**Predictive Classification for Alzheimer Disease**

Aaron Ramirez, Jaymin Patel, and Geoffrey Fadera

University of San Diego

Masters of Applied Artificial Intelligence

AAI-501: Introduction to Artificial Intelligence

Professor David Friesen

April 15 2024

**Background on Alzheimer Disease and Project Goal**

Alzheimer’s disease is a neurodegenerative disorder that affects cognitive functions like memory and behavior. It is mostly commonly caused by dementia. The cause of this disease is not fully understood but believed to be a factors of genetics, age and lifestyle factors (Alzheimer's disease fact sheet | National Institute on Aging 2023). Symptoms of this disease gradually worse to where memory loss can happen every couple of minutes, having difficulty completing simple tasks and have immediate mood changes. There is no cure for Alzheimer’s disease, the best there is, is medication to help manage symptoms. Early detection is vital to combating this disease. Finding it early can help create the right treatment plan to slow down the effects and help plan for the future of the disease.

Knowing the severity of this disease, this project aims to us AI techniques and algorithms to create a model to be able to detect for early signs of Alzheimer’s and determine the severity of the disease. There are four categories it could fall into, Non-Demented, Mild Demented, Very Mild Demented and Moderate Demented. There were a couple of models tested and evaluated against each other including, Convolution Neural Network (CNN), Deep Neural Network (DNN), CNN with graph-based features, K-nearest Neighbors (KNN), support vector machine SVM, Random Forest Classifier (RFC), RFC with Histogram of Gradient (HOG) feature extraction, and recurrent neural network (RNN).

**Related work**

This problem has been around for years, and scientists have tried to find solutions that work. There are a couple of papers out there that have proven models. Mr. Irshad Ahmad and his team did scientific research at Jouf University and were able to prove to achieve 99.35% recognition using principal component analysis with stepwise linear discriminant analysis and an artificial neural network (Ahmad et al., 2023). There was another team from China that proved prediction from 2.5D MRI images with around 80% accuracy (Lin et al., 2018). Finally, there is A.M El-Assy and his team from Mansoura Egypt who proved over 99% categorize and predict Alzheimer’s Disease from the MRI images (El-Assy et al., 2024). Overall, this proves the research for this problem is moving forward in the right direction and hopefully there will be a full proof method that follows.

**Data Set and Data Splitting**

For the testing that was done by our team, we used a Kaggle dataset that had a collection of MRI images from various websites. The data had four categories that could be use for classification previously stated. For the split, the data set had 6400 images that were split into test and train already for us. About 20% of the images were used for testing and the rest for training.

**Data Preprocessing**

For preprocessing, each model had specific requirements for their own but there was a general process created that each model started with. For the general process, the images were grayscale to create an even playing field. From there, each image was resized to the target size of 224 by 224 and performed thresholding and connected component analysis to remove background black space. This led it to where the brain outline would be touching the boundary of the image. The pictures also normalized the pixels to allow for more effective training. This was applied to both the Training data and the Testing data. The results at the end of this was having each image turned into arrays and ready to be used for our AI models. The shapes of the models were:

Train

Image shape: (5121, 224, 224)

Labels shape: (5121,)

Test

Image shape: (1279, 224, 224)

Labels shape: (1279,)

**Metrics definition**

There are 4 metrics we will be looking at, accuracy, precision, recall and f1-score.

* Accuracy: This represents the percentage of prediction that were correct
* Precision: This is defines as the accuracy of the specific classifier. This measure the proportion of predicted true positive among all the choices that got in that classifier.
* Recall: This is the sensitivity and is used to evaluate all the positives found within the classifiers (including false negatives). It answer the question of how many actual psotives did the model correctly identify.
* f1-score: This combines precision and recall for a balanced measure of performance.

**Models with metrics**

**CNN**

Some additional preprocessing was required. The labels were hot encoded for easier categorization. The model was sequential and 3 2D convolutional layers the was followed with Batch Normalization. Batch Normalization normalize the activation of the convolution layers to accelerate the training. Next the max pooling layers helped downsample and take only the maximum value within each region. From there the data was flattend into a 1D vector to be connected to the fully connected layer. This collapses the spatial dimensions of the image into a single dimension. The first the fully connected/dense layer start formulating the classification. Next is the dropout layer which helped prevent overfitting and finally the last dense layer will do the prediction.

To enable better learning we used the Adam optimizer and categorical cross entropy loss. We also included early stopping to stop training if validation does not improve. When training we also tested both Cyclic Learning Rate which tries to create a fluid learning rate as well as exponential decay learning. We found exponential did better with keeping a consistent model when discussing validation loss and validation accuracy increase.

