Modelling Parkinson's Disease Using MRI Images and Biomarkers

Pramoda Karnati
pkarnati@mit.edu

Ariel Skye Levy
aslevy@mit.edu

Erica Zhou
ezhou@mit.edu

***Abstract*— Parkinson's disease (PD) is a neurodegenerative disorder that causes the impairment of certain movement functions. It manifests itself heterogeneously across different patients, and it is also diagnosed and staged via relatively subjective processes, making it hard to study in an objective manner. In this paper, we investigate Parkinson's progression and stage Parkinson's using magnetic resonance imaging (MRI) images and critical biomarker data as more objective data sources. A convolutional neural network (CNN) extracts features from the imaging data, and k-means clustering groups together images with common features. By adding in biomarker data, we develop a fuller picture of the heterogeneity of this disease, but also of the commonalities that draw certain clusters together.**

I. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder that causes the impairment of certain movement functions. Although this disease is not fatal, it is chronic, and its symptoms progress gradually over time. Symptoms commonly include things like tremors, stiffness of movement, and loss of balance, but they vary greatly among patients. The strong variability in this disease and its treatment make it difficult to study, but it is also an extremely important disease, currently affecting more than 1.5 million Americans [1].

Parkinson's is difficult to diagnose and stage, generally relying on subjective evaluations of symptoms and surveys of the symptoms' effects on patients' daily lives. There is no objective way to diagnose PD, although MRI images have been shown to be effective in diagnosis [2], [3]. With regards to staging, Parkinson's is generally broken down into five stages based on severity of symptoms and level of impairment of daily life. Again, these stages are not defined from an objective measure, so there is a lot of variance in the stagings given by different clinicians.

In this paper, we use the Parkinson's Progression Markers Initiative (PPMI) data to model Parkinson's disease progression and staging using MRI scans as a more objective form of evaluation. Since symptoms are so varied and so subjective, having a more standardized way to stage PD has the potential to greatly improve research and treatment of the disease. Additionally, MRI scans have been shown to be successful at least in PD diagnosis, so we wish to see if the scans also give any information about progression of the disease. By running a convolutional neural network on the MRI images, we learn feature representations of the images that correspond to different stages of the disease. Given these representations and information about critical Parkinson's biomarkers, we generate groups of similar PD patients via a clustering algorithm. By analyzing these clusters, we develop a fuller picture of the commonalities that draw certain clusters or types of patients together, and we also gain an idea of the areas where the disease is most homogeneous.

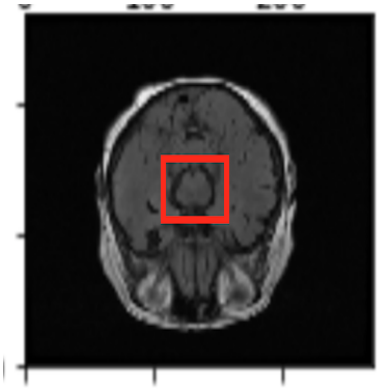


Fig. 1. Slice of T2 axial MRI scan of Parkinson's Disease patient highlighting the substantia nigra .

II. Related Work

Past analyses on Parkinson's disease have approached the problem from a few different angles. Biologically, it has been observed that PD primarily affects an area of the midbrain known as the *substantia nigra*, which generates the neurotransmitter dopamine and plays an important role in tasks like reward and movement. An image of the substantia nigra can be seen below in Figure 7.

From the machine learning perspective, researchers have looked at the ability to diagnose and stage PD from both important biomarkers and MRI images, and they have been fairly successful in doing so. On the side of images, there has been research about the substantia nigra's role in PD since the 1980s [4]. Tuite et al. have done work on the various uses of MRI in the diagnosis of PD [2]. One specific highlight of their research suggests that T2 imaging can be useful in PD diagnosis by recognizing the increased iron concentration and resulting relaxation time in the brains patients with PD. Lehericy et al. have demonstrated the potential of PD, although they have also acknowledged that there are definitely inconsistencies, and although it definitely has a lot of potential as a diagnosis method, MRI on its own is not a strong enough definer of Parkinson's to be a definitive clinical test [5].

With the advent of machine learning, work has been done on both the diagnosis and staging of PD using neural networks. Shinde et al. focused on neuromelanin-sensitive MRI data as a predictor for PD, reaching around 80%

accuracy in diagnosis and 86% in differentiating Parkinson's from atyipcal Parkinsonian symptoms [3]. Additionally, attempts have been made to stage Parkinson's from MRI data, although it has proven to be a fairly difficult task given the heterogeneity of the disease and the subjectivity even in its clinical staging. Schwarz et al. staged early and late-stage Parkinson's by analyzing the size reduction of the substantia nigra in T1 MRI scans, helping to confirm the importance of this brain region [6]. However, tasks like quantifying the amount of loss with stage or being able to predict stage have still proven to be a challenge to researchers.

