Turkish Journal of Electrical Engineering and Computer Sciences

Volume 28 | Number 1 Article 14

1-1-2020

Applying deep learning models to structural MRI for stage prediction of Alzheimer's disease

ALTUĞ YİĞİT

ZERRİN IŞIK

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2020) 28: 196 – 210

doi:10.3906/elk-1904-172

Applying deep learning models to structural MRI for stage prediction of Alzheimer's disease

Altuğ YİĞİT, Zerrin IŞIK*
Department of Computer Engineering, Faculty of Engineering, Dokuz Eylül University, İzmir, Turkey

Received: 24.04.2019 • **Accepted/Published Online:** 23.09.2019 • **Final Version:** 27.01.2020

Abstract: Alzheimer's disease is a brain disease that causes impaired cognitive abilities in memory, concentration, planning, and speaking. Alzheimer's disease is defined as the most common cause of dementia and changes different parts of the brain. Neuroimaging, cerebrospinal fluid, and some protein abnormalities are commonly used as clinical diagnostic biomarkers. In this study, neuroimaging biomarkers were applied for the diagnosis of Alzheimer's disease and dementia as a noninvasive method. Structural magnetic resonance (MR) brain images were used as input of the predictive model. T1 weighted volumetric MR images were reduced to two-dimensional space by several preprocessing methods for three different projections. Convolutional neural network (CNN) models took preprocessed brain images, and the training and testing of the CNN models were carried out with two different data sets. The CNN models achieved accuracy values around 0.8 for diagnosis of both Alzheimer's disease and mild cognitive impairment. The experimental results revealed that the diagnosis of patients with mild cognitive impairment was more difficult than that of patients with Alzheimer's disease. The proposed deep learning-based model might serve as an efficient and practical diagnostic tool when MRI data are integrated with other clinical tests.

Key words: Alzheimer's disease diagnosis, dementia diagnosis, convolutional neural networks, deep learning, structural MRI

1. Introduction

The causes of Alzheimer's disease are not still clearly known by scientists, but its diagnosis goes back to the past century. In 1906, Doctor Alois Alzheimer gave a remarkable lecture in which he described for the first time a form of dementia; later on it was named Alzheimer's disease. He introduced a 51-year-old woman patient from Frankfurt who had shown progressive cognitive impairment, focal symptoms, hallucinations, delusions, and psychosocial incompetence [1]. Depression, insomnia, incontinence, delusions, and hallucinations may manifest disease progression and neurological signs can occur such as sudden muscle contraction and gait disturbance [2]. Alzheimer's disease is the most common form of dementia, which is defined as sustained deterioration of intellectual functions in a patient. Operationally, because of diminished cognitive ability, a demented person conducts everyday activities worse compared to past performance. The formal diagnosis criteria of dementia include the development of multiple cognitive defects that include memory impairment and at least one of the following cognitive disturbances: aphasia, apraxia, agnosia, or a disturbance in executive functioning. Alzheimer's disease, with or without comorbid conditions, is by far the leading cause of dementia, observed in 75% or more of all pathological diagnoses of dementing disorders [3, 4]. The risk of having Alzheimer's disease

*Correspondence: zerrin@cs.deu.edu.tr

increases with age. Although most patients are older than 65 years of age, some of them might be younger than 65. In 2018, there were approximately 50 million Alzheimer's patients in the world and this number is expected to increase to 152 million in 2050 [1]. Accordingly, the estimated worldwide treatment costs were set at around $186 billion in 2018. This figure is expected to be about four times higher in 30 years [2]. Unfortunately, scientists have not yet developed a cure that can precisely treat and prevent Alzheimer's disease. The disease is divided into four stages: very mild dementia, mild dementia, moderate dementia, and severe dementia according to the accepted Clinical Dementia Rating (CDR) value. The treatment cost of very mild dementia patients and severe dementia patients is not the same; therefore, an early diagnosis of dementia diseases is important in terms of patient recovery and treatment cost. The initial diagnosis criteria of Alzheimer's disease were established in 1984 and focused on clinical symptoms only, since at this time Alzheimer's pathological changes in patients could not be measured in vivo, and so the disease could be definitively diagnosed only after death of a patient. With the advent of MRI and the discovery of cerebrospinal fluid (CSF) biomarkers and amyloid PET, the International Working Group (IWG) proposed new criteria in 2014, which formed the inspiration for a subsequent set of criteria by the National Institute of Aging and Alzheimer's Association (NIA–AA). Mild cognitive impairment is recognized as prodromal Alzheimer's disease in both the IWG and the NIA–AA criteria when it is supported by the presence of amyloid and neurodegeneration. Biomarker evidence can be used to attribute the clinical syndrome of dementia or mild cognitive impairment to underlying Alzheimer's pathological changes with high, intermediate, or low likelihood in the NIA–AA criteria [5, 6].

Images from for example computed tomography (CT), structural and functional magnetic resonance imaging (MRI), and positron-emission tomography (PET) are used as imaging biomarkers for assessment of Alzheimer's disease. MRI has been extensively investigated in Alzheimer's disease and, consistent with pathology, very early changes have been observed in the hippocampus and entorhinal cortex, but the most suitable structure for early diagnosis remains unclear. Both structural and functional imaging have made important contributions to the differential diagnosis of dementia and to understand its neurobiology, and showed great promise with regard to monitoring disease progression [7].

There are several publications about diagnosis of Alzheimer's disease using different imaging techniques. Feature-based morphometry (FBM) was used to determine patterns of patients and healthy individuals in volumetric MRI [8]. Another study proposed an image analysis method that reveals discriminative brain patterns associated with the presence of neurodegenerative diseases based on fusion strategy [9]. After the discriminative patterns were determined, the performance of classification with a support vector machine (SVM) was evaluated on different data sets. In another study, a classification method based on multilevel brain partitions was presented [10]. Histogram-based features obtained from MRI data were applied to classify different brain levels by using an SVM. A recent study provided a method for automatically detecting anatomical changes in the hippocampus, amygdala, planum temporale, and thalamus regions [11]. Differences and similarities between two individuals were identified according to gray level histogram values on MRI data. Unlike others, another recent study proposes to construct cascaded convolutional neural networks (CNNs) to learn the multilevel and multimodal features from MRI and PET brain images for Alzheimer's disease classification [12]. Most of the previous studies applied several feature extraction methods on image samples; the selection of these methods

[1]Statista (2018). Estimated number of people with dementia worldwide in 2018, 2030, and 2050 (in millions) [online]. Website https://www.statista.com/statistics/264951/number-of-people-with-dementia-from-2010-to-2050/ [accessed 01 November 2018].

