Biomedical Signal Processing and Control 71 (2022) 103217

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Alzheimer's disease classification using pre-trained deep networks

Jayanthi Venkatraman Shanmugam [a,*], Baskar Duraisamy [b], Blessy Chittattukarakkaran Simon [c], Preethi Bhaskaran [d]

[a] Department of Electronics and Communication Engineering, Rajagiri School of Engineering and Technology, Ernakulum, Kerala, India
[b] Electronics and Communication Engineering, Hindusthan College of Engineering and Technology, Coimbatore, Tamilnadu, India
[c] Department of Electronics Engineering, Cochin University of Science and Technology, Ernakulam, Kerala, India
[d] Department of Electronics and Communication Engineering, Rajagiri School of Engineering and Technology, Rajagiri Valley, Kakkanad, Ernakulam, Kerala, India

ARTICLE INFO

Keywords:
Alzheimer's disease
Deep learning
Transfer learning
Convolutional neural network
ADNI
Pre-trained networks

ABSTRACT

Alzheimer disease (AD) is a progressive neurologic disorder that causes the brain to shrink (atrophy) and brain cells to die. Alzheimer disease is the most common cause of dementia that slowly degrades the thinking and social interactions by destroying the brain cells. Although there is no treatment to reverse the progression of AD, detecting the onset of AD can be very effective in medical field. This paper focuses on early detection of various stages of cognitive impairement and AD by using neuroimages with Transfer Learning (TL). The Magnetic resonance imaging (MRI) images obtained from Alzheimer's Disease Neuroimaging (ADNI) database with various classes of Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Mild Cognitive Impairment (MCI), Late Mild Cognitive Impairment (LMCI) are classified using transfer learning approach. Three pre-trained networks, such as GoogLeNet, AlexNet and ResNet-18 are used in this classification which are trained and tested with 6000 images collected from ADNI database. The classification performance of these three networks is analyzed with the help of confusion matrix and its parameters. The overall accuracy of the GoogleNet, AlexNet and ResNet-18 in detecting AD are obtained as 96.39%, 94.08% and 97.51% respectively. The class wise performance of the pre-trained networks also analysed using confusion matrix parameters.

1. Introduction

Alzheimer's disease (AD) is chronic neurological disease that initiates with memory loss and gradually leads to death. AD is one of the most dangerous diseases; World Health Organization (WHO) published its report that the total number of AD cases is likely to reach 82 million within next 10 years. This count of patients can also extend to 152 million in another 20 years. This progression in disease rate is mostly reported in people above the age of 65 [1]. Even though the progression speeds of the disease differ with individuals, the expected life span diagnosis period varies depending on the person's health from three to nine years. The study [2] raises the importance of early diagnosis of this AD. Several types of abnormalities, like memory loss, mood swings deterioration in basic metabolism, behavioral problems are found during the progressive stage of AD.

Although AD is irreversible [3] discusses various prospective benefits of the early detection of the disease. An AD diagnosed patient can be provided with the treatment for cognitive losses, which is more useful if diagnosis at initial stage. The accurate diagnosis of AD is often misled due to the similarity of its symptoms to that of symptoms of normal aging. A perfectly skilful observation on brain cells is mandatory for an explicit diagnosis of AD, and some techniques for such an accurate diagnosis are discussed by Asramietal. Therefore, the decisive and non-invasive techniques for precise AD diagnosis is a mandatory in the current scenario [4–6].

Imaging techniques are the most common utilized method of AD diagnosis to find the disease. These techniques revolutionazied the process of disease diagnosis, because it facilitates non-invasive internal examination of the body. The history of AD treatment once faced a tough period when the disease could be found in life only after death. But, nowadays medical imaging techniques have a great role in AD diagnosis and treatment. Some of the imaging techniques used for AD diagnosis are X-ray computed tomography (CT), Magnetic Resonance Imaging (MRI), diffusion tensor image (DTI) and Positron emission tomography (PET). Among these, MRI and PET scanning are the most widely used imaging modalities in the works of DesaiandParmar, Duraisamy et al.,

* Corresponding author.
E-mail address: vsjayanthi0318@gmail.com (J.V. Shanmugam).

https://doi.org/10.1016/j.bspc.2021.103217
Received 24 March 2021; Received in revised form 8 September 2021; Accepted 30 September 2021
Available online 20 October 2021
1746-8094/

[7–9]. With the development and advancement of neuroimaging techniques, [10] discussed the use of features based on MRI and Fluorodeoxy-glucose positron emission tomography (FDG-PET) for evaluating rate of progress of disease. The diagnosis supported by various previously determined biomarkers, which determines the depth of disease. Fig. 1 depicts 5 biomarkers and the corresponding course of the disease. The MRI images obtained from ADNI database with various classes of Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Mild Cognitive Impairment (MCI), Late Mild Cognitive Impairment (LMCI) are classified using transfer learning approach.

Some of the existing researches on AD diagnosis using imaging modalities like that of Zhang et al., [11] did not consider physiological variations in the people. This is because these works follow a common procedure, which includes the usual scenario of pre-processing, region-of-interest (RoI) segmentation, feature extraction, selection of extracted features and classification. Although this procedure is challenging, it offers some major drawbacks on the results. It is clear from the results obtained by Liu et al., [12] that the above mentioned procedure for classification is time consuming and bring out unreliable results in the detection of brain regions using boundary detection.

