Introduction/Background

Cayman is a clean, responsive theme for GitHub Pages.

View on GitHub Download .zip Download .tar.gz

Introduction/Background

With current research indicating the strengths of various ML methods towards the purpose of neurodegenerative disease diagnosis, such as SVM, random forest classification, and CNN [1], we aim to experiment with various supervised learning techniques in order to accurately detect both Alzheimer's Disease (AD) and Parkinson's Disease (PD), as well as classify the relative progression of either disease.

The open-source dataset. we will be investigating consists of MRI images of brains subdivided into three groups: those with AD, those with PD, and those with no neurodegeneration. These images were sourced from various reputable sources, namely the National Institute of Health (NIH), the Alzheimer's Disease Neuroimaging Initiative (ADNI), and the Parkinson's Progression Markers Initiative (PPMI). Some data augmentation techniques were applied to the PD dataset in order to balance the overall dataset.

The second open-source dataset, consisting of black and white MRI images sourced from the ADNI, separates data into four subcategories: no dementia, very mild dementia, mild dementia, and moderate dementia. This dataset is somewhat unbalanced, as cases of moderate dementia are underrepresented.

Problem Definition

Neurodegenerative diseases are challenging to diagnose, incredibly costly to treat, and are becoming increasingly prevalent. As of 2018, AD, the most frequently occurring neurodegenerative disease, was the sixth leading cause of death in the U.S. [2]. With many economies worldwide struggling to support aging populations, it is estimated that "the number of individuals with AD is likely to reach 115 million worldwide by 2050". This poses a significant economic burden, as the global cost of AD was already nearing $1 trillion USD in 2016 [1].

Presently, no conclusive antemortem method of diagnosis nor effective treatment exists for either AD or PD [3]. Misdiagnosis of these conditions are common, with up to 25% of PD patients having been misdiagnosed during early stages [2]. However, thanks to the rapid development of modern

neuroimaging techniques, some have been able to leverage ML techniques using high dimensional brain imaging data to some success. As such, ML possesses potential in improving the accuracy, accessibility, and efficiency of clinical diagnosis [1].

Methods

Model 1: SVM (Support Vector Machine)

In order to enhance our ML model's performance in differentiating brain MRI scans by impairment, we implemented a preprocessing pipeline that includes feature scaling and dimensionality reduction. First, we flattened each 128x128 grayscale image into a 1D array of 16384 pixel values and normalized the pixel intensities, standardizing each feature to a mean of zero and a standard deviation of one. This reduces biases caused by brightness or contrast variations across images. Following this, we utilized Principal Component Analysis (PCA) to reduce the dimensionality, retaining 95% of the data's variance and effectively reducing the feature space to 1041 components. PCA not only improved computational efficiency, but also minimized noise and redundancy in the features, preventing overfitting. This preprocessing pipeline allowed us to retain critical information while optimizing both training efficiency and model interpretability for future impairment classification.

The method that we implemented for classification for this specific model was a multi-class SVM. SVMs are known to be effective in high-dimensional spaces and employ kernel functions. These two features are useful in our case since MRI scans often have a large number of features. The use of kernel functions were also a benefit for employing an SVM, since kernels can be used to transform non-linearly separable data into a higher dimensional space that can be linearly separable. The kernel we implemented in particular is the radial basis kernel (RBF), which reduced the amount of calculation tasks and simplified the comparison of complex features, allowing for classification in a more reasonable amount of time.

Model 2: Random Forest Classifier

Since decision trees are particularly well-suited to classification tasks, the second model we decided to implement was a Random Forest Classifier. Random Forest Classification is particularly useful when dealing with particularly large and complex datasets, as well as handling high-dimensional feature spaces. It is also highly robust in minimizing overfitting, which can pose a significant challenge during image analysis.

In order to optimize the classifier's performance, we modified the preprocessing steps applied to the first dataset inputted into the SVM model. We still applied PCA for dimensionality reduction, however we opted not to normalize the pixel intensities and instead applied Histogram Equalization, which enhances contrast in an image to ensure that the distribution of pixel

intensities was more uniform. This can be particularly helpful for differentiating certain visual patterns that correlate directly to brain impairments.

Model 3: CNN (Convolutional Neural Network)

We decided to implement a CNN as our third and final model. CNNs perform particularly well with image and visual data, and they're highly effective at detecting patterns in the data. The architecture of a CNN allows for relatively minimal preprocessing compared to the other two models. The convolutional layer(s) in a CNN apply filters to input in order to extract features, and the pooling layer(s) downsamples the input in order to facilitate computation and prevent overfitting. As such, the only preprocessing step we took for the data inputted into the CNN was PCA to reduce the total number of features and improve the time for fitting the model.

