JOURNAL OF
EMERGING INVESTIGATORS

Article

A novel approach for early detection of Alzheimer's disease using deep neural networks with magnetic resonance imaging

Aryan Ganesh[1] and Ganesh Vanamu[1]
[1]Solorsano Middle School, Gilroy, CA

SUMMARY

Since there is no cure for Alzheimer's disease (AD), early detection is essential to mitigate the symptoms. However, early detection is extremely challenging because brain MRI scans during these initial stages look very similar to normal MRI scans to the human eye. To assist in early disease detection, we developed a convolution neural network deep learning model using a transfer learning approach to extract features and a custom dense layer to detect and classify the early stages of AD. We collected MRI scans from Kaggle and ADNI institutes at several stages of AD (mild, moderate, severe) as well as healthy control scans. We used Resnet-50, VGG-16, and DenseNet-169 base models and compared their performances for classifying the stages of AD. The test area under the curve (AUC) for each of the base models Resnet-50, VGG-16 and DenseNet-169 were 0.8334, 0.9047 and 0.8898, respectively, based on 1279 Kaggle MRI images and 0.7332, 0.7383 and 0.7133, respectively, based on 132 ADNI MRI images. VGG-16 outperformed both DenseNet-169 and Resnet-50 models and showed more accurate results. Our deep learning approach also detected and extracted the region of interest (ROI) as the superior frontal gyrus (SFG) and the hippocampus for the Kaggle and ADNI brain axial MRI images, respectively.

INTRODUCTION

Alzheimer's disease (AD) is an irreversible chronic neurodegenerative disease that results in a progressive loss of behavioral and intellectual characteristics due to the deterioration of brain tissue (1). Common symptoms include memory decline, language and perception issues, problems with reasoning or judgment, disorientation, and difficulty in learning (1,2). It has recently been shown that AD is the leading cause of dementia; approximately 70% of dementias are due to AD and an estimated 5.4 million people are affected by AD (3). Some studies show that the changes in the brain may begin a decade or more before the onset of cognitive impairment, during which there is abnormal accumulation of amyloid and tau proteins in the brain (4). Though the affected people do not experience the symptoms, there are complex changes in the brain which slowly cause the malfunction of some neurons and gradually progress to the death of brain cells. The first identified risk factor is age, with most cases occurring in seniors over 65 years old, and the second identified risk factor is genetic predisposition, i.e., family history (5).

There are four different stages for the progression of AD, namely early mild cognitive impairment (EMCI), mild cognitive impairment (MCI), late mild cognitive impairment (LMCI), and AD (6). Someone in EMCI cannot be diagnosed for dementia because they do not exhibit enough symptoms that interfere with their everyday lives; therefore, it is essential to detect the disease in its early stages of MCI to reduce the progression of the disease, mitigate adverse symptoms, and improve quality of life (7). Common symptoms of MCI include forgetfulness as well as difficulty concentrating, managing finances, and completing tasks. In this stage, the affected person often goes into denial about their symptoms, but it is critical to detect this stage before the disease progresses to the final stage (7,8). Since the symptoms in the MCI stage could be easily mis-diagnosed as frontotemporal dementia (FTD), psychiatric disorders, vascular dementia, or Parkinson's disease, it is essential to distinguish AD progression at this stage (9). To overcome diagnostic difficulties, researchers are studying cerebrospinal fluid and using techniques such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) to detect early changes in the brains of people with MCI. Of these, MRI is the most used as it is completely noninvasive and widely available. MRI scans provide pictures of abnormal changes in the brain structures and detect shrinkage of areas of the brain. MRI theoretically provides the spatial resolution needed to resolve amyloid-β plaques (8-10).

Machine learning (ML) is a branch of artificial intelligence that allows computers to solve complicated real-world problems by using statistics, probability, and algorithms to learn from data of interest. ML algorithms with computer-aided diagnosis have been extensively used to develop high-performance medical image processing systems (10-11). ML techniques were found to be very useful for the diagnosis of AD, and many studies have used classical ML methods such as Random Forest (12) and support vector machine (SVM) (13) to analyze and interpret MRI scans, classify patterns, and model the data. There are some limitations to using ML in this area since these algorithms typically involve manual selection of pre-selected regions of interest (ROIs) on the brain MRI images. Manual selection of ROIs is labor-intensive, time-consuming, and leads to errors (14).

