Classification of Alzheimer's Disease Severity Using fMRI Images

Laura Banham

Problem Statement

Alzheimer's disease is a geriatric illness characterized by increasingly debilitating memory loss. Recent work has turned to high-quality brain scan technology (fMRI) to investigate the specific areas of the brain that are affected by this disease. In order to pinpoint the brain regions that are affected by Alzheimer's, fMRI scans of healthy and diseased individuals are taken and then compared for differences. Although fMRI produces high resolution observations, analysis of these scans is complicated by the fact that individuals can have slightly different patterns in their brain scans due to irrelevant causes (e.g., other illnesses), creating noise in the images. However, diseased and healthy people as a group have distinct signatures that can be identified by coalescing across individuals within each category. Once these signatures are identified, they can then be used to diagnose early stages of Alzheimer's disease in future patients.

Data Source and Structure

The dataset for this project came from Kaggle [1]. The images were collected from multiple websites. This dataset is composed of 6400 grayscale images (split into train and test datasets) from four groups of individuals with varying levels of dementia (number in parenthesis denotes number of images in each dataset): non-demented (2069 train, 491 test), very mildly demented (1425 train, 367 test), mildly demented (560 train, 157 test), and moderately demented (42 train, 10 test). Each image is made up of 208 x 176 pixels. Averaged images for each class are provided in Figure 1. From these images and the brain sections labelled in Figure 3, it appears that the size of the lateral ventricle is larger for more severely demented individuals.

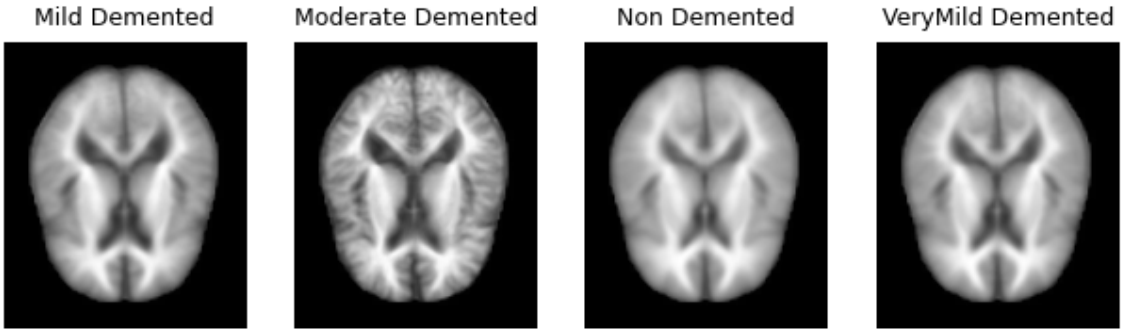

Figure 1. Averaged fMRI scans from each class.

An image representation of the standard deviations of the scans within each class are provided in Figure 2. Lighter sections of the standard deviation images indicate higher variability within the class. Unsurprisingly, from the brain sections labelled in Figure 3, we can see that the areas surrounding the lateral ventricle and the atrium of the lateral ventricle show the most variability within the groups.

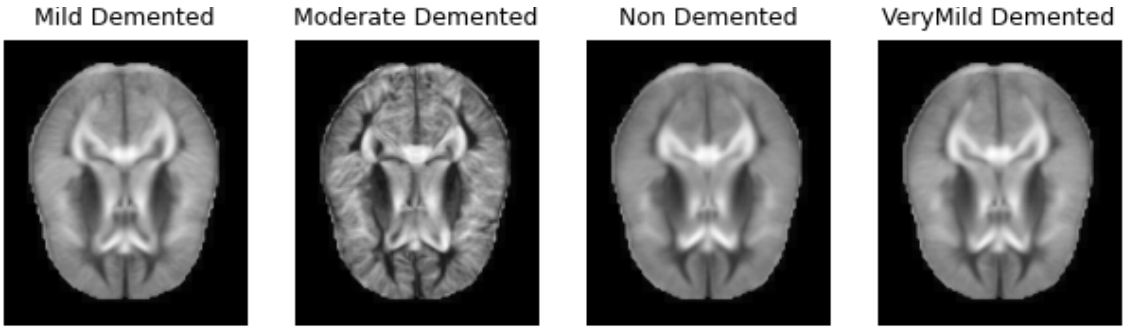

Figure 2. Standard deviation of fMRI scans from each class. Note: Lighter colors indicate higher variation within the class.