In [13,14], a survey was conducted in 26 centers in India, China, Latin America, Africa and South East Asia, which found that it was not very difficult to achieve a diagnostic strategy related to education and culture. The connections with Alzheimer's disease International (ADI), its worldwide influence, and the initiative of ADI volunteers are supported for this survey to examine the growth rate of Alzheimer's disease. The effort encouraged both the spreading of awareness and impact upon strategy, so the research proof is utilized by ADI and National Alzheimer's Associations to direct and promote advocacy [15,16].

Akgület et al., [17,18] mentioned the drawback of adaptive pruning technique that it failed to accurately sort out brain images. The reason for such reduced accuracy was that it lost some relevant training samples that had to be included in the feature space of the cluster. A solution for this issue was suggested by [19,20] in which the classification of brain images as normal and diseased classes was performed by a hybrid approach.

In [21], deep learning techniques applied like stacked auto encoder on brain images for feature representation. The region of interest (ROI) of brain images in different modalities, like PET, CSF and MRI were extracted and normalized. Another deep learning architecture designed [22] was built with stacked softmax output layer and auto-encoders, which was found to considerably reduce bottleneck effect.

All the above discussed works are undertaken for a common aim of AD diagnosis in the earliest with maximum perfection. Doctors insist for the detection of AD in the earliest possible stage, due to the fact that there are chances get saved from this irreversible disease by some preventive measures that can considerably slow down its progress. These preventive measures include physical activities, perfect sleep cycles, healthy social interactions and a specific diet plan named Mediterranean diet. Thus, to get saved from AD means to live healthy in all means.

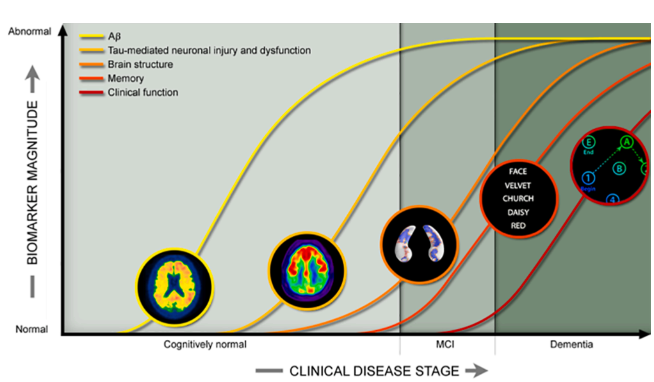

Fig. 1. Disease course corresponding to the level of 5 biomarkers.

People are more likely to get affected with a genetic history of AD. By physically as well as mentally healthy can save these people or atleast enable prominent decline in the disease development.

The complications in the diagnosis of AD and its immutable nature keep this disease only in initial stage of research for several decades. The support of neuro-imaging techniques was an inevitable part in the clinical and exploration developments in the case of AD. The possibility of AD diagnosis in life was possible only after these techniques. But, AD diagnosis still stood tough with higher degree of similarity in brain patterns. Thus accurate and precise classification algorithm has become a strict requirement in AD diagnosis. Deep learning is found to be the perfect solution for all these concerns and proved to give attractive results especially for classification tasks. Thus the classification of brain images in the different stages of disease can be efficiently performed using deep learning. The implementation of deep neural networks is made easy and effective with the help of latest softwares. Also, the facility of GPU high computing platform enables the training of a large data in significantly less time.

2. Relevance of deep learning in AD diagnosis

There are plenty of machine learning algorithms available for the classifications of brain images and to detect the development of AD. The feature extraction is the most prevailing requirement in classification using machine learning. Various classification algorithms work with different feature extraction methods. These include extraction of features from ROIs selected MRI voxels, cortical surface along vertex levels, cingulate cortex and hippocampus. The extraction of features diverge in many ways including a direct, STAND score and Atlas-based methods. The traditional feature extraction methods are time consuming and complex tasks, which affects the classification system in terms of performance metrics. The additional extracted features are the main cause for over fitting. The over fitting is avoided by less number of extracting high impact features.

The total input data is separated into two distinct set of training and testing classification model by following supervised learning. The detection of AD under its premature stage by the classification is typically done at three ways as CN v/s AD, CN v/s MCI and AD v/s MCI. The training set is utilized for training classifier and the performance of classification is assessed with the testing set. The confusion matrix terms of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) are calculated to find various parametric measures. The performance metrics are sensitivity, accuracy, negative predictive values and positive predictive values. An increase in data set size increases the testing samples; hence this results in higher classification accuracy on the cost of increased time of computation. It also demands high speed processors and huge memory. However, these challenges are supposed to be considered for practical implementation of the algorithms. The major shortcomings encountered in utilizing traditional ML algorithms are:

- Larger dataset
- Several stages of preprocessing
- Various levels of feature extraction
- Usage of large size of memory
- Analysis complexity
- Huge computation time

Deep learning (DL) deals with all these issues and offers a better convenient platform for medical image classification, specifically to detect AD in early stages by the classification of neural images of several classes. Deep learning algorithms offers an augmentation technique, which permits the increasing the number of data samples by augmenting single images to train the deep learning network. The DL has deeper architecture with more number of layers compared to the ML based neural networks with capability of feature extraction without external user level involvement. This simplifies the algorithm implementation and also improves the classification performance.

The researches utilizing deep learning on AD diagnosis have ended up in inspiring results that happens to be a key for further developments on this area. Various DL architectures have been experimented to achieve outstanding results for a wide variety of research problems. However, the convolutional neural network (CNN) stands out with the most extensively used architecture of deep learning that leads to better results in the simple manner. To select the appropriate architecture is always a challenging decision while implementing a deep learning model. The following discussions focus on some of the prior works that discuss the performance of CNN over other deep learning architectures for various purposes.