Deep learning is a subset of ML which allows us to solve

complex problems using artificial neural networks that train themselves on large datasets. Many studies have used deep learning convolutional neural network (CNN) models to automatically detect and classify the ROI using different neural network architectures (16-18), and there is evidence that CNNs can improve the learning process and provide better classification results for diagnosis of Alzheimer's disease (15). However, there are several drawbacks and issues in the existing studies. Firstly, several studies were not able to use pathologically proven datasets, meaning the data is not well-suited to achieve optimal performance (17-19). Secondly, few studies focus on identifying the ROI, which is crucial for medical diagnosis (20). Identifying the ROI helps provide a good understanding of the disease as well as increases the performance of the models. Lastly, many studies carry the issue of class imbalance, which refers to when a dataset has one class with many more instances than the other classes, leading to misinterpretation of the model (21).

Accuracy and area under the curve (AUC) are different ways to measure the performance of a ML model by distinguishing between different classes (22). Since accuracy measure does not consider probability of the prediction, AUC is a better measure of a machine learning model performance compared to accuracy under the same settings (23). Accuracy is not a proven potential metric for classification problems due to the fact that it does not focus on the class imbalance as well as the per-class performance. In a scenario where the number of data points for fraud detection is 10 and non-fraud detection is 90 and a model being trained on non-fraudulent activities more may tend to produce an accuracy of 85%. In cases where the model is not recognizing even 1 fraudulent activity out of 10 data points, the model is performing poorly even though it has 85% accuracy.

Transfer learning is all about gaining insights by addressing a problem and leveraging the knowledge gained and its application on a problem that is similar in nature. For example, parts of knowledge gained in recognizing one kind of automobile can be applied for recognizing all kinds of similar vehicles. This kind of transfer of knowledge and repurposing saves us from reinventing the wheel and has the potential to significantly improve the efficiency of the target output. Several problems related to computations, data availability, and analytics have taken advantage of this transfer learning method. Particularly for image processing and identification, the transfer learning model can gain potential knowledge by analyzing an image, which can easily be applied to a wide range of datasets and classes.

In this work, we used a transfer learning approach with pre-trained VGG-16, DenseNet-169, and Resnet-50 architectures for feature extraction and developed a custom dense layer for classification of AD MRI images. The most popular publicly available databases for MRI images are the Alzheimer‘s Disease Neuroimaging Initiative (ADNI) (24) and Open Access Series of Imaging Studies (OASIS) (25), also called Kaggle databases. We hypothesized that we could develop a deep learning model for detection of early stages of AD by analysis of MRI scans that works across both MRI imaging databases. Additionally, we predicted that our transfer learning approach using DenseNet-169 would have better performance compared to VGG-16 or Resnet-50 because it has more layers.

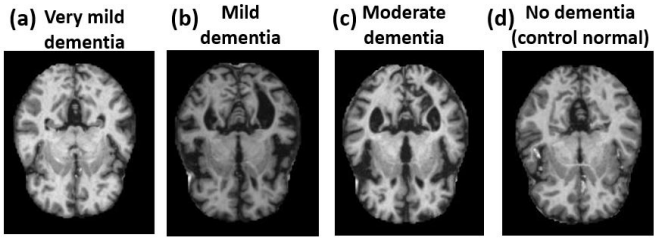

Figure 1: Example of Kaggle MRI images at various stages. (a) Very mild dementia (b) Mild dementia (c) Moderate dementia and (d) No dementia (control normal).

RESULTS

Architecture for the model

We implemented transfer learning using well established existing pre-trained VGG-16, Resnet-50, and DenseNet-169 base models in the image classification to extract the features and created a custom dense layer for classification of the stage of the AD (**Figure S1**). We tested our model using both ADNI and Kaggle databases. We used various techniques or layers like flattening, dropouts, batch normalization, dense layers and activation to build our model. Flattening is converting the 2-D vector matrix data into a 1-dimensional array for inputting it to the next layer for the dense layer to interpret correctly. Dropout is a way of cutting too much association among features by dropping the weights at a probability and preventing overfitting. Batch normalization is a technique for training very deep neural networks that standardizes the inputs to a layer for each mini-batch and helps stabilize the learning process and dramatically reduces the number of training epochs required to train deep networks. A dense layer feeds all outputs from the previous layer to all its neurons,

Kaggle data-set images	Very mild demented	Mild demented	Moderate demented	Non-demented
Training	1434 images	574 images	42 images	2048 images
Validation	358 images	143 images	10 images	512 images
Testing	448 images	179 images	12 images	640 images

Table 1: Kaggle dataset images for different classes in training, validation, and testing groups.

