We investigate those architectures which are the winner of this challenge for the classification of objects. In this study, we proposed a transfer learning model by customizing VGG family architecture (Mehmood et al., 2020). The reason behind choosing the VGG-19 architecture because it produced high accuracy results and more effective performance on computer-aided diagnosis problems. VGG-19 network which includes the 16 convolutional layers, 5 max-pooling layers with stride 2, and three fully-connected layers with a final softmax layer. We modify the last two fully connected layers and final classification layers as per our problem. These two fully connected layers are 1000 and 512 with binary classification. Secondly, we apply transfer learning to freeze the convolutional layers. In many applications during the transfer learning process only focus on trained fully connected layer on training data and convolutional layers are kept fixed. However, in our proposed model, we divide our model into two groups and progressively frozen the blocks of layers and training on with and without augmentation dataset. The proposed model as seen in Fig.2 and Fig.3. In GroupA, eight convolutional layers with three max-pooling, and in GroupB, twelve convolutional layers with four max-pooling layers are frozen. In the proposed model, we used different hyperparameters as seen in Table 2.

RESULTS

We check the performance of the proposed transfer learning model on six binary classifications, which include NC vs AD, NC vs EMCI, NC vs LMCI, EMCI vs LMCI, EMCI vs AD, and LMCI vs AD. We also evaluate our model with and without data augmentation. During the experiment, we divide each class of data into three steps. In the first step, we split data 20% for testing and retain remaining data further for training 80% and validation 20% as we have shown in the flow chart in Fig. 4. During data augmentation technique rotation range 10 degrees, width and height shift range 0.1 degrees, and shear range 0.15 degree, as shown in Fig. 1. We used Keras library for the implementation of our model on Z840 workstation Intel Xeon (R) E5-2630v3 @2.40 GHz*32 with 1 TB memory.

Finally, we check the performance of the proposed model through several measures sensitivity, specificity and accuracy are described in terms of True Positive (TP), True Negative (TN), False Negative (FN) and False Positive (FP).

$$Sensitivity = TP/(TP + FN) \quad (1)$$

$$Specificity = TN/(TN + FP) \quad (2)$$

$$Accuracy = (TN + TP)/(TN + TP + FN + FP) \quad (3)$$

Performance evaluation without augmentation

We applied our proposed transfer learning model on binary classification and established the results on testing data. In this section, we showed two methods of performance without augmentation in six binary classes. In GroupA, three blocks have been frozen, and GroupB froze the four blocks, as shown in Fig. 2 and Fig. 3. Our model achieved an accuracy result 93.83% on NC vs AD classification on GroupA, we also used the same number of images in GroupB and obtained the accuracy 95.33% (sensitivity 94.31% and specificity 96.26 %) on