3. Existing works on AD classification using transfer learning

Khan et al., [23] suggested transfer learning as a solution for challenges of deep learning algorithms that include the requirement of large datasets and careful optimizations on the network architectures. A VGG-16 CNN was modified and initialized with the pre-trained weights of the dataset that consists of natural images. The top network layers were trained with the dataset while the lower layers were kept hold, there by implementing a layer-wise learning methodology. These layers tried to reduce the data size by working on entropy to select the most informative slices of the image. The results showed that the transfer learning using VGG-16 achieved good performance in classification with 4% to 7% increase in accuracy over 3D CNN models combined with auto-encoders [24–25].

In [26], AD classification using fMRI images by making the advantages of transfer learning. The objective was to classify the brain MRI images into the sets of AD and healthy brain. One of the famous pre-trained networks, LeNet-5 was modified and trained to fulfill this requirement. The data is collected from ADNI that included the images of different age group. The created dataset was divided in such a way that 60% training, 20% validation and 20% testing. The results demonstrated that LeNet-5 yield the best results when modified and trained with fMRI images.

In [27], the detection of Alzheimer's disease cases was discussed, particularly at premature stages, with features derived as medical imaging. The whole brain image was transmitted with exception learning architecture transmission. The convolutional neural network (CNN) was built through the aid of separable convolution layers that may mechanically learn that common characteristics of the image data for classification.

In [28], an automatic detection of Alzheimer's disease was presented. As AD considerably involve that content and acoustics of spontaneous speech, natural language processing along with machine learning give hopeful systems for dependably noticing AD. The presented strategy helps the value of well-performed machine learning and linguistics-focused processing for noticing AD as speech and highlights that require to evaluate model efficiency on carefully balanced data sets with equal parameters of Consistent training and independent test data sets to decide that best Predictive Model Realization.

In [29], advanced machine learning methods was presented during the assessment and comparison of two machine learning discrimination among dementia of Alzheimer's-type (DAT) and non-DAT controls. Such two processes given probabilistic dementia score for this never-before-seen clinical data. The outcomes display better generalizability of two machine learning strategies, marking a significant step for translating previously trained machine learning models addicted to clinical practice.

In [30], a new pipeline based on deep convolutional neural network with spikes was presented to categorize AD with MRI scans. Particularly, an unsupervised convolutional spiking neural network (SNN) was pre-trained on MRI scans. At last, a supervised deep convolution neural network (CNN) was qualified on output of SNN classification tasks. The tests were conducted with Alzheimer's disease Neuroimaging Initiative (ADNI) dataset and successful outcomes were arrived for AD classification. The precision of the presented model through peak pre-training systems for the classification of three binaries as 90.15%, 87.30% and 83.90%. In [31], a strategy was presented to diagnose Alzheimer's disease with English speech. A series of linguistic characteristics like rate of nouns, rate of adjectives, rate of pronouns and rate of verbs were removed as recorded speech, along with its distribution of AD and control samples was utilized to train the classifier [32]. The presented strategy was recover automatic AD forecast with moderately little directed speech data set in the absence of utilizing expertly described linguistic characteristics. With sentence-level BERT $_{\text{Large}}$ through simple logistic regression classifier, precision and F1 scores of 88.08% and 87.23% were accomplished, improving leading-edge outcomes with 2.28% and 2.80%.

4. Transfer learning

Humans are always likely to apply a same knowledge for different tasks. The same information can be utilized to solve related tasks of different domains.

The transfer learning (TL) is defined as a method of adjusting a pre-trained network for the purpose of the user. In simple words, transfer learning is defined as the transfer of knowledge.

The mathematical aspect of TL [33] is described based on domain and task. D represents domain and T stands for the task to be performed. Every domain is contained with a feature space i.e. X. The probability distribution is given as $P(\boldsymbol{x})$,

where $\boldsymbol{x} = \{x_1, x_1, \cdots\cdots, x_1\} \in X$.

A specific domain can be represented in terms of feature space as,

$$D = \{X, P(\boldsymbol{x})\} \tag{1}$$

The task with two components, label space Y and an objective function$f(x)$,

$$T = \{Y, f(x)\} \tag{2}$$

The prognostic function learns from training data that consists of feature-label pairs,$\{x_i, y_i\}$

where

$$x_i \in X \text{ and } y_i \in Y \tag{3}$$

For a given source domain D_s, learning task T_s, target domain D_T along with learning task T_T, transfer learning improves the predictive function $f(x)$ in D_T with knowledge D_s and T_s up on the condition,

$$D_s \neq D_T or T_s \neq T_T \tag{4}$$

In deep learning, a model developed for one task is effectively reused for another task by making the necessary modifications on the existing model which is known as fine tuning. The type of fine tuning is applied to the pre-trained model is largely dependent on the new dataset with which the network has to be trained. The new dataset is similar to that of original dataset.

Lisa Torrey and Jude Shavlic discussed the three advantages that can be achieved by using transfer learning [34].

The following advantages of transfer learning from the training progress is given as below,

- Higher start value
- Larger slope
- Higher asymptote

Other notable advantages of transfer learning are,

- It saves the time to design a new model that suits for the new task.

- Even in the availability of less training data, can make use of the networks that are trained with huge amount of data to achieve the better results.