[2]Statista (2018). Costs of care for individuals in the US with Alzheimer's disease to Medicare and Medicaid from 2018 to 2050 (in billion US dollars). [online]. Website https://www.statista.com/statistics/643072/alzheimers-medicare-medicaid-care-costs-us/ [accessed 01 November 2018].

affects the performance of machine learning methods. Few studies utilized multiple data sets for training and evaluation of models; hence the limited number of samples might reduce the prediction performance of models. Compared to many studies on the diagnosis of Alzheimer's disease, fewer studies have attempted an automatic diagnosis of MCI patients.

In the present study, we aim to accurately determine Alzheimer's and dementia patients by analyzing MR image data with the help of deep learning approaches. The first novelty of this study is the learning of features automatically from MR image data by integrating a CNN model as the first phase of the diagnostic method. Another novelty is not applying a specific age limitation while annotating samples as Alzheimer's disease and dementia, which eventually generated a challenging prediction problem due to a wide range of age distribution. The study is divided into three main parts. The first part involves obtaining and processing three-dimensional MR data and converting it into two dimensions, the second involves preprocessing of the data, and the last part involves the application of deep learning models to the MR image data to classify Alzheimer's and dementia patients.

2. Materials and methods

In this section, we present general information about the study, descriptions of data sets, and stages of the study.

2.1. General framework

T1-weighted MR image samples were used to distinguish between healthy individuals, patients with Alzheimer's disease, and those with mild cognitive impairment. An overview of the study is shown in Figure 1. One of the MR data sets was used for the training process, while the other data set was used for testing. After obtaining the image data, MR slices were selected in three-dimensional space and reduced to two dimensions. Since the number of images for different projections was unbalanced, a data augmentation technique was applied. Some preprocessing methods were performed before images were given as input to the model. CNN models with different numbers of layers were created and then applied for classification of Alzheimer's disease types. Performance was compared in terms of the diagnostic power of the proposed model.

2.2. MRI data set

The Open Access Series of Imaging Studies (OASIS) data set was used for the training of the CNN models [13]. It presents MRI data sets of the brain free of charge for scientific studies. Patients older than 60 years were mostly diagnosed with mild to moderate Alzheimer's disease. It includes at least three acquired T1-weighted image records of patients who participated in MRI studies at the University of Washington. Subjects with serious head injury, history of clinically meaningful stroke, and use of psychoactive drugs were excluded, and subjects with gross anatomical abnormalities evident in their MRI images such as tumors were also eliminated. However, subjects with aged typical brain changes were accepted. The subjects consisted of 30 Alzheimer's patients, 70 MCI patients, and 316 healthy individuals. For each subject, three to four individual T1-weighted magnetization prepared rapid gradient echo (MP-RAGE) images were acquired on a 1.5-T Vision scanner (Siemens, Erlangen, Germany) in a single imaging session [14].

The Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) data set, which was not provided in the training phase, was used only for testing of the model [15]. Similar to OASIS, it is a series of longitudinal volumetric T1-weighted MRI scans of 46 mild moderate Alzheimer's subjects and 23 healthy

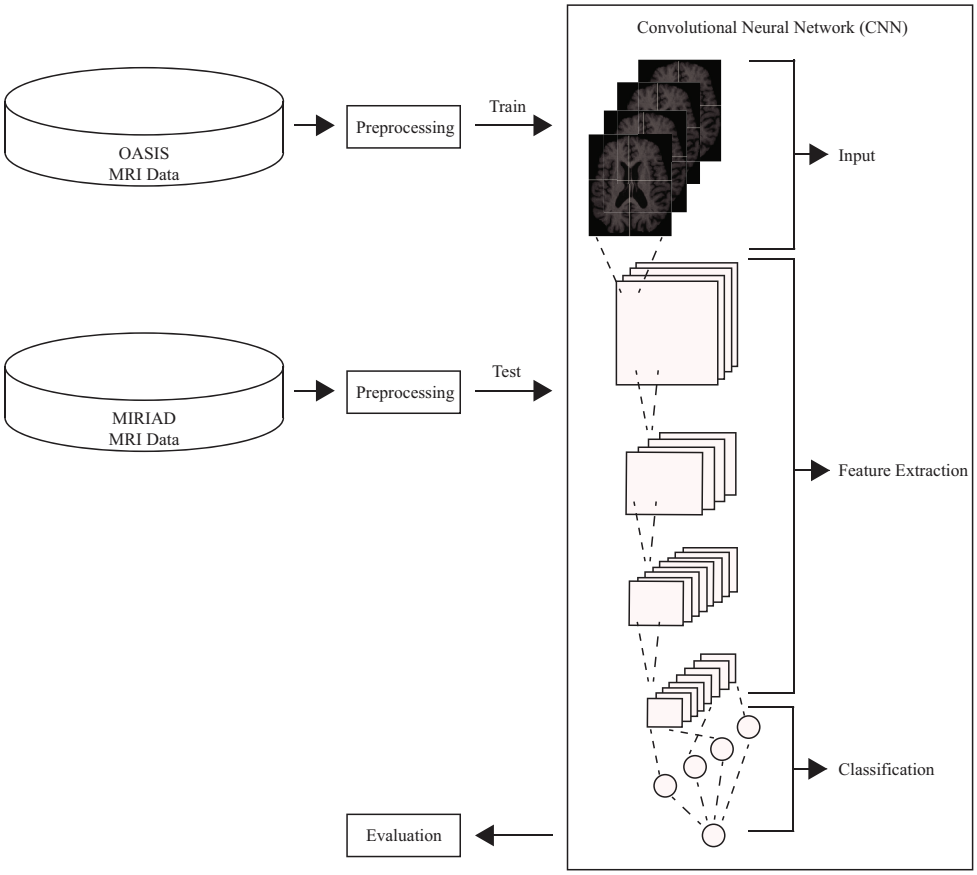

Figure 1. An overview of the study.

controls. It consists of 708 scans conducted on the same 1.5-T Signa MRI scanner (GE Medical systems, Milwaukee, WI, USA) and acquired by the same radiographer. Three-dimensional T1-weighted images were acquired with an IR-FSPGR (inversion recovery prepared fast spoiled gradient recalled) sequence, field of view 24 cm, 256 $\times$ 256 matrix, 124 1.5 mm coronal partitions, TR 15 ms, TE 5.4 ms, flip angle 15°, TI 650 ms. Sequences were collected at intervals of 2, 6, 14, 26, 38, and 52 weeks and 18 and 24 months from baseline, with accompanying information on sex, age and Mini Mental State Examination (MMSE) scores.

