Group 52: Early Detection of Alzheimer's Disease Through Machine Learning in MRI Scans

Zachary Burns, Zburns@eng.ucsd.edu
Derrick Cosmas, A53323988, Dcosmas@eng.ucsd.edu
Bryce Smith, A11724731, Bdsmith@eng.ucsd.edu
University of California San Diego, La Jolla, CA 92093-0238

Abstract

Since early diagnosis of Alzheimer's gives people the best chance for effective treatment, it is important to detect the disease as soon as possible[1]. One method to classify Alzheimer's is by identifying structural changes in the brain through magnetic resonance imaging (MRI). This image-based classification is a task well suited for modern computer vision. We propose that machine learning can assist in early detection, specifically using trained convolutional neural networks. The primary method of diagnosis is analysis of MRI images of the patient's brain. Our approach implements a VGG16 architecture trained on a series of these MRI images. This approach successfully classifies the MRI scans into four stages of Alzheimer's with 99% test accuracy, which is better than methods previously established in the Kaggle notebook[2].

Introduction

Alzheimer's disease (AD), is a chronic neurodegenerative disease that starts slowly and gradually worsens over time. The risk of developing Alzheimer's increases with age, and the number of people who will develop Alzheimer's disease is expected to skyrocket over the next few years as seventy-nine million adults reach retirement age[1]. The most evident characteristic of AD pathology is neuron loss, progressing from AD signature regions (e.g. hippocampus and amygdala) to the entire cortical region. This can be measured by a magnetic resonance imaging (MRI) scan[3]. The ability for machine learning to assist in AD's classification has been recognized but not widely adopted. We present a model that performs near perfect classification to show that machine learning can play an important role in helping people with this debilitating disease.

Our program takes in grayscale, 176 x 208 pixel T1 MRI images and predicts which of four clinical diagnoses they belong to. The four diagnosis classifications presented in this work are: non demented, very mild demented, mild demented, and moderate demented. The main contribution of our report is a VGG16-based deep neural network trained on a set of about 5000 of these images and their accompanying labels. This newly trained model predicts which class of Alzheimer's stage the MRI image belongs to, with accompanying conditional probabilities (between 0-1) for each class.

Related Works

As early as 1999[4], researchers have been employing various machine learning methods to detect AD onset from MRI data alone[5]. Researchers began by establishing biomarkers to be evaluated. A biomarker is a quantifiable indicator of a biological state. Hippocampal volume has been a key biomarker that doctors have used in the past, and it has shown high diagnostic accuracy for AD. However, hippocampal volume alone is considered clinically insufficient as a predictor of progression from mild cognitive impairment (MCI) to AD[6]. Hence there is a need to look at more features, and recent research has shown that important changes in many other cortical regions also occur[7]. Given the spatial features in the MRI scans, machine learning and deep learning methods tend to be good at aggregating such features for classification.

We have established that medical descriptors together with statistical and conventional machine learning methods have been widely used to aid the classification of AD. Deep learning tools have been gaining popularity as a machine learning tool because they enable richer and deeper representation of features. One work employed a Deep Neural Network containing autoencoders to combine features from Region of Interest (ROI) in PET and MRI scans. Plant et al. (2010) used feature selection to achieve 92% accuracy in a binary classification of AD[8]. Their paper demonstrates that principal component analysis, independent component analysis, structural equation modeling, and support vector machines have recently shown promise analyzing structural MRI to detect spatial patterns of atrophy associated with AD.

ROI-based methods such as this can extract representative features and partly reduce the feature dimensions, but usually the ROIs are too empirical as they are based on qualitative observation. This is not

sufficient to capture the more granular features associated with AD. An additional method proposed was to capture successive slices from MRI scans because it was deemed that each slice covers the significant areas for dementia detection. Wang et al.(2018) used DenseNet and ensemble methods to classify the entire 3D MRI scan[9], leading to a new state of the art three class classification (Alzheimer versus Mild Cognitive Impairment versus Cognitively Normal) accuracy of 97%.