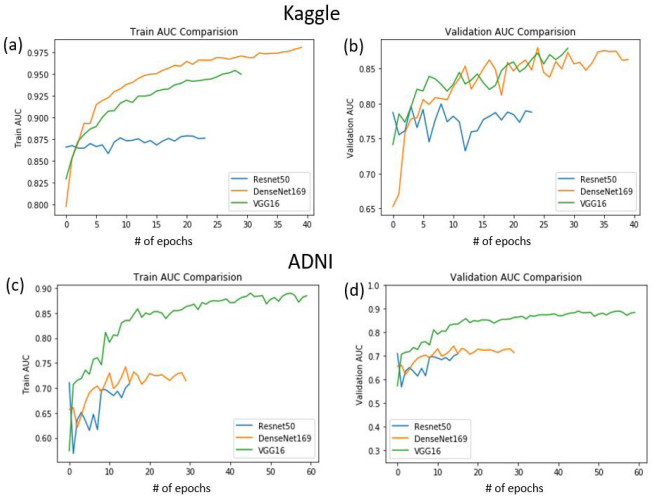

Figure 2: Training and validation AUC performance for each of the models using Kaggle and ADNI data sets. We split the Kaggle data into three sets: a training set consisting of 64% of the dataset (4098 images), a validation set consisting of 16% of the dataset (1023 images), and a testing set consisting of 20% of the dataset (1279 images). (a) The training Area Under the Curve (AUC) performance for each of the models DenseNet-169, VGG-16, and Resnet-50 were 0.98, 0.95, and 0.875, respectively, after 23, 30, and 39 epochs, respectively, using the Kaggle dataset. (b) The validation AUC performance for each of the models DenseNet-169, VGG-16, and Resnet-50 were 0.88, 0.87, and 0.78, respectively, after 22, 28, and 39 epochs, respectively, using the Kaggle dataset. We split the ADNI data into three sets: a training set consisting of 80% of the dataset (1035 images), a validation set consisting of 10% of the dataset (129 images), and a testing set consisting of 10% of the dataset (132 images). (c) The training Area Under the Curve (AUC) performance for each of the models DenseNet-169, VGG-16, and Resnet-50 were 0.705, 0.88, and 0.70, respectively, after 29, 59, and 16 epochs, respectively, using the ADNI dataset. (d) The validation AUC performance for each of the models DenseNet-169, VGG-16, and Resnet-50 were 0.705, 0.8905, and 0.70, respectively, after 29, 59, and 16 epochs, respectively, using the ADNI dataset.

with each neuron providing one output to the next layer. The activation function of a node defines the output of that node given an input or set of inputs and basically decides whether the neuron should be activated or not.

Test results using the Kaggle database

We obtained four classes of AD images using the Kaggle database, namely: "very mild demented," "mild demented," "moderate demented," and "non-demented" (**Figure 1**). The four classes represent four stages of the disease from non-demented (healthy or normal class) to moderately demented. We split the data into three sets: a training set consisting of 64% of the dataset (4098 images), a validation set consisting of 16% of the dataset (1023 images), and a testing set consisting of 20% of the dataset (1279 images). We used 1434 images, 574 images, 42 images, 2048 images for training, 358 images, 143 images, 10 images, 512 images for validation and 448 images, 179 images, 12 images and 640 images for testing for each of the classes very mild demented, mild demented, moderate demented and non-demented respectively. We used all the images that were available in the Kaggle dataset as shown in **Table 1**.

Model	Kaggle performance	ADNI performance
Resnet-50	0.8334	0.7332
VGG-16	0.9046	0.7383
DenseNet-169	0.8898	0.7113

Table 2: Performance (AUC) data for Resnet-50, VGG-16, and DenseNet-169 models using Kaggle and ADNI data.

Categorical cross entropy is a loss function that is used in multi-class classification tasks. These are tasks where a variable can only belong to one out of many possible categories, and the model must decide which one. Cross-entropy calculates the difference between two probability distributions. We used categorical cross entropy as the loss metric and AUC as the performance metric. Our dataset was highly imbalanced, and AUC has been shown as a potential metric to determine model performance for these types of cases (26). In the context of our results, the higher the AUC, the better the model is at predicting the classes and hence distinguishing between patients with disease and no disease. We used a transfer learning approach to train all the reference models (Resnet-50, VGG-16, and DESNENET-169).

An epoch is a process of training the neural network with the training data for one-cycle. It indicates the number of passes of the training dataset the ML algorithm completes. Initially, we use a random kernel and the performance was low. The number of epochs is the number of times that the algorithm will work through the entire training dataset. One epoch means that each sample in the training dataset has had an opportunity to update the internal model parameters. The performance of the model stops improving because the model detection rate is not getting better.

