



3D DenseNet Ensemble in 4-Way Classification of Alzheimer's Disease

Juan Ruiz¹ , Mufti Mahmud¹  , Md Modasshir², M. Shamim Kaiser³ ,
and for the Alzheimer's Disease Neuroimaging Initiative

¹ Department of Computing and Technology, Nottingham Trent University, Clifton,
Nottingham NG11 8NS, UK

mufti.mahmud@ntu.ac.uk, muftimahmud@gmail.com

² Department of Computer Science, University of South Carolina,
Columbia, SC, USA

³ IIT, Jahangirnagar University, Savar, 1342 Dhaka, Bangladesh

Abstract. One of the major causes of death in developing nations is the Alzheimer's Disease (AD). For the treatment of this illness, is crucial to early diagnose mild cognitive impairment (MCI) and AD, with the help of feature extraction from magnetic resonance images (MRI). This paper proposes a 4-way classification of 3D MRI images using an ensemble implementation of 3D Densely Connected Convolutional Networks (3D DenseNets) models. The research makes use of dense connections that improve the movement of data within the model, due to having each layer linked with all the subsequent layers in a block. Afterwards, a probability-based fusion method is employed to merge the probabilistic output of each unique individual classifier model. Available through the ADNI dataset, preprocessed 3D MR images from four subject groups (i.e., AD, healthy control, early MCI, and late MCI) were acquired to perform experiments. In the tests, the proposed approach yields better results than other state-of-the-art methods dealing with 3D MR images.

Keywords: Convolutional neural network · Deep learning · Magnetic resonance imaging · Machine learning · Neuroimaging

1 Introduction

Alzheimer's Disease (AD) is a neurodegenerative disease and the usual cause of dementia in adult life. It is characterized by the deterioration of neurons, affecting most of its functions, and producing the loss of immediate memory [17]. One study has shown that the approximate number of individuals affected by AD will duplicate in the next two decades, and by 2050, a diagnose of AD is anticipated to approximately be produced every half minute, forecasting almost one million new cases every year in the United States [3]. As a result of this, the cost of treating and taking care of AD patients will be increasing, so it becomes crucial to build computerized systems that can detect early AD accurately and slow down its progress.

Artificial intelligence, in particular, machine learning (ML) has gained unprecedented attention during the last decade with applications such as anomaly detection [7, 27, 28], assay detection [23], biological data mining [14, 16], disease diagnosis [2, 18, 19, 29], education [25], financial prediction [20], natural language processing [21], trust management [15] and urban services [9]. Several of these ML methods (e.g., random forest [6], and auto-encoders [11]) have been employed for this type of research recently. This difficult research can be solved with Deep learning (DL) models that can be fed with 3D images and learn features to perform better with enhance detection. Studies done recently have shown that convolutional neural networks (CNNs) yield better results than the traditional approaches in computerized prediction of AD from MR images [8].

This paper proposes a novel approach of probability-based fusion of several CNN models to diagnose AD stages using brain 3D MRI scans. This model is able to perform a 4-way classification between the healthy brains (CN), brains with early MCI (EMCI), late MCI (LMCI) and diseased brains with AD (AD), (CN vs. EMCI vs. LMCI vs. AD) on the ADNI dataset.

In the rest of this paper, Sect. 2 reviews the literature, Sect. 3 describes the proposed method, Sect. 4 reports and discusses the results, and Sect. 5 concludes the paper along with some possible future research directions.

2 Related Works

The automatic classification of AD is an issue that has been under research for more than a decade. In recent years, there has been considerable progress in the field with DL models achieving near-perfect accuracy scores [4], thanks to the progress in the robust DL models, more specifically CNN-based models have been widely used for medical diagnosis research.

When DL was started to be employed in medical imaging classification, Liu et al. [12] used a stacked auto-encoder, to classify the early stage of AD. Applying 10-fold cross-validation to measure the model and achieved 47.42% accuracy in classifying 4 classes. Their dataset was unbalanced which was a limitation for the approach; thus classifying some groups was more complex than others for the auto-encoder. Moreover, with the popularity of predesigned CNN architectures that performed well on the ImageNet Large-Scale Visual Recognition Challenge, researchers started to focus in the potential that transfer learning has for computational biology applications. Farooq et al. [8] proposed an approach using pre-designed CNNs and 2D segments of MRIs. The approach implemented complex CNN architectures from the ImageNet challenge, in this instance, two residual networks (Resnet-18 and ResNet-152) and Google’s LeNet achieved astonishing accuracy, 98.01%, 98.14% and 98.88% respectively in a 3-way classification. Many of the reported AD classification methods were applied on 2D segments of MRIs, which in nature are 3D, and these approaches usually need multiple processes for feature extraction that help in future phases of training the model. Korolev et al. [10] developed powerful and altered adaptations of the VGG and Residual network architecture to work with 3D images. Additional research using