Dataset and Features

We obtained our dataset from an online Kaggle challenge of MRI brain images[10]. The images are black and white with a standard dimension of 176 x 208 pixels. As given, there are about 5000 training images and 1200 test images. The dataset is separated into 4 classes: Non-Demented, Very Mildly Demented, Mildly Demented, and Moderately Demented. Each image is a 2D MRI cross section of the brain as shown in Figure 1. The images are planes taken at different heights of the brain. We chose this dataset for its relatively large set of images and the consistency of the images. All images came with the brain cross-section centered in the image and with the background cropped out. A disadvantage of this dataset was that it came with little information regarding the composition of the test and training set which later proved to be an issue.

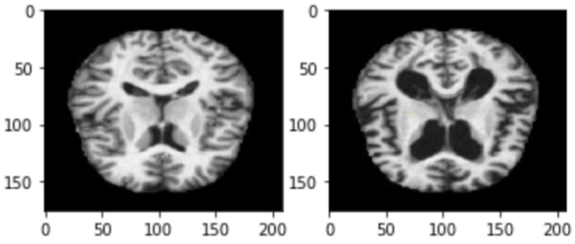


Figure 1: MRI Images from the Dataset at Different Heights in the Brain

After several rounds of training we found a discrepancy between the train and test datasets that needed to be fixed. This is problematic as the test set should generally be similar to the set the model was trained on or it will likely perform worse than the training set. In the test dataset each subject has only two slices in a certain region of the brain whereas in the train dataset there are slices at many heights. We fixed this problem by combining the test and train datasets, reshuffling them, and then splitting again into new test, train, and validation sets. After this modification our model performed much better.

There were several preprocessing steps taken with the data. All image values were normalized to between 0 and 1 from the original grayscale values of 0 to 255. Additionally, the single channel grayscale images were expanded to three channels by copying the same values to all three channels. This was done because the model we used, VGG16, was designed for color images and requires 3 channel inputs.

Several augmentations were made to the data to introduce diversity to the dataset and discourage the model from overfitting. These included randomly changing the magnification of the image and varying the overall brightness of the image. We did not perform any feature extraction but in future work we would be interested in performing a 2D Fourier transform on the images to see if better results could be achieved.

Methods

Creating and training a new deep neural network from scratch is computationally expensive, therefore we decided the best approach would be to utilize transfer learning from a pre-trained network. Many such networks for image classification exist and are readily available to download through Keras and Tensorflow with the weights trained on the ImageNet dataset. Since our dataset is different from the ImageNet set, we needed to retrain the fully connected layers at the output of the network for our classification problem. The weights for the convolutional layers, however, should not need as much modification. The earlier convolutional blocks in the model have learned to recognize very low-level features that are common to all natural images so they will already be optimized. The convolutional layers at the end will learn higher level features that are more specific to the dataset being used so these will likely need more optimization during training. In our case we allowed all the convolutional layers to be trained to further optimize the network.

We considered several models such as DenseNet[12] and Inception[13,15], but ultimately went with VGG16 as the model architecture is relatively simple and would be easier for us to debug. VGG16 is widely known as it was one of the first neural networks to perform well at the Large Scale Visual Recognition Challenge using a very deep network[14].

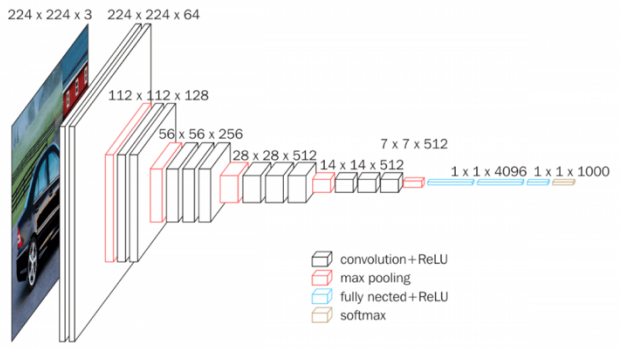


Figure 2: VGG16 Architecture[14]

Vgg16 is a standard CNN which consists of layers of convolutional blocks with max pooling layers in between and several fully connected layers at the end of the network[18]. The last fully connected layer has the same number of nodes as the number of classifiers with the softmax activation function. The softmax function is defined as:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \text{ for } i = 1, \ldots, K \text{ and } z = (z_1, \ldots, z_K) \in R^K \quad (1)$$

This formula computes the conditional probability of each class given the input (image) to the model. There is a conditional probability calculated for each of the classifications. The model predicts the image classification by picking the class with the highest conditional probability.