2.3. Preprocessing of brain MRI data

Volumetric brain MRI data were read and image processing procedures were applied. Three-dimensional data formats of Nifti (neuroimaging informatics technology initiative) and hdr (high dynamic range rendering) were used. Alzheimer's disease and dementia are neurological diseases that affect nerve cells and so it is wise to eliminate unnecessary data. Because the skull, eyes, fat, and muscle are not affected by these diseases, these regions are unnecessary data in raw images. The Brain Extraction Tool (BET) of the FMRIB's Software Library (FSL) was used for skull stripping [16, 17]. FSL is a comprehensive library of analysis tools for functional magnetic resonance imaging (FMRI), MRI, and diffusion tensor imaging (DTI) data. BET is an automated method for segmenting MRI images as brain and nonbrain regions. The histogram density is calculated from the input image. Once the threshold value is obtained, the image is expressed in binary to find the center of gravity

and the approximate head size. A triangular tessellation of a sphere's surface is initialized inside the brain and allowed a slow deformation one vertex at a time, following forces that keep the surface well spaced and smooth, while attempting to move toward the brain's edge. It is reprocessed with a higher smoothness constraint until a properly clean solution is achieved. As a result of this process, nonbrain elements are eliminated from the input image [18].

We selected ten slices from the axial, sagittal, and coronal projections for each patient's volumetric brain data. While selecting these slices, we paid attention to choose regions such as the hippocampus, thalamus, hypothalamus, amygdala, cerebellum, frontal lobe, parietal lobe, occipital lobe, and corpus callosum, which are affected in dementia and Alzheimer's disease. A data augmentation method was applied because of the unbalanced number of patient samples, which might lead to overfitting of the model. For this purpose, 10% right and left, up and down shifting, 20% zoom, and 20% shear operations were applied; zoomed and changed position images were produced. After an equal number of patient samples were obtained, the images were resized to $150 \times 150 \times 1$. In this process, bicubic interpolation over a 4×4 pixel neighborhood was used. Contrast limited adaptive histogram equalization (CLAHE) is a contrast enhancement method developed by Zuiderveld [19]. CLAHE was originally developed for medical imaging. In the present study, it was applied to prevent clustering in the histogram and to correct its distribution. Histogram clipping is performed while correcting the histogram distribution. Cropped pixels are redistributed equally to the entire histogram to keep the total number of histograms equal. This method solves the problems of global linear minimum maximum windowing and global histogram equalization without manual intervention.

It is better if there are no unnecessary details in the image detected by the computer. That is why a two-dimensional Gaussian smoothing filter was applied to reduce sharp pixel transitions between pixels and to soften the brain image. At the same time, the noise in the image was removed by applying this filter to the image.

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}. \tag{1}$$

Equation (1) shows the Gaussian function, where σ represents the standard deviation of the Gaussian distribution, and x is the distance from its position to the origin of the image. It is expected that the applied Gaussian function will not be zero at every point of the image. Figure 2 shows the stages of the processing to obtain the input brain images for the models. Images are generated from different projections to protect the volumetric information. After the images have been processed and produced, they are given as input of different CNN models.

2.4. Implementation of models

After preprocessing of the image data, we created CNNs and performed evaluations on these models. CNNs are deep artificial neural networks and commonly used in image related applications such as image classification, clustering, and interpretation. They are inspired by the layered vision mechanism of humans. The improvement of the hardware and the increased processing capacity of graphics processing units (GPUs) has allowed the training of deep networks on computers more efficiently.

Big data, which can be collected from many platforms, are the basis for the implementation of CNNs and other deep learning models. Because CNNs are designed as deep models, they can provide sufficient results in solving complex problems. One of the pioneering studies, the LeNet model, was developed by LeCun et al.

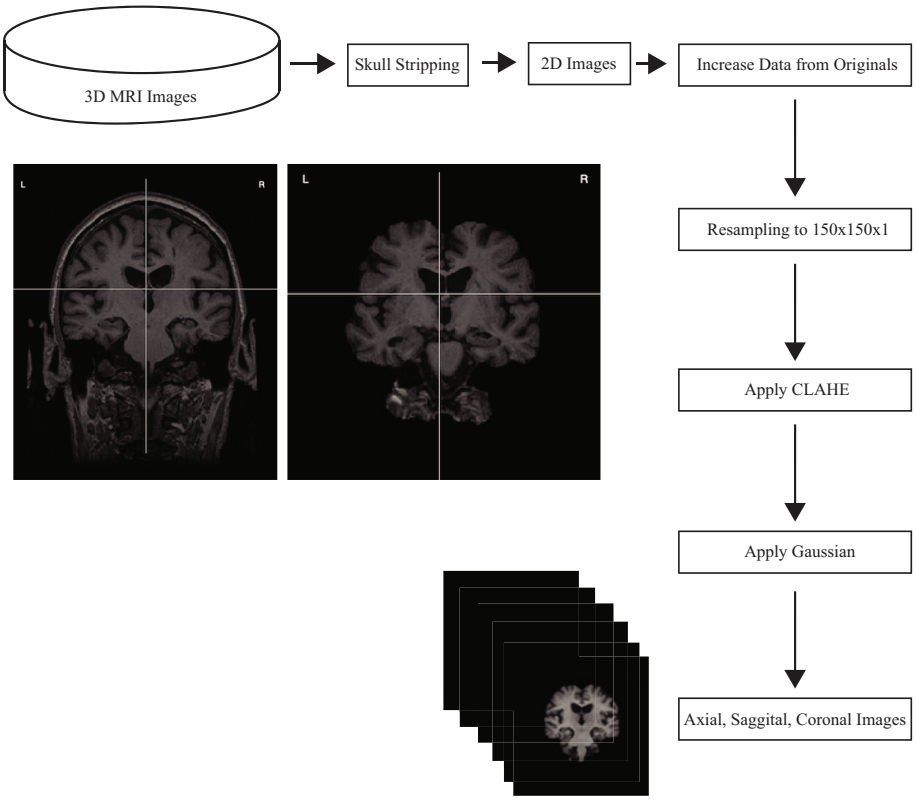

Figure 2. The stages of obtaining input images.

[20] in the 1990s and has achieved significant success in handwriting digit recognition. The model, which is trained with the back-propagated algorithm, could make predictions on the pixels of the image without feature extraction. In the following years, many successful deep models such as AlexNet, ZFNet, GoogleNet, VGGNet, and ResNet have been created [21–25].

The structure of CNNs consists of a convolutional layer, a pooling layer, and fully connected layers. Generally, many convolutional layers and pooling layers are stacked one after the other to produce a feature map, and the generated map is fed into the fully connected layer. In the convolutional layer, a filter with a determined size of n × n is applied to the image expressed as pixel matrices. The filter is applied by moving over the pixel matrices and navigating the entire image matrix. Depending on the type of filter applied, some features of the image are revealed. In the pooling layer, the size of the spatial dimension is reduced by performing the subsampling process. In this way, pooling layers provide ease of computation as well as providing a solution for overfitting. This layer can optionally be selected as maximum or average pooling. In the maximum pooling, the maximum pixel in the window is selected, while the average of the pixels in the window is obtained in the average pooling. After these operations, classification according to the feature map containing the extracted features is made by the fully connected layer, which might be multilayer perceptrons.

