



# Alzheimer disease diagnosis for magnetic resonance brain images using deep learning neural networks

Heba M. Ahmed<sup>1</sup> · Zeinab F. Elsharkawy<sup>2</sup>  · Ahmed S. Elkorany<sup>3</sup>

Received: 20 April 2021 / Revised: 21 April 2022 / Accepted: 27 October 2022 /  
Published online: 19 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

In this work, a Deep Convolutional Neural Network (DCNN) framework for Alzheimer's Disease (AD) diagnosis based on brain Magnetic Resonance Imaging (MRI) scans is presented. A multiclass DCNN classifier is used to discriminate between Normal Controls (NC), Mild Cognitive Impairment (MCI), and AD. The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset was used to train and test the proposed DCNN. Different train-test ratios have been examined. Average accuracies of 100% for AD/NC, 92.93% for NC/MCI, and 99.21% for AD/MCI were obtained. The proposed system achieved an average accuracy of 93.86% for a three-way AD/MCI/NC classification. To further examine the proposed system performance, Receiver Operation Characteristics (ROC) analysis and Confusion Matrix (CM) were also used. For certain scenarios, the Area Under ROC Curve (AUC) values of 1, 1, and 0.989 were obtained for AD, NC, and MCI, respectively. The results show higher metrics compared to previously published studies concerning AD diagnosis.

**Keywords** Deep CNN · ADNI · Alzheimer's disease · MRI · ROC

---

✉ Zeinab F. Elsharkawy  
zeinab\_elsharkawy@yahoo.com

Heba M. Ahmed  
eng.heba55@gmail.com

Ahmed S. Elkorany  
elkoranyahmed@yahoo.com

<sup>1</sup> Informatics Research Institute (IRI), City of Scientific Research and Technological Applications (SRTA-City), Alexandria, Egypt

<sup>2</sup> Engineering Department, Nuclear Research Center, Egyptian Atomic Energy Authority, Cairo, Egypt

<sup>3</sup> Department of Electronics and Electrical Communication Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

## 1 Introduction

Alzheimer's disease is a neurological brain disorder that results in the death of brain cells responsible for memorizing and thinking skills. Its consequences increase slowly and deteriorate over time making the patients unable to continue their ordinary daily activities. As stated by the Alzheimer's Association, AD is one of the top ten diseases in the United States which causes death [4]. It is caused by many factors among which age is the most significant. Elderly adults aged about 65 are at a high risk of suffering from this disease [15]. The key to curing Alzheimer's disease is early detection where appropriate treatment can slow the disease's progression. Therefore, the development of an early and accurate AD diagnosis is essential.

Pattern recognition, classification, and Machine Learning (ML) have gained a great attraction recently in emerging an automated diagnosis system of brain diseases with neuro-images such as MRI [2, 3], Diffusion Tensor Imaging (DTI) [9], functional MRI (fMRI) [26], Positron Emission Tomography (PET) [20]. There is large number of researchers focused on advanced ML models that utilize MRI data for AD diagnosis. These studies revealed that MRI scans are the most proper imaging modality in clinical diagnosis [30] and can be used to track different clinical AD phases [11].

Recently, advanced Deep Learning (DL) techniques especially Convolutional Neural Networks (CNN), have gained great interest in solving problems in automated medical imaging systems [5, 7]. The diagnosis of AD is a multi-tests assessment and requires a highly discriminative feature representation for automated classification. DL methods are capable of learning such representation from data. They can reveal latent or hidden features that can improve disease detection and classification; hence it has been used in automated AD diagnosis recently.

The current study aims to develop and validate a new DCNN framework capable of identifying individual diagnoses of AD, MCI, and NC based on brain MRI scans. The proposed classifier was trained and tested using the ADNI dataset. Specificity, Sensitivity, Accuracy, and Balanced Accuracy are used to evaluate the new DCNN classifier's performance. The suggested 3-class classification algorithm's performance is also evaluated using ROC analysis and CM. The performance of the given method was compared to previously published similar studies. The generated results outperform the published ones, demonstrating the robustness of the proposed DCNN framework.

The rest of the paper is structured as follows. Section 2 introduces the materials and methods that includes related AD diagnosis work and the proposed DCNN model. Section 3 states the experimental results. Finally, Section 4 reports the paper conclusion.

## 2 Materials and methods

### 2.1 Related work

Several studies have already been designed for early AD diagnosis. The majority of these AD studies are based on artificial intelligence and the ADNI dataset. PET, Genetics, Cognitive tests, MRI, Cerebrospinal fluid, and blood biomarkers are included in this dataset as predictors of the disease. Except for the study in [13], which used only the Open Access Series of

Imaging Studies (OASIS) dataset, the ADNI dataset was used in all of the studies in [1, 6, 8, 10, 12–14, 16–18, 21–25, 27–29, 31–36], and studies in [22–24] used both the ADNI and the OASIS datasets.

First, we introduce the related studies for binary classification of AD. Raza et.al [23] proposed a new ML-based screening and diagnosis of AD. The process of the AD diagnosis was achieved by the DL analysis of MRI scans, followed by an activity screening using body-worn inertial sensors. An accuracy of 95% was achieved to classify the patient's daily activities. Another classification approach for the detection of AD diagnosis, based on Mean Diffusivity (MD) extracted from DTI and Structural MRI (sMRI) brain images, has been introduced in [1]. An accuracy of 85.0%, 92.5%, and 80.0% respectively was accomplished for AD/MCI, AD/NC, and MCI/NC.

The study in [28] utilized a parameter-based DL method to discriminate the MCI patients who are susceptible to AD in 3 years and the MCI stable patients within the same period. The obtained 10-fold cross-validated accuracy, the area under the curve (AUC), specificity, and sensitivity are 86%, 0.925, 85%, and 87.5% respectively. A 3D DCNN architecture for brain MRI classification has been presented in [16]. Classification of AD versus MCI and NC was conducted on ADNI dataset. The study shows that similar performance can be obtained with reduced steps (i.e., skipping feature extraction steps). Hon and Khan [13] presented a transfer learning-based technique to identify AD using brain sMRI. Two DL techniques, i.e., Inception V4 and VGG16 were examined. A better performance was reached compared with current DL based approaches.