3D CNNs was conducted by Tang et al. [22] consisting of 1 ternary and 3 binary classification problems achieving 91.32% accuracy for the ternary classification, 88.43% for AD/MCI, 92.62% for MCI/NC and 96.81% for AD/NC. A different method researched by Wang et al. [24] achieved extraordinary performance in ternary classification with the application of an 3D ensemble approach. The approach consisted of merging the more efficient DenseNet classifiers that were trained individually and produced the probabilistic output through a softmax layer; lastly, the final classification was obtained by feeding the previous probability scores to the probability-based fusion approach.

3 Proposed Method

The proposed approach employs distinct 3D DenseNets that vary in their hyper-parameters and are fed with MR images that pass through the network and the networks classification probability goes to a probability-based fusion approach to make the last classification. Traditional network models comprises l layers, taking z_l as the output of the l^{th} layer, and every layer implements a non-linear transformation $H_l(\cdot)$, where l indexes the layer. To impede vanishing gradient and improve the information flow during the network training, the DenseNet employs the connections from a layer to all the following layers. In the approach

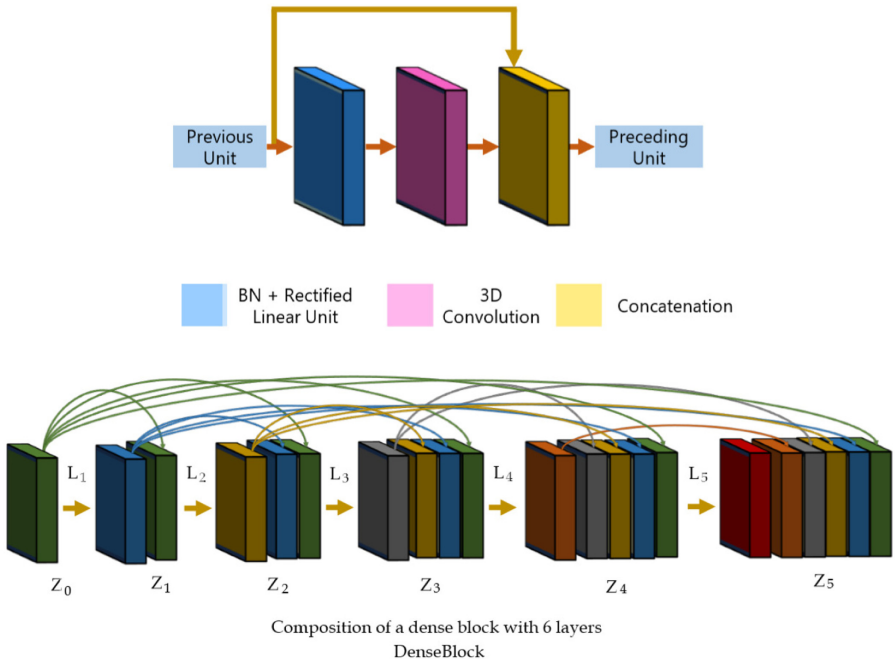


Fig. 1. Top: Architecture of a dense unit. Bottom: Composition of dense connectivity in a 6-layer dense block.

implemented for this study, the idea of dense connectivity is expanded to the 3D volumetric image classification task. Specifically, l is defined as:

$$z_l = H_l([z_0, z_1, z_2, \dots, z_{l-1}])$$

where z_0, z_1, \dots, z_{l-1} are 3D feature volumes produced in previous layers, [...] refers to the concatenation function. The structure of a dense unit is shown in Fig. 1 (top). The function $H_l(\cdot)$ has three main actions: a batch normalisation (BN) layer to decrease internal covariate transform, spatial convolution with k $3 \times 3 \times 3$ convolution kernels to produce 3D feature volumes, and to accelerate the training phase a rectified linear unit (ReLU). Figure 1 (bottom) shows a dense unit, that comprises one layer in a dense block, and each layer in the block connected with all the following layers. With dense connections between layers, feature utilisation is more efficient, and feature growth for each layer is lower than that of traditional CNNs. Thus, the models are compact and have less parameters than other networks.