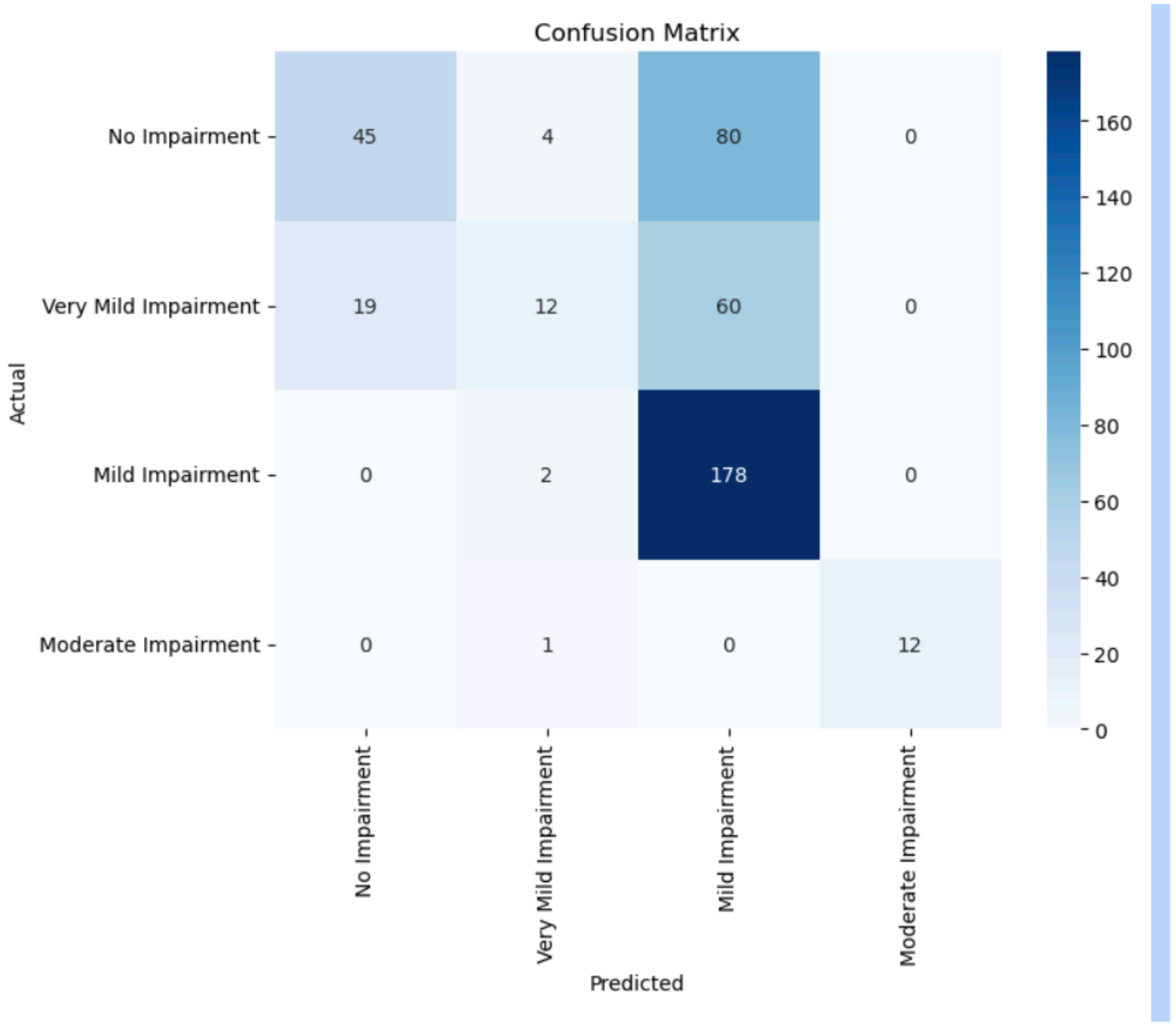
Results and Discussion

Model 1: SVM

Training our model took about 45 minutes for 10 iterations. With optimized parameters, we were able to obtain 59.806% accuracy.

```
The model is 59.806295399515 74% accurate
                    precision    recall  f1-score   support

       No Impairment       0.70      0.35      0.47       129
  Very Mild Impairment       0.63      0.13      0.22        91
      Mild Impairment       0.56      0.99      0.71       180
   Moderate Impairment       1.00      0.92      0.96        13

              accuracy                           0.60       413
             macro avg       0.72      0.60      0.59       413
          weighted avg       0.63      0.60      0.54       413
```

When looking at our confusion matrix, we observed that the best performing category for was Mild Impairment with 178 samples correctly classified out of 180.