Being one of the most reliable techniques for learning, transfer learning is widely utilized for various tasks. Pre-trained networks are the most useful parts of this requirement and these networks are designed in the deep learning framework called Convolutional Neural Network (CNN). Numerous pre-trained networks are available to train the user requirements. Some of the available pre-trained CNNs are AlexNet, GoogLeNet, ResNet, SqueezeNet, VGG-16, VGG-19, Dense Net, LeNet etc. All these networks are previously trained with the dataset named ImageNet, which contains 1000 classes of images. These pre-trained networks are winners of various challenges, which is trained with millions of images. Although, these data sounds with more benefits of transfer learning, the fine-tuned model is not obvious to produce the better results, until it has been tested and evaluated. The main aim of this work is to use the transfer learning for AD classification. The pre-trained networks are modified and trained with MRI images collected from ADNI database and its performance is analyzed.

Also, this work intends to justify the advantages of transfer learning by classifying the brain MRI images into various classes of AD.

5. Learning Methodology

The key aspect of DL models lay around three major elements: the data it uses, the network designed and the parameters set for training. Any deep learning model is based on the perfect balance of these three elements, so that the task is performed at maximum efficiency. The design of the model is confirmed after several trials by which the model is trained with some initial parameters and the performance is evaluated. If the performance is seen poor or less, then the dataset, network architecture or the training parameters are modified till preferred performance level accomplished. This paper proposes the utilization of pre-trained networks with simple modifications, therefore they can be effectively used to achieve maximum performance range in less efforts. The major stages involved in classification along with the description of the pre-trained networks used are discussed in the following sections.

5.1. Data preparation

The database contains DICOM images of brain acquired from MRI scanning. These images are gathered as Alzheimer's disease Neuroimage Initiative (ADNI) study, which initiated for the clinical trials related to AD treatment. The idea and progress behind ADNI is coordinated by Alzheimer's. Therapeutic Research Institute (ATRI) at University of Southern California. The first launch of the database was ADNI 1 that is a 5 years study. The further extension of 2 years resulted in the net version named ADNI GO and was then revised to ADNI 2 and ADNI 3.

The major intention of ADNI database is the development of biomarkers that enables to perform clinical trials on the early stages of AD. These biomarkers are used as predictors of cognitive decline. The images in ADNI are collected from different modalities, like MRI, fMRI, PET etc. By the current status of ADNI, it gathered thousands of data, such as brain scans, genetic profiles and cerebrospinal fluid biomarkers. In this work, the classification of the stages using transfer learning and structural MRI data are used along with the demographic biomarkers.

5.2. Network architecture

Transfer learning is performed by modifying pre-trained network based on the requirement and training its new dataset. In this work, the pre-trained networks are used to classify the brain images for AD diagnosis via transfer learning are GoogLeNet, ResNet-18 and AlexNet. These networks are selected due to the fact that they are designed in three different structures and unique features. The advantage of each of these designs is evaluated and modified for the classification of various phases of AD.

5.2.1. AlexNet

The smallest pre-trained network available for transfer learning is Alexnet which has 25 layers in total with a depth of 8 layers. The layer arrangement of AlexNet takes a sequence of convolution layer, normalization layer, pooling layer, Rectified Linear Unit (ReLU) and fully connected layer. The Alexnet refer simple among the existing pre-trained networks in terms of structure and classification accuracy.

5.2.2. GoogLeNet

GoogLeNet is a 144 layered network, which works on input images of size 224x224x3. The network has a depth of 22 layers. The inception module of GoogLeNet repeatedly connected one after another in a specific manner to form the complete structure. The inception module is the concatenation of the outputs of several convolution layers of different filter sizes acting parallel upon the previous layer. Thus the concatenated feature map comprises of the features of the same input acquired from different space of filters. This enhances the effectiveness of learning. An additional feature in GoogLeNet is the auxiliary classification output layers resulting in a deeper network.

5.2.3. ResNet-18

ResNet-18 architecture is one of the ResNet architectures that is formed by repeated addition of residual blocks. These blocks release the complexity in the network architecture, thereby improving its performance. The complexity of networks tends to increase when the number of convolution layers is increased which in turn makes the network deeper. This pattern of increasing the depth of networks leads to poor results as well as optimization issues. Residual modules of ResNet architectures have proven as the best solution for this concern. ResNet-50 and ResNet-101 are some other networks that follow ResNet architecture, propagate the added result of the input and the convolution layer, which is termed as residue. This reduces the overhead of the propagation of larger number of features. Thus, the over-fitting problem is reduced for faster optimization of larger networks.

5.3. Performance analysis

The trained network performance is calculated with accuracy and then compared with the other pre- trained networks. The network architecture, preprocessing stages and the training options are modified to reach an improved accuracy in classification. The confusion matrix parameter of TP, TN, FP and FN helpful in analyzing that performance.

Accuracy comments on correctly made network predictions and it is calculated from the ratio of true estimates to total estimates. This is because accuracy is not trustworthy if the performance of the network for all the classes are in equal level.

As per the simple definition, error is the inverse of accuracy in the sense, $Error = 1 - Accuracy$. It signifies the number of misclassified instances by taking a ration of the wrong predictions to total number of predictions.

False positive rate otherwise called fall-out, comments on the rate of positive prediction made by the network. It is given as the ratio of wrongly predicted positive under total number of negatives.

Precision gives an overview of how many instances in the prediction are actually positive among the total number of positive predictions. Thus, it is also known as positive predictive value (PPV). The best value of precision is given as 1.0 or 100% and the worst is 0.

Recall or sensitivity otherwise called as true positive rate that is expressed as $Truepositiverate = 1 - Falsepositiverate$. It is the suitably classified positives among the total positive instances,

$$F1 - score = \frac{2.Precision.Recall}{Precision + Recall} \tag{5}$$

F1-score is also called the harmonic mean for precision and recall, because the F1-score is useful when the model has low precision and high recall or vice versa. It is the most utilized performance parameter when the performance of the network on every class is not in the same level.