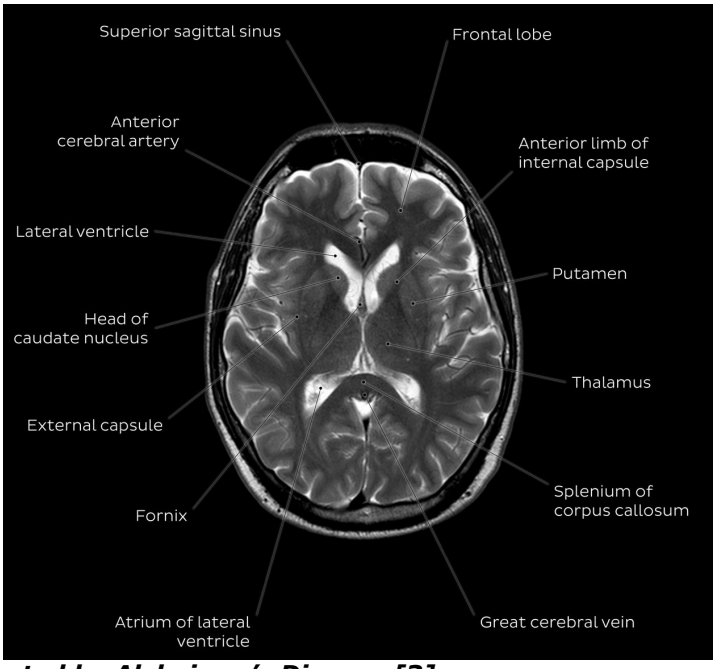

Figure 3. Brain regions affected by Alzheimer's Disease [3].

Confirming previous observations, an image representation of the difference between moderately demented and non-demented individuals (see Figure 4) clearly highlights the role of the lateral ventricle in signaling dementia.

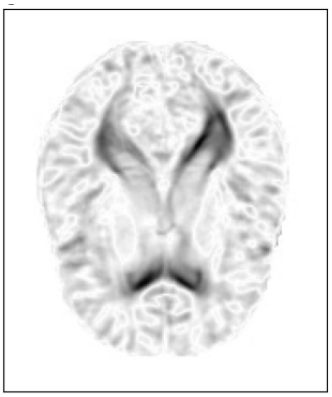

Figure 4. Difference between average images of moderately and non-demented classes. Note: Darker colors indicate larger differences.

Methodology
The focal analysis for this project is to classify brain scan images by dementia level (multiclass classification). In addition to fitting a classification algorithm, several steps will be taken to collect and prepare the dataset. All data cleaning, preparation and modeling will be conducted in Python. Due to the time cost associated with the computation, I may only use a subset of the full dataset for the final analysis. An outline of the steps for the full analysis is provided below:

1. Read the training and testing images into two large, labeled datasets.
2. Use cross-validation to select optimal hyperparameters for each model (described in Step 3).
3. Fit several classification algorithms with the optimal hyperparameters from Step 2 on the training dataset:
    a. *Multinomial Logistic Regression*: This method extends the popular Logistic Regression (i.e., Logit) model to three or more classes. It uses a linear combination of weighted factors to estimate the probability of an observation belonging to a specific class.
    b. *Decision Tree*: This model sequentially branches the dataset into binary decision groups using influential factors. The final decision tree is then used to predict future classes.
    c. *Naïve Bayes:* This model assumes that the features are independently distributed and then uses the distribution of the given observations to determines the likelihood that a future observation belongs to each class. The predicted class is assumed to be the class that has the greatest likelihood of being correct.
    d. *K Nearest Neighbors (KNN*): This method uses the data points nearest to a new data point (in the feature space) to determine the most likely class of the new observation.
    e. *Random Forest*: This ensemble method extends the decision tree model above building decision trees and then averaging them to create a final, higher performing classifier. The class of a future observation is assumed to be the class predicted by the majority of the decision trees constructed.
    f. *Adaptive Boosting*: This ensemble method sequentially constructs a linear combination of models using the misclassified observations in prior models to weight the training dataset of subsequent models.

Model Evaluation and Results
After fitting each model, I calculated the *accuracy* (the proportion of all cases that were classified correctly), *precision* (the proportion of the predicted positive cases that were correctly classified as positive; since this is a multiclass model, precision will be calculated and reported as the average of the precision for each class), *recall* (the proportion of the true positive cases that were correctly calculated as positive; as noted above the reported value will be the average recall across the four classes), and *f1 score* (the harmonic mean of precision and recall) of each using the test dataset. Results are displayed in Table 1 and Figure 5 below.

Model	**Train Accuracy**	**Test Accuracy**	**Precision**	**Recall**	**F1-Score**
Logistic Regression	1.0000	0.9824	0.98	0.98	0.98
SVM	1.0000	0.9844	0.98	0.98	0.98

KNN	1.0000	0.9980	1.00	1.00	1.00
Naïve Bayes	0.5083	0.4829	0.51	0.48	0.48
Decision Tree	1.0000	0.7278	0.72	0.73	0.72
Random Forest	1.0000	0.9083	0.91	0.91	0.91
AdaBoost (Logit)	1.0000	0.9805	0.98	0.98	0.98
*AdaBoost (Logit, SVM)**	1.0000	0.9844	0.98	0.98	0.98

Table 1. Metrics for fitted models. *This model added after evaluating previous models.