Automatic AD and MCI classifier that is based on brain MRI and deep neural networks was designed by Basaia et.al [6]. The system performance was examined in identifying AD, converters MCI (c-MCI) and Stable MCI (s-MCI). High accuracy levels were obtained in all cases. In [24], DL technique was presented for AD diagnosis and monitoring of the AD patient. The technique demonstrated the importance of high-accuracy AD diagnosis with physical activities, as well as how these activities can be logged with high accuracy without human intervention. Janghel et al. [14] applied the VGG 16 architecture of DCNN for features extraction and SVM for the AD detection task. The CNN performance was improved by performing some preprocessing on the image dataset before feature extraction. An average accuracy of 99.95% and 73.46% was achieved respectively for the fMRI dataset and the PET classification.

Xiao et al. [34] used Sparse logistic regression (SLR) for early AD diagnosis of 197 cases. To impose a sparsity constraint on logistic regression, SLR utilized  $L_{1/2}$  regularization. When compared to other classical techniques, experimental results showed that the SLR improved the AD/MCI classification performance. Several methods are also used for AD diagnoses such as a multi-model DL framework in [18, 36]. In [18], a 3D Densely Connected Convolutional Networks (3D DenseNet) was constructed to learn features of the 3D patches then these learned features were combined to classify disease status. To extract the 3D patches features, a 3D Densely Connected Convolutional Network (3D DenseNet) was built, and then these features were utilized for disease status classification. The proposed multi-model procedure outscored single-model procedures as well as a number of other competing methods. Two independent CNNs were used in [36] to obtain PET and MRI image features. The results showed that the presented multi-modal auxiliary diagnosis achieved a superb efficiency.

Concerning the three-way AD/MCI/NC classification, Gupta et al. [12] proposed an approach for computerized AD diagnosis in which cross-domain features are used to

represent MRI data. High classification performance was achieved using this approach. Payan et al. [21] proposed a classification MRI-based technique that utilizes 3D CNNs and sparse autoencoders to predict patient disease status. The results outperform the ordinary 3D CNNs algorithm in comparison to several other classifiers in the literature. Cárdenas-peña et al. [10] presented a DL model based on central kernel alignment. They compared the supervised pre-training method to two unsupervised initialization approaches. The experiment showed that artificial neural network (ANN) pretraining outperforms the contrasted algorithms and reduces the class biasing. Hence, better MCI discrimination is obtained. Discrimination between AD, MCI, and NC individuals using 3D MRI and neuropsychological measures (NM) was introduced in [27]. A fusion pipeline combines data from multiple modalities including volumetric MRI and NM features. The results show the effectiveness of the presented fusion pipeline compared to the 3D CNN architectures. A robust automatic technique of High-level Layer Concatenation Autoencoder (HiLCAE) and 3D-VGG16 are utilized to detect AD using MRI and PET images was introduced by Vu et al. [32]. The experiment was conducted on an ADNI dataset and classified into one of three classes: NC, MCI, and AD.

Wang et al. [33] proposed an ensemble of 3D-DenseNet for MCI and AD classification. Many experiments were performed to analyze the performance of the system with different architectures and hyper-parameters. The proposed model outperforms the other model using the ADNI dataset. Xiao et al. [35] introduce a new classification framework to classify AD or MCI from NC. Three different features (i.e., gray-level co-occurrence matrix (GLMC), Gabor feature, and gray-matter volume) were used. The results show that the multi-feature fusion improves system performance. Tong et al. [29] presented a new data fusion method from multiple modalities for AD classification. The AUC of receiver-operator characteristic (ROC) is 82.4% between MCI patients and NC, 98.1% between AD patients and NC, and 77.9% in a 3-way classification. Several techniques were also used for AD diagnoses such as DL technology [8, 22], representation of regional abnormality via DL [17], universum SVM [25], and hippocampal atrophy technique [31].

The related study with speech aspects was introduced in [19] where a new ML technique used the spectrogram patient's speech features for AD detection. The patients' speech data from NC and AD was collected and examined. The results presented that, among the used models, the Logistic Regression CV model had accomplished the highest performance. As mentioned before, deep learning methods are a new candidate in the medical imaging field. Therefore, a CNN-based framework is examined in this paper for the MRI classification into one of the three output groups.

## 2.2 Proposed work

The proposed network is a new DCNN that uses an MRI brain scan to automate AD classification. For feature extraction, the proposed framework includes three convolutional blocks. A convolutional layer, batch normalization layer, rectified linear unit (ReLU), and a max-pooling (MAX P) layer make up each convolutional block (Conv). A stride of 2 and a max-pooling of  $2 \times 2$  was utilized in each layer. After the convolutional blocks, a fully connected (FC) layer was applied for classification. The FC layer followed by Softmax layer that has three different output categories results in the predicted probability distribution of three classes (NC, MCI and AD). Figure 1 shows the proposed 16 layers of CNN architecture.

