ORIGINAL PAPER

Dementia diagnosis by ensemble deep neural networks using FDG-PET scans

Altuğ Yiğit[1,2] · Yalın Baştanlar[2] · Zerrin Işık[3]

Abstract

Dementia is a type of brain disease that affects the mental abilities. Various studies utilize PET features or some two-dimensional brain perspectives to diagnose dementia. In this study, we have proposed an ensemble approach, which employs volumetric and axial perspective features for the diagnosis of Alzheimer's disease and the patients with mild cognitive impairment. We have employed deep learning models and constructed two disparate networks. The first network evaluates volumetric features, and the second network assesses grid-based brain scan features. Decisions of these networks were combined by an adaptive majority voting algorithm to create an ensemble learner. In the evaluations, we compared ensemble networks with single ones as well as feature fusion networks to identify possible improvement; as a result, the ensemble method turned out to be promising for making a diagnostic decision. The proposed ensemble network achieved an average accuracy of 91.83% for the diagnosis of Alzheimer's disease; to the best of our knowledge, it is the highest diagnosis performance in the literature.

1 Introduction

Dementia is a neurological disorder that causes structural and functional changes in the brain. The most common form of dementia is Alzheimer's disease. It is a neurodegenerative disease that affects cognitive functions by disrupting the functioning of neurons [1]. [18F]-fluorodeoxyglucose Positron Emission Tomography (FDG-PET), amyloid, tau, Magnetic Resonance Imaging (MRI), CT scans are employed in the diagnosis of Alzheimer's and other dementia diseases.

The examination of such medical scans allows us to obtain many structural, molecular or functional information about the brain with a non-invasive technique. PET scanning is an imaging technique that provides information about diseases by imaging tissues in vivo. In order to obtain this image, a radioactive tracer is injected into the patient. Cell and molecular information about organs and tissues is obtained by examining the image. Different PET images can be obtained with different radioactive tracers administered to the patient [2]. FDG-PET scans provide information about glucose metabolism disorders. Glucose is an important energy source of the brain, and it is provided to understand the functional disorders caused by neurological diseases by detecting disorders in glucose consumption. Since functional disorders can be detected with PET images, it is important to detect brain-related disorders before the anatomical changes occur.

In the PET scans, a decrease in blood flow and glucose metabolism in the temporal and parietal cortex indicates Alzheimer's disease [3]. It is estimated that neurons begin to degenerate 20 to 30 years before the clinical course of Alzheimer's disease. In this process, the amount of plaque and tangles accumulated in the patient's brain increases, and

✉ Altuğ Yiğit
  altugyigit@iyte.edu.tr

  Yalın Baştanlar
  yalinbastanlar@iyte.edu.tr

  Zerrin Işık
  zerrin@cs.deu.edu.tr

1 The Graduate School of Natural and Applied Sciences, Dokuz Eylul University, 35390 Buca,Izmir, Turkey

2 Department of Computer Engineering, Izmir Institute of Technology, 35430 Urla,Izmir, Turkey

3 Department of Computer Engineering, Dokuz Eylul University, 35390 Buca,Izmir, Turkey

the first symptoms begin to appear, some cognitive disorders may be detected, but this patient is not in a condition to be considered as an Alzheimer's patient. This stage is often referred to as mild cognitive impairment (MCI) [4]. In a study, MCI patients exhibited the same behaviors as healthy individuals rather than Alzheimer's patients, although they achieved lower values in some tests such as full-scale IQ and Controlled Verbal Word Combination Test in general cognitive measures [5]. For these reasons, it is more difficult to diagnose MCI patients than Alzheimer's patients, but the diagnosis of a patient at this stage is very important for the treatment of early symptoms.

In this study, deep neural networks were combined to create an ensemble model for computerized diagnosis of dementia diseases. Volumetric or two-dimensional perspective features obtained from PET scans have been used in the literature. However, there is a lack of information on which type of features is appropriate to develop a more effective diagnostic system. Disparate deep neural networks have been constructed and utilized to make ensemble decisions for obtaining more robust predictions. The first network is a type of 3D convolutional neural network (CNN) which focuses on volumetric features, while the other is a type of 2D CNN which assesses grid-based two-dimensional features. The proposed ensemble network outperformed studies in the literature on diagnosis of Alzheimer's disease with cross-validation. It achieved an accuracy of 91.83 that is slightly better than the single models. The main contribution of this study is designing and experimenting an efficient ensemble method that takes advantage of different features to improve diagnostic accuracy of Alzheimer's disease.