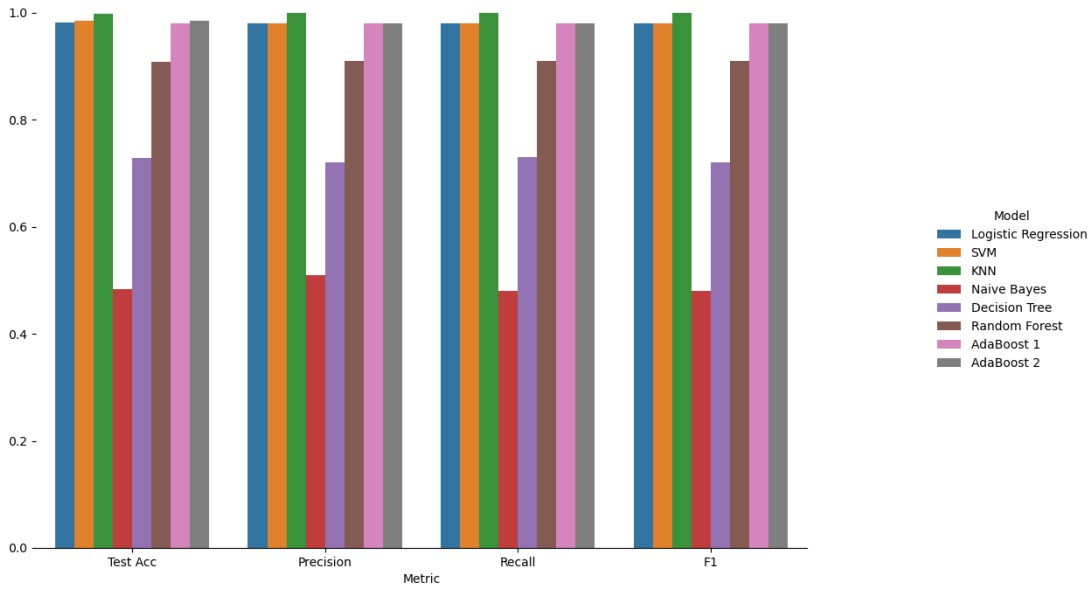

Figure 5. Plot of metrics for fitted models.

From Table 1 and Figure 5, it is evident that the Naïve Bayes model performed extremely poorly. This is likely because the Naïve Bayes algorithm assumes independence among the input features, which is not the case when the input features are pixels in the same region of a brain scan (logically, pixels that are near each other should be somewhat correlated). The tree-based approaches (Decision Tree and Random Forest) also performed relatively poorly which indicates that treating each pixel independently (and making binary decisions based on each one) does not improve performance. In addition, one can clearly see that the KNN and SVM models performed very well. Both models utilize a spatial (e.g., Euclidean distance) framework for classification. This result is somewhat intuitive because images that are visually similar to one another should lie close to each other when plotted as a point in high-dimensional space. Interestingly, the Logistic Regression model performed well. This approach uses each feature (pixel) individually and outputs a probability of belonging to a class. Similar to the Naïve Bayes model, the Logit model also assumes that the factors are not correlated (i.e., non-multicollinearity), however, the model performed similarly well to the SVM and KNN models which is surprising.

Using the results of the first seven models, I then built an eighth model that was a mixture of the Logistic Regression and SVM models to see if this could provide a better prediction than either of them individually. The new AdaBoost model only performed as well as the SVM model alone, however.

Conclusions

Limitations

The current project was primarily limited by computational power. More complex models (such as a stochastic gradient boosting model or a neural network model) were not fitted due to a lack of hardware that could run such computationally expensive algorithms.

Future Directions

The KNN model was very effective in predicting dementia, however, due to the structure of the model it cannot be incorporated into a AdaBoost algorithm. Recent work has attempted to capture the strengths of both types of algorithms by fitting a AdaBoost algorithms on the training dataset and then using a KNN algorithm to make predictions from those AdaBoost classifiers [4]. This approach combines the high accuracy of the AdaBoost algorithm with the simplicity of the KNN algorithm to create a better model. In addition, further models could investigate combining the statistical approach of the Logit model with the spatial approach of the KNN model to see if a better ensemble model could be created.

References

[1] S. Dubey, "Alzheimer's Dataset", *Kaggle.com*, Dec. 2019.
https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images

[2] E. Byeon, "Exploratory Data Analysis Ideas for Image Classification", *towardsdatascience.com*, Sept. 2020.
https://towardsdatascience.com/exploratory-data-analysis-ideas-for-image-classification-d3fc6bbfb2d2

[3] J. Vasković, "Normal brain MRI", *kenhub.com*, Nov. 2022.
https://www.kenhub.com/en/library/anatomy/normal-brain-mri

[4] Y. Gao, J-y. Pan, and F. Gao, "Improved boosting algorithm through weighted k-nearest neighbors classifier", *2010 3rd International Conference on Computer Science and Information Technology*, pp. 36-40, 2010.
https://ieeexplore.ieee.org/document/5563551/authors#authors