Test Loss: 1.3505847454071045

Test Accuracy: 0.7208756804466248

precision recall f1-score support

MildDemented 0.69 0.45 0.55 179

ModerateDemented 1.00 0.33 0.50 12

NonDemented 0.77 0.81 0.79 640

VeryMildDemented 0.66 0.71 0.68 448

accuracy 0.72 1279

macro avg 0.78 0.58 0.63 1279

weighted avg 0.72 0.72 0.72 1279

**DNN**

Similar to CNN the images were train to include batch\_size and preserve the original batch. The model was sequential and started with flattening the data and then having two dense layers tied with batch normalization. Dropout was included but with a lower rate. We had the same optimize as well as cross entropy and exponential decay learning scheduler. The new addition that helped improve this model was an Image Data Generator. This was used to augment parameters of the picture and have more random shift in the features to help capture better learning.

Test Loss: 1.1579606533050537

Test Accuracy: 0.5215011835098267

precision recall f1-score support

MildDemented 0.00 0.00 0.00 179

ModerateDemented 0.00 0.00 0.00 12

NonDemented 0.62 0.69 0.65 640

VeryMildDemented 0.40 0.50 0.44 448

accuracy 0.52 1279

macro avg 0.25 0.30 0.27 1279

weighted avg 0.45 0.52 0.48 1279

**CNN with Graph base features**

The images are preprocessed by putting the images into batches and processed with the mobilenet\_v2 model set to better be used for feature extract. Next the new processed images are put our feature extraction function to pull the extracted features into a list and have principal component analysis applied to simplify the complexity of the features. The features list is then put through the compute feature similarity function to get the similarity between the features and vectors using cosine. Next we construct the adjacency matrix on the feature similarity. We put both our test and training set through this process. We also added the grayscale index for the model.

The model itself is a CNN model very similar to the previous stated CNN model. The difference is there is an image input, graph input and adjacent input as well as a concatenate flattened layer before fully connecting. Cyclic learning rate scheduler is also applied as it resulted in better training results.

Test Loss: 1.9398

Test accuracy: 0.6450

precision recall f1-score support

MildDemented 0.51 0.39 0.44 179

ModerateDemented 1.00 0.25 0.40 12

NonDemented 0.65 0.91 0.76 640

VeryMildDemented 0.70 0.38 0.50 448

accuracy 0.65 1279

macro avg 0.72 0.48 0.52 1279

weighted avg 0.65 0.65 0.62 1279

**KNN**

The additional preprocessing, the images are flattened for compatibility. For KNN it is tested for 3 neighbors and for neighbor ranges from 1 – 14 and report back the results. The way it works is classify by majority votes to its X amount of neighbor. The test with 3 neighbors proved superior compared to grouping 14 variations of neighbors

Test Accuracy: 0.73

precision recall f1-score support

MildDemented 0.64 0.59 0.61 179

ModerateDemented 0.50 0.58 0.54 12

NonDemented 0.79 0.81 0.80 640

VeryMildDemented 0.69 0.68 0.68 448

accuracy 0.73 1279

macro avg 0.66 0.67 0.66 1279

weighted avg 0.73 0.73 0.73 1279

**SVM**

No additional preprocessing. The regularization parameter set to one as this was found to be the optimal hyperplane separation for best results.

Accuracy: 0.6387802971071149

Classification Report:

precision recall f1-score support

MildDemented 0.75 0.30 0.42 179

ModerateDemented 1.00 0.33 0.50 12

NonDemented 0.67 0.80 0.73 640

VeryMildDemented 0.57 0.56 0.56 448

accuracy 0.64 1279

macro avg 0.75 0.50 0.55 1279

weighted avg 0.65 0.64 0.62 1279

**RFC**

For RFC the data created class weights to try and balance the data since Moderate Demented has very little data. RFC will iterate through each parameter combination and record the performance. From there it will create classifiers and try to fit by using the training data. It will calculate the accuracy and then repeat the process trying to find the best validation accuracy. There was testing with contrasting equalization as well as Histogram of Gradient (HOG). Both provided subpar results compared to traditional RFC. The HOG performed best with a cell 40x40 but still was considered subpar.