Specificity or selectivity of the network is the ratio of correctly identified negatives among actual negatives, which is labelled as true negative rate.

$$Mathews\ Correlation\ Coefficient\ (MCC) = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (6)$$

MCC is the most effective way to quantify the quality of classification if the number of images in the classes is very different. The output of MCC falls in the range of -1 to $+1$ where -1 stands for misclassification and $+1$ denotes perfect classification.

$$Kappa = \frac{Accuracy - Randomaccuracy}{1 - Randomaccuracy} \quad (7)$$

where

$$Randomaccuracy = \frac{(TN + FP)(TN + FN) + (FN + TP)(FP + TP)}{(TP + TN + FP + FN)^2} \quad (8)$$

Kappa gives the comparison of accuracy of the system with that of an expected accuracy. The expected accuracy or random accuracy is given as the ratio of sum of actual and predicted pairs to the square of the total number of instances.

6. Implementation and results

6.1. Data acquisition and preprocessing

The image samples are obtained from the ADNI database with various classes of AD. The images are of the MRI modality in which the subjects come under the age category of 55 to 100 years. All of the observed subjects are in the age above 50 years because the chances of being affected by Alzheimer's disease are more in this age. It is also mentioned in surveys that all cases of Alzheimer's disease reported are above 50 years of age. Every class contains the images of male and female subjects. The images collected from 5 classes were used in transferring learning as image data store, which is separated from training and testing. In this 6000 images utilized for training and 1800 images utilized for testing the DNNs therefore total 7800 images utilized for deep neural network classifiers.

The images collected from each class are used in deep learning as image data store. The images are placed into data store and resized to meet the input dimensional requirement of the network input layer. The pre-processing stages are done by using the augmented image data store function. The size required for AlexNet was 227 × 227 × 3 and the requirement of all other networks was224 × 224 × 3. Before training and testing the image samples of the target domain, they undergo the preprocessing process. MRI scanners through the training process can suffer degradation, like low contrast, based on poor luminance cause with optical devices. For recovering the visual feature of images, image development systems, like linear contrast stretching, are created to recover that distribution of pixels through a broad range of intensities. The data store is separated as training and testing with the ratio 7:3 in such a way that the images for both the sets are selected randomly. It is also ensured that the images in testing set are exclusive from that of training set for effective validation of the classification.

The detailed class-wise demographic information is given on Table 1.

Table 1
Demographic Data.

Disease Class	Age	Number of female subjects	Number of male subjects	Total number of subjects
AD	**55 – 90**	**180**	**97**	**277**
LMCI	**60 – 95**	**128**	**106**	**234**
MCI	**55 – 100**	**117**	**111**	**228**
EMCI	**60 – 95**	**192**	**72**	**264**
CN	**55 – 95**	**97**	**65**	**162**

6.1.1. DNN (Deep neural network)

The images from the ADNI database are categorized with DNN architecture by transfer learning of binary and multiple class issue. The detection of Alzheimer's disease is the great importance of research under medical sciences. However, helps their treatment at the time of diagnosis their stages, which further maximizes the importance of multi-class problem. Shallow convolutional layers of previously trained AlexNet model are transferred over the data network image, composed with minimum-level characteristics removed more than 1,000,000 images. Through, the transfer of the convolutional layers, the layers of the AlexNet architecture are precisely matched with the complete and segmented brain scanners. The final three layers are configured for training options for the novel classification issue. The GM, WM and CSF segments, together with non-segmented image, as four data sets on DNN model is trained to learn that particular characteristics of the task. The training and testing processes of every data set were recorded time duration of 100 epochs and its equivalent evaluation outcomes were arrived under the form of confusion matrix. The proposed system is trained and experimented for binary and multi-class Alzheimer's classification. This outcomes portrays that classification outcomes on NC, EMCI, MCI, LMCI and AD segments segmented as individual segments.

Data augmentation under data analysis systems utilized to raise the amount of data by involving somewhat modified copies of existing data or recently generated synthetic data as of existing data.

In this 6,000 images utilized for training and 1800 images utilized for testing the DNNs therefore total 7800 images utilized for deep neural network classifiers.

6.2. Networks used

Transfer learning is performed using pre-trained networks; GoogLeNet, ResNet-18 and AlexNet and the performance of each network is analyzed. It was found that the time required for training increases proportionally with the complexity of networks used. The comparison of performance of these three pre-trained networks is done with parameters calculated from confusion matrix.

6.3. Training options

The training options and hyper parameters assigned for training the networks are listed in Table 2. Here, a same set of training options is used to compare the performance of different architectures. Stochastic Gradient Descent Method (SGDM) is the optimization algorithm used in the training process. The momentum is set as 0.9. The networks are

Table 2
Hyper-parameters.

Parameters	Value
Optimization algorithm	**SGDM**
Momentum	**0.9**
Initial learning rate	$\mathbf{10^{-4}}$
Maximum number of epochs	**100**
Mini-batch size	**32**
Validation frequency	**3**

trained at the learning rate of 10^{10} for 100 epochs. The dataset is divided into mini-batches of size 32 and they are shuffled during the training epochs. The model was trained for 100 epochs. The horizontal axis refers the number of epochs and vertical axis implies an error rate. It could terminate training; while an error rate of validation data is minimal. As a result, it maximizes the number of epochs, will have over fitted model. At the age of deep learning, it is not so common to have premature stop.