We calculated the performance of the models across epochs using the custom dense layer we developed (**Figure 2 A-B**). The training AUC performances for each of the models DenseNet-169, VGG-16, and Resnet-50 were 0.98, 0.95, and 0.875 after 23, 30, and 39 epochs, respectively. Both DenseNet-169 and VGG-16 showed an increasing trend in the AUC performance metric until they reached their final values. Resnet-50 was flat even after 20 epochs which means that the model training is not improving with epochs (**Figure 1**). The validation AUC performances for each of the models DenseNet-169, VGG-16, and Resnet-50 were 0.88, 0.87, and 0.78 after 22, 28, and 39 epochs, respectively. Similarly, both DenseNet-169 and VGG-16 showed an increasing trend in performance until they reached their final values. Resnet-50 started off at around 0.87 and was flat even after 20 epochs. The testing AUC performances for each of the models Resnet-50, VGG-16, and DenseNet-169 were

ADNI data-set images	EMCI	MCI	LMCI	AD	CN
Training	192 images	186 images	57 images	136 images	464 images
Validation	24 images	23 images	7 images	17 images	58 images
Testing	24 images	24 images	8 images	18 images	58 images

Table 3: ADNI dataset images for different classes in training, validation, and testing groups

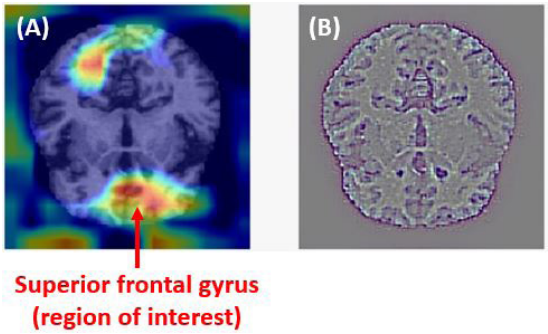

Figure 3: Region of interest (ROI) using Kaggle data showing the superior frontal gyrus (SFG) on the brain axial MRI images of the VGG-16 data. (A) with ROI and (B) without ROI. Red color shows the accurate location of the ROI.

0.8334, 0.9047, and 0.8898, respectively, based on 1279 MRI test images (**Table 2**). VGG-16 had the best performance for testing, training and validation compared to other models. Additionally, our deep learning approach also determined the ROI to be the superior frontal gyrus (SFG) (**Figure 3**).

Test results using the Kaggle database

We obtained five classes of AD images using the ADNI database, namely: early mild cognitive impairment (EMCI), mild cognitive impairment (MCI), late mild cognitive impairment (LMCI), AD (AD), and control normal (CN) (**Figure 4**). The five classes represent five stages of the disease from CN (healthy or normal class) to AD classes. We split the data into three sets: a training set consisting of 80% of the dataset (1035 images), a validation set consisting of 10% of the dataset (129 images), and a testing set consisting of 10% of the dataset (132 images). We used 192 images, 186 images, 57 images, 136 images, 464 images for training, 24 images, 23 images, 7 images, 17 images, 58 images for validation and 24 images, 24 images, 8 images, 18 images and 58 images for testing for each of the classes EMCI, MCI, LMCI, AD and CN respectively as shown in Table 3. Generally, T2 star is not the best image to use because it typically has a lower resolution and the delineation between the white matter and cortex is not well defined. We used T2 images in our study because in people with AD, you will see cortical thinning clearly and hence T2 images are better suitable to be used in the early AD detection studies.

The training AUC performances for each of the models DenseNet-169, VGG-16, and Resnet-50 were 0.705, 0.88, and 0.70 after 29, 59, and 16 epochs, respectively (**Figure 2 C-D**). VGG-16 showed an increasing AUC trend until reaching its final values. Both DenseNet-169 and Resnet-50 did not show a significant increase in the performance and soon saturated after the initial 6-8 epochs which means that the model training is not improving with epochs (**Figure 1**). The validation AUC performances for each of the models DenseNet-169, VGG-16, and Resnet-50 were 0.705, 0.8905, and 0.70 after 29, 59, and 16 epochs, respectively (**Figure 2 C-D**). VGG-16 showed an increasing AUC trend until the model reached its final value. Both DenseNet-169 and Resnet-50 did not show much increase in the performance and saturated after the initial 6-8 epochs.

The testing AUC for each of the models Resnet-50, VGG-16, and DenseNet-169 were 0.7332, 0.7383, and 0.7133, respectively, based on 132 test MRI images (**Table 1**). Overall, VGG-16 had the best performance for testing, training and validation compared to other models. Our deep