In the present study, three CNN models of similar structure were constructed. The type of pooling layer was determined as maximum pooling. A rectified linear unit (ReLU), which is widely used in modern deep learning models, was used as the activation function. Let x and $f(x)$ be the number of inputs and an activation function, respectively. The ReLU is defined as $f(x) = max(0, x)$. Moreover, a binary cross-entropy

function was employed as a loss function. The binary cross-entropy function is defined as in Equation (2).

$$L = -\frac{1}{N}\sum_{i=1}^{N} y_i.\log(p(y_i)) + (1-y_i).\log(1-p(y_i)), \tag{2}$$

where N is the number of entries, i indicates iteration, y_i shows i's element label, and $p(y_i)$ is the predicted label. This function also called log loss due to logarithmic operations. Different models were obtained by changing the total number and positions of the layers. The designed CNNs and their layers, filters, and dropout rates are presented in Figure 3. Model 1 consists of three convolutional layers, two pooling layers, and two fully connected layers. It has 16,938,561 parameters in total. Model 2 is a smaller model, with two convolutional layers, two pooling layers, and three fully connected layers. It has 1,311,697 parameters. Model 3 is a larger model, with five convolutional layers, three pooling layers, and four fully connected layers. It is based on Alexnet with 36,446,129 parameters in total. Since the input of the original network receives three channel color images, the network was modified to take grayscale images. The dropout layer randomly blocks neurons by the determined percentage during training; thus it helps to avoid overfitting. The initial learning rate was determined as 0.0002 and the preferred schedule decay value as 0.004. Nesterov-accelerated adaptive moment estimation (Nadam) was used as an optimizer. Nadam is a different version of the Adam optimizer, which is widely used in deep learning models. Nadam contains an extra momentum term so as to not stuck in local minimum points.

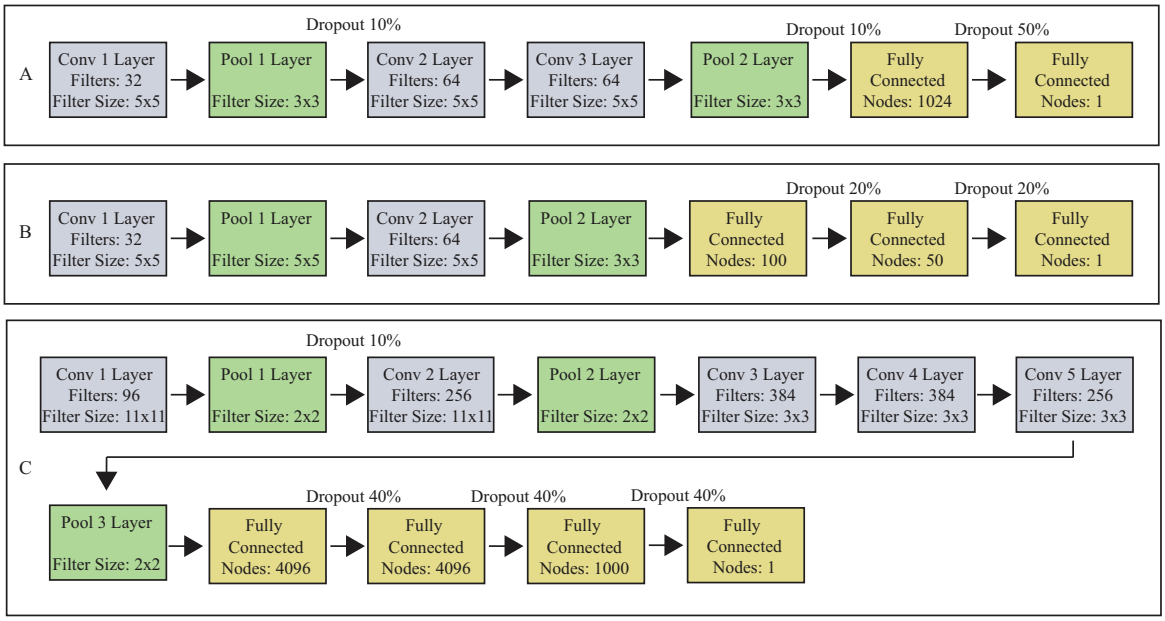

Figure 3. Architecture of Model 1 (A), Model 2 (B), and Model 3 (C).

3. Results and discussion

In this section we present evaluations of the proposed models. In the experimental evaluation, two types of dementia (mild cognitive impairment and Alzheimer's disease) were diagnosed. The entire workflow was implemented using the Python programming language; training processes and evaluations were conducted on a DELL PowerEdge R740 Linux server, which has a 2.2 GHz Intel Xeon central processing unit (CPU) with

32 cores and 128 GB RAM capacity. The models were trained with different projections and several evaluation metrics were calculated.

3.1. Performance in Alzheimer's disease prediction

Table 1 reports the performance of various projections on different models for discrimination of Alzheimer's patients and healthy individuals. Axial, sagittal, and coronal represent the brain projection taken from the image. The type "All" indicates training of a model with all brain projections together. On the other hand, "All - Axial", "All - Sagittal", and "All - Coronal" indicate that a model is trained with all the projections and evaluated with only axial, sagittal, or coronal projections, respectively. These results show that there are minor differences between the performances of Model 1, Model 2, and Model 3.

Table 1. The model performances for distinguishing healthy individuals from Alzheimer's patients*.

Data type	Model	Epoch	Time (h)	AUC	ACC	SENS	SP
Axial	Model 1	100	12:51	0.81	0.78	0.95	0.65
Axial	Model 2	100	02:43	0.78	0.75	0.91	0.62
Axial	Model 3	100	12:40	0.81	**0.81**	0.86	0.75
Sagittal	Model 1	100	12:51	0.63	0.70	0.69	0.71
Sagittal	Model 2	100	02:43	0.62	0.67	0.70	0.60
Sagittal	Model 3	100	12:40	0.63	0.70	0.70	0.69
Coronal	Model 1	100	12:51	0.79	0.76	0.92	0.63
Coronal	Model 2	100	02:43	0.75	0.75	0.82	0.66
Coronal	Model 3	100	12:40	0.74	0.79	0.77	0.82
All	Model 1	105	40:20	0.73	0.74	0.64	0.81
All - Axial	Model 1	105	–	0.85	**0.83**	0.72	0.94
All - Sagittal	Model 1	105	–	0.60	0.68	0.69	0.68
All - Coronal	Model 1	105	–	0.76	0.71	0.57	0.94
All	Model 2	100	07:30	0.71	0.73	0.78	0.64
All - Axial	Model 2	100	–	0.78	0.78	0.86	0.68
All - Sagittal	Model 2	100	–	0.59	0.64	0.68	0.54
All - Coronal	Model 2	100	–	0.77	0.77	0.84	0.67
All	Model 3	100	37:30	0.67	0.73	0.73	0.73
All - Axial	Model 3	100	–	0.81	**0.82**	0.86	0.75
All - Sagittal	Model 3	100	–	0.50	0.61	0.62	0.37
All - Coronal	Model 3	100	–	0.71	0.75	0.75	0.74