In previous research it was shown that the hyper-parameters of the 3D DenseNet affects the performance [24]. Multiple tests were conducted with diverse hyper-parameter sets to produce individual networks with unique compositions, and that were able to extract different features. One sensible hyper-parameter demonstrated to enhance the outputs of the network was the growth rate. If each function H_l generates g volume-features, it means that the l^{th} layer has $g_0 + g \times (l - 1)$ input volume-features, where g_0 is the number of channels in the input layer. The 3D DenseNet can have compact layers, e.g., $g = 12$, where g is the growth rate of the network. Every layer appends g feature-maps of its own to the state given that every layer has access to all the previous volume-features in its block.

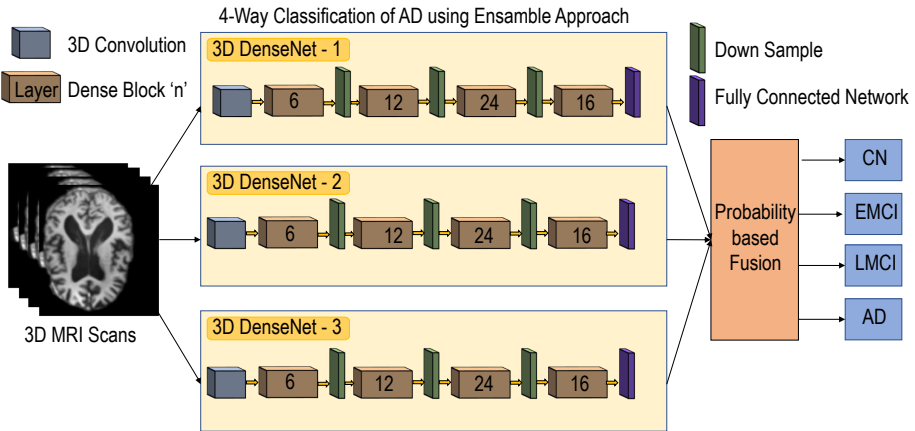


Fig. 2. Architecture of the proposed ensemble 3D DenseNet framework for 4-way AD classification.

The proposed method consists in the implementation of a probability-based fusion ensemble approach [26], having the probabilistic outcome of the last classification layer from the varying individual networks are combined (see Fig. 2). Compared to the usual majority voting method that uses as labels, the outcome that appeared the most in the models, in this ensemble method, every model is individually trained, thus the error margin among the different classifiers are insignificant, making the results of the approach superior compared to one single classifier. The error margin could rise for simple classifiers if the subject classification is complicated and there’s incertitude among the distinct classes. As an example, take three classifiers, the output probabilities of the classifier layer for CN, EMCI, LMCI, and AD are: (1)0.8, 0.1, 0.1, 0.0 (2)0.4, 0.5, 0.0, 0.1 (3)0.3, 0.5, 0.1, 0.1, respectively. Making use of a majority voting approach, the classification result is Early MCI. On the other hand, this is not the most accurate answer, considering that the classification of the prediction model 1 is more certain in the prediction, while 2 and 3 had incertitude in theirs. The probability-based fusion approach will take the sum of the probabilistic output for each class of all the classifiers and then make a more certain prediction. For this research, u individual models were picked, and the probabilities of 3D DenseNet $_u$ assigned to classes on testing set were:

$$\Omega^u = (\beta_1^u, \beta_2^u, \beta_3^u, \beta_4^u)$$

where β_n^u indicates the probabilities of the class n . Then Ω^u is normalized by:

$$\Omega^k = \frac{Y^u}{\max[\beta_1^u, \beta_2^u, \beta_3^u, \beta_4^u]}$$

when outputs of the c -based 3D DenseNets have been calculated, the final prediction label is determined by the probability-based fusion method as:

$$a = \arg \max \left(\prod_{u=1}^c \beta_1^u, \prod_{u=1}^c \beta_2^u, \prod_{u=1}^c \beta_3^u, \prod_{u=1}^c \beta_4^u \right)$$

3.1 Experimentation

MRI Data. Structural brain MRI scans from the ADNI dataset (<http://adni.loni.usc.edu/>) were used (n=600 images) in this study. Preprocessed MRI scans (e.g., mask, intensity normalisation, reorientation, and spatial normalisation) were downloaded in NIfTI file format from ADNI2 and ADNIGO. For all the experiments the dataset was divided into 80% training and 20% validation, hence the training set consisted of 480 brain scans. With the goal of having an optimal dataset for the network, both the training and validation sets were balanced.