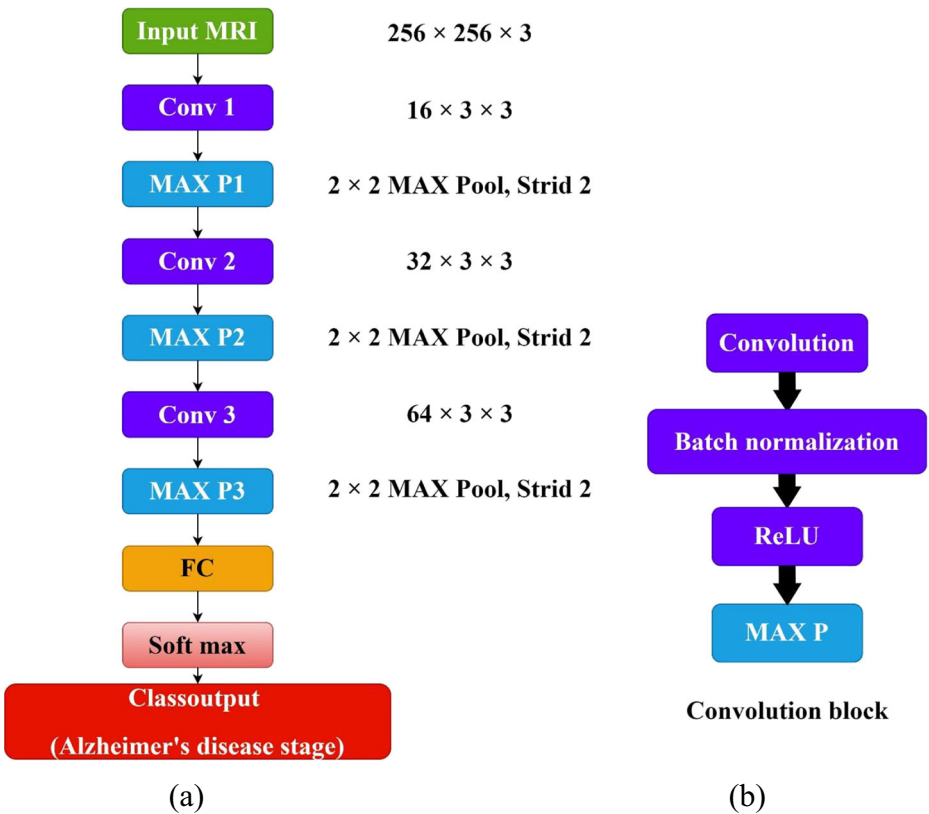


Fig. 1 a) The proposed CNN architecture b) The detailed convolution blocks

The fundamental block of DCNN is the convolution layer which contains kernels. These kernels are convolved to detect patterns and features over input images. The kernel is a filter of different windows (i.e.,  $3 \times 3$ ,  $7 \times 7$ , and  $13 \times 13$ ) that are convolved over each input to extract a feature map.

For example, if the convolutional layer is 1, then the feature map of this layer is  $j$  and it can be obtained as follows:

$$x_j^l = b_j^l + \sum_{i \in N_j} x_i^{l-1} * k_{ij}^l \tag{1}$$

where

- $N_j$  the set of the input feature map.
- $x_i^{l-1}$  the input feature map of the layer 1 (the output feature map of the previous layer).
- $k_{ij}^l$  the partial input feature map convolutional kernel.
- $b_j^l$  the bias offset of the feature map  $j$  after convolution.
- $*$  the convolutional calculation.

A nonlinear activation function called ReLU is applied after the convolution layer to obtain nonlinear feature maps. ReLU select the maximum value between zero and the input  $I_x$ ; therefore, the nonlinear activation map  $f(x)$  is calculated as:

$$f(x) = \max(0, I_x). \quad (2)$$

A DCNN gives translation-invariant feature maps via the use of the MAX P layer. Due to the pooling layer, the feature representation is stable and more concise and reduces the next stage computational burden. The max-pooling is computed as follows:

$$pool_{i,j} = \max_p f(x)_{i+p, j+p} \quad (3)$$

where the spatial positions are  $i$  and  $j$ , and  $p$  is the pooling window size.

In the end, the network normally has FC layers, where each feature map pixel is a neuron and forward to every neuron in the FC layers. One FC layer is used in our architecture. The softmax layer has three different output categories: AD, NC, and MCI. Depending on the feature representation, any input image was classified into one of these three categories. Furthermore, using the new proposed DCNN framework, the overall classification accuracy is improved. A CNN has neurons with weights and biases, just like a traditional NN. During the training task, the CNN learns these values and upgrades them with each new training task. However, in the case of CNNs, all hidden neurons in a layer have the same weights and bias values.

### 3 Experimental analysis

#### 3.1 Dataset

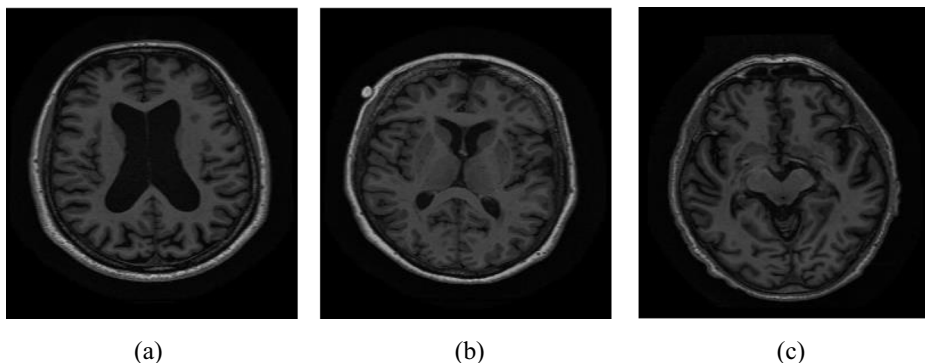
The employed dataset used is the structural brain MRI scans which are provided by the ADNI dataset (<http://adni.loni.usc.edu/>). A total number of 199 patients were reported, in which there were 42 AD, 97 MCI, and 60 NC cases. Table 1 describes the demographic characteristics of the used samples, including age, gender, and number. The employed dataset samples are shown in Fig. 2.

