Transfer Learning for Alzheimer's Disease Detection on MRI Images

Amir Ebrahimi-Ghahnavieh
School of Electrical Engineering and Computing
The University of Newcastle
Callaghan, NSW 2308, Australia
amir.ebrahimighahnavieh@uon.edu.au

Suhuai Luo
School of Electrical Engineering and Computing
The University of Newcastle
Callaghan, NSW 2308, Australia
suhuai.luo@newcastle.edu.au

Raymond Chiong
School of Electrical Engineering and Computing
The University of Newcastle
Callaghan, NSW 2308, Australia
raymond.chiong@newcastle.edu.au

***Abstract*— In this paper, we focus on Alzheimer's disease detection on Magnetic Resonance Imaging (MRI) scans using deep learning techniques. The lack of sufficient data for training a deep model is a major challenge along this line of research. From our literature review, we realised that one of the current trends is using transfer learning for 2D convolutional neural networks to classify subjects with Alzheimer's disease. In this way, each 3D MRI volume is divided into 2D image slices and a pre-trained 2D convolutional neural network can be re-trained to classify image slices independently. One issue here, however, is that the 2D convolutional neural network would not be able to consider the relationship between 2D image slices in an MRI volume and make decisions on them independently. To address this issue, we propose to use a recurrent neural network after a convolutional neural network to understand the relationship between sequences of images for each subject and make a decision based on all input slices instead of each of the slices. Our results show that training the recurrent neural network on features extracted from a convolutional neural network can improve the accuracy of the whole system.**

Keywords*— *deep learning, transfer learning, Alzheimer's disease, convolutional neural networks, recurrent neural networks

I. Introduction

With the development of new technologies, computer-aided algorithms can be used to improve diagnosis in hospitals in terms of accuracy and speed. Advances in medical imaging and medical image analysis have provided powerful tools to generate and extract valuable neuroimaging information for monitoring neurodegeneration, and there is a great enthusiasm to make use of images for diagnosis and prediction. One area where computer algorithms might even be able to provide more accurate assessment compared to that of medical experts (e.g., radiologists) is Alzheimer's Disease (AD) detection [1]. AD is a progressive neurodegenerative disorder starting in the middle or old age, first affecting memory functions by causing deterioration in memory, then gradually affecting all cognitive functions and eventually causing death. Patients suffering from AD may forget about their life and how to perform daily activities. They may also not recognise their family members. Diagnosing AD and different stages of the disease is a big challenge because signs for identifying AD can be found in normal healthy brains of older people too [2].

In clinical practice, AD detection is based on information provided through some careful clinical examination, a thorough interview of the patient and relatives, and taking risk factors such as old age, family history and the presence of related genes in a person's genome into consideration. In 2010, the number of people over 60 years of age living with dementia was estimated at 35.6 million worldwide and 310,000 in Australasia. This number is expected to (almost) double every 20 years, which would make it 115.38 million worldwide and 790,000 in Australasia by 2050 [3]. Dementia is also the second leading cause of death in Australia, with 13,126 reported deaths in 2016 [4]. AD is the most common cause of dementia, which accounts for an estimated 60% to 80% of the cases [5]. The cost of care for patients with AD and other types of dementia is expected to increase dramatically, making it one of the costliest chronic diseases [5, 6]. Although many treatment strategies have been explored to prevent and slow down the disease, limited success is reported [7]. Due to the high cost and low success rate of AD treatments, early and accurate diagnosis of AD is not only desirable but crucial. Early diagnosis, in this case, enables cost reduction related to care and better living arrangements, as it gives patients access to supportive therapies that can help them maintain their independence for longer and delay institutionalisation [8, 9]. Early diagnosis may also support new research into understanding more of the disease and developing new treatments.

In this paper, we propose a new method to classify subjects with AD from healthy Normal Controls (NCs) based on a combination of transfer learning on 2D Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. Transfer learning is about training networks on a specific (even irrelevant) dataset as initialisation, and re-training them on a new dataset by only fine-tuning the networks. Here, pre-trained CNNs are used to extract discriminative features from 2D slices of Magnetic Resonance Imaging (MRI) scans, and LSTM is used to include spatial dependencies across the MRI slices in the classification process. Although several studies took advantage of 2D CNNs, our contribution is to combine them with LSTM to improve the classification performance. In the next section, we review recent studies on AD detection. Subsequently, our proposed method is explained in detail. Experimental results are then discussed before we conclude the study.