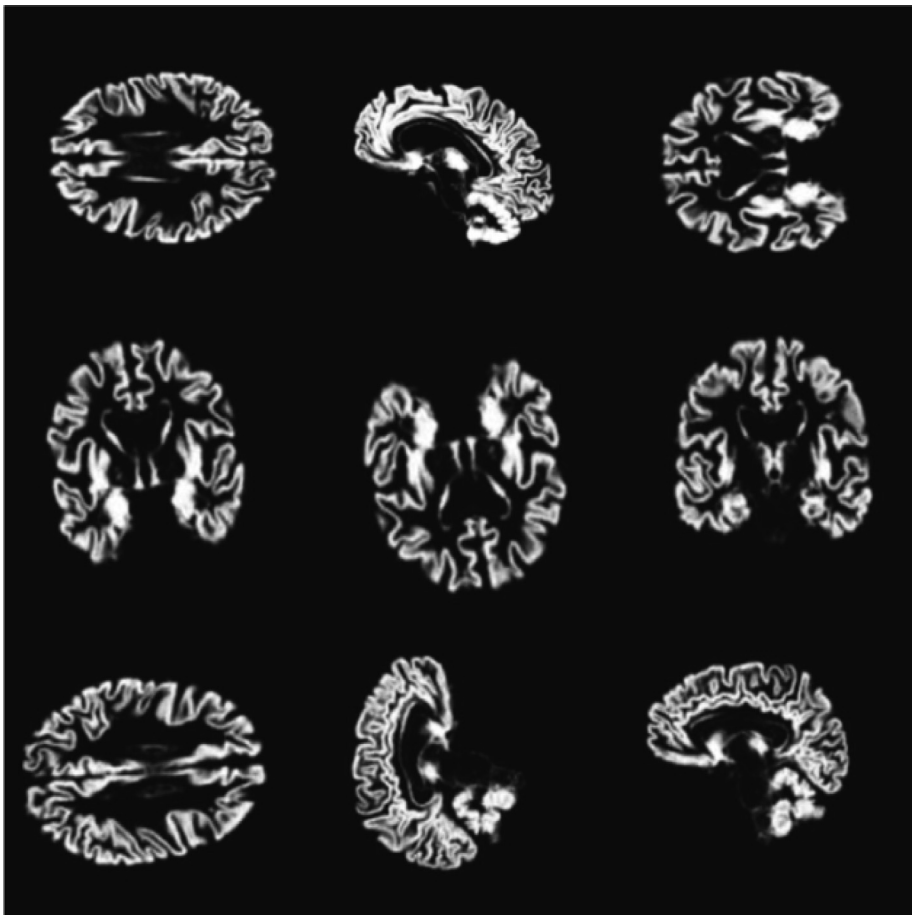

Fig. 1. Data for the proposed model in the form of gray matter. The first row showed the three views axial, coronal, and sagittal without augmentation. The second and third rows show data after augmentation, used for the classification.

NC vs AD classification. Furthermore, when we compare the classification results of NC vs LMCI with two groups, then GroupB performs the 5% comparatively higher performance. As shown in Table 3, a proposed technique can discriminate between EMCI vs LMCI accuracy and specificity of more than 83% in GroupB. However, on the other hand, in GroupA EMCI vs AD produced optimal results 82.34% as compare to GroupB in terms of accuracy and sensitivity. Detailed results of both methods are shown in Fig. 5.

Performance evaluation with augmentation

Table 4 shows the effect of two groups based on transfer learning with the augmentation dataset. Here, we see that the proposed model has shown promising results on NC vs AD classification and obtained 98.73% accuracy shows in GroupB. Next, we examine in Table 4, EMCI vs LMCI results in almost the same on both groups and accuracy performance more than 81%. For the pair of NC vs LMCI, the best accuracy result was achieved by