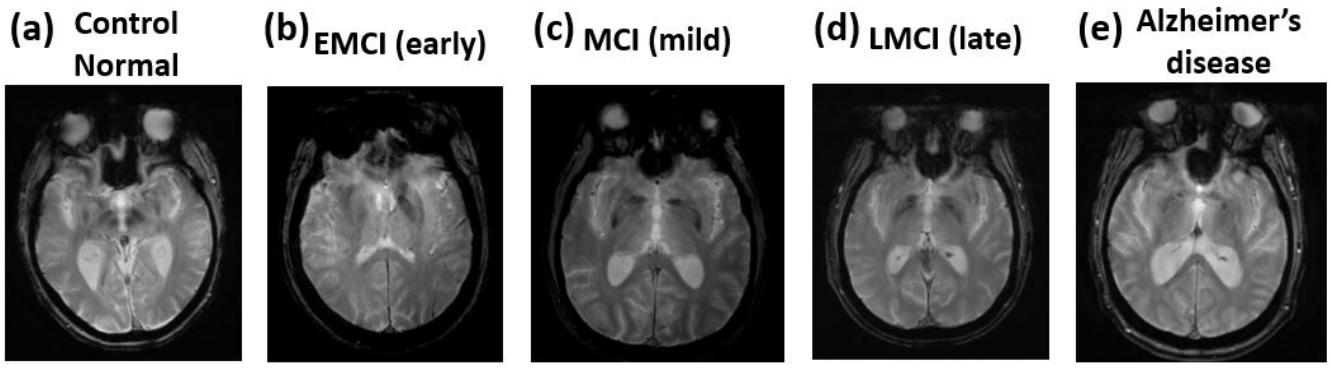

Figure 4: Example of ADNI MRI images at various stages. (a) Control Normal (b) Early Mild Cognitive Impairment (EMCI) (c) Mild Cognitive Impairment (MCI) (d) Late Mild Cognitive Impairment (LMCI), and (e) AD (AD).

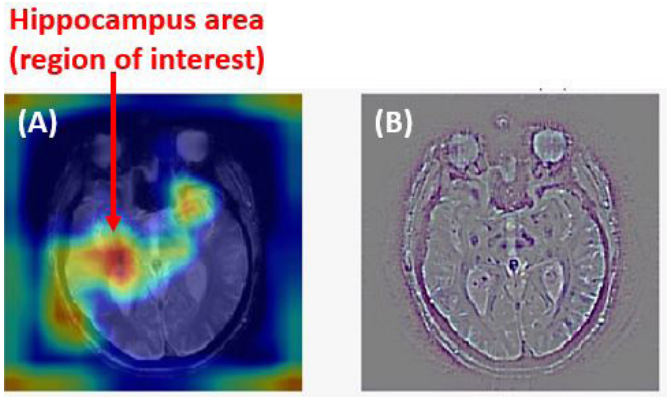

Figure 5: Region of interest (ROI) using ADNI data showing the hippocampus area on the brain axial MRI images of the VGG-16 data. (A) with ROI and (B) without ROI. Red color shows the

learning approach also determined the ROI for the VGG-16 model to be the hippocampus area on the brain axial MRI images (**Figure 5**).

DISCUSSION

In this work, we compared different models and datasets to improve early detection of AD using MRI images. By determining the testing dataset model performances, we can conclude that the VGG-16 model outperformed both DenseNet-169 and Resnet-50. Though Resnet-50 has been previously shown to have better performance on the ImageNet dataset (27-28), a large dataset of annotated photographs intended for use in development of visual object recognition software, relative to other models, our data showed Resnet-50 performed poorly on the medical imaging datasets. This is probably because Resnet-50 has a skip connection architecture unlike the other two which have straight forward architecture. Thus, straight forward architectures may better fit our dataset. Of the two models using a straightforward architecture, VGG-16 performed better than DenseNet-169. It is possible that DenseNet-169 may have too many layers and our images may not have enough features to extract, leading to overfitting of the dataset. Instead, we may just need straightforward architecture with a moderate number of layers for exemplary performance. For the Kaggle data, both DenseNet-169 and VGG-16 showed an increasing trend in the AUC performance metric on training and validation data until they reached their final values which means the training performance is improving with adding more data and with more epochs. Resnet-50 did not show any improvement with an increasing number of epochs suggesting the training performance not responding to epochs. All the models were stopped at 15 epochs if the performance did not improve.

One interesting observation was that the DenseNet-169 model using the Kaggle dataset showed better training performance compared to the VGG-16 model, though it was not better for validation and testing. This discrepancy may be due to the presence of many layers in the DenseNet-169 model which helped during the training phase, but overfitting may have affected the validation and testing performance (27). This can be prevented by human intervention such as labeling, acquiring more data, and other methods (27-30).

To the best of our knowledge, our transfer learning model performance (AUC = 0.906) is the highest compared to other studies using transfer learning and Kaggle datasets (31-32) as shown in Table-2. Our AUC performance using the Kaggle dataset was also higher compared to performance metrics on the ADNI dataset, likely due to the fact that the Kaggle dataset had 2.5 times the number of images as the ADNI dataset. Due to limited axial images in the ADNI dataset and the degree of manual involvement in the processing of these ADNI MRI images, we were not able to get as many ADNI images as we were from the Kaggle dataset in this study.