The current dataset is imbalanced, which may explain why the images demonstrating very mild or no cognitive impairment were misclassified much more frequently than those displaying mild or moderate impairment. It may also be possible the data has extreme values for a particular label.

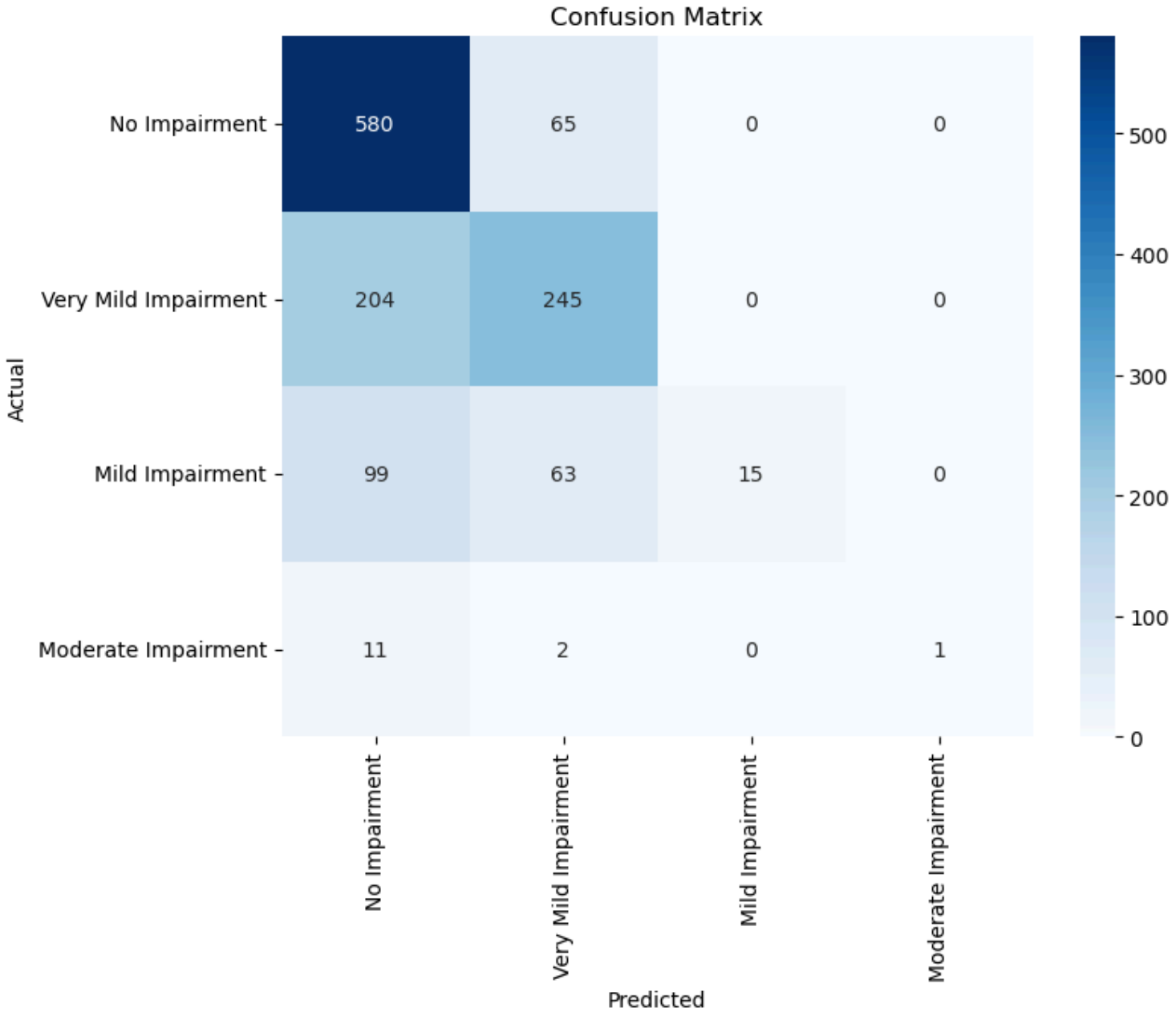
Model 2: Random Forest Classifier

We predicted a Random Forest classifier would perform better than SVM, since decision trees are particularly suited to classification tasks. Our results affirm this prediction: after tuning the hyperparameters, we were able to achieve around a 63% accuracy.

```
The model is 63.50194552529182% accurate
                    precision    recall  f1-score   support

       No Impairment       0.60      0.97      0.74       645
  Very Mild Impairment       0.78      0.40      0.53       449
      Mild Impairment       0.77      0.06      0.11       177
  Moderate Impairment       0.00      0.00      0.00        14

            accuracy                           0.64      1285
           macro avg       0.54      0.36      0.34      1285
        weighted avg       0.68      0.64      0.57      1285
```