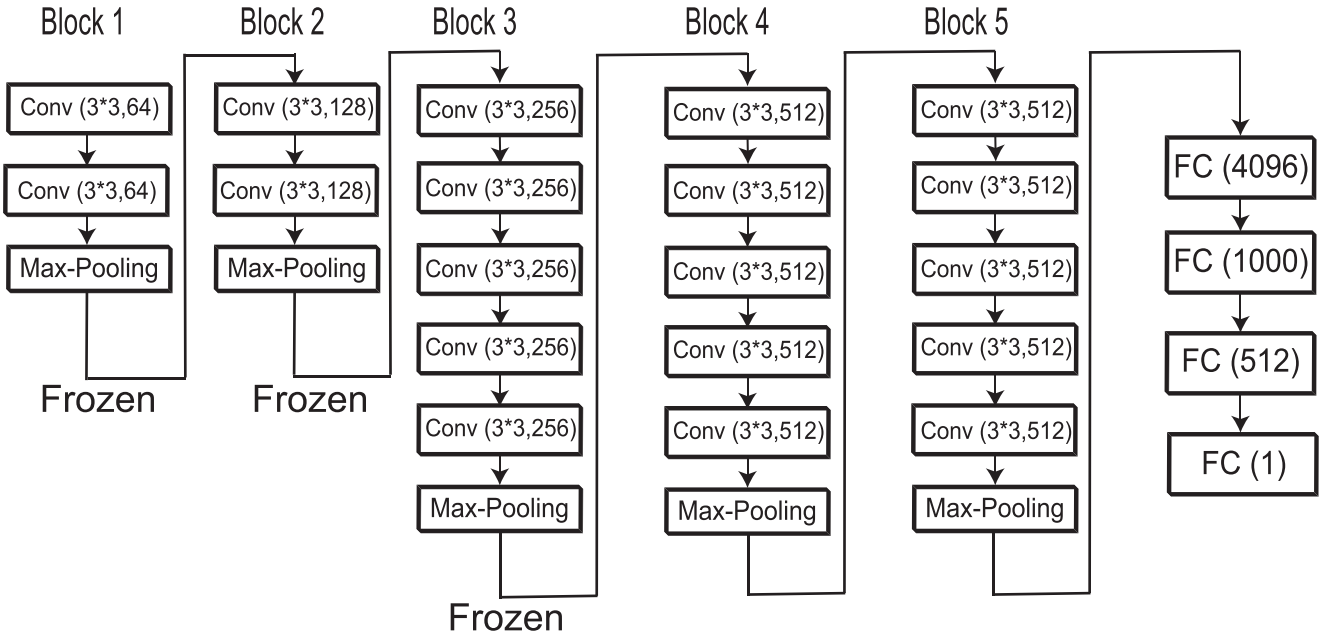

Fig. 2. The architecture of the proposed network with frozen of block1, block2, and block3. The first two blocks have four conv layers with two max-pooling, and block three have four conv layers with one max-pooling layer. The kernel size of all conv layers is kept 3*3. Finally, FC layers are used to attain results. (Conv:Convolution; FC: Fully connected layer).

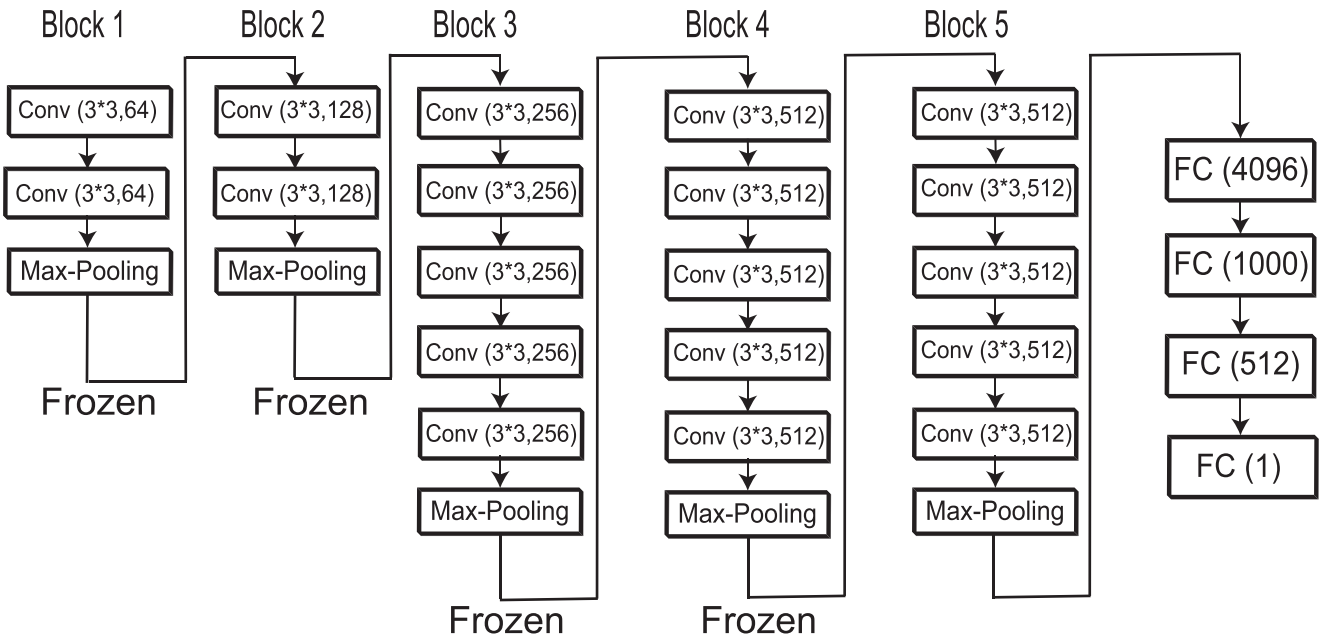

Fig. 3. The architecture of the proposed network with frozen of block1, block2, block3, and block4. The first two blocks have four conv layers with two max-pooling and block3, block4 have eight conv layer with two max-pooling layers. kernal size of all conv layers is kepr 3*3. The filters for all the blocks are: 64, 128, 256,512. Finally FC layers is used to obtain the results (Conv:Convolution; FC: Fully connected layer).

Table 2. Hyper-parameters for the proposed method, used during training and testing, ReLU (rectified linear unit)

HYPERPARAMETERS	
Activation Function	ReLU Sigmoid
Base Learning Rate	$1e^{-5}$
Epochs	20
Batch Size	32
Optimizer	Adam
Loss Fucntion	Binary Cross Entropy

GroupB with an accuracy of 89.15% with an augmentation approach. On the other hand, if we see the EMCI vs AD classification performance in terms of sensitivity, the GroupA obtained the highest value of 83.69%. Detailed results of both methods are shown in Fig. 6.