Parameter Selection. Multiple tests on the 3D DenseNet were carried out and the network hyper-parameters were optimised to obtain best results on the 4-way classification task. The following hyper-parameter settings were used during the training:

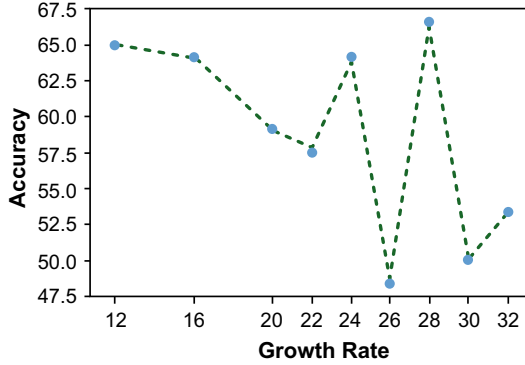


Fig. 3. Comparison of different growth rates.

- Adam stochastic optimisation algorithm;
- Pytorch’s Cross-Entropy loss function;
- *Learning rate* = 0.0001;
- *Batch size* = 4;
- *Dropout* = 0.5;
- *Epochs* = 100.

Growth Rate Analysis. The number of new features incremented at each layer is determined by the hyper-parameter g known as the growth rate of the model. Figure 3 shows the considerable change in accuracy of the models depending of the number assigned to g . When $g = 28$, the model achieved the best performance; nonetheless, it can also be observed that with $g = 12$, the accuracy is near the result best performance. Previous research [24] shown that 3D DenseNet with low growth rate was incapable of learning crucial features for prediction and consequently, did not achieve good performance.

DenseNet Network Depth Selection. Different network depth configurations of DenseNet, specifically, the 121, 169 and 201 were compared for time and accuracy while making the executing the 4-way classification task. As shown in Fig. 4, DenseNet-121 was the most efficient network depth in both parameters, and hence was chosen to be the base classifier for this task.

3.2 Selecting Optimal Number of Models

The combination of different individual classifiers can reduce the error margin. Principally due to the probability-based fusion approach being able to combine the probabilistic output of different classifier models and produce a more certain decision based on more robust and reliable data, instead of producing predictions based on only one classifier or having a majority voting method (see Sect. 3).

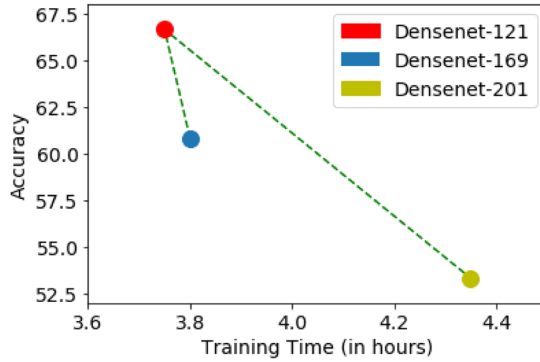


Fig. 4. Comparison of different network depths of the DenseNet model.

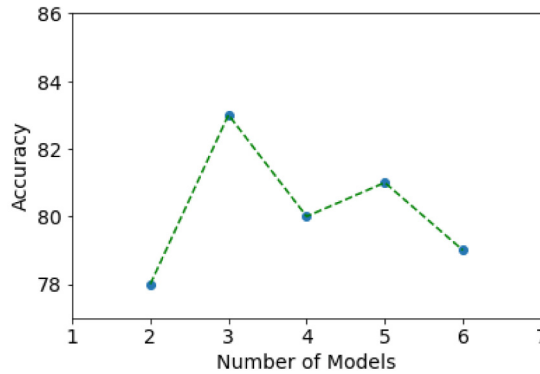


Fig. 5. Comparison of different number of models in the ensemble.

Various tests were carried out to probe the optimal number of models in creating the ensemble. As shown in Fig. 5, the ensemble with three models achieved the best accuracy. These experiments suggest that the quality of the models, i.e., how good is an individual model in predicting a specific class, is what actually is going to determine the number of models that produce the optimal performance.