In order to further optimize the model's performance, we decided to create a mixture of models, consisting of a Random Forest Classifier, a Gradient Boosting Classifier, and a Logistic Regression model with normalized input. By feeding all of these models into a Voting Classifier, which aggregates the predicted label based on the majority vote (much like a Random Forest Classifier takes the majority vote of a collection of decision trees), we were able to achieve a slightly higher accuracy of 65.44%.

```
The model took this long to train: 0:08:19.34
The model is 65.44747081712062% accurate
                    precision    recall  f1-score   support

       No Impairment       0.65      0.90      0.75       645
  Very Mild Impairment       0.65      0.55      0.59       449
      Mild Impairment       1.00      0.08      0.16       177
  Moderate Impairment       1.00      0.07      0.13        14

            accuracy                           0.65      1285
           macro avg       0.83      0.40      0.41      1285
        weighted avg       0.70      0.65      0.61      1285
```

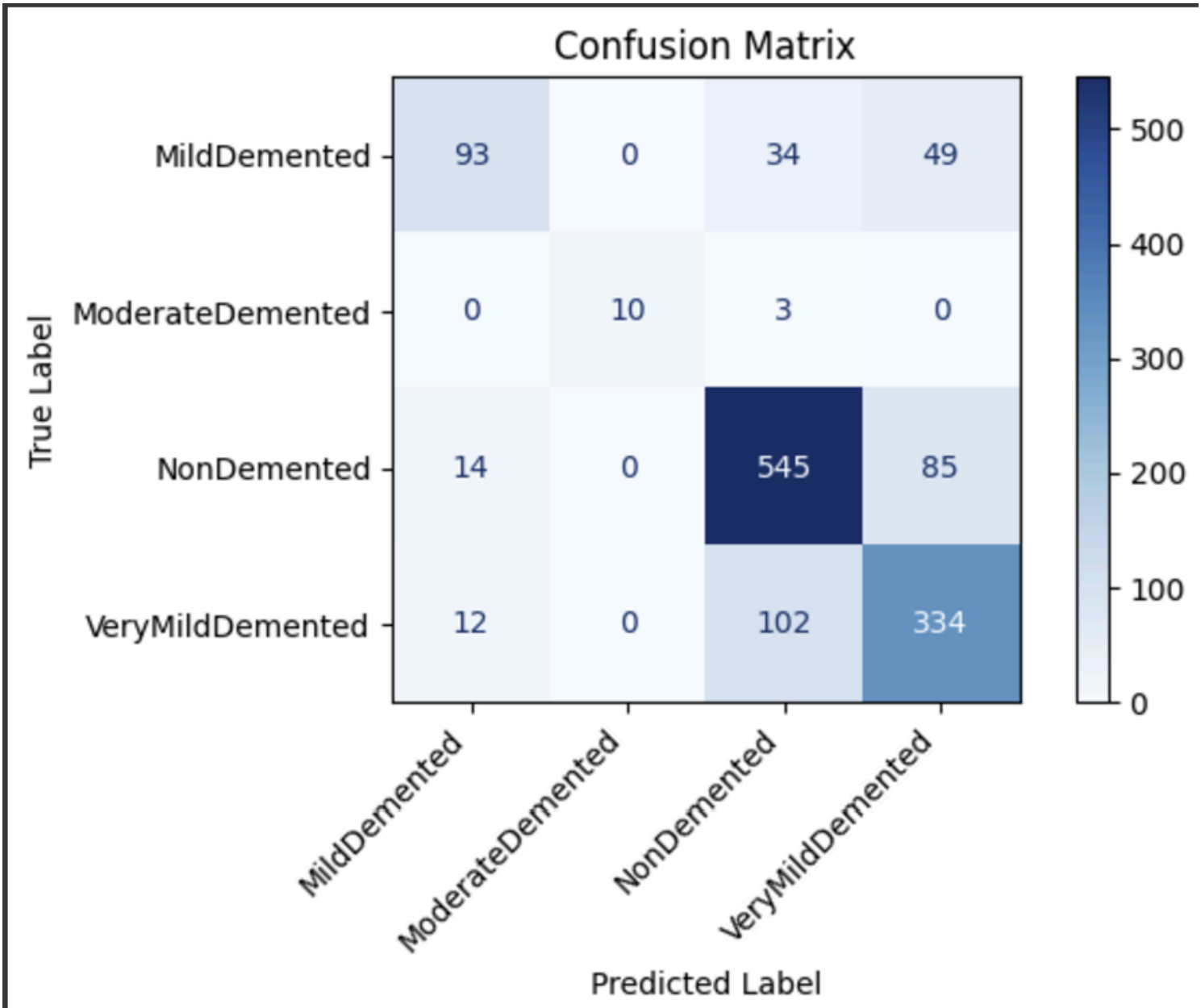


Based on the confusion matrix, the mixture model best classified images illustrating no impairment, but struggled to accurately identify mild, very mild, and moderate levels of neurodegeneration. The differences between the categories of very mild and mild neurodegeneration are very subtle and often missed by medical professionals, so the model's underperformance in these two categories aren't necessarily indicative of a particular flaw in the model itself. The model did, however, perform very poorly in identifying cases of moderate impairment, which is likely due to the significant underrepresentation of images displaying moderate neurological impairment within the dataset.

Model 3: CNN

Because CNNs are especially effective in detecting patterns in image data, we expected this model to perform the best. This model achieved a 76% accuracy, outperforming both of our other models.

We implemented the CNN using pytorch and used pytorch's CrossEntropyLoss in placement of softmax as this function automatically softmaxes the output. After testing with different convolutional and layer sizes, the combination that got us the accuracy of 76% is with two convolutional layers and three fully connected layers. This was the amount of layers that worked best along with 10 epochs since an increase in epochs from 10 didn't yield any significant amount of change and stagnated. The lack of accuracy in the "Moderate Demented" category is most likely due to the unbalanced dataset that was used to train the model as well as just a lack of data for moderate dementation. The way we try to get the model to minimize these problems is by adjusting the weights per class. Using this technique along with other trial and error with convolutional and layer testing, we were able to get above a 75% accuracy with our model in the end.