6.4. Training progress

The networks are trained with a set of training data after deciding the network architecture, training options and data set. The performance comparison is made on GoogLeNet, ResNet-18 and AlexNet. After the training performance of these networks are validates based on accuracy and other metrics. The altered pre-trained networks are trained with similar training parameters with a fixed amount of epochs as shown in Fig. 2.

It is apparent from the plot that AlexNet and ResNet-18 achieved better optimization than GoogLeNet in which ResNet – 18 performed better that that of AlexNet. Since GoogLeNet plot is not optimized in 40 epochs, it can be inferred that 40 epochs are not sufficient for GoogLeNet to achieve optimization. Thus, it is noticed that bigger networks need more epochs to reach the optimum level.

6.5. AD classification using Pre-trained networks

The Table 3 lists out the classification performance of three pre-trained networks. These results show that the performance of ResNet – 18 predominates the other two networks and GoogLenet is the least performing network among these three networks.

6.6. Class-wise performance assessment

To further authenticate that performance of networks, an evaluation is done by comparing the performance of all the three networks classifying in every class AD using. This makes use of confusion matrix of all the three networks. This enables us to assess the performance of the networks and analyze whether the network performs uniformly for all the classes. This helps in deciding better training parameters with which the network has to be trained to perform well in classifying different classes of AD.

6.6.1. Confusion matrix

The confusion matrix of Alex Net, ResNet-18 and GoogLeNet, is shown in Table 4. It is apparent that a larger number of correctly predicted elements are in the class AD.

Table 3
Overall network performance.

Performance	GoogLeNet	Alexnet	ResnNet-18
Accuracy	97.56	96.19	98.63
Error	2.44	3.81	1.37
Sensitivity/Recall	93.77	90.10	96.88
Specificity	98.45	97.57	99.15
Precision	94.44	90.99	96.33
False Positive Rate	1.55	2.43	0.85
F1 –score	93.97	90.44	96.56
Kappa	92.53	88.11	95.73

6.6.2. Performance parameters of every class

The class-wise performance of AlexNet is displayed in Table 5. It is observed that AlexNet performs well for the class EMCI and it is least performed for AD. But it is apparent as the table of all classes is categorized with a similar range of performance regarding all the parameters.

6.6.3. Class-wise performance parameters of AlexNet

6.6.3.1. Class-wise performance parameters of ResNet-18. The class-wise performance of ResNet-18 given in Table 6 shows that the brain MRI images are the most correctly classified for the class CN and is least correctly classified for the class MCI. But it can be observed that the class CN is very well classified with 97.01% of accuracy and the accuracy of other classes are less compared to the CN. The least accuracy of classification is 83.33% of MCI, which is not in a near range of CN. Thus, in the case of ResNet-18, cannot depend only on accuracy to analyze overall network performance. The value of specificity, precision, sensitivity and MCC should also be considered.

6.6.4. Class-wise performance parameters of GoogLeNet

Finally, the class-wise performance parameters of GoogLeNet calculated from the values are shown in Table 7. CN is classified with comparatively high accuracy and the least performance is for MCI. Similar to AlexNet, the classification accuracy of GoogLeNet on all the five classes is in a near range with less variation in between.

On comparing the performance of these three pretrained networks, the accuracy in detecting the early development is higher when the ResNet-18 is used. The deeper network structure contributes the improved precision in the early development of the cognitive impairment and AD.

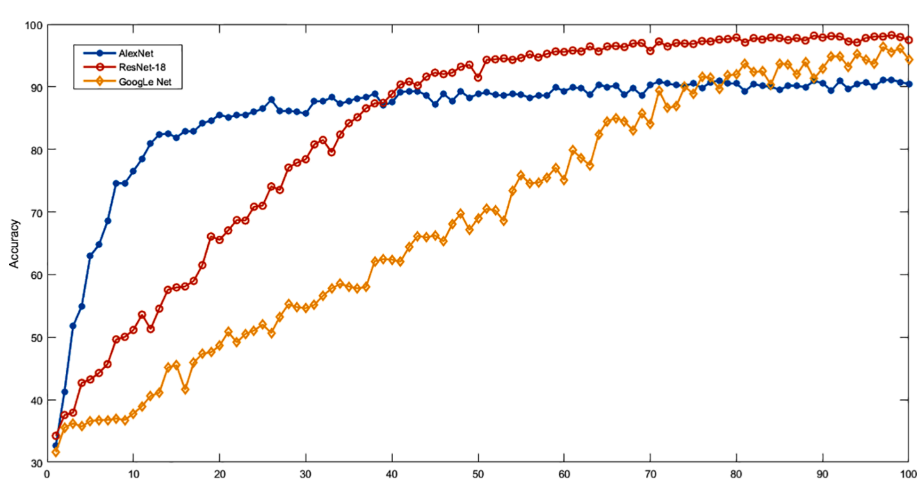

Fig. 2. Training progress.

Table 4
The five instances of confusion matrix of AlexNet, ResNet-18, GoogLeNet.

	Alex Net				ResNet-18				GoogLeNet			
Accuracy	TP	FP	FN	TN	TP	FP	FN	TN	TP	FP	FN	TN
AD	230	11	10	109	230	9	1	120	194	5	8	153
CN	200	4	6	150	183	1	3	173	153	6	5	196
EMCI	154	5	5	196	230	2	1	127	253	1	5	101
LMCI	252	8	9	91	150	2	2	206	220	4	5	131
MCI	235	6	5	114	253	3	1	103	260	5	2	93

Table 5
Class-wise performance of AlexNet.