4 Results and Discussion

4.1 Individual Classifier Performance

The test findings for the independent classifier models and their parameters are shown in Table 1. The best results out of the three was produced when $g = 28$, insinuating that 28 is the optimal growth rate to achieve higher results in 4-way classification with the DenseNet implementation of this study. Figure 6 shows that although classifier 1 was only 53.33% accurate in this task, it performed

fine predicting EMCI, while the other two struggled. One justification to why the classifier 1 could extract features to predict EMCI subjects could be its growth rate. With $g = 32$, the classifier 1 was the one with the bigger number of parameters, and this gave the model some leverage to extract more complex features. This being said, classifier 1 struggled when it comes to predict more simple groups like CN, on which the other two performed better; this might occur when the classifier is too complex for the training data.

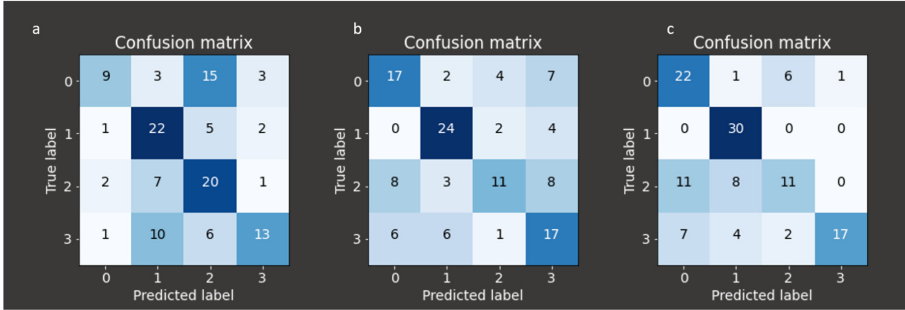


Fig. 6. Confusion matrices (a) for classifier-1 (accuracy = 53.33%), (b) for classifier-2 (accuracy = 57.50%), and (c) for classifier-3 (accuracy = 66.67%). Labels: 0 = CN, 1 = AD, 2 = EMCI, 3 = LMCI.

Table 1. Parameter comparison of different network structures.

Model	Growth rate	Convolutional kernel size	Parameters(n)	Layers	Training duration	Accuracy(%)
Classifier 1	32	$3 \times 3 \times 3$	11,226,500	121	3 h:35 min	53.33
Classifier 2	22	$7 \times 7 \times 7$	5,392,134	121	3 h:03 min	57.50
Classifier 3	28	$7 \times 7 \times 7$	8,649,224	121	3 h:45 min	66.67

4.2 Comparison with Residual Network

The results of the comparison between the DenseNet-121 and the ResNet-18 are presented in Fig. 7. The experiment demonstrated that DenseNet-18 has the quality to be trained faster and achieve more accuracy compared to the ResNet-18. The longer training time is probably due to the ResNet-18 network having around 108 million parameters to train compared to the 8.6 million of the DenseNet-121.

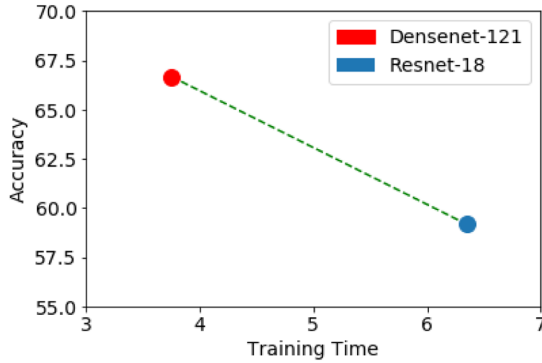


Fig. 7. Comparison of ResNet-18 and DenseNet-121.

Table 2. Test outcome of our approach compared with different methods in 4-way classification of AD.