The sections of paper are organized as follows. In Sect. 2, previous work on computerized diagnosis using PET and ensemble models are presented. In Sect. 3, data preparation steps and proposed method are explained in detail. Experimental results are shown and compared with other relevant studies in the literature in Sect. 4. Finally, our conclusion and future work are presented in Sect. 5.

2 Related work

Nowadays, computer-aided diagnosis algorithms have started to be used instead of performing the manual evaluation with an observer thanks to the improvement in the processing power of machines. In the literature, machine learning models that take PET scans as input and perform diagnosis were reported. Cabral *et al.* [6] used voxel densities of FDG-PET images obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset to detect conversion of MCI to Alzheimer's disease. PET images were used in the computerized diagnosis of many neurological diseases.

Mudali *et al.* [7] proposed a model to detect Parkinson's syndrome with FDG-PET. Kerr *et al.* [8] obtained FDG-PET features by extracting 47 regions for the detection of epilepsy patients.

Deep learning algorithms, as opposed to classical machine learning methods, are very successful in detecting many diseases. CNN models that take the images as input produce the results by performing the classification process. In computerized support systems developed with CNN models, medical image processing is performed directly on the data without any feature engineering work. In order to detect Alzheimer's and other dementia diseases, deep learning models are divided into two-dimensional and three-dimensional approaches for processing images. The developed models commonly use MRI and PET images. Two-dimensional approaches are based on the processing of sections obtained from volumetric brain images. Kang *et al.* [9] proposed a CNN model to predict positive and negative β-Amyloid states in patients with Alzheimer's and MCI. Some studies [10] reported that the two-dimensional CNN model that receives brain surface perfusion images performs a more efficient classification than the three-dimensional CNN that receives brain images containing the whole brain. On the contrary, three-dimensional approaches have been widely used as well [16–18]. In these studies, volumetric brain images are input, and processing is performed on voxels. It also reported that using multiple imaging modalities improved the performance [18]. Vu *et al.* [11] have developed the Sparse Automatic Encoder (SAE) and CNN models that input MRI and PET images for the diagnosis of dementia.

Instead of just following the decision obtained from a single supervised machine learning model, employing ensemble models that evaluate more than one decision might be more successful. In various studies [12–15], the performances of the composed ensemble models were compared with the single models, and better classification performances have been reported. In our study, an ensemble method that combines CNN models is proposed for the detection of dementia diseases. Voxel intensities and grid-based pixel intensities were given into three-dimensional and two-dimensional CNN models, separately. An ensemble method has been proposed to create robust decision models. We performed classification on these models to identify Alzheimer's patients and individuals with MCI.

3 Materials and methods

In this section, we present the data preparation steps, followed by the explanation of the proposed method. Overview of the method is shown in Fig. 1.
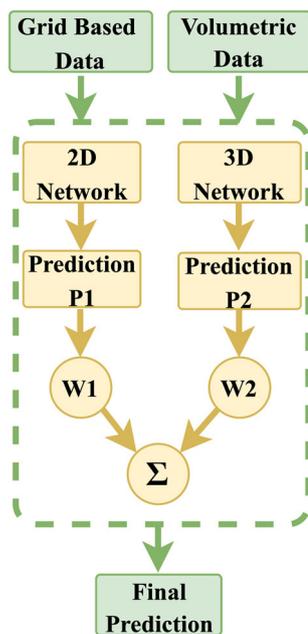
Fig. 1 Overview of the proposed ensemble method

Fig. 2 FDG-PET scans of an Alzheimer's patient. This is the 16-image grid input for the 2D CNN

3.1 Obtaining data

The data required to train and test the proposed deep learning model were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) project [1]. ADNI-3 is the latest version of the ADNI project and is currently being updated [19]. Clinical data such as age, gender, cognitive test values of each participant, and different medical imaging techniques such as MRI, FDG-PET are available in ADNI. In this study, FDG-PET scans were preferred for the diagnosis of dementia disease; for this reason, FDG-PET scans from all versions were obtained from the dataset. 985 FDG-PET scans (296 Alzheimer's, 373 MCI, 316 healthy controls) were employed to perform diagnosis.