We modified the original VGG16 network by inserting new fully connected layers at the end of the network that were untrained. We then restrained the entire network using our training set. For our loss function we used categorical cross-entropy as it is widely used for multiclass classification problems. The categorical-cross entropy is defined as:

$$H(p,q) = -\sum_{x} p(x) \log(q(x)) \quad (2)$$

Where p is the true distribution and q is the computed distribution.

We had originally used an Adam optimizer, which is an adaptive moment optimizer. This has the benefits of being straightforward to implement, is computationally efficient, and its hyper-parameters have intuitive interpretation, typically requiring little tuning. However, we found through research that an Adam optimizer might not be well suited for our needs as they function best for data which has non-stationary targets or non-convex optimization problems[11] Our classification problem has stationary targets and a generally convex solution.

Our research indicated that stochastic gradient descent (SGD) with Nesterov momentum is well suited for image classification, particularly medical images. Since our function is relatively smooth, this iterative form of optimization is extremely effective. The Nesterov momentum helps the optimization to move through local minima which may be encountered on the way to the true weights. We saw an immediate improvement of several percent in test accuracy when we switched from Adam to SGD with Nesterov. We experimented with different momentum values (from .1-.99) and the best values were achieved with the momentum set at 0.9.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 176, 208, 3)]	0
block1_conv1 (Conv2D)	(None, 176, 208, 64)	1792
block1_conv2 (Conv2D)	(None, 176, 208, 64)	36928
block1_pool (MaxPooling2D)	(None, 88, 104, 64)	0
block2_conv1 (Conv2D)	(None, 88, 104, 128)	73856
block2_conv2 (Conv2D)	(None, 88, 104, 128)	147584
block2_pool (MaxPooling2D)	(None, 44, 52, 128)	0
block3_conv1 (Conv2D)	(None, 44, 52, 256)	295168
block3_conv2 (Conv2D)	(None, 44, 52, 256)	590080
block3_conv3 (Conv2D)	(None, 44, 52, 256)	590080
block3_pool (MaxPooling2D)	(None, 22, 26, 256)	0
block4_conv1 (Conv2D)	(None, 22, 26, 512)	1180160
block4_conv2 (Conv2D)	(None, 22, 26, 512)	2359808
block4_conv3 (Conv2D)	(None, 22, 26, 512)	2359808
block4_pool (MaxPooling2D)	(None, 11, 13, 512)	0
block5_conv1 (Conv2D)	(None, 11, 13, 512)	2359808
block5_conv2 (Conv2D)	(None, 11, 13, 512)	2359808
block5_conv3 (Conv2D)	(None, 11, 13, 512)	2359808
block5_pool (MaxPooling2D)	(None, 5, 6, 512)	0
global_max_pooling2d (Global	(None, 512)	0
flatten (Flatten)	(None, 512)	0
dense (Dense)	(None, 1024)	525312
dropout (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 4)	4100

Total params: 15,244,100
Trainable params: 15,244,100
Non-trainable params: 0

Figure 3: Our Model Parameters

The role of hidden units in neural networks is to approximate a 'function' efficiently from the available data-samples which can be generalized to unseen data. The problem is that the available data often has non-linear underlying patterns which can only be extracted by using a non-linear approximation function. So, when the hidden units try to approximate a function for these samples, they tend to fit a higher order approximator (they try to create a function that accurately approximates each instance in the training set) and by doing so, they overfit to the data-samples.

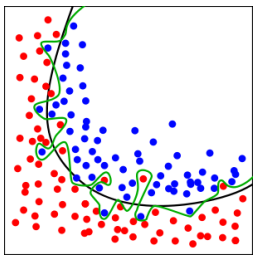


Figure 4: Overfitting[16]

From Figure 4 we can see how overfitting can cause a model to perform well on a training set but be unable to recognize similar patterns in new data. Dropout layers were also added to the network model. The drop out technique turns off some of the hidden units in previous layers essentially 'dropping out' some of the parameters during the learning process. This is done so that the model does

not learn every detail in the training set. The ideal model would only learn the most efficient representations of the dataset so that it can be accurately applied to new scenarios.