Performance Metric	AD	CN	EMCI	LMCI	MCI
Accuracy	94.08	97.34	97.51	95.19	96.82
Error	5.92	2.66	2.49	4.81	3.18
Recall	90.61	97.01	91.67	87.88	83.33
Specificity	95.16	97.42	98.93	97.34	99.00
Precision	85.37	90.44	95.43	90.63	93.10
False Positive Rate	4.84	2.58	1.07	2.66	1.00
F1 Score	87.92	93.61	93.51	89.23	87.95
Matthews Correlation Coefficient	84.06	92.02	92.00	86.15	86.30
Kappa	84.00	91.93	91.97	86.14	86.13

Table 6
Class-wise performance of ResNet-18.

Performance Metric	AD	CN	EMCI	LMCI	MCI
Accuracy	97.51	98.88	99.14	98.88	98.71
Error	2.49	1.12	0.86	1.12	1.29
Recall	92.78	99.15	96.49	96.59	99.38
Specificity	98.99	98.82	99.79	99.56	98.60
Precision	96.62	95.47	99.10	98.46	92.00
False Positive Rate	1.01	1.18	0.21	0.44	1.40
F1 Score	94.66	97.27	97.78	97.51	95.55

Table 7
Class-wise performance of GoogLeNet.

Performance Metric	AD	CN	EMCI	LMCI	MCI
Accuracy	96.39	97.17	98.28	97.60	98.37
Error	3.61	2.83	1.72	2.40	1.63
Recall	94.22	98.72	95.61	90.15	90.12
Specificity	97.07	96.78	98.93	99.78	99.70
Precision	90.94	88.51	95.61	99.17	97.99
False Positive Rate	2.93	3.22	1.07	0.22	0.30
F1 Score	92.55	93.33	95.61	94.44	93.89

7. Conclusion

In this paper, three different pre-trained deep learning models through transfer learning using GoogLeNet, ResNet-18 and AlexNet is proposed. These models are trained with 6000 MRI images obtained from the ADNI database to classify five different classes of NC, EMCI, MCI, LMCI and AD. Originally, the deep learning networks AlexNet, ResNet-18 and GoogLeNet are designed to classify 1000 classes of a database Image net. The architecture of these networks is modified through transfer learning to classify the five different AD classes obtained from ADNI database. The overall performance of the networks and the class-wise evaluation is carried out. The outcomes portray that efficiency of networks slowly reduces as the complexity of the structure increases. Based on the performance analysis, it is evident that all these three networks are efficiently classifying various stages of AD with high accuracy. The AlexNet relatively performs well using transfer learning. The future tentative work will focus on designing deep learning networks exclusively for classifying the AD classes.

Funding information

None.

CRediT authorship contribution statement

Jayanthi Venkatraman Shanmugam: Conceptualization, Methodology, Writing - original draft. **Baskar Duraisamy:** Supervision, Writing - review & editing. **Blessy Chittattukarakkaran Simon:** Writing - review & editing. **Preethi Bhaskaran:** Writing - review & editing.

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.

References

[1] R. Rosas-Romero, Classification of Alzheimer ' s disease subjects from MRI using hippocampal visual features To cite this version : HAL Id : hal-00993379, Comput. Med. Imaging Graph. 44 (1) (2015) 1–6.
[2] G. Manley, Public Access NIH Public Access, 71 (2), 233–236.
[3] G. Lee, K. Nho, B. Kang, K.A. Sohn, D. Kim, Predicting Alzheimer's disease progression using multi-modal deep learning approach, Sci. Rep. 9 (1) (2019) 1–2.
[4] F.F. Asrami, Alzheimer's Disease Classification using K-OPLS and MRI, Linköping University, 2012.
[5] D. Baskar, V.S. Jayanthi, A.N. Jayanthi, An efficient classification approach for detection of Alzheimer's disease from biomedical imaging modalities, Multimedia Tools Appl. 78 (10) (2019) 12883–12915.
[6] P. Geetha, V.S. Jayanthi, A.N. Jayanthi, Multiple share creation based visual cryptographic scheme using diffusion method with a combination of chaotic maps for multimedia applications, Multimedia Tools Appl. 78 (13) (2019) 18503–18530.
[7] A. Sayeed, M. Petrou, N. Spyrou, A. Kadyrov, T. Spinks, Diagnostic features of Alzheimer's disease extracted from PET sinograms, Phys. Med. Biol. 47 (1) (2002) 137–148.
[8] K.D. Desai, S. Parmar, Effective early detection of Alzheimer's and Dementia disease using Brain MRI Scan Images, Int. J. Emerg. Technol. Adv. Eng. 2 (4) (2012).
[9] B. Duraisamy, J.V. Shanmugam, J. Annamalai, Alzheimer disease detection from structural MR images using FCM based weighted probabilistic neural network, Brain Imaging Behav. 13 (1) (2019) 87–110.
[10] J. Rieke, F. Eitel, M. Weygandt, J.D. Haynes, K. Ritter, Visualizing convolutional networks for MRI-based diagnosis of Alzheimer's disease, in: Understanding and Interpreting Machine Learning in Medical Image Computing Applications, Springer, Cham, 2018, pp. 24–31.
[11] Y. Zhang, S. Wang, P. Phillips, Z. Dong, G. Ji, J. Yang, Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC, Biomed. Signal Process. Control 21 (2015) 58–73.
[12] M. Liu, J. Zhang, E. Adeli, D. Shen, Landmark-based deep multi-instance learning for brain disease diagnosis, Med. Image Anal. 43 (2018) 157–168.
[13] R.N. Kalaria, G.E. Maestre, R. Arizaga, R.P. Friedland, D. Galasko, K. Hall, J. A. Luchsinger, A. Ogunniyi, E.K. Perry, F. Potocnik, M. Prince, R. Stewart, A. Wimo, Z.-X. Zhang, P. Antuono, Alzheimer's disease and vascular dementia in developing countries: prevalence, management, and risk factors, Lancet Neurol. 7 (9) (2008) 812–826.
[14] S. Mythili, K. Thiyagarajah, P. Rajesh, F.H. Shajin, Ideal position and size selection of unified power flow controllers (UPFCs) to upgrade the dynamic stability of systems: an antlion optimiser and invasive weed optimisation algorithm, HKIE Trans. 27 (1) (2020) 25–37.
[15] P. Rajesh, F.H. Shajin, A Multi-Objective Hybrid Algorithm for Planning Electrical Distribution System, Eur. J. Electr. Eng. 22 (4-5) (2020) 224–509.
[16] F.H. Shajin, P. Rajesh, Trusted Secure Geographic Routing Protocol: outsider attack detection in mobile ad hoc networks by adopting trusted secure geographic routing