DISCUSSIONS

Researchers have recently conducted many studies on the early diagnosis of AD using machine learning and deep learning approaches. Therefore, many researchers developed the computer-aided system, which helps to diagnose the early stage of AD, especially in deep learning. CNN produced promising results in video and image processing (Barros et al., 2018). It is a fully trainable system that did not require the experts to manipulate the datasets because of CNN, which can automatically extract the features. Max pooling is the main part of CNN, which reduces the size of the feature map (Krizhevsky and Sutskever, n.d). However, the lack of

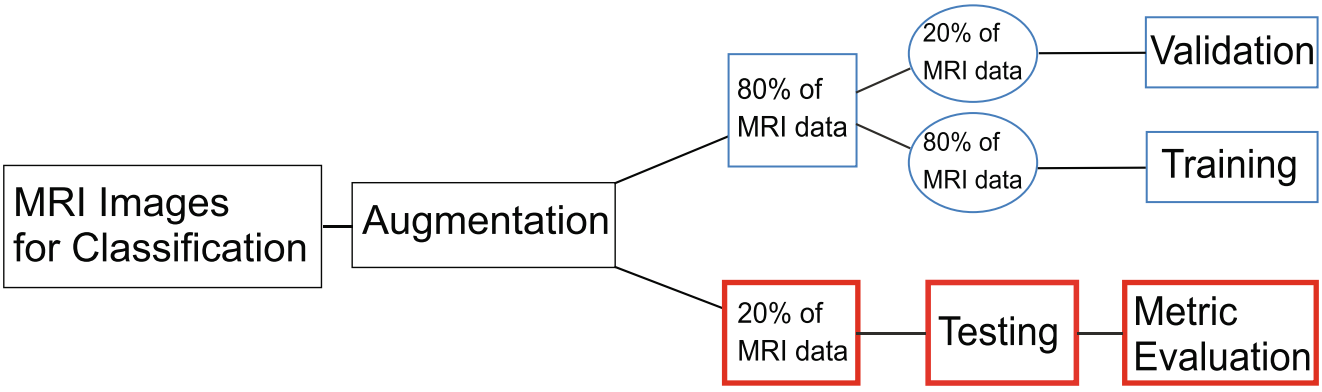

Fig. 4. Framework of the proposed method on MRI data of each classification task (NC vs AD, NC vs EMCI, NC vs LMCI, EMCI vs LMCI, EMCI vs AD, LMCI vs AD). Augmentation is applied to all data samples, i.e, 80% for training, and 20% testing. NC = normal control, EMCI = early mild cognitive impairment, LMCI = late mild cognitive impairment, and AD = Alzheimer's disease.

Table 3. Evaluation metric on testing data for GroupA and GroupB without data augmentation. These two groups showed the accuracy, sensitivity, and specificity rate on six binary classes as shown in Fig. 5

	GroupA			GroupB		
Image Classes	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
NC vs AD	93.83	92.15	95.13	95.33	94.31	96.26
NC vs EMCI	81.20	80.25	82.15	85.16	84.29	85.98
NC vs LMCI	82.72	81.63	83.81	87.91	86.61	89.01
EMCI vs LMCI	79.5	79.05	81.11	83.72	82.09	85.13
EMCI vs AD	82.34	81.22	83.17	81.93	81.63	81.98
LMCI vs AD	74.22	73.15	75.33	82.31	82.18	82.03