3.2 Preprocessing of data

Volumetric images in the dataset were obtained and processed in Neuroimaging Information Technology Initiative (NIfTI) format. PET scans were co-registered and averaged from their baseline PET scan and generated images with 1.5 mm cubic voxels. Spatial orientation and intensity normalization were applied to the scans. Scanner-specific isotropic Gaussian smoothing was applied to create scans with uniform isotropic resolution. These stages were implemented by ADNI; we acquired these scans to evaluate all patients under the same conditions. We have composed two types of input images from the PET scans. The first one contains
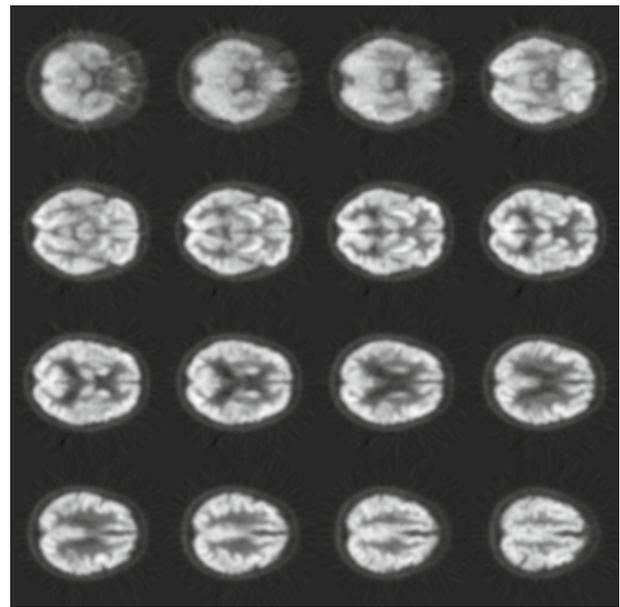
voxel intensities; these values have been normalized into 0-1 range. Normalization was performed on each perspectives of the volumetric brain scans. The PET scan has been visualized with the brain viewer application from the Statistics Online Computational Resource (SOCR) community[2]. Since the input of 3D network is volumetric data, it contains all directions of the brain.

In order to create 2D network input, some axial plane slices have been selected from volumetric scans. Sixteen slices were composed by shifting the index value by three to the right and left, starting from the middle of the data. All images have been composed into a 4x4 matrix grid form. Matrix is in grayscale color space and normalized into 0-1 scale similar to the first input. The creation of the 2D input was inspired by the study performed at the University of California [20]. They did not apply any method to select brain voxels. Grid matrices were composed of axial slices that have been chosen according to interval value; therefore, many brain regions have been included in the form of a grid. Matrices were created for each patient or healthy individual. Generated grid-based input of an Alzheimer's patient is shown in Fig. 2, which is resized to 320x320 pixels prior to feeding the network.

On the other hand, volumetric inputs were prepared for feeding 3D network. On that account, volumetric convolution operation was performed on the input. This operation was applied to all feature volumes that obtained from previous layer. Subsequently, the new feature volumes have
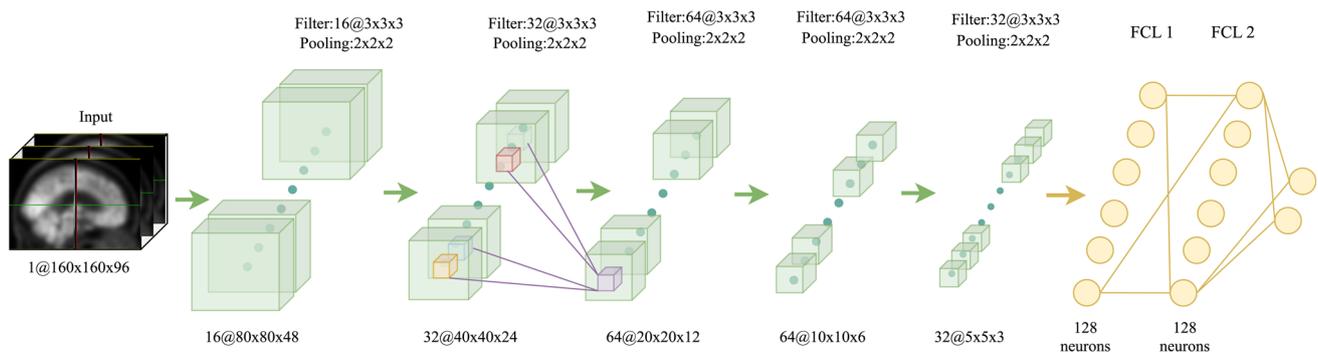

Fig. 3 Structure of the proposed 3D network

been generated by adding convolution values with bias. Network structure is shown in Fig. 3. The size of this input is 160x160x96; it comprises of all perspectives of brain, i.e., axial, coronal, sagittal. After obtaining volumetric and grid scans, we split data into six-fold; each fold comprises two sets as roughly 20% test and 80% train. Python programming language was preferred to develop and implement all pre-processing methods.