Experiment and Results

Several important hyperparameters were chosen during the training process through small adjustment and iterative testing. We set our final learning rate at 1e-4 because this rate showed fast convergence without excessive overfitting. A small batch size was chosen to introduce noise which offers a regularizing effect, and also to make training the data less memory intensive. We experimented with different sizes and the best performance was achieved at a batch size of 20.

During the training process, we can see below that the model did not take a linear path to this result, with some noise visible in the validation accuracy and loss. It is in these areas that the SGD with Nesterov Momentum can be seen to be working, bringing the model out of local minima and back in the correct direction, before settling into stability around 50 epochs. Accuracy of the model is defined as how many images are fitted into the correct classes, namely the "true positives."

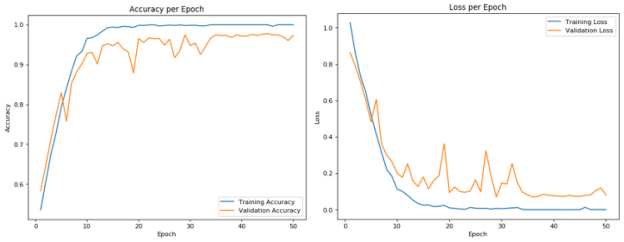
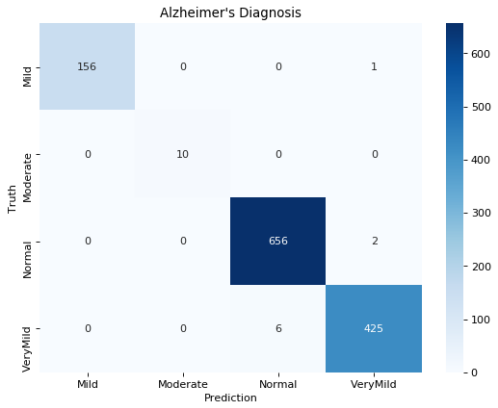


Figure 3: Training/Testing Accuracy and Loss

Our final results can be seen in figure 6; a validation and testing accuracy of over 99% with a loss of ~0.034. This is an exceptional result that we were very proud of. We used a confusion matrix to see which classes, if any, were being misclassified more. We can see in the matrix that there is almost no misclassification occurring at all, with the few that do occur happening mostly between the "Normal" brains and the "Very Mildly Demented" brains. This result makes sense intuitively as the distinction between normal and mild dementia is the smallest relative to each of the other classes.

Conclusion and Future Work

In this report, we utilized the VGG16 baseline model for detection of Alzheimer's from MRI scans. VGG16, despite being originally trained on ImageNet dataset, can be repurposed for disease recognition by transfer learning the network on a dataset consisting of MRI images. To address the problem of limited dataset, augmentation techniques on the dataset are used such as random zooming and brightness variation. Since the dataset consists of MRI from slices of different brain regions, the testing set is modified to include samples of different brain slices. Performing training on the augmented data was able to achieve greater than 99% test accuracy with very small misclassifications on normal and very mild demented images.




```
138/138 [==============================] - 278s 2s/step - loss: 0.0305 - accuracy: 0.9927
20/20 [==============================] - 38s 2s/step - loss: 0.0190 - accuracy: 0.9920
40/40 [==============================] - 81s 2s/step - loss: 0.0339 - accuracy: 0.9928
Train Accuracy: 99.27%
Validation Accuracy: 99.20%
Test Accuracy: 99.28%
```

Figure 4: Confusion Matrix and Accuracy

Our current methodology performs inference over the entire MRI scan rather than only specific regions in the brain. The excellent results produced by this method has supported the hypothesis that features of Alzheimer's diseases are not discerned in specific brain regions but rather aggregate and hidden features present throughout the brain.

For future work, the team would like to explore feature extraction on select brain regions. More data from more patients are also required in the future to ensure the accuracy of the model stays consistently high. Several methods in literature are used to address the problem of limited training data which the team would like to explore more in the future.