II. Related Work

There are many published methods in the field of machine learning since 2000, and particularly in deep learning since 2013. After 2013, the exploration of novel architectures in neural networks took off and deeper models have become popular especially in medical image processing [10]. That was when the importance of deep learning in AD detection was revealed and published papers in this area increased rapidly in 2017, as can be seen in Fig. 1. In recent years, deep learning technologies have demonstrated revolutionary performance in

978-1-7281-3745-2/19/$31.00 ©2019 IEEE

several areas including but not limited to visual object recognition, human action recognition, object tracking, image restoration, de-noising, segmentation tasks, audio classification, and brain-computer interaction [11]. With the success of deep learning in classifying 2D natural images, more and more studies have attempted to take advantage of the use of deep learning in the domain of medical images [12, 13]. For neuroimaging data, deep learning models especially CNNs can discover the latent or hidden representation, find links between different parts of an image, produce the overall cognition, and efficiently capture disease-related pathologies [10]. Due to the complexity of AD-related medical images, however, there is still a long way ahead for researchers to apply deep learning techniques for AD detection.

Deep learning models have been successfully applied for different medical image analyses such as structural MRI (referred to as MRI in the rest of this paper), functional MRI (fMRI), Positron Emission Tomography (PET), and Diffusion Tensor Imaging (DTI). The prevalence of neuroimaging modalities in single-modality studies is shown in Fig 2. According to our literature review, the most prominent diagnostic modality for AD detection is MRI, and that motivated us to focus on MRI scans in this paper. Although training a deep neural network from scratch has been done in many studies, it is often not feasible to use deep models because convergence can take too long and a dataset of sufficient size is required [14]. While image classification datasets used for object classification have millions of images, neuroimaging datasets typically only have hundreds of images, and the limited number of training data yields overfitting issues. In practice, it is common to use proven pre-trained CNNs on one domain-specific task as the initialisation step, and re-train them for new tasks by only fine-tuning their final layers [15, 16]. This is because the lower layers of CNNs contain more generic features that are useful to many tasks and can be transferred from one application domain to another. This concept, known as "transfer learning", has been a powerful tool to enable training of a large network without overfitting. It has been reported that transfer learning, even for cross-domain tasks, is faster and can achieve better performance compared to training from scratch [17-19].

A first transfer learning approach for AD detection using deep learning can be found in the work of Gupta et al. [20], where they utilised a 2D CNN with one convolutional layer and a max-pooling layer and eventually a neural network with a single hidden layer for classification after feature extraction with a sparse auto-encoder. They demonstrated that using natural images to train the auto-encoder can enhance classification performance in the following layers. Another transfer learning method was proposed by Aderghal et al. [21], where they trained three 2D CNNs with two convolutional layers (one CNN for each view of the brain scan) on only three slices at the centre of the hippocampal region of some MRI images. With a limited amount of DTI images, instead of training from scratch they employed transfer learning of models that have been trained on the MRI images to the target DTI dataset. Eventually, they combined all the networks to make the final decision with a majority voting strategy.

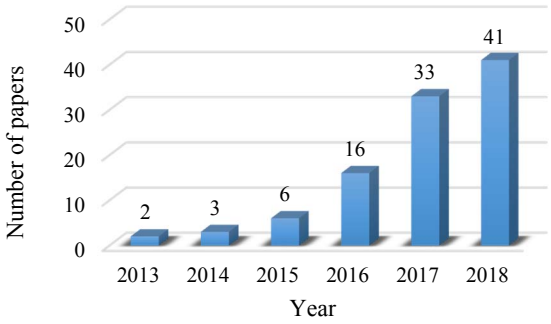


Fig. 1. The number of deep learning papers on AD detection using neuroimaging modalities.