Validation Accuracy: 0.6778733385457388

Classification Report:

precision recall f1-score support

MildDemented 1.00 0.12 0.22 179

ModerateDemented 0.00 0.00 0.00 12

NonDemented 0.74 0.85 0.79 640

VeryMildDemented 0.57 0.67 0.62 448

accuracy 0.68 1279

macro avg 0.58 0.41 0.41 1279

weighted avg 0.71 0.68 0.64 1279

**RNN**

The labels are encoded into integers and class weights are added again. The gate recurrent layer is 128 unites and has batch normalization tied right after. There is a dropout to try and improve learning but with only .2. The model is complied with Adam optimizer and uses sparse categorical cross entropy due to how we process the labels. The model was also tested with no class weight and the results between the two are so similar that the difference is negligible. Below are the results without the class weights.

Validation Accuracy: 0.68

precision recall f1-score support

MildDemented 0.74 0.36 0.49 179

ModerateDemented 1.00 0.25 0.40 12

NonDemented 0.70 0.85 0.77 640

VeryMildDemented 0.65 0.58 0.62 448

accuracy 0.68 1279

macro avg 0.77 0.51 0.57 1279

weighted avg 0.69 0.68 0.67 1279

**Results**

Comparing all the results and metrics below is the order of best performance:

1. **K-Nearest Neighbors (KNN):**
   * Test Accuracy: 0.73
   * Precision, Recall, and F1-score: Balanced performance across all classes.
   * KNN shows good performance in predicting Alzheimer's disease.
2. **CNN:**
   * Test Accuracy: 0.72
   * Precision, Recall, and F1-score: Moderate performance with a slight emphasis on NonDemented and VeryMildDemented classes.
   * CNN demonstrates decent performance in identifying Alzheimer's disease from MRI scans.
3. **CNN with Graph:**
   * Test Accuracy: 0.65
   * Precision, Recall, and F1-score: Moderate performance across all classes.
   * CNN with Graph provides reasonable accuracy in classifying Alzheimer's disease.
4. **Recurrent Neural Network (RNN):**
   * Validation Accuracy: 0.68
   * Precision, Recall, and F1-score: Balanced performance, but slightly lower recall for MildDemented class.
   * RNN shows promising results but could be further optimized for better accuracy.
5. **Random Forest Classifier (RFC):**
   * Validation Accuracy: 0.68
   * Precision, Recall, and F1-score: Moderate performance, but struggles in predicting the ModerateDemented class.
   * RFC shows acceptable performance but needs improvement in predicting certain classes.
6. **Support Vector Machine (SVM):**
   * Accuracy: 0.64
   * Precision, Recall, and F1-score: Moderate performance with an emphasis on NonDemented class.
   * SVM performs reasonably well but could improve in predicting other classes.
7. **DNN:**
   * Test Accuracy: 0.52
   * Precision, Recall, and F1-score: Poor performance, especially in predicting the MildDemented and ModerateDemented classes.
   * DNN performs inadequately compared to other models and requires significant improvement.

Overall KNN exhibits best performance with CNN with graph features following. Further finetuning and optimization are still needed for all models to improve their performance. None are perfect models yet.

**Conclusion and Continue Work**

In summary, this research went through 7 different models to try and find the best classification method for MRI images that could be showing demented symptoms. The results showed promising support for KNN and both CNNs. RNN showed promise too but work needs to be done to improve the recall on mild demented and getting scored properly. The next steps would be fine tuning KNN and both CNNs. For KNN there can be work done to do more testing on distance metrics and feature scaling. For CNN with graph features, this can be converted into a GCN. After that it would need to be tuned with its hyperparameters and layers. For CNN There are multiple ways to improve it. Looking at research outside the scope of this project, there are models that are ensembling two or more CNNs with different hyperparameters and metrics that are able to get very effective results. Tuning on any of these will continue to improve these models. These models will also need to be validated with bigger datasets. Once both of these are complete, this can help contribute to research for Alzheimer’s disease detection and start exploring clinical settings.

**References**

U.S. Department of Health and Human Services. (2023). *Alzheimer’s disease fact sheet*. National Institute on Aging. https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet

Ahmad, I., Siddiqi, M. H., Alhujaili, S. F., & Alrowaili, Z. A. (2023). Improving alzheimer’s disease classification in Brain MRI images using a neural network model enhanced with PCA and Swlda. *Healthcare*, *11*(18), 2551. <https://doi.org/10.3390/healthcare11182551>

Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., Guo, G., Xiao, M., Du, M., & Qu, X. (2018). Convolutional neural networks-based MRI image analysis for the alzheimer’s disease prediction from mild cognitive impairment. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00777

El-Assy, A. M., Amer, H. M., Ibrahim, H. M., & Mohamed, M. A. (2024b). A novel CNN Architecture for accurate early detection and classification of alzheimer’s disease using MRI Data. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-53733-6

Dubey, S. (2019, December 26). *Alzheimer’s dataset ( 4 class of images)*. Kaggle. https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images