#### 3.2 Performance measurement

The performance examination was done using the following parameters i.e., Specificity, Sensitivity, Accuracy, and Balanced Accuracy. This is done to evaluate the overall performance of the proposed classification algorithm. True negative cases are referred to as specificity, which is defined as the proportion of negative cases correctly classified as true negative cases. True positive cases are referred to as sensitivity, which refers to the proportion of actual true cases correctly classified. Accuracy is a term that describes how close a measurement is to the actual value of a quantity. The average of each class's proportion

**Table 1** The dataset demographic information

| Category      | AD                | NC                | MCI               |
|---------------|-------------------|-------------------|-------------------|
| Number        | 42                | 60                | 97                |
| Female / Male | 27/15             | 38/22             | 40/57             |
| Age(Mean±SD)  | 74.21429±8.306729 | 75.23333±3.761386 | 74.47423±8.064961 |



**Fig. 2** Samples of the utilized dataset (a) AD (b) NC (c) MCI

corrects is used to calculate Balanced Accuracy. Table 2 illustrates these parameters. Where, TN (True Negative), and TP (True Positive) are the accurately classified cases. Also, the FN (False Negative), and FP (False Positive) are the inaccurately classified cases. The ROC analysis and CM are also used to evaluate the performance of the proposed 3-class classification algorithm.

### 3.3 Experimental setup and evaluation

To evaluate the classifier performance initially, the database image set was divided into two subsets i.e., testing set and training set. Different cases of dataset combinations are used (i.e., 90% of the images for training and 10% for testing the network, 80% for training and 20% for testing, and finally, 70% for training and 30% for testing).

**Table 2** The classifier evaluation parameters

| Parameters              | Mathematical equation                             |
|-------------------------|---|
| Sensitivity (SEN)       | $\frac{TP}{TP+FN}$                                |
| Specificity (SPE)       | $\frac{TN}{TN+FP}$                                |
| Accuracy (ACC)          | $\frac{TP+TN}{TP+TN+FP+FN}$                       |
| Balanced Accuracy (BAC) | $0.5 * (\text{Sensitivity} + \text{Specificity})$ |

**Table 3** Classification performance for (NC vs. AD)

|                         | ACC (%) | SPE (%) | SEN (%) | BAC (%) |
|-------------------------|---------|---------|---------|---------|
| 90% training-10%testing | 100     | 100     | 100     | 100     |
| 80% training-20%testing | 100     | 100     | 100     | 100     |
| 70% training-30%testing | 100     | 100     | 100     | 100     |
| Average                 | 100     | 100     | 100     | 100     |

**Table 4** Classification performance for (NC vs. MCI)

|                         | ACC (%) | SPE (%) | SEN (%) | BAC (%) |
|-------------------------|---------|---------|---------|---------|
| 90% training-10%testing | 93.75   | 90      | 100     | 95      |
| 80% training-20%testing | 93.55   | 100     | 83.33   | 91.67   |
| 70% training-30%testing | 91.49   | 93.10   | 88.89   | 90.996  |
| Average                 | 92.93   | 94.37   | 90.74   | 92.56   |

**Table 5** Classification performance for (AD vs MCI)

|                         | ACC (%) | SPE (%) | SEN (%) | BAC (%) |
|-------------------------|---------|---------|---------|---------|
| 90% training-10%testing | 100     | 100     | 100     | 100     |
| 80% training-20%testing | 100     | 100     | 100     | 100     |
| 70% training-30%testing | 97.62   | 100     | 92.31   | 96.154  |
| Average                 | 99.21   | 100     | 97.44   | 98.72   |

## 4 Result discussion

The DCNN Performance was validated with four classifications: Patients with NC vs. AD, NC vs. MCI, MCI vs. AD, and finally three-class classification (NC, MCI, and AD).

### 4.1 NC vs. AD

In comparing the NC and AD, a 100% accuracy has been obtained. Table 3 shows the Classification performance of (NC vs. AD). As can be seen in Table 3, the introduced method is effective in classifying NC and AD, with the highest classification rate (100%), a sensitivity of 100%, a specificity of 100%, and balanced accuracy of 100%.

### 4.2 NC vs. MCI

To further examine the system in terms of individual classes (NC vs. MCI), Table 4 reports the Classification performance for (MCI vs NC). In distinguishing MCI from NC subjects, the proposed method reached an accuracy of about 92.93%, a specificity of 94.37%, a sensitivity of 90.74%, and balanced accuracy of 92.56%.

**Table 6** Classification performance for three - class classification (AD, NC, and MCI)

|                         | ACC (%) | SPE (%) | SEN (%) | BAC (%) |
|-------------------------|---------|---------|---------|---------|
| 90% training-10%testing | 95      | 97.62   | 96.67   | 97.15   |
| 80% training-20%testing | 94.9    | 96.67   | 93.06   | 94.87   |
| 70% training-30%testing | 91.67   | 95.47   | 92.13   | 93.80   |
| Average                 | 93.86   | 96.60   | 93.95   | 95.27   |



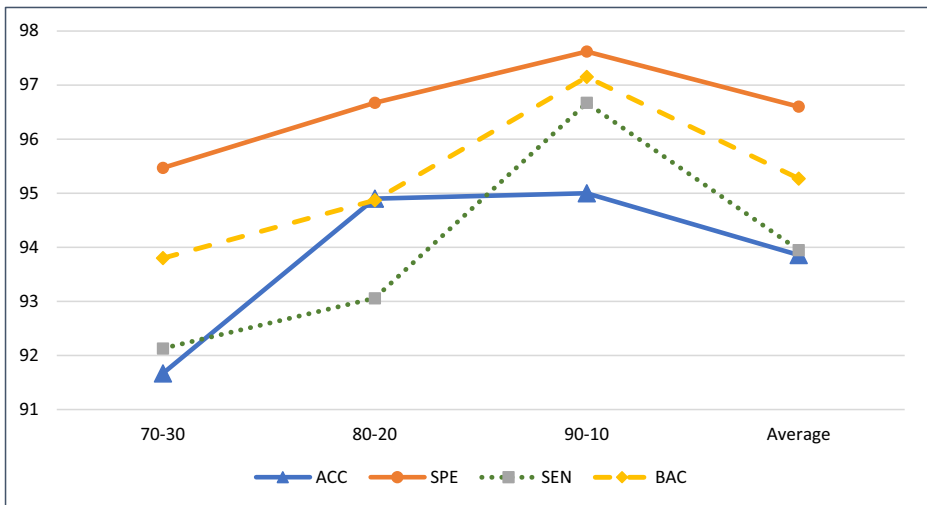


Fig. 3 The proposed DCNN framework for 3-class classification

### 4.3 MCI vs AD

The experimental results of AD and MCI classification are reported in Table 5. As shown, the highest accuracy and specificity (higher than 99%) are obtained. Moreover, sensitivity and balanced accuracy obtained are about 97.44% and 98.72% respectively.