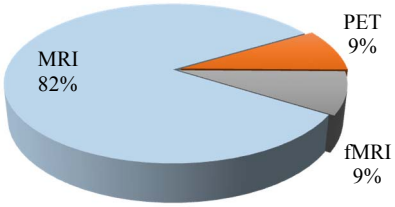


Fig. 2. Prevalence of neuroimaging modalities in single-modality studies

Focusing on well-known architectures and transfer learning from the ImageNet dataset [22], researchers have attempted to build 2D CNNs that take image slices as input based on VGGNet-16 for 20 mid-coronal slices of MRI [23], ResNet-18 for the median axial slice of MRI [17], Inception-V3 for 3 axial slices of MRI [19], Inception-V4 for the whole of MRI slices [24], VGGNet-16 and Inception-V4 for 32 most informative slices (using an entropy-based sorting mechanism) from the axial plane of each MRI scan [18], and GoogLeNet and ResNet-152 for 166 axial slices of MRI [25]. An AlexNet was fine-tuned for 2D images at the centre of each Region-Of-Interest (ROI) of PET scans in [26]. Then, an ensemble of 30% of well-performed AlexNets on all ROIs with a majority voting strategy was selected as the classifier. An ensemble of three DenseNet-styled models with different depths (121-161-169) was applied with transfer learning and the final decision was made by majority voting in [27]. For three selected axial MRI slices, which include the occipital horn, frontal horn of lateral ventricle, thalamus, lateral ventricles, inferior temporal, and middle temporal cortices, VGGNet-11 architectures were fine-tuned and then the predictions from these models were combined in [28]. More details on these structures are available in [15, 16]. Generally speaking, to be able to use these well-known architectures for transfer learning, the problem needs to be simplified from a 3D volume to 2D image slices. Slice-based architectures are usually designed based on the assumption that 2D image slices can encode certain properties of the brain. However, with faster training, it is possible that spatial relations between slices may be lost during this process. According to [19], models based on transfer learning can achieve the same accuracy as a 3D-CNN model trained from scratch.

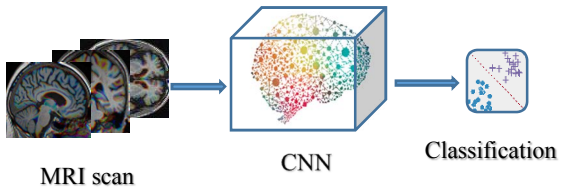
978-1-7281-3745-2/19/$31.00 ©2019 IEEE

III. METHODS

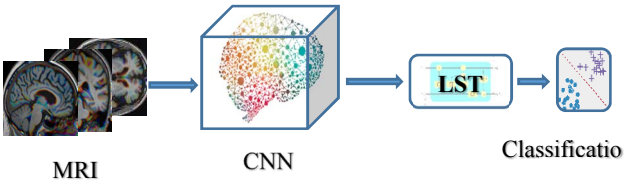
In this work, our aim is to include spatial relations among 2D slices of MRI by combining 2D CNNs with LSTM. The dataset used was taken from the ADNI[1] (Alzheimer's Disease Neuroimaging Initiative) study. After an initial stage of feature extraction employing a CNN model pre-trained on ImageNet and re-trained on the ADNI dataset, we used an LSTM architecture to capture helpful information for AD detection on a sequence of images. Traditional recurrent neural networks are not as deep as CNNs in terms of the number of layers. They may also have problems in long-term input data and require large datasets [29]. LSTM is a kind of recurrent neural network but has a more complex structure. It contains three gate units, namely the forgetting gate, input gate and output gate, and a memory cell unit. To gauge the performance of each CNN, we ran experiments on different models with different combinations of CNN and LSTM to see which model is more accurate on our dataset. We also tried both single-view and multi-view approaches to see which view is more discriminative in AD detection, and if the multi-view approach would be more accurate. A diagram of these setups is shown in Fig. 3.

We opt for the ADNI dataset as it is one of the most commonly used in the literature [30]. The primary goal of ADNI has been to test whether serial MRI, PET, or other biological markers, clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment and early AD. MRI scans used in this study were limited to AD and NC subjects, where a total of 132 subjects for each class (only baseline or screening scans) were utilised in our experiments.