Contributions

All team members performed well, with rigid roles not being adopted. We communicated and coordinated almost daily to ensure that our project was meeting milestone-deadlines which we set for ourselves. Each member had a significant hand in all aspects of the project. With that said, some primary roles are outlined below:

- Bryce Smith worked primarily on data processing and model training
- Zachary Burns worked primarily on data processing and model training
- Derrick Cosmas worked primarily on literature review and document writing

Critique Responses

Critiques by group 43:

- *Group 52 did a good job of covering the history and background for their specific problem. The rationale for why they were pursuing this problem was well developed and their thought process was clearly laid out and easy to understand. I especially appreciated the discussion of other deep learning methods and the mention of the complex relationship between anatomical structure and disease. In your presentation, the difficulty in differentiating Alzheimer's disease and Mild Cognitive Impairment was touched upon. Were you or have any research groups been able to find any visual demarcation to differentiate the two? If not, would it make more sense to exclude those images from the data set?*

Response: We did not find an exact visual demarcation between MCI and AD. As of today, researchers are still looking at the exact differences between MCI and AD from looking purely at MRI scans. Current dataset is composed of 4 classes, Normal, Very Mild, Mild and Moderate demented. Based on literature review, we treat Very Mild cases as similar to MCI cases. For the purpose of this project, MCI cases are referred to as Very Mild demented cases.

- *We also had a few questions about the dataset discussed in the presentation. You mentioned that slices were taken from different regions of the brain, where were the slices located and what was their significance?*

Response: MRI slices are located at different heights of the brain beginning from the base of the brain to upper areas of the brain. This is an example of a MRI slice from lower and upper areas of the brain

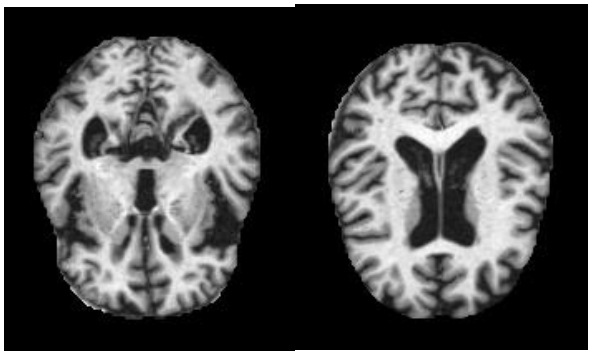


Figure 7: MRI Scans (Lower Brain Left, Upper Brain Right)

- *Similarly, would this model benefit from having more 'redundant data', such as a cut in a different orientation/planes?*

Response: We felt that additional data with different MRI slices at different brain orientations would not be useful to our current endeavor. First, the test data available to use are brain scans of the orientation looking at the brain from top. Other types of MRI scans at different orientations can be made available by scanning from the side view of the brain. However, since the orientations of the test data are only from one view, adding side view information would not be useful for current methodology.

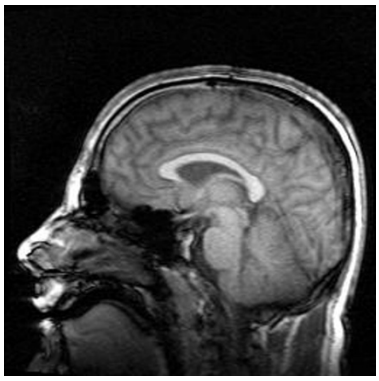


Fig. 8 MRI Brain Scan from Side View[17]

Second, unlike traditional object classification tasks in a scene, there is no issue of occlusion or translation present in the problem. All MRI scans follow certain medical standards in the orientation and look because MRI machines are set up in certain ways and consistently produce scans with the same orientation. Hence, neither translation nor rotation are performed on training data for augmentation because all test cases follow the same orientation as training case.

- *Pertaining to the preprocessing of the data, I would be interested in knowing if all data was taken from the same model MRI, and whether the pixel sizes matched well across images.*

Response: Currently, there is no way of 100% certainty in knowing whether the images are taken from the same model of MRI machine. With regards to pixel sizes, all are matched at sizes of 176x208.