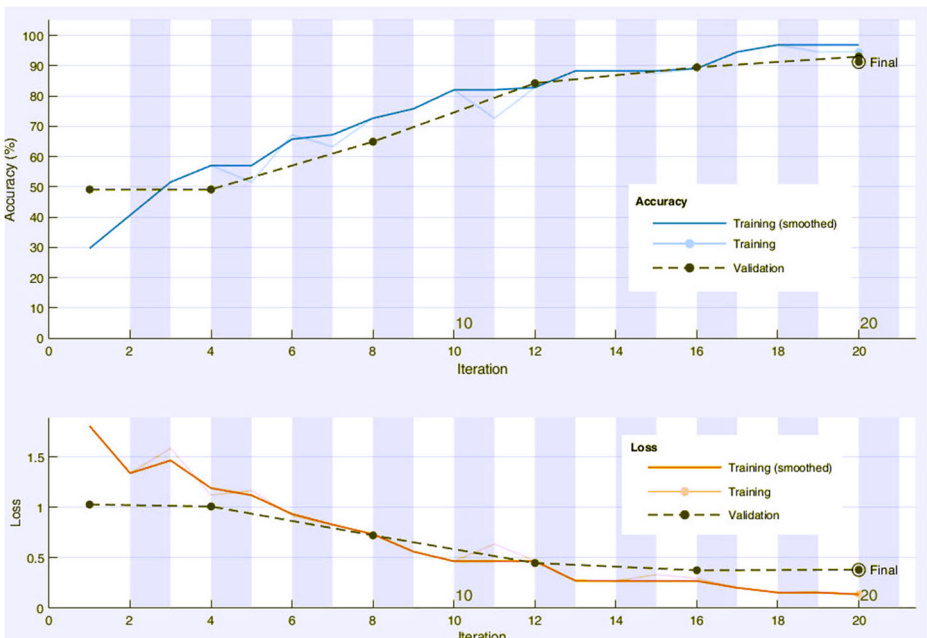


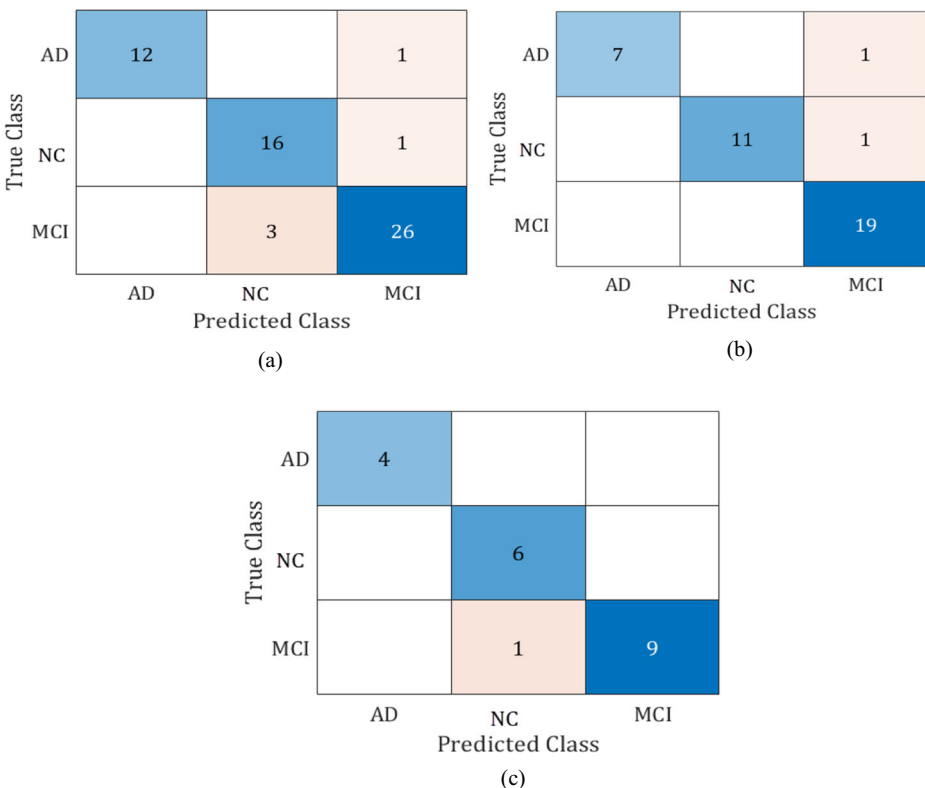
Fig. 4 The DCNN training behavior

#### 4.4 Three-class classification (classification of AD, NC, and MCI)

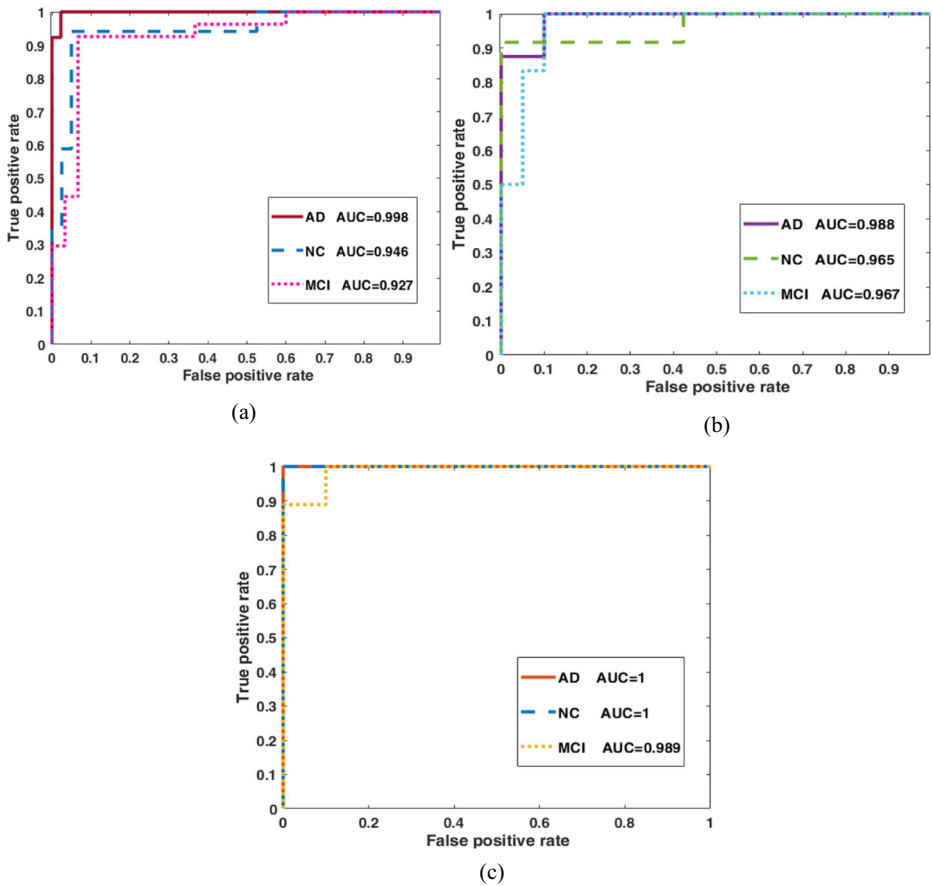
These are the results of the three-class classification. As reported in Table 6 and Fig. 3, the proposed method reached an accuracy of about 93.86%, a sensitivity of 93.95%, a specificity of 96.60%, and balanced accuracy of 95.27%. According to the experimental results, one can find that the proposed method is still effective and robust in the 3-class classification.

Figure 4 shows the behavior of accuracy and loss during the training and evaluation processes. The loss and accuracy values that obtained while training the proposed framework revealed its efficiency on training and validation data. 70% of the data is used for training while the rest is used for validating process. The findings show that as the loss value decreases, the framework improves its accuracy.

The performance of the 3-class classification is depicted in the form of CMs in Fig. 5 and ROC curves in Fig. 6 with different training-testing ratios (i.e., 70% - 30%, 80% - 20%, and 90% - 10%). In (90% - 10%) classification ratio, 20 images are tested including 4 AD, 6 NC and 10 MCI. It is clear that only one image of MCI was misclassified and all AD and NC images are correctly classified as described in Fig. 5c. The AUC values of 1, 1, and 0.989 were obtained for AD, NC, and MCI, respectively. In the two other classification cases, 59 and 39 images are tested in (70% - 30%) and (80% - 20%), respectively and one image of AD and NC