* ACC: Accuracy, SENS: Sensitivity, SP: Specificity

Figure 4 shows some brain images of Alzheimer's patients. The group indicated by a presents the correctly classified Alzheimer's patients and group b covers the images of misclassified patients. The best performance in the classification of healthy individuals and those with Alzheimer's disease was obtained when the model was trained with all brain projections. This model gave the best result (0.83) during the evaluation of axial images. On the other hand, the poorest metric value (0.61) was obtained while evaluating the sagittal images on the model trained according to all projections. Since Alzheimer's disease is associated with sex, the sex ratios of misclassified patients were also examined. Out of the 172 male patients, 66 of them were incorrectly estimated,

while 22 of the 223 females were incorrectly predicted. In general, the model led to more mistakes in male patients than in female patients.

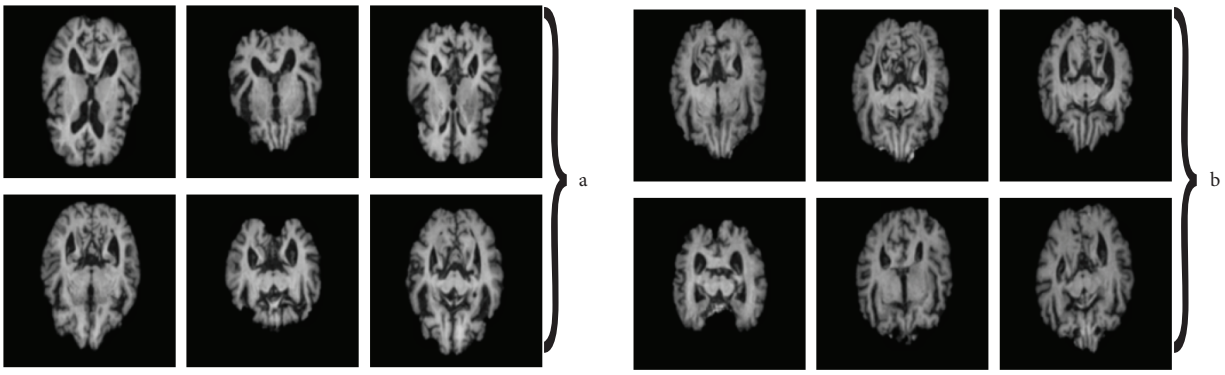

Figure 4. Some examples for correctly classified (a) and misclassified (b) images of Alzheimer's patients.

3.2. Performance in mild cognitive impairment prediction

Table 2 presents classification performance results for determining the difference between individuals with mild cognitive impairment and healthy individuals. The accuracy values are reported according to different models and projections. Similar to the diagnosis of Alzheimer's disease, there are minor performance differences between the models in estimating MCI patients except for the sagittal projection. The results indicate that the capability of discrimination of healthy individuals with mild cognitive impairment is lower than that of healthy individuals with the Alzheimer's disease classification. The images of MCI patients who were misclassified by the best predictive model are shown in Figure 5. Group a represents the images of patients diagnosed correctly, while group b represents images of misdiagnosed patients. Since MCI is the earlier stage of dementia it does not show dramatic brain region differentiation. When the results were examined in terms of projections, although the best success was obtained with axial and coronal images, more unsuccessful results were obtained with sagittal images. When the results in Figure 6 (a) are examined in terms of brain projections, the models trained with axial images were more successful than those trained with sagittal and coronal projections to identify Alzheimer's patients. Compared to other models, Model 3 is more successful on the sagittal projection to identify MCI patients as shown in Figure 6 (b). The model trained with axial images achieved 0.82 accuracy and higher than 0.9 sensitivity value. Sex ratios also were examined in MCI patients. Unlike Alzheimer's patients, more misclassifications were observed in the prediction of female patients with MCI. While there was only 1 error in 16 male patients, 21 false predictions were made in 54 female patients.

In addition to obtaining sex ratios, the age distribution of all patients was analyzed, since dementia diseases were mostly associated with aging. The age distribution histograms for Alzheimer's and MCI patients are shown in Figure 7. Patient ages are between 57 and 87 years. In addition, the age distribution of misdiagnosed patients is given in the plot by orange and red bars. The age distribution of the patients who are incorrectly predicted from Alzheimer's patients is more widespread between 57 and 87; those who are misdiagnosed with MCI are generally older than 66. According to these results, it is more difficult to estimate MCI patients among elderly individuals.

Table 2. The model performances for distinguishing healthy individuals from those with mild cognitive impairment*.

Data type	Model	Epoch	Time (h)	AUC	ACC	SENS	SP
Axial	Model 1	100	12:51	0.60	**0.82**	0.93	0.81
Axial	Model 2	100	02:43	0.53	0.78	0.55	0.79
Axial	Model 3	100	12:40	0.54	0.78	0.57	0.79
Sagittal	Model 1	100	12:51	0.53	0.72	0.30	0.79
Sagittal	Model 2	100	02:43	0.49	0.69	0.20	0.77
Sagittal	Model 3	100	12:40	0.50	0.77	0.20	0.78
Coronal	Model 1	100	12:51	0.56	**0.80**	0.71	0.80
Coronal	Model 2	100	02:43	0.58	0.79	0.56	0.81
Coronal	Model 3	100	12:40	0.54	0.79	0.67	0.79
All	Model 1	105	40:20	0.52	0.74	0.30	0.78
All - Axial	Model 1	105	–	0.51	0.78	1.00	0.78
All - Sagittal	Model 1	105	–	0.52	0.65	0.25	0.79
All - Coronal	Model 1	105	–	0.53	0.78	0.63	0.79
All	Model 2	100	07:30	0.60	0.74	0.41	0.82
All - Axial	Model 2	100	–	0.59	0.76	0.48	0.81
All - Sagittal	Model 2	100	–	0.61	0.67	0.34	0.84
All - Coronal	Model 2	100	–	0.59	0.79	0.57	0.81
All	Model 3	100	37:30	0.51	0.78	0.50	0.78
All - Axial	Model 3	100	–	0.50	0.78	0.00	0.78
All - Sagittal	Model 3	100	–	0.50	0.77	0.25	0.78
All - Coronal	Model 3	100	–	0.51	0.78	1.00	0.78