3.3 Ensemble of deep neural networks

In this study, two deep classification networks that take disparate type of data input have been constructed. The structure of the first network, which takes voxel intensities as input, is shown in Fig. 3. After 5 sets of 3D-convolution and pooling layers, data are flattened and followed by two fully connected layers and a softmax classification layer. We have preferred 3x3x3 convolution filters which do not change the size (with padding) and 2x2x2 max-pooling to decrease the volume of each unit. L2 regularization and batch normalization were employed after three-dimensional max pooling operations to avoid overfitting and reduce the number of epochs. The second classifier network is shown in Fig. 4. It takes the grid-based axial brain slices obtained from FDG-PET scans. This network has been created by inspiration of the VGG-16 model [21]. In this model, multiple convolution operations were carried out one after the other and transferred to the pooling layers. Two fully connected layers have 4096 neurons each followed by a 2-class softmax layer.

Ensemble learning is a powerful method suggesting that combined multiple base learners are better than the single weak learner. After creating the two basic learners (2D and 3D networks), these were combined in parallel to create meta learners. In this way, instead of evaluating only voxel intensities of FDG-PET scans, meta learners also take into account grid-based axial brain slices. A combination was performed with majority voting; all the more, a customized dynamically weighted majority voting algorithm was applied to classify Alzheimer's and MCI patients.

Since we have a limited number of samples in the dataset, we preferred to apply transfer learning. Namely, ImageNet weights were used to initialize our Alzheimer's disease classification network. On the other hand, obtained weights after fine-tuning of Alzheimer's disease network were applied to train the MCI network in a similar manner with [29]. The reason is that some patients with MCI may have similar glucose metabolism with Alzheimer's patients.

For all networks, Rectified Linear Unit (ReLU) activation function has been employed in the output of layers except the last layer. In the last layer, softmax activation function has been applied for two neurons to perform final classification.
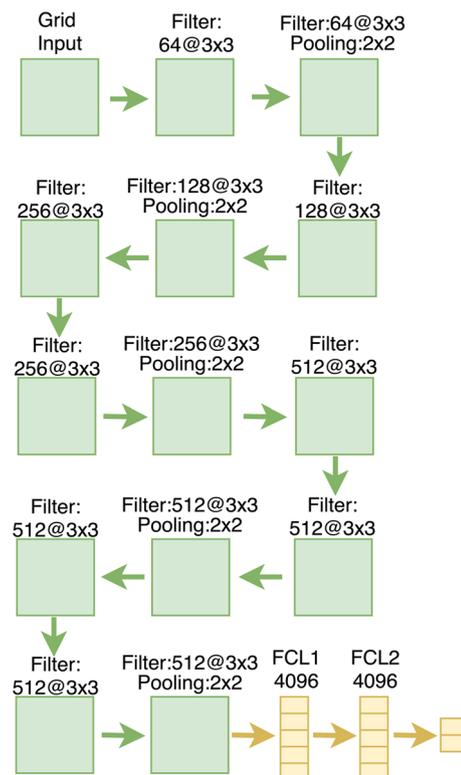

Fig. 4 Structure of the 2D network

We preferred to utilize categorical cross-entropy loss as the cost function:

$$L(y_i, \hat{y}_i) = -(y_i \cdot log(\hat{y}_i) + (1 - y_i) \cdot log(1 - \hat{y}_i)) \qquad (1)$$

where y_i is ground truth label value of ith sample and $\hat{y}_i$ is predicted value of ith sample. Besides, L2 regularization loss is added to the cost function for all networks in order to avoid overfitting:

$$L_r = \frac{1}{K} \sum_{i=1}^{K} L(y_i, \hat{y}_i) + \frac{\lambda}{2 \cdot K} \sum_{j=1}^{M} W_j^2 \qquad (2)$$

where L is the loss function given in Eq. 1, K is the size of data, λ is the regularization parameter, M is the number of features, and W_j is the sum of jth weight vector. The Adam optimization approach has been applied to optimize the cost function, and the initial learning rate values were 1e-5 and 1e-4 for the 2D and 3D networks, respectively.