**Fig. 5** The Confusion matrices of the proposed 3-class classification model with different training-testing ratios (a) 70% - 30% (b) 80% - 20% (c) 90% - 10%



**Fig. 6** ROC curves of the proposed 3-class classification model with different training-testing ratio (a) 70% - 30% (b) 80% - 20% (c) 90% - 10%

was misclassified as shown in Fig. 5b and c. The AUC values of the other classification cases are clarified in Fig. 6b and c. The results declare that the (90% - 10%) classification case obtained the best AD detection with a classification accuracy of 95%.

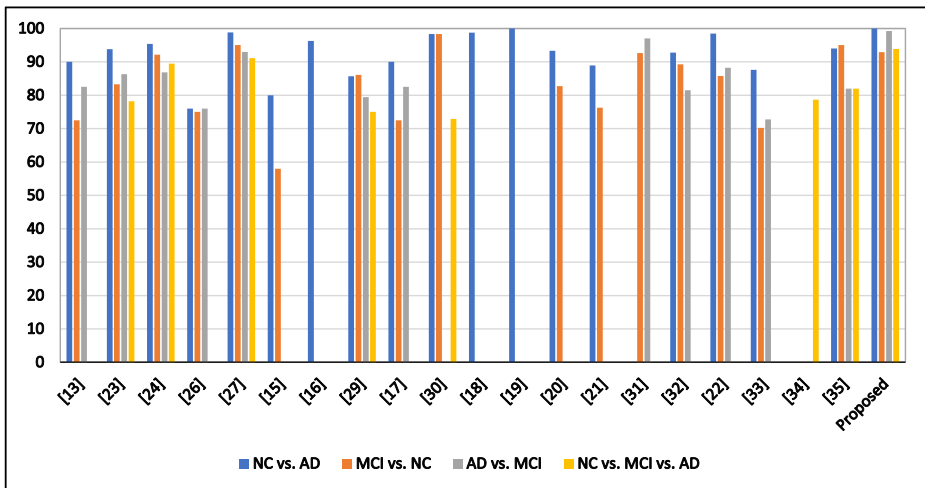
The method used to divide the database image set into subsets for training and testing has a large impact on the results. The larger the images used in training the DCNN, the higher the accuracy obtained. As a result, one of the most important reasons for improving the results is to use a database with the largest images possible in the train phase, which increases the DCNN classifier’s ability to discriminate between the different classes. The average accuracy for NC versus MCI is relatively low, as can be seen from the results. As a result, the three classifications (AD, NC, and MCI) have a relatively low accuracy. This is due to the use of a small number of dataset images and the MCI being a stage between NC and AD. MCI and NC have a high similarity ratio.