In this study and unlike previously published work, we employed automatic detection and extraction of ROI. Our deep learning model determined that the ROI was the superior frontal gyrus (SFG) using Kaggle data and the hippocampus area using the ADNI brain axial MRI images. The SFG contributes to higher cognitive functions and particularly to working memory. The hippocampus area mainly participates in short-term memory and sometimes episodic memory (33-34). Notably, the prediction of the ROI by our model using Kaggle and ADNI datasets was different. This was mostly likely because the axial scans of the MRI scans were not matched across the datasets (ADNI vs Kaggle), and they were sliced at different locations axially. Our model detected other possible ROIs in the Kaggle and ADNI images (Figures 4 and 7, yellowish-green areas), but these were not as prominent as the hippocampus and superior frontal gyrus areas in the ADNI and Kaggle images, respectively.

We have improved upon one of the issues (of not using pathologically proven dataset) other researchers have experienced by using benchmark datasets like ADNI and Kaggle (17-19). Our study also identified the ROI, which identifies the location of the region which could be the root cause of the disease and hence provides good understanding of the disease as well as increases the performance of the models. Additionally, we overcame the class imbalance issue using a data augmentation technique. Lastly, we rotated the images to certain angles to make the model more generalizable.

MATERIALS AND METHODS

Data acquisition and pre-processing

Data from the ADNI and Kaggle databases were fed into the model after data preprocessing (Figure S2). In the preprocessing step, we re-sized all the images to be consistent at 224 x 224 pixels to match the academic standard. We used the axial T1 star images of the axial brain MRI scans for Kaggle and T2 star images for ADNI to classify the stage of the AD and to find the region of interest. The ADNI database is jointly funded by the National Institutes of

Health (NIH) and industry via the Foundation for the NIH. The OASIS database, also called Kaggle, is freely available to the scientific community. Kaggle compiles a multi-modal dataset generated by the Knight Alzheimer Disease Research Center (Knight ADRC) and its affiliated studies.

The data augmentation was very crucial in our project as the ADNI dataset was much smaller than the Kaggle dataset, and only 1038 images were available for training, consisting of five classes. Therefore, we used some augmentation techniques i.e., rotation, flipping, etc. which helped the model to be more generalizable and reduce over-fitting of the dataset.

Model details

We used categorical cross entropy as the loss metric and AUC (area under the curve) as the performance metric. The AUC performance metric uses recall, sensitivity, specificity, and precision internally to determine the model performance. The two approaches we used to compare the models are transfer learning approach and custom model building in tensor Flow. We also used two different datasets to evaluate the benchmarking datasets namely ADNI and Kaggle.

Custom dense layer details

Rectified Linear Unit (ReLU) is a type of activation function in neural networks that returns 0 if the input is negative, but for any positive input, it returns that value back. We used a rectified linear unit (Relu), which rules out the negative valued output from the previous layer and only considers the positive valued output from the previous layer. We made a dense architecture with two dense layers consisting of 2048 and 1024 nodes, respectively (Figure S1). We used batch normalization for each dense layer output to redistribute the output values to a mean of 0 with a standard deviation of 1 such that the network will be stable throughout the training. We used dropout for regularization to only consider a certain number of nodes for each epoch to ensure proper training happens for each node.

ACKNOWLEDGMENTS

We would like to acknowledge Mr. Madhu Charan for guiding us through the process of building the neural network model. Furthermore, we wish to thank the reviewers and editors for giving us valuable, constructive feedback, so we could improve our report.