2D and 3D models were combined with a custom dynamically weighted majority voting algorithm to create meta learners for the classification of patients with Alzheimer's and MCI as shown in Fig. 1. In essence, the contributions of 2D and 3D networks may not be equal to each other. When necessary, we apply weights for increasing or decreasing the impact of the networks and building robust meta learners. Let K be the number of classes, and W_β be the weight of a base learner which changes dynamically. Softmax function and ensemble equation are presented in Eq. 3 and Eq. 4. There are two base learners in our study, one weight parameter is enough. When W_β is the weight of the first base learner (s_1), $1 - W_\beta$ becomes the weight of the second one (s_2).

$$s(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad for \quad j = 1, ..., K \qquad (3)$$

$$E_{out} = \begin{bmatrix} W_\beta \cdot s_1(z_1) + (1 - W_\beta) \cdot s_2(z_1) \\ W_\beta \cdot s_1(z_2) + (1 - W_\beta) \cdot s_2(z_2) \end{bmatrix} \qquad (4)$$

The dynamic change of weights is based on the confidence scores. The idea is that if the decisions of the two base learners are different and one of them is overconfident, then the weight of the overconfident learner is decreased. Otherwise, a majority voting with equal weights is applied.

The overconfidence problem has been associated with networks using ReLU activation functions and softmax output layers [22–25]. In brief, it is known that neural networks are good at making predictions but not good at telling when these predictions are reliable. For instance, an out-of-distribution sample can have a very high confidence just because scores of other classes are low. Researchers have proposed solutions that are based on Bayesian approximations [24], calibration with temperature scaling [22], and a similar technique called

Table 1 Structures of feature fusion models

Feature/task	Number of neurons					
	Input	1st	2rd	3rd	4rd	Output
2D/AD	4096	256	256	256	256	2
3D/AD	2400	256	256	256	256	2
2D/MCI	4096	256	256	–	–	2
3D/MCI	2400	256	256	–	–	2

relaxed softmax [23]. We attacked the overconfidence problem not within the same network but while merging the scores of base learners. In other words, we penalized the overconfident decisions by decreasing their weights (W_β) with respect to the other base learner. In softmax layer, values that are greater than 0.9 were accepted as overconfident decisions. In the diagnosis of Alzheimer's and MCI, W_β values are set to 0.3 and 0.1 for penalizing current overconfident decisions as shown in Eq. 4.

3.4 Feature fusion

In order to emphasize the achievement of ensemble methods, several feature-level fusion models have been constructed. A serial fusion was applied on the features that extracted from 2D and 3D networks (Fig. 3 and 4). 2400 features from 3D network were extracted by flattening the output of last convolution and pooling operations, whereas 4096 features from 2D network were acquired from the first fully connected layer. All features were extracted automatically after the networks are trained without any prior feature engineering. Features were normalized prior feeding to a multilayer perceptron (MLP) model. The experimental results obtained from the fusion methods have been compared with ensemble methods. Initially, diverse machine learning methods on fused features such that SVM with linear and radial basis kernels, decision tree, random forest, and MLP were applied. The best performing models were the MLPs with four and two layers for AD and MCI, respectively (Table 1). A large number of deep and shallow MLP models were designed to compare performances with each other.

4 Experimental results

In this part, we present our results and compare them with the studies in the literature. After obtaining pre-processed brain scans, 2D and 3D single networks, feature fusion networks, as well as ensemble models have been employed to perform two main binary classifications. These are Alzheimer's disease (AD) versus healthy controls (HC) and MCI versus healthy controls (HC). We applied six-fold cross-validation. In each fold, dataset is split into six subsets which consist of 80%

Table 2 Comparison of the performances of our models

Task	Network	Acc	Max Acc
AD/HC	2D	90.03±2.9	94.12
AD/HC	3D	91.66±3.7	97.06
AD/HC	Fusion	**89.05±4.1**	**94.12**
AD/HC	**Ensemble**	91.83±3.0	96.08
MCI/HC	2D	70.39±2.3	73.91
MCI/HC	3D	69.37±3.5	73.91
MCI/HC	Fusion	**68.93±2.8**	**72.17**
MCI/HC	**Ensemble**	71.26±3.3	74.78