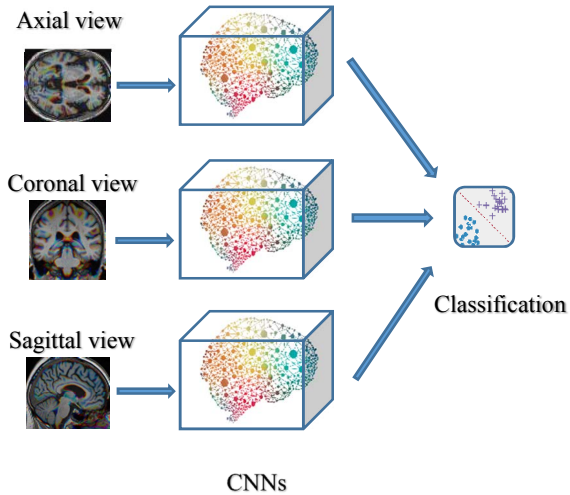
According to our literature review, image pre-processing steps are necessary and the accuracy of a classification method strongly depends on them. With the widespread use of deep learning techniques, some of the pre-processing steps have become less critical [31, 32]. However, many studies still use different pre-processing techniques on raw data including intensity normalisation, registration, tissue segmentation, skull stripping, and motion correction [33]. Among them, intensity normalisation and registration are the two most commonly utilised techniques. Normalisation is the process of mapping intensities of all images into a reference scale. Registration is a kind of spatial alignment of the images to a common anatomical space, and it is necessary due to the complexity of brain structures and the differences between brains of different subjects. Inter-patient image registration aids in standardising the neuroimaging modalities with respect to a common fixed-size template. In this study, the only pre-processing steps we have done were intensity normalisation (zero-centre) and registration to the Montreal Neurological Institute space [34] using the SPM12 toolbox [35].



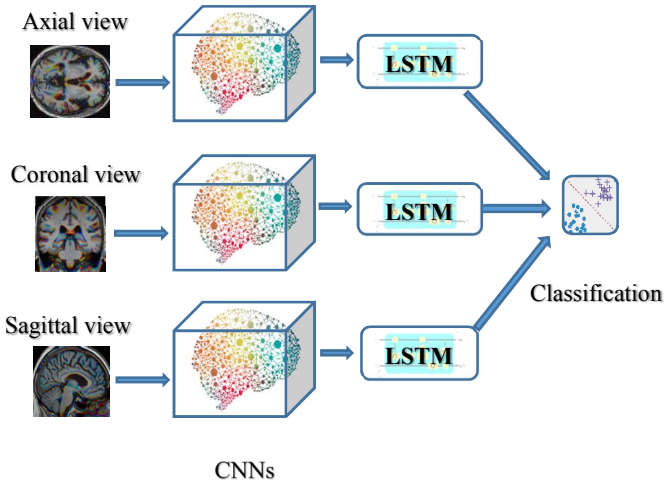
(a) Single-view CNN



(c) Single-view CNN + LSTM



(b) Multi-view CNNs



(d) Multi-view CNNs + LSTM

Fig. 3. Architectures of the tested approaches.

[1] http://adni.loni.usc.edu/

After intensity normalisation and registration, we extracted 2D image slices of each view (axial, coronal and sagittal) by discarding the first and the last slices of MRI with the dimension of $79 \times 95 \times 79$. In total, there were 48 axial, 72 coronal and 57 sagittal slices that were further divided into 16 RGB axial images, 24 RGB coronal images and 19 RGB sagittal images, respectively. The reason behind using RGB images was that all pre-trained 2D CNN structures took RGB images as input (just like the ImageNet dataset).

Subsequently, each 2D CNN structure was modified, where the last fully-connected layer was replaced with a fully-connected layer of two outputs and a weight/bias learning rate factor 10 times more than the previous layers (equal to 0.003 here). Setting this learning rate would allow us to train the new fully-connected layer faster than the other layers (the other layers were already trained on ImageNet). Also, each RGB image was resized to the size of the input layer of each CNN. Since the number of subjects was not too many for training a deep model, data augmentation was followed. Reflecting axial and coronal views in addition to ± 5 pixel translation and $\pm 5\%$ scaling were done on our training set only. We expect the CNN models to be able to extract discriminative features related to AD, especially brain shrinkage, from each slice and make a slice-based decision for each subject. A diagram of our single-view structure can be seen in Fig. 3(a). From this figure, we see that a 2D CNN is trained on single-view (axial, coronal or sagittal) image slices of all subjects using a training set. Then, for a test subject, the network classifies all image slices and the final decision is made by majority voting on all image slices. A similar procedure is followed on the multi-view structure shown in Fig. 3(b), except that in addition to majority voting on single-view image slices, another majority voting is applied over all the three views to make the final decision.