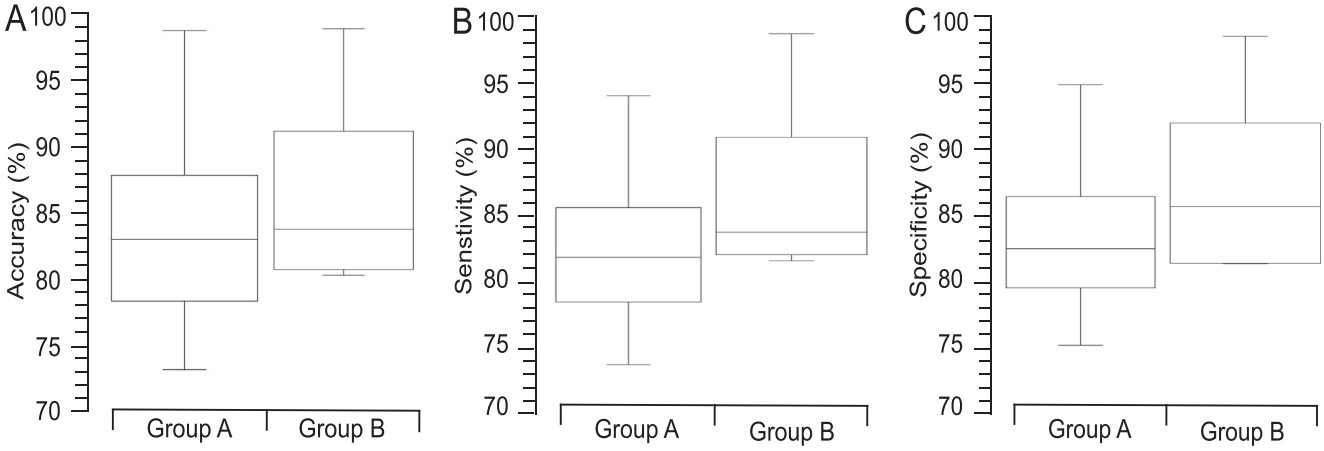

Fig. 5. Proposed model performance in term of accuracy, sensitivity and specificity on six binary classes without data augmentation. In above figure,from left to right in box plot A, B and C the overall performance achieved by Group B which are 95.33%, 94.31% and 96.26% respectively.

Table 4. Evaluation metric on testing data for GroupA and GroupB with data augmentation. These two groups showed the accuracy, sensitivity and specificity rate on six binary classes. NC vs Ad attained the highest rate 98.73% in term of accuracy as shown in Fig. 6

	GroupA			GroupB		
Image Classes	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
NC vs AD	95.38	95.93	94.61	98.73	98.19	99.09
NC vs EMCI	85.14	84.61	85.42	87.06	86.61	86.63
NC vs LMCI	85.89	86.17	85.39	89.15	89.24	88.86
EMCI vs LMCI	81.73	80.12	83.07	81.06	80.61	81.52
EMCI vs AD	83.69	83.64	83.43	84.15	83.76	83.12
LMCI vs AD	76.73	77.31	75.78	82.07	81.39	82.41

an annotated dataset to train the model on scratch is a big problem. This study has developed a transfer learning or fine-tuning approach with MRI images to attain the automatic detection of different stages of Alzheimer's disease. In order to resolve the overfitting issue on a small dataset, augmentation plays a key role in the transfer learning model.

The designed model tests the performance based on three parameters such as accuracy, sensitivity, and specificity. In the clinical field, these measures help correctly classify healthy and ill patients. MRI is a potent modality for the Alzheimer's patient's classification and helps the doctors to diagnose this at an early stage. MCI is a critical stage for Alzheimer's patients (Liu et al., 2018). MCI is divided further into two stages showing the conversion of patients on early-stage or late stages. EMCI has demonstrated the early stage of AD and provides the option of treatment to overcome the dementia risk factor. In aging research fields, many CAD systems are developed for the classification of AD stages. This research work focused on segregation, such as NC people, EMCI, LMCI, and AD patients. These prediction results focused on the specific gray matter (GM) region, which is more useful in predicting AD's early diagnosis. In the first step, the method prediction of average performance without augmentation on GroupA 82.58% and 86.18%, which showed the effectiveness of the proposed model. Moreover, we investigate our model prediction with augmentation, and we attain the 98.73% performance accuracy for NC vs AD and averages accuracy of Group 1, is 84.76% and 87.06% for Group-2, which is the highest performance amongst the proposed models.