Ref.	Architecture	Dataset (n)				Accuracy(%)
		CN	[s/E]MCI	[p/c/L]MCI	AD	
[12]	3D SAE	52	56	43	51	47.42
[5]	RF	60	60	60	60	61.9
[13]	MSDNN	360	409	217	238	75.44
[1]	3D ResNet	237	245	189	157	83.01
This work	3D DenseNet En	120	120	120	120	83.33

Legend: s = stable, p = progressive, c = converting, E = early, L = late, RF = Random Forest, MSDNN = Multi-scale Deep Neural Network, En. = Ensemble

4.3 Comparison with the State-of-the-Art

Test outcomes for this research approach compared to other similar research are shown in Table 2. While comparing with other models which use 3D MRI as input, our proposed model achieved the best performance (83.33%) which is 0.32% more than the test outcomes shown in [1]. However it becomes crucial to note here that this is not comparable in a straightforward fashion given that the other studies made use of stable MCI and converting or progressive MCI as two distinct phases of MCI, while in this research early MCI and late MCI were used, which seems to be different in literature.

Test results in the study show that the ensemble approach can lead to higher classification performance. Primarily because this approach can accumulate the probabilistic output of different classifier models and produce better predictions employing more robust and reliable data, instead of classifying based on only one classifier model.

Matched against the independent classifiers, a substantial increment in the accuracy is shown on the classification on both phases of MCI; as a result of

merging the output from classifier 1 that performed well predicting EMCI and the other classifier's predictions that were accurate classifying the other groups.

5 Conclusion

This study presented an ensemble of multiple 3D densely connected convolutional neural networks to predict AD as well as two critical stages of MCI (known as early MCI and late MCI) utilising MR images. The prediction and discrimination between early MCI, late MCI and AD can aid in the recognition of different dementia's phases and allow the early treatment in those early life phases. With the goal of figuring out the problem of having a limited number of MR images for the training phase, the proposed approach was implemented. The 3D DenseNet is more simple to train with its lower number of parameters due to having the type of connections that enhance the flow of gradients and data throughout the network. Various tests were conducted to study the performance of the model with different parameters. Furthermore, these tests produced individual classifier models with diverse parameters and structures that could be used for the ensemble. A probability-based fusion approach was used to merge the probabilistic outputs from these models. The model's accuracy was enhanced while using the probability-based fusion approach, obtaining a final accuracy of 83.33%, in comparison to the individual member classifiers. This proves that the approach of this study is a robust and reliable method in 4-way prediction tasks, while also outperforming some previous studies. In further research of this work, we would like to implement an increment in the training dataset, to test the classifier 1 of this research, and the ensemble approach to enhance the results. Otherwise further study could include implement less training data and finding an approach that yields the same or higher accuracy with the goal in mind of having a real-world environment where the data could be scarce.

Acknowledgment. *Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu/>). For up-to-date information, see <http://adni-info.org/>.

References

1. Abrol, A., Bhattarai, M., Fedorov, A., Du, Y., Plis, S., Calhoun, V.: Deep residual learning for neuroimaging: an application to predict progression to Alzheimer's disease. *J. Neurosci. Methods* **339**, 108701 (2020)
2. Ali, H.M., Kaiser, M.S., Mahmud, M.: Application of convolutional neural network in segmenting brain regions from MRI data. In: Liang, P., Goel, V., Shan, C. (eds.) *BI 2019. LNCS*, vol. 11976, pp. 136–146. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-37078-7_14
3. Association, A.: 2014 Alzheimer's disease facts and figures. *Alzheimer's Dement.* **10**(2), e47–e92 (2014)
4. Basaia, S., et al.: Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage Clin.* **21**, 101645 (2019)