- *Early in the presentation, MRI image segmentation was mentioned, or at least a picture showing it can be seen in slide 3. Would it make sense to feed the model only regions of interest, or is the whole brain scan more useful? I can see both methods having merit.*

Response: Literature review mentioned that selective regions of interest of the brain can be useful in predicting Alzheimer's. At the same time, another hypothesis is present as an argument against using specific regions because Alzheimer's is not indicated by deformation of specific regions. We decided that it is more recommended in using the powerful capabilities of a deep network in collecting the aggregate information present throughout the brain. Our excellent results of high test accuracy definitely supported the latter hypothesis. We can foresee feeding the model images of specific brain regions to produce interesting results, but at the moment inference over the entire MRI Scan is preferred.

- *For the model, transfer learning was a very good and time efficient method that we did not think of for our group. Could you guys elaborate on why those specific weights were chosen. Were they trained on Alzheimer's data previously? We may have missed the background to this design choice. It seems that using a model pre trained on a task related to brain imaging may have enhanced your results.*

Response: Model pre trained on brain scans can definitely enhance the results or at least shorten the training process. However, our work presented that VGG16 pretrained on ImageNet is shown to be versatile enough that transfer learning can be performed on the baseline model for MRI classification.

- *The final proposed model using multistep learning sounded very promising. Finally, the results were cohesive and made sense for the circumstance of the model. The confusion matrix is a good way of demonstrating the results. For the final model, including a confusion matrix for another network trying to achieve the same goal would be helpful in quantifying your success.*

Response: Indeed the confusion matrix is highly useful for multiclass classification tasks. It enables the team to identify the areas where classification is failing and eventually fix.

- I look forward to seeing how well the final model performs and agree with many of the conclusions group 52 made. Overall, the presentation was well put together and the group did a great job presenting their findings.

Critiques by group 58:

- *The choice of the topic Early Detection of Alzheimer Through Deep Learning Analysis of MRI Images is well presented. The presentation is precise. The potential models discussed are present-day. The incremental advantages over the conventional Machine Learning was explained deftly and is possible because of the additional datasets. Since our topics align, we understand the importance of the dataset analysis. Explaining/Studying Datasets a little more would have been more insightful.*

Response: Thank you for your thoughtful consideration. We hope that our exploration of the dataset is more thorough in this final report.

- *Experimenting with different hyper-parameters, focusing to prevent overfitting and randomizing the model seems fair. Adding/Removing the FC output layers and relating the consequences of the results with the dataset features will probably give us constructive feedback to improve the model. All in all, the presentation was precise and the points are driven home in a very clear way. All the best!*

Response: Thank you again. Removing the fully connected output layers had not occurred to us, as we thought this would hamper classification. We will try that and see if we can further improve our model.

Critiques by group 80.

- *Group 52 completely demonstrated an Alzheimer detection algorithm based on the VGG16 model. The project's motivation and background is well organized and introduced. They have a very clear goal and compare different models' tradeoffs before applying deep learning method to achieve their goal.*

Response: Thank you for the positive assessment.

- *For the data preprocessing part, since the hippocampus volume is a critical Alzheime feature related, did the slices cut from different directions or depths of the brain affect the hippocampus region inside the MRI image and finally affect the final prediction results?*

Response: That is a very interesting question. For our experiments we used the entire training set so never trained only on certain regions of the brain. However, our model was able to correctly predict images in the validation and test datasets that were at many different slice locations so we doubt that there is an effect. If there was, we should see a larger amount of misclassification, but we see almost none in our final result.

- *Group 52 did an impressive job on image classification based on VGG16, but since the algorithm only needs to distinguish four groups of images, how does the VGG model compare to other traditional classification methods like k-means, etc.?*

Response: This is interesting to consider. We did not compare VGG16 with "traditional" methods. Given more time we could have tried this to compare. However, we doubt that a method such as k-means would be as effective. K-means is a relatively simple method that just compares how similar a new input is to the mean of the classifications. A convolutional neural network is much more powerful and learns spatial features very well. Also, the four images classifications are visually very similar so even 4 classifications is relatively challenging.

- *Although the dataset and labels are from Kaggle, it is still not well distributed, some categories have significantly more data than others. From the results maybe 3 classes rather than 4 can help improve the results?*

Response: This is something we considered. For example, in our video we consider going to a binary classifier first to see if we can get better results as there will not be as large class imbalances. However, we did not do this as we first discovered the discrepancy between the train and test sets and fixed it. After this change we got very good results, so we did not need to go to fewer classifications.