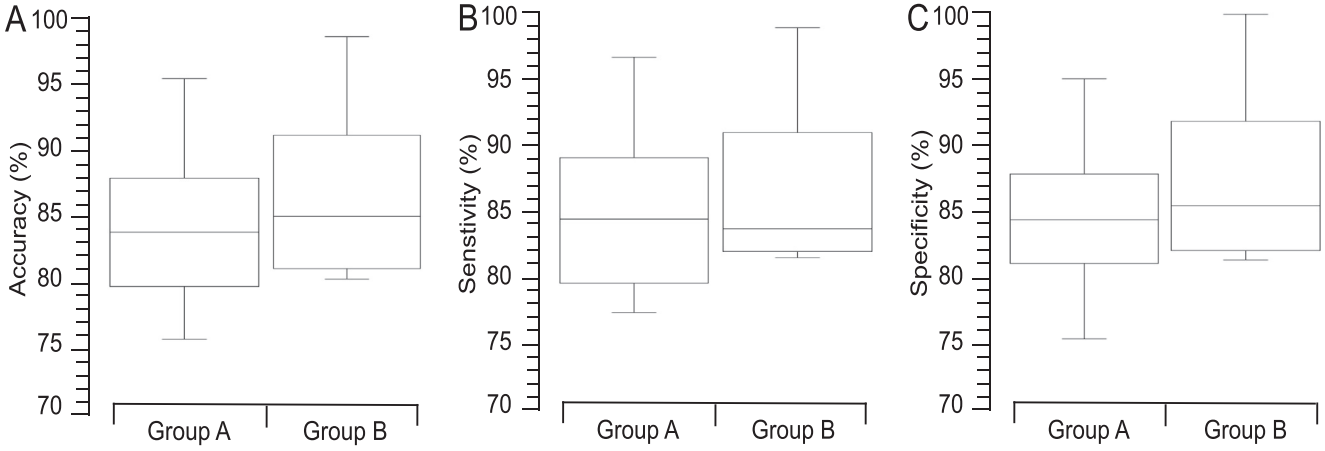

Fig. 6. Proposed model performance in term of accuracy, sensitivity and specificity on six binary classes data with augmentation. In above figure, from left to right in box plot A, B and C the overall performance achieved by Group B which are 98.73%, 98.19% and 99.09% respectively.

We observed the model which used the 3D convolutional neural network to classify without skull stripping data. Secondly, to improve the effectiveness and performance accuracy using transfer learning on ADNI datasets attained 99.33% results during the binary classification of normal control and AD patients (Hosseini-Asl et al., 2016). MCI conversion is linked with the number of risk factors that affect to convert in AD. In terms of gray matter, density showed a clear difference between healthy subjects and AD. Therefore, the MCI conversion part shown gray matter intensity reduction compared to the non-conversion part, which is useful for diagnoses (Yang and Liu, 2020). However, the pre-trained model helps diagnose the disease in daily clinical practice because it took less time to process and produce high-performance results on a less annotated dataset. In this approach, they introduced the convolutional network to achieve a state of the art results (Wu et al., 2018). The major advantage produced by the model to reduce the parameters which directly impact in term of regularization and improve the results of classification.

Tables 5 and 6 shows the comparison with several studies that have investigated the early diagnosis of AD patients. However, the proposed model produces the best results in terms of accuracy on AD vs NC 98.73% score and EMCI vs LMCI 83.72%. In researcher (Basaia et al., 2019) produced the 98.10% results in terms of sensitivity for AD vs NC. In addition, our model also outperforms for the remaining four classification tasks, such as

Table 5. Evaluation results coincide with the AD/NC classification. Our proposed model results compared with different five studies in term of accuracy, sensitivity and specificity

Methods	Accuracy	Sensitivity	Specificity
Ortiz et al. [45]	90.09	86.12	94.1
Wee et al. [1]	92.35	90.35	94.31
Khedher et al. [7]	87.53	88.65	86.17
Basaia et al. [46]	98.2	98.1	98.3
Ahmed et al. [10]	90.2	82.92	97.2
Proposed model	**98.73**	**98.19**	**99.09**

Table 6. Evaluation results coincide with the EMCI/LMCI classification. Our proposed model results compared with different four studies in term of accuracy, sensitivity and specificity. It attained the highest classification results 83.72%

Methods	Accuracy	Sensitivity	Specificity
Wee et al. [1]	75.05	63.5	84.41
Basaia et al. [46]	75.1	75.8	74.1
Lei et al. [47]	78.05	81.58	75
Yang et al. [48]	72.19	73.82	73.05
Proposed model	**83.72**	**82.09**	**85.13**

NC vs EMCI 87.06%, NC vs LMCI 89.15%, EMCI vs AD 84.15%, and LMCI vs AD 82.31% in term of accuracy. It can be seeming transfer learning extract the more useful features for the classification of AD on the brain's GM segmentation.