protocol, Int. J. Pervasive Comput. Commun. (2020), https://doi.org/10.1108/IJPCC-09-2020-0136.
[17] M.K. Thota, F.H. Shajin, P. Rajesh, Survey on software defect prediction techniques, Int. J. Appl. Sci. Eng. 17 (2020) 331–344.
[18] P.S. Mathuranath, R. Menon, N. Ranjith, P.S. Sarma, J. Verghese, A. George, S. Justus, M.S. Kumar, Incidence of Alzheimer's disease in India: A 10 yearsfollow-up study, Neurol. India 60 (6) (2012) 625, https://doi.org/10.4103/0028-3886.105198.
[19] C.B. Akgül, D. Ünay, A. Ekin, Automated diagnosis of Alzheimer's disease using image similarity and user feedback, in: Proceedings of the ACM International Conference on Image and Video Retrieval, 2009, pp. 1–8.
[20] E.-S. El-Dahshan, T. Hosny, A.-B. Salem, Hybrid intelligent techniques for MRI brain images classification, Digital Signal Process. 20 (2) (2010) 433–441.
[21] X. Zhu, H.I. Suk, D. Shen, Matrix-similarity based loss function and feature selection for Alzheimer's disease diagnosis, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3089–3096.
[22] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, D. Feng, Early diagnosis of Alzheimer's disease with deep learning, in: 2014 IEEE 11th international symposium on biomedical imaging (ISBI), IEEE, 2014, pp. 1015–1018.
[23] Z. Li, Y. Zhao, S. Wang, Brain MRI image classification based on transfer learning and support vector machine, in: Optoelectronic Imaging and Multimedia Technology VI, International Society for Optics and Photonics, 2019, p. 111871.
[24] K. Oh, Y.C. Chung, K.W. Kim, W.S. Kim, I.S. Oh, Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning, Sci. Rep. 9 (1) (2019) 1–16.
[25] Y. Kazemi, S. Houghten, A deep learning pipeline to classify different stages of Alzheimer's disease from fMRI data, in: 2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), IEEE, 2018, pp. 1–8.
[26] Y.P. Lin, T.P. Jung, Improving EEG-based emotion classification using conditional transfer learning, Front. Hum. Neurosci. 11 (2017) 334.
[27] Almadhoun, R. Husam, Samy S. Abu-Naser. Classification of Alzheimer's Disease Using Traditional Classifiers with Pre-Trained CNN. In 2021 International Journal of Academic Health and Medical Research (IJAHMR) 5.4.
[28] A. Balagopalan, B. Eyre, J. Robin, F. Rudzicz, J. Novikova, Comparing Pre-trained and Feature-Based Models for Prediction of Alzheimer's Disease Based on Speech, Front. Aging Neurosci. 13 (2021) 189.
[29] D.a. Ma, E. Yee, J.K. Stocks, L.M. Jenkins, K. Popuri, G. Chausse, L. Wang, S. Probst, M.F. Beg, K. Anstey, Blinded Clinical Evaluation for Dementia of Alzheimer's Type Classification Using FDG-PET: A Comparison Between Feature-Engineered and Non-Feature-Engineered Machine Learning Methods, J. Alzheimers Dis. 80 (2) (2021) 715–726.
[30] R.E. Turkson, H. Qu, C.B. Mawuli, M.J. Eghan, Classification of Alzheimer's Disease Using Deep Convolutional Spiking Neural Network, Neural Process. Lett. 53 (4) (2021) 2649–2663.
[31] Roshanzamir, Alireza, A. Hamid, S.B. Mahdieh, Transformer-based deep neural network language models for Alzheimer's disease risk assessment from targeted speech. In 2021 BMC Medical Informatics and Decision Making 21.1: 1-14.
[32] E. Hosseini-Asl, R. Keynton, A. El-Baz, Alzheimer's disease diagnostics by adaptation of 3D convolutional network, in: 2016 IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pp. 126–130.
[33] L. Torrey, J. Shavlik, E.S. Olivas, J.M. Guerrero, M.M. Sober, J.M. Benedito, A. S. Lopez, Handbook of Research on Machine Learning Applications and Trends, 2010.
[34] B.C. Simon, D. Baskar, V.S. Jayanthi, Alzheimer's Disease Classification Using Deep Convolutional Neural Network, in: 2019 9th International Conference on Advances in Computing and Communication (ICACC), IEEE, 2019, pp. 204–208.