* ACC: Accuracy, SENS: Sensitivity, SP: Specificity

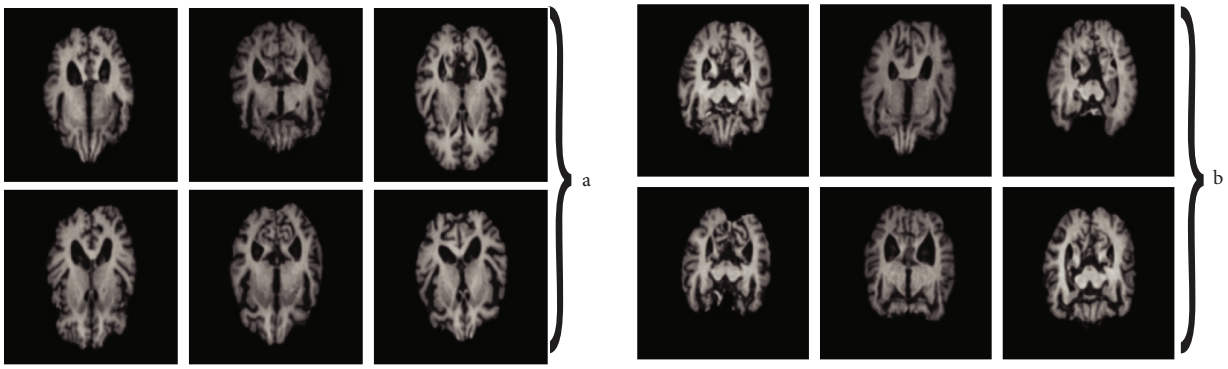

Figure 5. Some examples for correctly classified (a) and misclassified (b) images of MCI patients.

3.3. Comparison with other studies

When the results of this study are compared with the literature, the observations are in the same direction. In previous studies, generally a higher performance in the diagnosis of Alzheimer's patients has been reported; however, diagnosis is more difficult in MCI patients. SVM is a widely preferred machine learning technique for

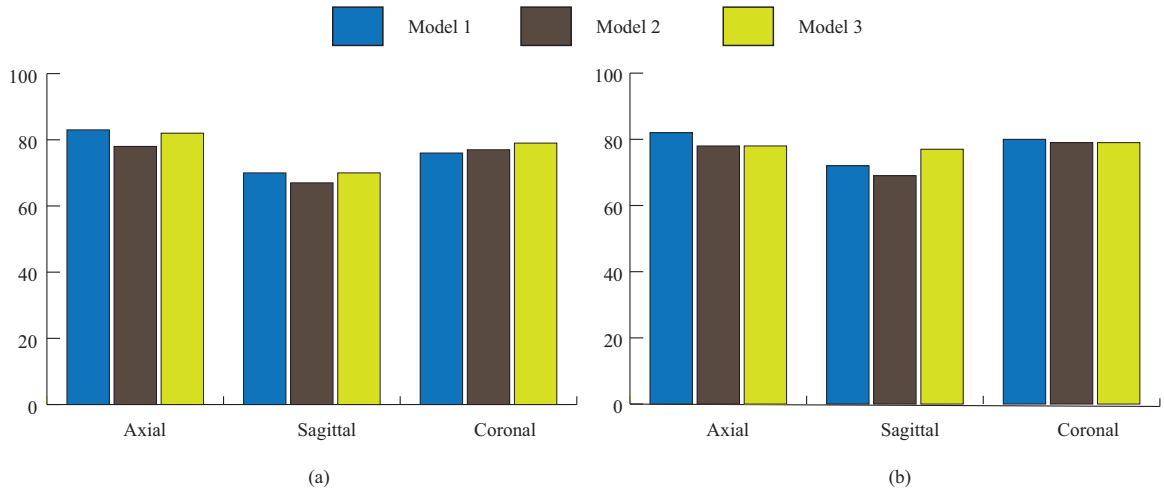

Figure 6. Alzheimer's disease (a) and MCI (b) diagnosis accuracy performances according to different projections (axial, sagittal, coronal).

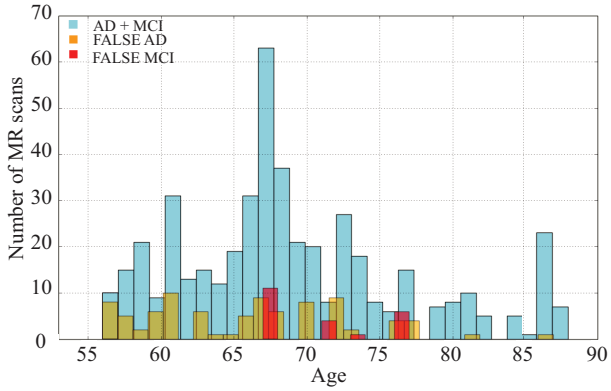

Figure 7. Age distribution of Alzheimer's and MCI patients and misclassified individuals.

diagnosing patients due to its high achievement in classification problems in medical applications. However, serious feature engineering should be carried out to develop a successful SVM model. Poor diagnosis results might be obtained using wrong image features. Since many structural changes occur in the brain in dementia and other brain diseases, the data should be examined for atrophy of many regions, not for a single region. Popular deep learning methods are also used in medical applications and their usage will become more widespread. The novelty of the current study is the design of the image features, since two-dimensional slice selections were made from three-dimensional images and automatic extraction was obtained by learning of features directly from image data. Deep learning methods have been applied to different types of image data (e.g., MRI, CT, PET scan) and multimodal models have been developed by combining these methods. In the present study, the achievement of CNN models on brain MRI data was measured.

Table 3 provides a comparison of previous studies developed for Alzheimer's disease and MCI diagnosis. Rueda et al. [9] generated a saliency map from brain MRI data and applied fusion optimization for identifying anatomical patterns. Eighteen intensity- and contrast-based features were evaluated on Alzheimer's disease; the classification performance was assessed using an SVM model on the OASIS and MIRIAD data sets. Although the same data sets were also used in the current study, the setup in the study by Rueda et al. [9] is different due to separate evaluation of data sets. They obtained 86% accuracy on the OASIS data set for the identification

of Alzheimer's patients aged 60–80 years; they reported 80% accuracy in the test set with ages of 60–96 and 70% accuracy was found in the determination of disease stages. The proposed CNN model obtained superiority with 83% accuracy in elderly individuals up to 87 years of age. On the other hand, Rueda et al. [9] achieved 84% accuracy on the MIRIAD data set, which covers individuals aged 62–76 years. In the current study, no age limitation was applied in any type of evaluation. Although a similar accuracy value was obtained with the MIRIAD data set in the current study, the age range was higher, which creates a different diagnostic problem. In other metric values, Rueda et al. [9] reported 0.83 and 0.82 specificity values for MIRIAD and OASIS, respectively; the proposed model showed a superior success with a 0.94 specificity value. In another study, Li et al. [10] presented a method that can identify Alzheimer's and Parkinson's patients based on multilevel partition. They extracted Scale-Invariant Feature Transform (SIFT) features from the MRI data and constructed Bag-of-visual-words (BOVW) to express the obtained features. In contrast, in the current study, the selected brain slices were evaluated as a whole; the brain was not divided into partitions. They employed an SVM classifier to identify the Alzheimer's patients from the OASIS data set and various accuracy values have been reported based on brain partition levels. They obtained the highest accuracy of 75% with 7 levels of unity. In the current study, the proposed model demonstrated a higher accuracy with 83% on the whole brain images.