In this study, we propose a layer-wise transfer learning approach for six classification tasks. We selected the approved architecture capable of early-stage diagnosis of AD. The distinction between EMCI and LMCI to help out the experts to treat on time of dementia disease. To overcome this issue on a small number of annotated data, we apply transfer learning with the augmentation technique and improve performance accuracy. We investigate proposed models with detailed experiments on 300 ADNI subjects with six binary classes. We also checked the effects on performance after frozen the number of blocks in our model. Furthermore, we compared our proposed technique results with state-of-the-art methods. We discovered that our model significantly outperforms on AD vs NC classification and achieved 98.73% in terms of testing accuracy. Future implication includes applying the proposed model for lungs and breast cancer detection.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships

that could have appeared to influence the work reported in this paper.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (Nos. 61771380, 61906145, U1730109, 91438103, 61771376, 61703328, 91438201, U1701267, 61703328); the Equipment pre-research project of the 13th Five-Years Plan (Nos. 6140137050206, 414120101026, 6140312010103, 6141A020223, 6141B06160301, 6141B07090102), the Major Research Plan in Shaanxi Province of China (Nos.2017ZDXM-GY-103, 017ZDCXL -GY-03-02), the Foundation of the State Key Laboratory of CEMEE (Nos.2017K0202B, 2018K0101B 2019K0203B, 2019Z0101B), the Science Basis Research Program in Shaanxi Province of China (Nos.16JK1823, 2017JM6086, 2019JQ-663). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

Aderghal K, Khvostikov A, Krylov A, Benois-Pineau J, Afdel K, Catheline G (2018) Classification of Alzheimer disease on imaging modalities with deep cnns using cross-modal transfer learning. In: 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS). p. 345–350.

Ahmed OB, Benois-Pineau J, Allard M, Catheline G, Amar CB, Initiative ADN, et al. (2017) Recognition of Alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and multiple kernel learning. Neurocomputing 220:98–110.

Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S (2016) Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imaging 35(5):1207–1216.

Badakhshannoory H, Saeedi P (2011) A model-based validation scheme for organ segmentation in ct scan volumes. IEEE Trans Biomed Eng 58(9):2681–2693.

Barros MT, Silva W, Regis CDM (2018) The multi-scale impact of the Alzheimer's disease on the topology diversity of astrocytes molecular communications nanonetworks. IEEE Access 6:78904–78917.

Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, Filippi M, Initiative ADN, et al. (2019) Automated classification of Alzheimer's disease and mild cognitive impairment using a single Mri and deep neural networks. NeuroImage: Clinical 21 101645.

Bi X, Li S, Xiao B, Li Y, Wang G, Ma X (2019) Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology. Neurocomputing 21:1232–1245.

Chaddad A, Desrosiers C, Niazi T (2018) Deep radiomic analysis of Mri related to Alzheimer's disease. IEEE Access 6:58213–58221.

Chen H, Ni D, Qin J, Li S, Yang X, Wang T, Heng PA (2015) Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE J Biomed Health Inform 19(5):1627–1636.

Chougrad H, Zouaki H, Alheyane O (2019) Multi-label transfer learning for the early diagnosis of breast cancer. Neurocomputing 11:835–847.

Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. p. 248–255.

Feng C, Elazab A, Yang P, Wang T, Zhou F, Hu H, Xiao X, Lei B (2019) Deep learning framework for Alzheimer's disease diagnosis via 3d-cnn and fsbi-lstm. IEEE Access 7:63605–63618.

Gupta A, Ayhan M, Maida A (2013) Natural image bases to represent neuroimaging data. International Conference on Machine Learning:987–994.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. 2015. arXiv preprint arXiv:1512.03385. .

Hernández-García A, König P (2018) Further advantages of data augmentation on convolutional neural networks. In: International Conference on Artificial Neural Networks. p. 95–103.

Hosseini-Asl E, Keynton R, El-Baz A (2016) Alzheimer's disease diagnostics by adaptation of 3d convolutional network. In: 2016 IEEE International Conference on Image Processing (ICIP). p. 126–130.

Ieracitano C, Mammone N, Bramanti A, Hussain A, Morabito FC (2019) A convolutional neural network approach for classification of dementia stages based on 2d-spectral representation of eeg recordings. Neurocomputing 323:96–107.

Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al., 2008. The Alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 27(4), 685–691..

Khan NM, Abraham N, Hon M (2019) Transfer learning with intelligent training data selection for prediction of Alzheimer's disease. IEEE Access 7:72726–72735.

Khedher, L., Ramrez, J., Grriz, J.M., Brahim, A., Segovia, F., s Disease Neuroimaging Initiative, A., et al., 2015. Early diagnosis of Alzheimer disease based on partial least squares, principal component analysis and support vector machine using segmented Mri images. Neurocomputing, 151, 139–150..

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. .

Krizhevsky, A., Sutskever, I., n.d.. Ge + lqwrqł, pdjh1hwfodvvlilfdwlrq with deep convolutional neural network. Communications of the ACM, 60(6), 84a–90. .