* Acc: Average accuracy, Max Acc: Best fold accuracy

Table 3 Comparison of AD diagnosis with literature

Study	Method	Acc	AD/HC
Hinrichs *et al.* [27]	LPB	84.00	89/94
Gray *et al.* [28]	SVM	88.00	50/55
Gray *et al.* [31]	SVM	81.60	71/69
Cheng and Liu [16]	CNN	87.13	93/100
Liu *et al.* [29]	CNN	88.08	93/100
Kim *et al.* [30]	GAP	91.02	141/348
Liu *et al.* [32]	CNN+RNN	91.20	93/100
Proposed	Ensemble	91.83	296/316

* Acc: Average accuracy

Table 4 Comparison of MCI diagnosis with literature

Study	Method	Acc	MCI/HC
Gray *et al.* [31]	SVM	70.02	147/69
Liu *et al.* [32]	CNN	78.09	146/100
Proposed	Ensemble	71.26	373/316

* Acc: Average accuracy

train and 20% test samples. 2D and 3D networks were trained up to 150 and 100 epochs, respectively. Since 3D networks require more computation power and approximately twice as much training time than 2D networks, it has been trained with fewer epochs. Table 2 summarizes average classification performances as well as the maximum accuracy which indicates the best fold accuracy for classification AD versus HC and MCI versus HC. Detailed performance of each fold is provided in supplementary material for both AD and MCI. All experiments were performed on a GPU (Nvidia Tesla V100) accelerated Dell Server.

Our experiments show that the performance of AD diagnosis is better than the MCI diagnosis. Since MCI is an early stage of dementia, it is hard to differentiate from normal cognitive, this situation is also discussed in the literature [26]. Although there is not a huge difference, the 3D network showed slightly better performance than the 2D network in the classification of AD. Nevertheless, it requires more computation power and more training time. It seems that the proposed grid-based image processing method is pretty useful for representing volumetric data with a lower cost. Thanks to this advantage, it can be preferred as an alternative to volumetric inputs for deep neural networks.

Our adaptive majority voting ensemble algorithm, which is utilized in the ensemble method for AD classification, regulated decisions of base learners, and decisions were improved for avoiding over-confident estimations. As a result, the ensemble method achieved higher performance than the single and feature fusion networks according to average accuracy values. In MCI detection, the grid-based 2D network is performed better than the 3D network. It is not only a cost-effective method but also achieves better accuracy value than its 3D network. Although the performance is not significantly different, the ensemble method in identification of MCI patients performed slightly better than the compared base models. It is clearly seen that serial feature fusion is not as effective as our proposed adaptive ensemble method.

There are diverse studies in the literature that proposed various solutions to detect AD and MCI. Comparisons with

these studies are summarized in Tables 3 and 4. All these studies have employed cross-validation in their experimental setup. Although it takes a long time with deep networks, we also carried out cross-validation instead of using one random test set since it is the correct way to estimate generalization performance. With the proposed ensemble method, 91.83% accuracy has been achieved to detect AD. We also achieved the best fold accuracy of 97.06% with a deep 3D network. These values are superior to the studies available in the literature. Another advantage is that in our study FDG-PET scans are given as input to the deep networks after some preprocessing steps without a separate feature extraction step. Feature extraction is carried out automatically with convolution and max pooling operations unlike [27,28,31].

We achieved 71.26% accuracy to detect MCI patients; experiments were carried out without dividing MCI patients into sub-categories such as stable or progressive. The number of participants covered in our study is higher than the most of studies available in the literature as shown in Tables 3 and 4. In terms of patient sample sizes, our study reveals more realistic performances of deep neural models for diagnosis of AD and MCI.

5 Conclusion and future work

In this study, we investigated whether the representation of the volumetric data in a two-dimensional grid-based approach is contributing to the final decision with ensemble learning to improve the diagnostic accuracy and early

diagnosis of AD. We proposed a method that exploits both the volumetric data and its 2D version, which is obtained by converting data into a grid image matrix; in addition, we presented a serial fusion method to perform diagnostic tasks. Subsequently, we introduced an ensemble learning algorithm to avoid overconfidence and thereby improve classification performance of neural networks. Our main contribution is designing an efficient ensemble method that takes advantages of diverse features.