Table 3. Comparisons of previous studies and the proposed model*.

Author	Diagnosis type	Data set	AUC	ACC	SENS	SP
Rueda et al. [9]	Alzheimer's (MRI)	OASIS	–	0.80	0.75	0.82
Rueda et al. [9]	Alzheimer's (MRI)	MIRIAD	–	0.84	0.85	0.83
Li et al. [10]	Alzheimer's (MRI)	OASIS	–	0.75	0.74	0.78
Giraldo et al. [11]	Alzheimer's (MRI)	OASIS+MIRIAD	0.92	–	0.90	0.90
Zhang et al. [26]	Alzheimer's (MRI)	ADNI	–	0.85	–	–
Ahmed et al. [27]	Alzheimer's (MRI)	ADNI	–	0.84	0.79	0.88
Ahmed et al. [27]	Alzheimer's (MRI)	BORDEAUX-3CITY	–	0.78	0.80	0.75
Demirhan [28]	Alzheimer's (MRI)	OASIS	–	0.80	0.80	0.80
Proposed Method	Alzheimer's (MRI)	OASIS+MIRIAD	0.85	0.83	0.72	0.94
Giraldo et al. [11]	MCI (MRI)	OASIS+MIRIAD	0.74	–	0.85	0.91
Zhang et al. [26]	MCI (MRI)	ADNI	–	0.74	–	–
Ahmed et al. [27]	MCI (MRI)	ADNI	–	0.69	0.63	0.75
Demirhan [28]	MCI (MRI)	ADNI	–	0.82	0.84	0.80
Proposed Method	MCI (MRI)	OASIS+MIRIAD	0.60	0.82	0.93	0.81

* ACC: Accuracy, SENS: Sensitivity, SP: Specificity, ADNI: Alzheimer's Disease Neuroimaging Initiative database (adni.loni.usc.edu)

A limited number of studies in the literature use separate data sets for training and evaluation purposes. In the current study, the OASIS data set was used only for the training stage and the MIRIAD data set used only in the evaluation stage. This setup was not applied in recent studies except for in the study by Giraldo et al. [11], who used a special method to extract brain regions, which were represented by histograms of gray levels, and the differences between subjects were determined by using Earth mover's distance. They employed a Random Unit Sampling Boost method to perform diagnosis of patients. In the current study, the proposed CNN method achieved a close accuracy value (0.83) and a higher specificity value (0.94) in the Alzheimer's disease diagnosis. They did not provide an accuracy value for model performance. Although the proposed CNN

model has a lower AUC value (0.85) compared to the study by Giraldo et al. [11] (0.92), the proposed model does not explicitly extract MRI features (i.e. brain parcellation process on MRI); instead it uses automatically computed features provided by convolutional and pooling layers of the CNN model. The proposed CNN model has the best performance in terms of accuracy values for diagnosis of MCI. The other study [11] reported a 0.74 AUC value, which is higher than the 0.6 AUC value of the proposed CNN model. However, the feature engineering applied in the other study [11] should be evaluated carefully while comparing the performance of the two studies. A higher AUC performance in the other study [11] could have originated due to focusing on specific brain regions (hippocampus, amygdala, planum temporale, etc.), which eventually might lead to more accurate diagnosis, especially in MCI patients.

4. Conclusion

Dementia causes difficult times for both patients and caregivers. Unfortunately, there is no known proper treatment for the prevention of Alzheimer's disease. Therefore, diagnosis of the disease at an early stage by utilizing the possibilities of technology can prevent the progression of the disease. Besides the benefit of early diagnosis to the patients, it also contributes to the national economy. This economic contribution is related to the lower cost of treatments applied in early stages of dementia compared to in later stages.

The use of imaging techniques such as MRI, CT, and PET provides crucial data about neurological diseases. The image should be interpreted correctly to determine a stage or to diagnose any disease by using an advanced image analysis technique. From this point of view, it is necessary to have a human expert to interpret these images in an appropriate way. Deformations might occur in some parts of brain with aging. When examining brain images, it is more difficult to determine dementia in elderly individuals. In such cases, the use of an expert system can definitely help physicians. An expert system can process images by taking raw inputs and a decision about the health condition of the patient is returned according to the previously developed expert model. The present study investigates the diagnostic biomarkers in MR images of patients by using deep learning and image analysis methods. The discrimination of mild cognitive impairment individuals from Alzheimer's disease patients is not a trivial task in the clinic. Therefore, in addition to the diagnosis of Alzheimer's disease patients, our study also aims to identify individuals with mild cognitive impairment. The diagnosis of patients in the cognitive impairment stage of dementia will help for slowing disease progression by decreasing the earlier symptoms with appropriate medication.

MR images were evaluated by an expert machine learning system for assisting in diagnosis within the scope of the study. MR images in three-dimensional space were reduced to two dimensions to limit the total running time of the system. It was ensured that the objects in images other than the brain region were eliminated using a well-known FSL library. In order to obtain a large amount of patient samples, images were reproduced by applying some data augmentation methods. In addition, focusing on multiple regions was achieved by producing images from different projections of the brain. For this purpose, axial, coronal, and sagittal images were produced separately for each patient. As with most medical image processing applications, the brain region in the image was highlighted with filters to enhance contrast and to remove noise. Deep learning models are used in the diagnosis of many brain diseases with high accuracy when the data size is appropriate. Therefore, CNN models were applied in the interpretation of brain images. The OASIS and the MIRIAD data sets were used together to enlarge patient MRI samples. The experiments revealed accuracy values of greater than 0.8 with deep learning models for the diagnosis of Alzheimer's disease using T1 weighted MRI data. The diagnosis of mild cognitive impairment was found to be more difficult than that of Alzheimer's disease. The comparison

with previous studies revealed that the proposed CNN-based model has great potential for the diagnosis of both types of dementia. The proposed model does not require extra feature extraction or feature selection procedures that should be applied to the MRI data. Once an optimum CNN model is constructed and its parameters set, this model can be applied to new patient data without human intervention. These advantages will increase the practical usage of the proposed model as an automatic clinical diagnostic tool in the near future.