Received: March 17, 2021
Accepted: April 24, 2021
Published: March 20, 2022

REFERENCES

1. A. Macdonald and D. Pritchard, “A Mathematical Model of Alzheimer's Disease and the APOE Gene” Astin Bulletin, vol. 30, no.1 , 2000, pp. 69-110.
2. Thompson, C.A., Spilsbury, K., Hall, J., Birks, Y., Barnes, C., Adamson “Systematic review of information and support interventions for caregivers of people with dementia” BMC Geriatr., 2007, Vol. 7, no. 18, pp-7.18. doi: http://dx.doi.org/10.1186/1471-2318-7-18.
3. Lori L. Beason-Held, “Changes in Brain Function Occur Years before the Onset of Cognitive Impairment” J Neurosci., Vol 33, No. 46, 2013, pp. 18008–18014. doi: 10.1186/1471-2318-7-18.
4. Bateman *et al*., “Clinical and biomarker changes in dominantly inherited Alzheimer's disease.” N Engl J Med., Vol. 367, 2012; pp. 795–804. doi: 10.1056/NEJMoa1202753.
5. Ujjwala *et al*., “Alzheimer: A disease of brain”, International Journal of Engineering and Creative Science, Vol. 1, No. 1, 2018, pp. 10-14
6. Xia-an Bi *et al.,* “Analysis of Progression Toward AD Based on Evolutionary Weighted Random Support Vector Machine Cluster”, Front Neurosci, Vol. 12, 2018, pp. 1-11. doi: 10.3389/fnins.2018.00716
7. Ronald *et al.,* “Mild Cognitive Impairment”, Continuum (Minneap Minn), Vol. 22, No. 2, 2016, pp. 404–418. doi: 10.1212/CON.0000000000000313
8. Subramanyam et al, “Mild cognitive decline: Concept, types, presentation, and management.” Journal of Geriatric Mental Health, Vol. No.1, 2015, pp. 10-20. DOI: 10.4103/2348-9995.181910
9. Finger *et al*., Frontotemporal Dementias, Continuum (Minneap Minn). 2016 Apr; 22(2 Dementia): 464–489. doi: 10.1212/CON.0000000000000300
10. Santos *et al*., “Artificial intelligence, ML, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine”, Radiol Bras., Vol. 52, No. 6, 2019, pp. 387–396, doi: 10.1590/0100-3984.2019.0049
11. Khan *et al*., “Early diagnosis of Alzheimer's disease using ML techniques: A review paper.” 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), Lisbon, Portugal, 2015, pp. 380-387.
12. Tanveer *et al*., “ML Techniques for the Diagnosis of AD: A Review”, ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 16, No. 1, 2020, pp. 1-35 doi: 10.1145/3344998
13. Tripoliti, E. E., Fotiadis, D. I., and Argyropoulou, M. “A supervised method to assist the diagnosis and monitor progression of AD using data from an fMRI experiment.” Artif. Intell. Med. Vol. 53, No. 1, 2011, pp. 35–45. doi: 10.1016/j.artmed.2011.05.005
14. Leemput, K., Van Maes, F., Vandermeulen, D., and Suetens, P. “Automated model-based tissue classification of MR images of the brain.” IEEE Trans. Med. Imaging, Vol. 18, No. 10, 2002, 897–908. doi: 10.1109/42.811270
15. Silva, I. R., Silva, G. S., de Souza, R. G., dos Santos, W. P., & Roberta, A. D. A. (2019, July). Model Based on Deep Feature Extraction for Diagnosis of AD. In 2019 International Joint Conference on Neural Networks

(IJCNN), Budapest, Hungary, 2019, pp. 1-7, doi: 10.1109/IJCNN.2019.8852138.
16. Li F, Liu M; "Alzheimer's Disease Neuroimaging Initiative. Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks." Comput Med Imaging Graph., Vol. 70, 2018, 70, pp. 101-110. doi: 10.1016/j.compmedimag.2018.09.009. Epub 2018 Oct 2. PMID: 30340094.
17. Liu, M., Zhang, J., Adeli, E., & Shen, D. "Landmark-based deep multi-instance learning for brain disease diagnosis." Medical image analysis, Vol. 43, 2018, pp. 157-168. doi: 10.1016/j.media.2017.10.005. Epub 2017 Oct 27. PMID: 29107865; PMCID: PMC6203325.
18. Ebrahimighahnavieh, M. A., Luo, S., & Chiong, R. "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review." Computer Methods and Programs in Biomedicine, Vol. 187, 2020, 105242. doi: https://doi.org/10.1016/j.cmpb.2019.105242.
19. Islam *et al.*, "Brain MRI analysis for AD diagnosis using an ensemble system of deep convolutional neural networks." Brain Inform., Vol. 5, No. 2, 2018, pp. 1-14. doi: 10.1186/s40708-018-0080-3
20. Mehmood *et al.*, "A Deep Siamese Convolution Neural Network for Multi-Class Classification of Alzheimer Disease." Brain Sci., Vol 10, No. 84, 2020, pp. 1-15. doi: 10.3390/brainsci10020084
21. Kanghan *et al.*, "Classification and Visualization of AD using Volumetric Convolutional Neural Network and Transfer Learning." Scientific Reports. Vol 9, No. 18150, 2019, pp. 1-15. doi: https://doi.org/10.1038/s41598-019-54548-6
22. Charles X. Ling, Jin Huang, and Harry Zhang, "AUC: a Statistically Consistent and more Discriminating Measure than Accuracy", IJCAI'03: Proceedings of the 18th international joint conference on Artificial intelligence, 2003 Pages 519–524
23. Charles X. Ling, Jin Huang, and Harry Zhang, "AUC: A Better Measure than Accuracy in Comparing Learning Algorithms", Advances in Artificial Intelligence, 16th Conference of the Canadian Society for Computational Studies of Intelligence, 2003, pp. 329–341.
24. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, Jack CR Jr, Jagust WJ, Shaw LM, Toga AW, Trojanowski JQ, Weiner MW. "Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization." Neurology., Vol. 74, No. 3, 2010, pp. 201-209. doi: 10.1212/WNL.0b013e3181cb3e25.
25. Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. "Open Access Series of Imaging Studies (OASIS): Cross sectional MRI data in young, middle aged, nondemented, and demented older adults." Journal of Cognitive Neuroscience, Vol. 19, No. 9, 2007, pp. 1498-507. doi: 10.1162/jocn.2007.19.9.1498.
26. André *et al.*, "A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms", BMC Medical Informatics and Decision Making, Vol. 20, No. 4, 2020.
27. Liu *et al.*, "AD detection using depth wise separable convolutional neural networks", Computer Methods and Programs in Biomedicine, Vol. 203, 2021, 106032. https://doi.org/10.1016/j.cmpb.2021.106032
28. Sun Y., Wang X, Tang X. (2014) Deep Learning Face Representation from Predicting 10,000 Classes. Computer Vision and Pattern Recognition. IEEE, pp.891-1898.
29. Karystinos G N, Pados D A. (2000) On overfitting, generalization, and randomly expanded training sets. IEEE Transactions on Neural Networks, 11(5):1050.
30. Yip K Y, Gerstein M. "Training set expansion: an approach to improving the reconstruction of biological networks from limited and uneven reliable interactions", Bioinformatics, 15;25 (2), 2009, pp.243-250.
31. Liu *et al.*, "On the design of convolutional neural networks for automatic detection of AD", Proceedings of ML Research, Vol. 116, 2020, pp. 184-201.
32. Folego *et al.*, "AD Detection Through Whole-Brain 3D-CNN MRI", Frontiers in Bioengineering and Biotechnology, Vol. 8, Article 2020, pp. 1-14. doi: 10.3389/fbioe.2020.534592
33. Tom *et al.*, "The hippocampus is required for short-term topographical memory in humans", Hippocampus,17(1), 2007, pp: 34–48. doi: 10.1002/hipo.20240
34. Nicole *et al.*, "Frontal Structural Neural Correlates of Working Memory Performance in Older Adults", Front Aging Neurosci., 2016; 8: 328. doi: 10.3389/fnagi.2016.00328