Deep neural networks were employed to treat volumetric data and grid matrices. A network with three-dimensional convolutions was applied to volumetric data, whereas networks with two-dimensional convolutions were performed on grid image matrices. The results show that the grid processing approach can be preferred to evaluate medical image data as an alternative of using whole volumetric data. In this approach, axial images were combined to create whole matrices; however, it can be improved by adding other brain projections and creating a network that evaluates these matrices. We applied batch normalization and transfer learning to avoid overfitting and reduce the training time of deep networks. ImageNet parameters were put upon the two-dimensional deep network for the classification of AD and HC. Although content in the ImageNet data and our FDG-PET data is not similar, pre-trained parameters made a significant contribution to improve the network performance. In addition, parameters learned after training of the AD network have been employed to initialize parameters of the MCI network, since it shows better performances than the random initialization of parameters.

Our ensemble learning method which comprises 2D and 3D networks exhibited promising results for the classification of AD with an accuracy of 91.83% and MCI with an accuracy of 71.26%. Despite that grid processing approach showed encouraging results with two-dimensional networks, close results have been achieved with whole volumetric brain scans on three-dimensional networks. Further that, fewer pre-processing steps were applied to obtain whole volumetric scans. Dementia diseases are brain-related diseases that cause volume loss or atrophy. In this context, it can be an advantage to evaluate atrophy in the brain regions with a three-dimensional approach. It may be considered to optimize multidimensional deep networks to build up efficient models. We reported the results with axial projections of the brain and the whole volumetric PET scans. Therefore, there was no segmentation operation to extract brain regions; instead that allowed learning significant biomarkers from the data. The disadvantage is that it is troublesome for the visual explanation of the biomarkers by the human experts. Nevertheless, deep neural networks can capture important volumetric features as biomarkers of brain diseases especially when these are hard to catch by human eyes.

As future work, we intended to develop deep neural networks that are able to evaluate not only functional changes in the brain metabolism but also structural properties related to dementia diseases. Although it is important to detect functional changes in the brain at the early stages of dementia, after disease progresses, structural properties can show significant signs about the health status and abilities of the patient. In this sense, combining the advantages of functional and structural properties of the brain can be beneficial to evaluate the stage or progression of dementia diseases. For this purpose, FDG-PET scans exhibit functional features; on the other hand, MRI scans provide useful information about structural changes in the brain. In this context, we planned to construct multi-modal deep models that are able to evaluate FDG-PET and MRI scans as multiple features for the diagnosis of dementia diseases.

References

1. Alzheimers, A.: 2015 Alzheimers disease facts and figures. Alzheimers and dementia: the journal of the Alzheimers Association **11**(3), 322 (2015)
2. Gambhir, S.S.: Molecular imaging of cancer with positron emission tomography. Nature Rev. Cancer **2**(9), 683 (2002)
3. Jagust, W.J., Eberling, J.L., Reed, B.R., Mathis, C.A., Budinger, T.F.: Clinical studies of cerebral blood flow in Alzheimer's disease. Ann. New York Acad. Sci. **826**(1), 254–262 (1997)
4. Jellinger, K.A.: Criteria for the neuropathological diagnosis of dementing disorders: routes out of the swamp? Acta Neuropathologica **117**(2), 101–110 (2009)
5. Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E.: Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. **56**(3), 303–308 (1999)
6. Cabral, C., Morgado, P.M., Costa, D.C., Silveira, M.: Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages. Comput. Biol. Med. Alzheimer's Dis. Neuroimaging Initiatie **58**, 101–109 (2015)
7. Mudali, D., Teune, L.K., Renken, R.J., Leenders, K.L., Rerdink, J.B.T.M.: Classification of Parkinsonian Syndromes from FDG-PET Brain Data Using Decision Trees with SSM/PCA Features. Comput. Math. Methods Med. **2015**, 10 (2015)
8. Kerr, W.T., Nguyen, S.T., Cho, A.Y., Lau, E.P., Silverman, D.H., Douglas, P.K., Stern, J.M.: Computer-aided diagnosis and localization of lateralized temporal lobe epilepsy using interictal FDG-PET. Front. Neurol. **4**, 31 (2013)