- *Transfer learning part is fascinating. How did you get the intuition?*

Response: We were inspired in two parts to do this. One source was the online Stanford lectures where transfer learning is mentioned. Another was Brian Whiteaker who suggested we look into it when we asked for help after trying to train a VGG16 network from scratch.

Libraries

Scikit-learn
Tensorflow-Keras
Numpy
Matplotlib

References

1. "Why Get Checked?" *Alzheimer's Association*, 2020,
   https://www.alz.org/alzheimers-dementia/diagnosis/why-get-checked
   https://pubmed.ncbi.nlm.nih.gov/7164002/#:~:text=The%20breathing%20pattern%20during%20all,of%20the%20level%20during%20wakefulness.

2. Johnson, Fox, Sperling, Klunk. "Brain Imaging and Alzheimer's Disease" *Cold Spring Harbor Perspectives in Medicine,* Apr 2012,
   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3312396/

3. Visser, P.J., Scheltens, P., Verhey, F.R.J., Schmand, B., Launer, L.J., Jolles, J., Jonker, C., 1999. Medial temporal lobe atrophy and memory dysfunction as predictors for dementia in subjects with mild cognitive impairment. J. Neurol. 246, 477–485.

4. Rathore, Habes, Iftikhar, Shacklett, Davatzikos. "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages" *NeuroImage, 15, Jul, 2018,*
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5511557/

5. Csernansky, J.G., Wang, L., Swank, J., Miller, J.P., Gado, M., McKeel, D., et al., 2005. Preclinical detection of AD: hippocampal shape and volume predict dementia onset in the elderly. Neuroimage 25, 783–792.

6. Frisoni, Fox, Jack, Scheltens, Thompson. "The Clinical Use of Structural MRI in Alzheimer Disease" *Nature Reviews Neurology, Feb, 2010*
https://www.nature.com/articles/nrneurol.2009.215

7. Plant, Teipel, Oswald, Bohm, Meindl, Mourao-Miranda, Bokde, Hampel, Ewers. "Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease" *NeuroImage, Mar, 2010,*
https://www.sciencedirect.com/science/article/pii/S1053811909012312?via%3Dihub

8. S. Wang, H. Wang, Y. Shen and X. Wang, "Automatic Recognition of Mild Cognitive Impairment and Alzheimers Disease Using Ensemble based 3D Densely Connected Convolutional Networks," *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, 2018, pp. 517-523, doi: 10.1109/ICMLA.2018.00083.

9. Dubey, Sarvesh. "Alzheimer's Dataset (4 Class of Image)" *Kaggle, Dec, 2019,*
https://www.kaggle.com/tourist55/alzheimers-dataset-4-class-of-images

10. Kloper, Neil. "Weekly Kaggle-Task Wiki!" *Kaggle 21, Mar, 2020,*
https://github.com/Neil-Kloper/Weekly-Kaggle-Task/wiki

11. Khandelwal, Renu. "Overview of Different Optimizers for Neural Networks" Medium, Feb, 2019,
https://medium.com/datadriveninvestor/overview-of-different-optimizers-for-neural-networks-e0ed119440c3

12. Huang, G., et al. "Densely Connected Convolutional Networks" arXiv (2016)

13. Szegedy, C. et al "Rethinking the Inception Architecture for Computer Vision" arXiv (2015)

14. Shaikh, F. "10 Advanced Deep Learning Architectures Data Scientists Should Know!" analyticsvidhya.com (2017)

15. Szegedy, C., et al. "Going Deeper with Convolutions" CVPR (2015).

16. Kumar, A. "Machine Learning - How to Diagnose Underfitting/Overfitting of Learning Algorithm" (2015) https://vitalflux.com/machine-learning-diagnose-underfittingoverfitting-learning-algorithm/

17. Albrow, M. “Using tech to peek inside your body” (2018) https://news.fnal.gov/2018/02/seeing-inside-your-body/

18. Simonyan, K., Zisserman, A. “Very Deep Convolutional Networks for Large-Scale Image Recognition” arXiv (2014).