The comparison of the average accuracy of the proposed method and the recently published studies using the same ADNI datasets is shown in Table 7 and summarized in Fig. 7. From the table and the figure, it is clear that the proposed classifier achieved the highest performance compared to the literature. However, the CA of the proposed method for classification of (MCI

**Table 7** The comparison of the proposed classifier average accuracy with literature

| Study    | Year | Method                          | Average Accuracy (%) |            |            |                   |
|----------|------|---------------------------------|----------------------|------------|------------|-------------------|
|          |      |                                 | NC vs. AD            | MCI vs. NC | AD vs. MCI | NC vs. MCI vs. AD |
| [13]     | 2017 | Inception V4+VGG16              | 96.25                | –          | –          | –                 |
| [24]     | 2019 | DL                              | 98.74                | –          | –          | –                 |
| [14]     | 2020 | VGG 16+SVM                      | 99.95                | –          | –          | –                 |
| [16]     | 2017 | 3D DCNN                         | 80                   | 58         | –          | –                 |
| [34]     | 2020 | SLR                             | 93.33                | 82.75      | –          | –                 |
| [18]     | 2020 | 3D DenseNet                     | 88.90%               | 76.25      | –          | –                 |
| [8]      | 2020 | PCANet                          | –                    | 92.6       | 97.01      | –                 |
| [27]     | 2018 | NM                              | 76                   | 75         | 76         | –                 |
| [1]      | 2018 | DTI+MD                          | 90.00                | 72.50      | 82.50      | –                 |
| [6]      | 2019 | deep neural networks            | 99.2                 | 76.1       | 85.9       | –                 |
| [17]     | 2019 | Regional abnormality via DL     | 92.75                | 89.22      | 81.46      | –                 |
| [36]     | 2019 | Two independent CNNs            | 98.47                | 85.74      | 88.20      | –                 |
| [25]     | 2020 | Universum SVM                   | 87.6                 | 70.2       | 72.73      | –                 |
| [22]     | 2020 | DL using transfer learning      | –                    | –          | –          | 78.64             |
| [29]     | 2017 | Data fusion method              | 98.3                 | 98.3       | –          | 72.9              |
| [12]     | 2013 | Cross-domain features           | 93.80                | 83.30      | 86.30      | 78.20             |
| [21]     | 2015 | 3D CNNs and sparse autoencoders | 95.39                | 92.11      | 86.84      | 89.47             |
| [32]     | 2018 | HiLCAE and 3D-VGG16             | 98.8                 | 95         | 93         | 91.13             |
| [35]     | 2017 | multi-feature fusion            | 85.71                | 86.11      | 79.44      | 75.00             |
| [31]     | 2020 | hippocampal atrophy             | 94                   | 95         | 82         | 82                |
| Proposed |      | DCNN framework                  | 100                  | 92.93      | 99.21      | 93.86             |

vs. NC) is less than of [31, 32], the proposed new DCNN achieved the highest CA of other classification scenarios (i.e., NC vs. AD, MCI vs. AD, and AD vs. NC vs. MCI).



**Fig. 7** Comparison of the proposed approach with literature

## 5 Conclusion

In this study, a new efficient automated AD diagnosis DCNN based framework using brain MRI scans was introduced and examined. DCNN performance was tested in distinguishing NC, MCI, and AD. Experimental data are obtained using the ADNI dataset that includes 199 subjects. Four classification metrics were measured for four different situations of dataset combinations i.e., NC vs. AD, NC vs. MCI, MCI vs. AD, and the most difficult scenario of 3-class classification (NC, MCI, and AD). The experimental results are compared with the published results. The proposed framework results outperform the literature in all four classification situations in terms of Accuracy, Specificity, Sensitivity, and Balanced Accuracy. In classifying NC and AD, the highest classification rate of 100% is obtained for Accuracy, Sensitivity, Specificity, and Balanced accuracy. In distinguishing MCI from NC subjects, the proposed method reached an accuracy of about 92.93%, a specificity of 94.37%, a sensitivity of 90.74%, and balanced accuracy of 92.56%. In the case of AD and MCI classification the highest accuracy and specificity (higher than 99%) are obtained. Moreover, sensitivity and balanced accuracy obtained are about 97.44% and 98.72% respectively. Finally for the three-class classification, the proposed method reached an accuracy of about 93.86%, a sensitivity of 93.95%, a specificity of 96.60%, and balanced accuracy of 95.27%. The small variations between NC, MCI, and AD arises the need for more than 199 subjects to further improve the framework performance. In the last situation, ROC and CM are also used. The AUC values of 1, 1, and 0.989 are achieved for AD, NC, and MCI, respectively.

## 6 Future work

The usage of another larger MRI dataset with pretrained CNN will be the upcoming challenge to achieve better accuracy.