9. Kang, H., Kim, W.G., Yang, G.S., Kim, H.W., Jeong, J.E., Yoon, H.J., Kang, D.Y.: VGG-based BAPL score classification of 18F-Florbetaben Amyloid Brain PET. Biomed. Sci. Lett. **24**(4), 418–425 (2018)

10. Lizuka, T., Fukasawa, M., Kameyama, M.: Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies. Sci. Rep. **9**(1), 8944 (2019)

11. Vu, T.D., Yang, H.J., Nguyen, V.Q., Oh, A.R., Kim, M.S.: Multimodal learning using convolution neural network and sparse autoencoder. In: 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), 309–312 (2017)

12. Liu, M., Zhang, D., Shen, D.: Ensemble sparse classification of Alzheimer's disease. NeuroImage **60**(2), 1106–1116 (2012)

13. Ortiz, A., Munilla, J., Gorriz, J.M., Ramirez, J.: Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. Int. J. Neural Syst. **26**(07), 1650025 (2016)

14. Yao, D., Calhoun, V.D., Fu, Z., Du, Y., Sui, J.: An ensemble learning system for a 4-way classification of Alzheimer's disease and mild cognitive impairment. J. Neurosci. Methods **302**, 75–81 (2018)

15. An, N., Ding, H., Yang, J., Au, R., Ang, T.F.: Deep ensemble learning for Alzheimer's disease classification. J. Biomed. Inform. **105**, 103411 (2020)

16. Cheng, D., Liu, M.: Classification of Alzheimer's disease by cascaded convolutional neural networks using PET images. In: International Workshop on Machine Learning in Medical Imaging, Springer, Cham, 106–113 (2017)

17. Choi, H., Jin, K.H.: Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. Behav. Brain Res. **344**, 103–109 (2018)

18. Huang, Y., Xu, J., Zhou, Y., Tong, T., Zhuang, X.: Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network. Front. Neurosci. 13 (2019)

19. Weiner, M., Veitch, D., Aisen, P., Beckett, L., Cairns, N.: The Alzheimer's disease neuroimaging initiative 3: continued innovation for clinical trial improvement. Alzheimer's and Dementia **13**(5), 561–571 (2017)

20. Ding, Y., Sohn, J.H., Kawczynski, M.G., Trivedi, H., Harnish, R., Jenkins, N.W., Behr, S.C.: A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain. Radiology **290**(2), 456–464 (2019)

21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations (ICLR) (2014)

22. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: International Conference on Machine Learning, 1321–1330 (2017)

23. Neumann, L., Zisserman, A., Vedaldi, A.: Relaxed softmax: efficient confidence auto-calibration for safe pedestrian detection. In: NIPS Workshop on Machine Learning for Intelligent Transportation Systems (2018)

24. Kristiadi, A., Hein, M., Hennig, P.: Being bayesian, even just a bit, fixes overconfidence in relu networks. In: International Conference on Machine Learning (ICML), 5436–5446 (2020)

25. Nguyen, A., Yosinski, J.: Clune J. High confidence predictions for unrecognizable images, In CVPR, Deep neural networks are easily fooled (2015)

26. Bertens, D., Vos, S., Kehoe, P., et al.: Use of mild cognitive impairment and prodromal AD/MCI due to AD in clinical care: a European survey. Alz Res Therapy **11**, 74 (2019)

27. Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M.K., Johnson, S.C.: Alzheimer's disease neuroimaging initiative, spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage **48**(1), 138–149 (2009)

28. Gray, K.R., Wolz, R., Heckemann, R.A., Aljabar, P., Hammers, A., Rueckert, D.: Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. Neuroimage **60**, 221–229 (2012)

29. Liu, M., Cheng, D., Wang, K., Wang, Y.: Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. Neuroinformatics **16**(3–4), 295–308 (2018)

30. Kim, H.W., Lee, H.E., Oh, K., et al.: Multi-slice representational learning of convolutional neural network for Alzheimer's disease classification using positron emission tomography. BioMed Eng OnLine **19**, 70 (2020)

31. Gray, K. R., Wolz, R., Keihaninejad, S., Heckemann, R. A., Aljabar, P., Hammers, A.: *et al.*, Regional analysis of FDG-PET for use in the classification of Alzheimer's disease. In: Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro San Francisco, (2011)

32. Liu, M., Cheng, D., Yan, W.: Alzheimer's disease neuroimaging initiative, classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. Front. Neuroinform. **12**, 35 (2018)