A novel approach for early detection of Alzheimer's disease using deep neural networks with magnetic resonance imaging- Supplemental

Aryan Ganesh[1] and Ganesh Vanamu[1]

[1]Solorsano Middle School, Gilroy, CA

Supplemental:

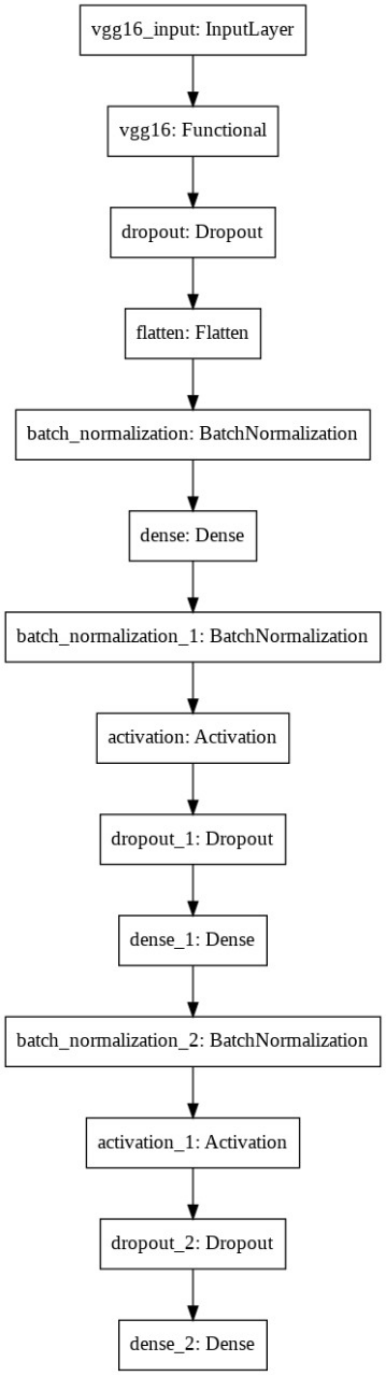

Figure S1. CNN deep learning model architecture using transfer learning (using VGG-16, Resnet-50, or DenseNet-169) for feature extraction and custom dense layer for classification.

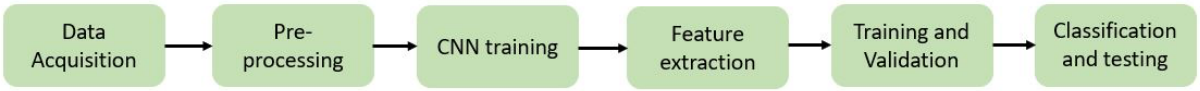

Figure S2. Block diagram showing the various steps involved in the deep learning CNN model