Li F, Tran L, Thung K-H, Ji S, Shen D, Li J (2015) A robust deep model for improved classification of ad/mci patients. IEEE J Biomed Health Inform 19(5):1610–1616.
Liu M, Cheng D, Wang K, Wang Y, Initiative ADN, et al. (2018) Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. Neuroinformatics 16(3–4):295–308.
Liu X, Wang C, Bai J, Liao G (2019) Fine-tuning pre-trained convolutional neural networks for gastric precancerous disease classification on magnification narrow-band imaging images. Neurocomputing 9:7030–7039.
López M, Ramrez J, Górriz JM, Álvarez I, Salas-Gonzalez D, Segovia F, Chaves R, Padilla P, Gómez-Río M, Initiative ADN, et al. (2011) Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease. Neurocomputing 74(8):1260–1271.
Lyndon D, Kumar A, Kim J, Leong PHW, Feng D (2015) Convolutional neural networks for medical clustering. CLEF (Working Notes).
Maqsood M, Nazir F, Khan U, Aadil F, Jamal H, Mehmood I, Song O-Y (2019) Transfer learning assisted classification and detection of Alzheimer's disease stages using 3d Mri scans. Sensors 19 (11):2645.
McGeer PL (1986) Brain imaging in bluealzheimer'sdisease. British Med Bull 42(1):24–28.
Mehmood A, Maqsood M, Bashir M, Shuyuan Y (2020) A deep siamese convolution neural network for multi-class classification of Alzheimer disease. Brain Sci 10(2):84.
Oh K, Chung Y-C, Kim KW, Kim W-S, Oh I-S (2019) Classification and visualization of blue alzheimer's disease using volumetric convolutional neural network and transfer learning. Sci Rep 9 (1):1–16.
Payan, A., Montana, G., 2015. Predicting Alzheimer's disease: A neuroimaging study with 3d convolutional neural networks. arXiv preprint arXiv:1502.02506. .
Phong TD, Duong HN, Nguyen HT, Trong NT, Nguyen VH, Van Hoa T, Snasel V (2017) Brain hemorrhage diagnosis by using deep learning. In: Proceedings of the 2017 International Conference on Machine Learning and Soft Computing. p. 34–39.
Sezer A, Sezer HB (2019) Convolutional neural network based diagnosis of bone pathologies of proximal humerus. Neurocomputing 19:1929–1938.
Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 35(5):1285–1298.
Suk, H.-I., Lee, S.-W., Shen, D., n.d. Alzheimer's disease, and neuroimaging i.(2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. Neuroimage, 101, 569–582..
Wang G, Forsyth D, Hoiem D (2010) Comparative object similarity for improved recognition with few or no examples. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. p. 3525–3532.
Wang H, Shen Y, Wang S, Xiao T, Deng L, Wang X, Zhao X (2019) Ensemble of 3d densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease. Neurocomputing 333:145–156.
Wee C-Y, Yap P-T, Shen D, Initiative ADN (2013) Prediction of alzheimer's disease and mild cognitive impairment using cortical morphological patterns. Human Brain Mapping 34 (12):3411–3425.
Wu C, Guo S, Hong Y, Xiao B, Wu Y, Zhang Q, Initiative ADN, et al. (2018) Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks. Quantitative Imaging Med Surgery 8(10):992.
Xu L, Yao Z, Li J, Lv C, Zhang H, Hu B (2019) Sparse feature learning with label information for Alzheimer's disease classification based on magnetic resonance imaging. IEEE Access 7:26157–26167.
Yang Z, Liu Z (2020) The risk prediction of Alzheimer's disease based on the deep learning model of brain 18f-fdg positron emission tomography. Saudi J Biolog Sci 27(2):659–665.
Young J, Modat M, Cardoso MJ, Mendelson A, Cash D, Ourselin S, Initiative ADN, et al. (2013) Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. NeuroImage: Clinical 2:735–745.
Yue L, Gong X, Li J, Ji H, Li M, Nandi AK (2019) Hierarchical feature extraction for early Alzheimer's disease diagnosis. IEEE Access 7:93752–93760.
Zeng N, Qiu H, Wang Z, Liu W, Zhang H, Li Y (2018) A new switching-delayed-pso-based optimized svm algorithm for diagnosis of Alzheimer's disease. Neurocomputing 320:195–202.

(Received 14 June 2020, Accepted 3 January 2021)
(Available online 17 January 2021)