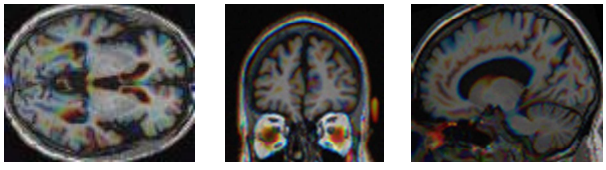
For the LSTM network, two LSTM layers, each with 100 hidden units, were designed by trial and error on our dataset. The LSTM network was supposed to be fed by features extracted after the second last layer of the CNN (i.e., before its final layer). We anticipated that the LSTM network would be able to understand the relation between sequences of images for each subject and make a decision based on all input slices instead of each slice. As shown in Fig. 3(c), a 2D CNN extracts features from single-view image slices of each subject and an LSTM network takes features of these image slices and handles them as a sequence. The same procedure is followed on the multi-view structure shown in Fig. 3(d), except again majority voting is applied over all the three views to make the final decision.

IV. Experimental Setup and Results

As aforementioned, the dataset used in our work was taken from the ADNI study. For our experiments, we divided this dataset into training, validation and test sets, with 100, 16 and 16 scans per class, respectively. We used the same training parameters for all CNNs, which included a mini batch size of 64, an initial learning rate of 0.0003 and L2 regularisation of 0.0005. The optimiser was stochastic gradient descent with a momentum of 0.9, and the maximum number of epochs was 100 with early stopping according to the validation set. Also, the same training parameters were used for all LSTM networks, which included a mini batch size of 64, an initial learning rate of 0.01 and L2 regularisation of 0.0001. The optimiser was also stochastic gradient descent with a momentum of 0.9, and the maximum number of epochs was 50 with early stopping according to the validation set. The training set was shuffled every epoch and data augmentation was conducted at the same time for CNNs. Therefore, in each epoch, the network would see slightly different input data in a different order. We utilised the MATLAB deep learning toolbox to implement and train our models on a GPU. To have fair comparisons, the same training, validation and test sets were used for all the models.

Table I shows the results of classification performance of different CNNs on axial, coronal and sagittal views. Each CNN decided on all slices of the related view independently and the final decision was made by majority voting on the whole stack of images for a specific subject. This means that a CNN would make 16, 24 and 19 decisions on axial, coronal and sagittal views, respectively. The most discriminative image slices of all three views are shown in Fig. 4. According to the results from Table I, we see that there are no considerable differences among all the views, but the sagittal view is slightly better on average. The last column of the table shows the results of combining the decisions on each view with majority voting for multi-view classification. As can be seen, on average multi-view models are more reliable and accurate compared to single-view models. For example, SqueezeNet has 90.62% of accuracy, 81.25% of sensitivity and 100% of specificity. Sometimes, however, the multi-view approach would lead to ambiguity and reduce the accuracy of the deep model (e.g., with GoogleNet the sagittal view has a higher accuracy rate of 84.38 compared to multi-view's 81.25).

Table II shows the results of classification performance of different CNNs + LSTM on axial, coronal and sagittal views. Each CNN was responsible for feature extraction from each image slice and the LSTM network was responsible to find the relation between sequences of images for each subject and make a decision based on all input slices instead of each slice. This means that the LSTM network would make a single decision on either axial, coronal or sagittal view. Again, from Table II we see that there are no considerable differences among all the views, but the coronal view is slightly better on average. The last column of the table shows results of combining the decision on each view for multi-view classification. Comparing models with and without LSTM, we can see about 2% of improvement on the accuracy on average.



(a) Axial (a) Coronal (a) Sagittal

Fig. 4. The most discriminative image slices of different views.