```
Classification Report:
                precision    recall  f1-score   support

    MildDemented       0.78      0.53      0.63       176
ModerateDemented       1.00      0.77      0.87        13
     NonDemented       0.80      0.85      0.82       644
VeryMildDemented       0.71      0.75      0.73       448

        accuracy                           0.77      1281
       macro avg       0.82      0.72      0.76      1281
    weighted avg       0.77      0.77      0.76      1281


Overall Accuracy: 76.66%
```

Unlike our previous two models, the CNN was able to correctly classify most of the images showing moderate levels of neurodegeneration, despite its underrepresentation in the dataset. Similarly, it was relatively accurate in classifying MRI images demonstrating no dementia, but it struggled to differentiate between mild and very mild instances.

In conclusion, we have demonstrated the performance of three different supervised learning methods applied towards the detection of neurodegenerative diseases using brain imaging data. We have found that, of our three selected models, a CNN yielded the most accurate results, followed by a mixture model aggregating the results of a Random Forest Classifier, a Gradient Boosting Classifier, and a Logistic Regression model with normalized input, then followed by a simple Random Forest Classifier, and last of all, SVM. We furthermore identify our dataset's imbalance as a significant obstacle to achieving higher performing models, demonstrating that most of our models showed particularly poor performance classifying categories of data underrepresented in the overall dataset.

References

[1] M. A. Myszczynska et al., "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases," Nature Reviews Neurology, vol. 16, pp. 440–456, August 2020. [Accessed 2 October 2024].

[2] A. Tagaris, D. Kollias, and A. Stafylopatis, "Machine Learning for Neurodegenerative Disorder Diagnosis — Survey of Practices and Launch of Benchmark Dataset," International Journal on Artificial Intelligence Tools, vol. 27, no. 3, May 2018. [Accessed 2 October 2024].

[3] F. Khaliq, J. Oberhauser, D. Wakhloo, and S. Mahajani, "Decoding degeneration: the implementation of machine learning for clinical detection of neurodegenerative disorders," Neural Regeneration Research, vol. 18, pp. 1235-1242, June 2023. [Accessed 2 October 2024].

Contribution Table

Member	Contribution
Mickey	Models 2+3 Data Sourcing and Cleaning, Model 3 Selection, Model 3 Coding, Final Report
Snehil	Models 2+3 Data Sourcing and Cleaning, Models 2+3 Data Pre-Processing, Model 3 Coding, Final Recording, Final Report

Member	Contribution
Anais	Model 2 Selection, Model 2 Coding, Final Recording, Final Report
Shana	Model 3 Coding, Final Report
Diana	Model 2 Coding, FInal Recording, Final Report

GanttChart