In future work, for more accurate prediction of dementia and Alzheimer's disease, it is planned to develop a volumetric deep learning model that accepts different types of brain images (PET, MRI, etc.) as 3D input. It is also desirable to use a large data set with more patient samples for better generalization performance. Besides the image techniques, other factors such as the patient's history, age, socioeconomic status, and MMSE test score are used in the clinic for the diagnosis of dementia. An ideal diagnostic tool should consider all of these factors during the training of an intelligent system to improve its diagnosis ability. The diagnostic consistency and interpretation of such a system will be more satisfactory, since it consults different aspects of the disease.

Acknowledgments

A. Yiğit is supported by the 100/2000 CoHE Doctoral Scholarship Project. Assistance provided by M.D.Ö. Koska is greatly appreciated. Data used in the preparation of this article were obtained from the Open Access Series of Imaging Studies (OASIS) project (http://oasis-brains.org/) and the Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) database (http://miriad.drc.ion.ucl.ac.uk). The MIRIAD investigators did not participate in the analysis or writing of this report. The MIRIAD data set is made available through the support of the UK Alzheimer's Society (Grant RF116). The original data collection was funded through an unrestricted educational grant from GlaxoSmithKline (Grant 6GKC).

References

[1] Maurer K, Volk S, Gerbaldo H. Auguste D and Alzheimer's disease. Lancet 1997; 349 (9064): 1546-1549. doi: 10.1016/S0140-6736(96)10203-8

[2] Richards SS, Hendrie HC. Diagnosis, management, and treatment of Alzheimer's disease: a guide for the internist. Archives of Internal Medicine American Medical Association 1999; 159 (8): 789–798. doi: 10.1001/archinte.159.8.789

[3] Morris JC. Classification of dementia and Alzheimer disease. Acta Neurologica Scandinavica 1996; 94 (S165): 41–50. doi: 10.1111/j.1600-0404.1996.tb05871.x

[4] Jellinger K, Danielczyk W, Fischer P, Gabriel E. Clinicopathological analysis of dementia disorders in the elderly. Journal of the Neurological Sciences 1990; 95 (2): 239-258. doi: 10.1016/0022-510X(90)90072-U

[5] Scheltens P, Blennow K, Breteler MM, de Strooper B, Frisoni GB et al. Alzheimer's disease. Lancet 2016; 388 (10043): 505–517. doi: 10.1016/S0140-6736(15)01124-1

[6] O'Brien JT. Role of imaging techniques in the diagnosis of dementia. The British Journal of Radiology 2007; 80 (2): 71–77. doi: 10.1259/bjr/33117326

[7] Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. The Lancet Neurology 2014; 13 (6): 614–629. doi: 10.1016/S1474-4422(14)70090-0

[8] Toews M, Wells W, Collins DL, Arbel T. Feature-based morphometry: discovering group-related anatomical patterns. NeuroImage 2009; 49 (3): 2318-2327. doi: 10.1016/j.neuroimage.2009.10.032

[9] Rueda A, Gonzalez FA, Romero E. Extracting salient brain patterns for imaging-based classification of neurodegenerative diseases. IEEE Transactions on Medical Imaging 2014; 33 (6): 1262-1274. doi: 10.1109/TMI.2014.2308999

[10] Li T, Zhang W. Classification of brain disease from magnetic resonance images based on multi-level brain partitions. In: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Florida, USA; 2016. pp. 5933-5936.

[11] Giraldo DL, García-Arteaga JD, Cárdenas-Robledo S, Romero E. Characterization of brain anatomical patterns by comparing region intensity distributions: Applications to the description of Alzheimer's disease. Brain and Behavior 2018; 8 (4): e00942. doi: 10.1002/brb3.942

[12] Liu M, Cheng D, Wang K, Wang Y. Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. Neuroinformatics 2018; 16 (3-4): 295-308. doi: 10.1007/s12021-018-9370-4

[13] Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC et al. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience, 2007; 19 (9): 1498-1507. doi: 10.1162/jocn.2007.19.9.1498

[14] Mugler, JP, Brookeman, JR. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). Magnetic Resonance in Medicine 1990; 15 (1): 152–157. doi: 10.1002/mrm.1910150117

[15] Malone IB, Cash D, Ridgway GR, MacManus DG, Ourselin S et al. MIRIAD—Public release of a multiple time point Alzheimer's MR imaging dataset. Neuroimage 2013; 70: 33-36. doi: 10.1016/j.neuroimage.2012.12.044

[16] Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE et al. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 2004; 23 (S1): 208-319. doi: 10.1016/j.neuroimage.2004.07.051

[17] Jenkinson M, Pechaud M, Smith S. BET2: MR-based estimation of brain, skull and scalp surfaces. In: Eleventh Annual Meeting of the Organization for Human Brain Mapping; Toronto, Ontario, Canada; 2005. pp. 167.

[18] Smith SM. Fast robust automated brain extraction. Human Brain Mapping 2002; 17 (3): 143-155. doi: 10.1002/hbm.10062

[19] Zuiderveld K. Contrast limited adaptive histograph equalization. Graphic Gems IV. San Diego, CA, USA: Academic Press Professional, 1994, pp. 474–485.

[20] LeCun Y, Boser BE, Denker JS, Henderson D, Howard RE et al. Handwritten digit recognition with a back-propagation network. In: 2nd International Conference on Neural Information Processing Systems; Denver, CO, USA; 1990. pp. 396-404.

[21] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. In: 25th International Conference on Neural Information Processing Systems; Stateline, NV, USA; 2012. pp. 1097–1105.

[22] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: 13th European Conference Computer Vision; Zurich, Switzerland; 2014. pp. 818–833.

[23] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S et al. Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition; Boston, MA, USA; 2015. pp. 1-9. doi: 10.1109/CVPR.2015.7298594.

[24] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations; San Diego, CA, USA; 2015. pp. 1409-1556.

[25] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition; Las Vegas, NV, USA; 2016. pp. 770-778. doi: 10.1109/CVPR.2016.90

[26] Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage 2011; 59 (2): 895–907. doi: 10.1016/j.neuroimage.2011.09.069

[27] Ahmed OB, Mizotin M, Benois-Pineau J, Allard M, Catheline G et al. Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex. Computerized Medical Imaging and Graphics 2015; 44: 13-25. doi: 10.1016/j.compmedimag.2015.04.007

[28] Demirhan A. Classification of structural MRI for detecting Alzheimer's disease. International Journal of Intelligent Systems and Applications in Engineering 2016; 195-198. doi: 10.18201/ijisae.2016SpecialIssue-146973