TABLE I. ACCURACIES OF DIFFERENT CNN MODELS (SINGLE-VIEW AND MULTI-VIEW)

CNN model	Axial	Coronal	Sagittal	Multi-view
GoogleNet	71.88	81.25	84.38	81.25
AlexNet	75	81.25	84.18	84.38
VGGNet-16	84.38	81.25	84.38	84.38
VGGNet-19	84.38	81.25	84.38	78.12
SqueezeNet	84.38	81.25	87.5	90.62
Resnet18	87.5	81.25	81.25	84.38
Resnet50	81.25	81.25	78.12	87.5
Resnet101	81.25	78.12	75	78.12
Inceptionv3	81.25	78.12	78.12	78.12

TABLE II. ACCURACIES OF DIFFERENT CNN MODELS + LSTM (SINGLE-VIEW AND MULTI-VIEW)

Deep model	Axial	Coronal	Sagittal	Multi-view
GoogleNet + LSTM	87.5	87.5	75	87.50
AlexNet + LSTM	84.38	90.62	81.25	84.38
VGGNet-16 + LSTM	81.25	84.38	81.25	84.38
VGGNet-19 + LSTM	81.25	84.38	87.5	84.38
SqueezeNet + LSTM	87.5	90.62	90.62	90.62
Resnet18 + LSTM	87.5	81.25	78.12	81.25
Resnet50 + LSTM	87.5	87.5	81.25	87.50
Resnet101 + LSTM	81.25	81.25	81.25	78.12
Inceptionv3 + LSTM	75	84.38	78.12	81.25

V. CONCLUSION

In this paper, we utilised transfer learning on MRI scans to detect AD. Different CNNs were trained on the same dataset in single-view and multi-view modes. In our first approach, CNNs classified 2D image slices according to disease-related patterns in each slice and the final decision was made by a combination of slice-based decisions. In our second approach, we used features extracted from CNN models to feed a recurrent neural network. Recurrent neural networks are typically used for time-sequence problems. Here, we used them for an image-sequence problem, where they were expected to understand disease-related patterns across image slices of an MRI scan. Our results showed that recurrent neural networks can improve the accuracy of CNNs in general. The main challenge in our work being that brain shrinkage happens both due to aging and AD. It is therefore very difficult to classify AD subjects from healthy old people only based on MRI scans. For our future work, what we need to do is to include other neuroimaging modalities (especially PET scans) or features (such as memory tests) in our system, taking different aspects of AD into consideration.

REFERENCES

[1] S. Klöppel *et al.*, "Accuracy of dementia diagnosis—A direct comparison between radiologists and a computerized method," *Brain,* vol. 131, no. 11, pp. 2969-2974, 2008.

[2] Y. Kazemi and S. K. Houghten, "A deep learning pipeline to classify different stages of Alzheimer's disease from fMRI data," in *Proceedings of the IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2018 pp. 1-8.

[3] M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri, "The global prevalence of dementia: A systematic review and metaanalysis," *Alzheimer's & Dementia,* vol. 9, no. 1, pp. 63-75. , 2013.

[4] A. B. o. Statistics, "Causes of death, Australia, 2015," Australian Bureau of Statistics Canberra, 2016.

[5] A. s. Association, "2018 Alzheimer's disease facts and figures: Includes a special report on the financial and personal benefits of early diagnosis," Alzheimer's Association, 2018.

[6] M. D. Hurd, P. Martorell, A. Delavande, K. J. Mullen, and K. M. Langa, "Monetary costs of dementia in the United States," *New England Journal of Medicine,* vol. 368, no. 14, pp. 1326-1334, 2013.

[7] F. Mangialasche, A. Solomon, B. Winblad, P. Mecocci, and M. Kivipelto, "Alzheimer's disease: clinical trials and drug development," *The Lancet Neurology,* vol. 9, no. 7, pp. 702-716, 2010.

[8] M. Prince, R. Bryce, and C. Ferri, "World Alzheimer Report 2011: The benefits of early diagnosis and intervention," Alzheimer's Disease International, 2011.

[9] S. Paquerault, "Battle against Alzheimer's disease: The scope and potential value of magnetic resonance imaging biomarkers," *Academic Radiology,* vol. 19, no. 5, pp. 509-511, 2012.

[10] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis,* vol. 42, pp. 60-88, 2017.