## Declarations

**Competing interests** The authors report no declarations of interest.

## References

1. 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 *IEEE International Symposium on Computer-Based Medical Systems (IEEE)*, pp 345–350
2. Ahmed HM, Youssef BAB, Elkorany AS, Elsharkawy ZF, Saleeb AA, El-Samie FA (2019) Hybridized classification approach for magnetic resonance brain images using gray wolf optimizer and support vector machine. *Multimed Tools Appl* 78:27983–28002
3. Ahmed HM, Youssef BAB, Elkorany AS, Saleeb AA, Abd El-Samie F (2018) Hybrid gray wolf optimizer–artificial neural network classification approach for magnetic resonance brain images. *Appl Opt* 57:B25–B31
4. Alzheimer Association (2019) 2019 Alzheimer’s disease facts and figures. *Alzheimers Dement* 15:321–387
5. Backstrom K, Nazari M, Gu IYH, Jakola AS (2018) An efficient 3D deep convolutional network for Alzheimer’s disease diagnosis using MR images. *IEEE International Symposium on Biomedical Imaging (IEEE)*, pp 149–153
6. Basaia S, Agosta F, Wagner L, Canu E, Magnani G (2019) Automated classification of Alzheimer ‘s disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage Clin* 21

7. Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, Filippi M (2019) Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage: Clinical* 21(2019):101645. <https://doi.org/10.1016/j.nicl.2018.101645>
8. Bi X, Li S, Xiao B, Li Y, Wang G, Ma X (2020) Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology. *Neurocomputing* 392:296–304
9. Buono VL, Palmeri R, Corallo F, Allone C, Pria D, Bramanti P, Marino S (2020) Diffusion tensor imaging of white matter degeneration in early stage of Alzheimer's disease: a review. *Int J Neurosci* 130:243–250
10. Cárdenas-peña D, Collazos-huertas D, Castellanos-domínguez G (2016) Centered kernel alignment enhancing neural network Pretraining for MRI-based dementia diagnosis. *Comput Math Methods Med* 2016:1–10
11. Goryawala M, Zhou Q, Barker W, Loewenstein DA, Duara R, Adjouadi M (2015) Inclusion of neuropsychological scores in atrophy models improves diagnostic classification of Alzheimer's disease and mild cognitive impairment. *Comput Intell Neurosci* 2015:1–14
12. Gupta A, Ayhan MS, Maida AS (2013) Natural image bases to represent neuroimaging data. In *International Conference on International Conference on Machine Learning*, pp 987–994
13. Hon M, Khan NM (2017) Towards Alzheimer's disease classification through transfer learning. In *IEEE International Conference on Bioinformatics and Biomedicine (IEEE)*, pp 1166–1169
14. Janghel RR, Rathore YK (2020) Deep convolution neural network based system for early diagnosis of Alzheimer's disease. *IRBM*. <https://doi.org/10.1016/j.irbm.2020.06.006>
15. Korolev IO (2014) Alzheimer 's disease: a clinical and basic science review. *Med Student Res J* 4:24–33
16. Korolev S, Safiullin A, Belyaev M, Dodonova Y (2017) Residual and plain convolutional neural networks for 3D brain MRI classification. In *IEEE International Symposium on Biomedical Imaging (IEEE)*, pp 835–838
17. Lee E, Choi J-S, Kim M, Suk H-I (2019) Toward an interpretable Alzheimer's disease diagnostic model with regional abnormality representation via deep learning. *NeuroImage* 202:116113
18. Liu M, Li F, Yan H, Wang K, Ma Y, Shen L, Xu M, Initiative AsDN (2020) A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. *NeuroImage* 208:116459
19. Liu L, Zhao S, Chen H, Wang A (2020) A new machine learning method for identifying Alzheimer's disease. *Simul Model Pract Theory* 99:102023
20. Nordberg A, Rinne JO, Kadir A, Lngström B (2010) The use of PET in Alzheimer disease. *Nat Rev Neurol* 6:78–87
21. Payan A, Montana G (2015) Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks, computer, arXiv [Preprint]. Available at: <https://arxiv.org/abs/1502.02506v1>
22. Puente-Castro A, Fernandez-Blanco E, Pazos A, Munteanu CR (2020) Automatic assessment of Alzheimer's disease diagnosis based on deep learning technique. *Comput Biol Med* 120:103764
23. Raza M, Awais M, Ellahi W, Aslam N, Nguyen HX, Le-minh H (2019) Diagnosis and monitoring of Alzheimer 's patients using classical and deep learning techniques. *Expert Syst Appl* 136:353–364
24. Raza M, Awais M, Ellahi W, Aslam N, Nguyen HX, Le-Minh H (2019) Diagnosis and monitoring of Alzheimer's patients using classical and deep learning techniques. *Expert Syst Appl* 136:353–364
25. Richhariya B, Tanveer M, Rashid AH (2020) Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). *Biomed Signal Process Control* 59:101903
26. Sarraf S, Tofighi G Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data. In *2016 Future Technologies Conference (FTC, 2016)*, pp 816–820
27. Senanayake U, Dawes L, Sowmya A (2018) Deep fusion pipeline for mild cognitive impairment diagnosis. In *IEEE International Symposium on Biomedical Imaging (IEEE)*, pp 1394–1397
28. Spasov S, Passamonti L, Duggento A, Lio P, Toschi N (2019) A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer 's disease. *NeuroImage* 189:276–287
29. Tong T, Gray K, Gao Q, Chen L, Rueckert D (2017) Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. *Pattern Recogn* 63:171–181
30. Tong T, Wolz R, Gao Q, Guerrero R, Hajnal JV, Rueckert D (2014) Multiple instance learning for classification of dementia in brain MRI. *Med Image Anal* 18:808–818
31. Uysal G, Ozturk M (2020) Hippocampal atrophy based Alzheimer's disease diagnosis via machine learning methods. *J Neurosci Methods* 337:108669
32. Vu T, Ho N, Yang H, Kim J, Song H (2018) Non-white matter tissue extraction and deep convolutional neural network for Alzheimer 's disease detection. *Soft Comput* 22:6825–6833
33. 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
34. Xiao R, Cui X, Qiao H, Zheng X, Zhang Y (2020) Early diagnosis model of Alzheimer's Disease based on sparse logistic regression. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-020-09738-0>

35. Xiao Z, Ding Y, Lan T, Zhang C, Luo C, Qin Z (2017) Brain MR image classification for Alzheimer 's disease diagnosis based on multifeature fusion. *Comput Math Methods Med* 2017:1–13
36. Zhang F, Li Z, Zhang B, Du H, Wang B, Zhang X (2019) Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease. *Neurocomputing* 361:185–195

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.