5. Dimitriadis, S.I., Liparas, D., Tsolaki, M.N., Initiative, A.D.N., et al.: Random forest feature selection, fusion and ensemble strategy: combining multiple morphological MRI measures to discriminate among healthy elderly, MCI, cMCI and Alzheimer's disease patients: from the Alzheimer's disease neuroimaging initiative (ADNI) database. *J. Neurosci. Methods* **302**, 14–23 (2018)
6. Dongren, Y., Calhoun, V., 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)
7. Fabietti, M., et al.: Neural network-based artifact detection in local field potentials recorded from chronically implanted neural probes. In: *Proceedings of the IJCNN*, pp. 1–8 (2020)
8. Farooq, A., Anwar, S., Awais, M., Rehman, S.: A deep CNN based multi-class classification of Alzheimer's disease using MRI. In: *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 1–6 (2017)
9. Kaiser, M.S., et al.: Advances in crowd analysis for urban applications through urban event detection. *IEEE Trans. Intell. Transp. Syst.* **19**(10), 3092–3112 (2018)
10. Korolev, S., Safiullin, A., Belyaev, M., Dodonova, Y.: Residual and plain convolutional neural networks for 3D brain MRI classification. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 835–838 (2017)
11. Liu, S., et al.: ADNI: multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Trans. Biomed. Eng.* **62**(4), 1132–1140 (2015)
12. Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., Feng, D.: Early diagnosis of Alzheimer's disease with deep learning. In: *2014 IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014*, pp. 1015–1018, January 2014
13. Lu, D., Popuri, K., Ding, G.W., Balachandar, R., Beg, M.F.: Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural mr and FDG-PET images. *Sci. Rep.* **8**(1), 1–13 (2018)
14. Mahmud, M., Kaiser, M.S., Hussain, A.: Deep learning in mining biological data, pp. 1–36. [arXiv:2003.00108](https://arxiv.org/abs/2003.00108) [cs, q-bio, stat], February 2020
15. Mahmud, M., Kaiser, M.S., Rahman, M.M., Rahman, M.A., Shabut, A., Al-Mamun, S., Hussain, A.: A brain-inspired trust management model to assure security in a cloud based IoT framework for neuroscience applications. *Cogn. Comput.* **10**(5), 864–873 (2018)
16. Mahmud, M., Kaiser, M.S., Hussain, A., Vassanelli, S.: Applications of deep learning and reinforcement learning to biological data. *IEEE Trans. Neural Netw. Learn. Syst.* **29**(6), 2063–2079 (2018)
17. McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.: Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology* **34**(7), 939–944 (1984)
18. Miah, Y., Prima, C.N.E., Seema, S.J., Mahmud, M., Kaiser, M.S.: Performance comparison of machine learning techniques in identifying dementia from open access clinical datasets. In: *Proceedings of the ICACIn*, pp. 69–78. Springer, Singapore (2020)
19. Noor, M.B.T., Zenia, N.Z., Kaiser, M.S., Mahmud, M., Al Mamun, S.: Detecting neurodegenerative disease from MRI: a brief review on a deep learning perspective. In: Liang, P., Goel, V., Shan, C. (eds.) *BI 2019. LNCS*, vol. 11976, pp. 115–125. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-37078-7_12
20. Orojo, O., Tepper, J., McGinnity, T., Mahmud, M.: A multi-recurrent network for crude oil price prediction. In: *Proceedings of the SSCI*, pp. 2940–2945, December 2019

21. Rabby, G., Azad, S., Mahmud, M., Zamli, K.Z., Rahman, M.M.: TeKET: a tree-based unsupervised keyphrase extraction technique. *Cogn. Comput.* (2020). <https://doi.org/10.1007/s12559-019-09706-3>
22. Tang, H., Yao, E., Tan, G., Guo, X.: A fast and accurate 3D fine-tuning convolutional neural network for Alzheimer's disease diagnosis. In: Zhou, Z.-H., Yang, Q., Gao, Y., Zheng, Yu. (eds.) *ICAI 2018. CCIS*, vol. 888, pp. 115–126. Springer, Singapore (2018). https://doi.org/10.1007/978-981-13-2122-1_9
23. Tania, M.H., et al.: Assay type detection using advanced machine learning algorithms. In: *Proceedings of the SKIMA*, pp. 1–8 (2019)
24. Wang, H., et al.: Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease. *Neurocomputing* **333**, 145–156 (2019)
25. Watkins, J., Fabietti, M., Mahmud, M.: SENSE: a student performance quantifier using sentiment analysis. In: *Proceedings of the IJCNN*, pp. 1–6 (2020)
26. Wen, G., Hou, Z., Li, H., Li, D., Jiang, L., Xun, E.: Ensemble of deep neural networks with probability-based fusion for facial expression recognition. *Cogn. Comput.* **9**, 597–610 (2017). <https://doi.org/10.1007/s12559-017-9472-6>
27. Yahaya, S.W., Lotfi, A., Mahmud, M.: A consensus novelty detection ensemble approach for anomaly detection in activities of daily living. *Appl. Soft Comput.* **83**, 105613 (2019)
28. Yahaya, S.W., Lotfi, A., Mahmud, M., Machado, P., Kubota, N.: Gesture recognition intermediary robot for abnormality detection in human activities. In: *Proceedings of the SSCI*, pp. 1415–1421, December 2019
29. Zohora, M.F., et al.: Forecasting the risk of type II diabetes using reinforcement learning. In: *Proceedings of the ICIEV*, pp. 1–6 (2020)