[11] P. V. Rouast, M. Adam, and R. Chiong, "Deep learning for human affect recognition: Insights and new developments," *IEEE Transactions on Affective Computing,* 2019. DOI: 10.1109/TAFFC.2018.2890471.

[12] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access,* vol. 6, pp. 9375-9389, 2018.

[13] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering,* vol. 19, pp. 221-248, 2017.

[14] L. Gao *et al.*, "Brain disease diagnosis using deep learning features from longitudinal MR images," in *Proceedings of Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*, 2018, pp. 327-339.

[15] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857,* 2017.

[16] D. Su, H. Zhang, H. Chen, J. Yi, P.-Y. Chen, and Y. Gao, "Is robustness the cost of accuracy? A comprehensive study on the robustness of 18 deep image classification models," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 631-648.

[17] A. Valliani and A. Soni, "Deep residual nets for improved Alzheimer's diagnosis," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2017, pp. 615-615.

[18] M. Hon and N. Khan, "Towards Alzheimer's disease classification through transfer learning," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017, pp. 1166-1169.

[19] V. Wegmayr and D. Haziza, "Alzheimer Classification with MR images: Exploration of CNN Performance Factors," in *Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL)*, 2018, pp. 1-7.

[20] A. Gupta, M. Ayhan, and A. Maida, "Natural image bases to represent neuroimaging data," in *Proceedings of the International Conference on Machine Learning*, 2013, pp. 987-994.

[21] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, "Classification of Alzheimer disease on imaging modalities with deep CNNs using cross-modal transfer learning," in *Proceedings of the IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, 2018, pp. 345-350.

[22] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision,* vol. 115, no. 3, pp. 211-252, 2015.

[23] C. D. Billones, O. J. L. D. Demetria, D. E. D. Hostallero, and P. C. Naval, "DemNet: A convolutional neural network for the detection of Alzheimer's disease and mild cognitive impairment," in *Proceedings of the IEEE Region 10 Conference (TENCON)*, 2016, pp. 3724-3727.

[24] J. Islam and Y. Zhang, "A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data," in *Proceedings of the International Conference on Brain Informatics*, 2017, pp. 213-222.

[25] A. Farooq, S. Anwar, M. Awais, and M. Alnowami, "Artificial intelligence based smart diagnosis of alzheimer's disease and mild cognitive impairment," in *Proceedings of the International Smart Cities Conference (ISC2),* , 2017, pp. 1-4.

[26] C. Zheng, Y. Xia, Y. Chen, X. Yin, and Y. Zhang, "Early diagnosis of Alzheimer's disease by ensemble deep learning using FDG-PET," in *Proceeding of the International Conference on Intelligent Science and Big Data Engineering*, 2018, pp. 614-622.

[27] J. Islam and Y. Zhang, "An ensemble of deep convolutional neural networks for Alzheimer's disease detection and classification," in *NIPS 2017 Workshop on Machine Learning for Health, arXiv preprint arXiv:1712.01675*, 2017.

[28] S. Qiu, G. H. Chang, M. Panagia, D. M. Gopal, R. Au, and V. B. Kolachalama, "Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring,* vol. 10, pp. 737-749, 2018.

[29] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Classification in BioApps. Lecture Notes in Computational Vision and Biomechanics*, vol. 26: Springer, 2018, pp. 323-350.

[30] C. R. Jack Jr *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine,* vol. 27, no. 4, pp. 685-691, 2008.

[31] M. Liu, D. Cheng, K. Wang, Y. Wang, and A. s. D. N. Initiative, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics,* pp. 1-14, 2018.

[32] D. Cheng and M. Liu, "Classification of Alzheimer's disease by cascaded convolutional neural Networks using PET images," in *Proceedings of the International Workshop on Machine Learning in Medical Imaging*, 2017, pp. 106-113.

[33] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain MRI segmentation: State of the art and future directions," *Journal of Digital Imaging,* vol. 30, no. 4, pp. 449-459, 2017.

[34] V. Fonov *et al.*, "Unbiased average age-appropriate atlases for pediatric studies," *Neuroimage,* vol. 54, no. 1, pp. 313-327, 2011.

[35] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, 2011.

978-1-7281-3745-2/19/$31.00 ©2019 IEEE