Previous work with biomarkers alone has attempted to produce more effective staging mechanisms for assessment of the disease itself. The two most commonly clinically used tools to evaluate the state of the disease are the Movement Disorder Society-Unified Parkinsons Disease Rating Scale (MDS-UPDRS) and the Hoehn and Yahr (HY) scale. The MDS-UPDRS tracks the development of symptoms of the disease while HY quantitatively stages the disease. Due to flaws in both of these mechanisms, Roy et al. created a model that combines MDS-UPDRS ratings and HY scale information through predictive modeling into a novel PD staging measure. These predictive models gave an accuracy as high as 97.46% on predicting the new stage mechanism. [7]

Previous work by Vasquez-Correa et al. has also attempted to use multimodal symptom determinations from various different inputs, including speech, handwriting, and gait to classify Parkinsons patients in different stages of the disease. A deep learning approach was highly accurate to determine control subjects from Parkinsons patients and suitable to categorize the neurological state of patients into initial, intermediate, and late stages of the disease. [8]

Previous work by Faghri et al. has used unsupervised and supervised machine learning approaches to first create patient subtypes and then project which subtype the patient belonged to four years based on baseline measurements. The unsupervised model identified three progression subtypes of patients, slow, moderate, and fast disease progressors. A supervised model was then run on baseline measurements of the patients to predict which category of progressor they belonged to, which achieved an AUC of 0.93. [9]

III. Dataset

The dataset used for this project comes from the Parkinson's Progression Makers Initiative (PPMI) database, a longitudinal study aiming to elucidate the biology of PD progression utilizing imaging data, biological samples, and clinical and behavioral assessment data. The PPMI dataset consists of both biomarker data and image scans for registered patients, both Parkinson's (PD) patients and healthy controls, at various time points over a period of four years.

A. Patient Population

Table 1 illustrates the distribution of the visit entry data on demographic and stage information.

Type of Subject	
PD Subject	550 (60.7%)
Control	356 (69.3%)
Stage	
0	349 (38.5%)
1	177 (19.5%)
2	365 (40.3%)
3	15 (1.7%)
Gender	
Male	624 (68.9%)
Female	282 (31.1%)
Race	
White	853 (94.2%)
Black	24 (2.6%)
Asian	3 (0.3%)
Other	26 (2.9%)
Age	
< 56	277 (30.6%)
56 - 65	353 (39.0%)
> 65	276 (30.4%)

Table 1: Distribution of PPMI patients analyzed

B. Imaging Data

The database contains 6,904 various MRI images of both healthy and PD patients. These images vary in type of MRI (e.g. T1,T2), the angle of image (e.g. Axial, Coronal, Sagittal), and overall technique of imaging. Due to the heterogeneity of data sources and clinical techniques, the image scans required a significant amount of preprocessing before they could be used to learn anything about the disease. In this project, only T2 axial images of the brain were considered. From each of these images, a sample of 10 slices from the section of the midbrain containing the substantia nigra (estimated to be around 40% of the way from the bottom of the brain) was collected for modeling. Preprocessing of the images includes: extracting the scans from the files (in `.dcm` format) using the tool `PyDicom`, normalizing the inputs, standardizing sizes. The images were then fed into a convolutional neural network.

1) Imaging Data Preprocessing: MRI image files are stored as a series of DICOM (Digital Imaging and Communications in Medicine) files each representing a slice of the brain. We use the Axial view for each patient. Each image is fairly large, consisting of up to hundreds of `512*512` pixel slices, depending on the resolution of the scan. Additionally images were fairly heterogeneous, with small scans being `256*204` pixels with as few as only around 10 slices. MRI scans are also fairly large in size and require a lot of memory and processing power. Therefore, an initial challenge for our group was downloading images and having a platform to store them. We originally used Google Cloud but switched to AWS as our main processing platform. Each series of DICOM files was stored in a complicated hierarchical structure for each patient; therefore, we had to first load all scans for each patient and maintain a link between individual image slices to the patient and the time at which it was taken. We

correlated the stage at which the patient was in to the each MRI scan by comparing the timestamp of the patient visit and the MRI scan.

The `PyDicom` package in Python was used to extract the images from the DICOM files into a human-viewable format. We can visualize an example of the slices of an individual patient using `pyplot` below in 2 and 3. From these image slices, the rough location of the substantia nigra was determined to be around 40% of the way up from the bottom of the brain. For each patient, this location represents the area of interest of the MRI for Parkinson's disease.

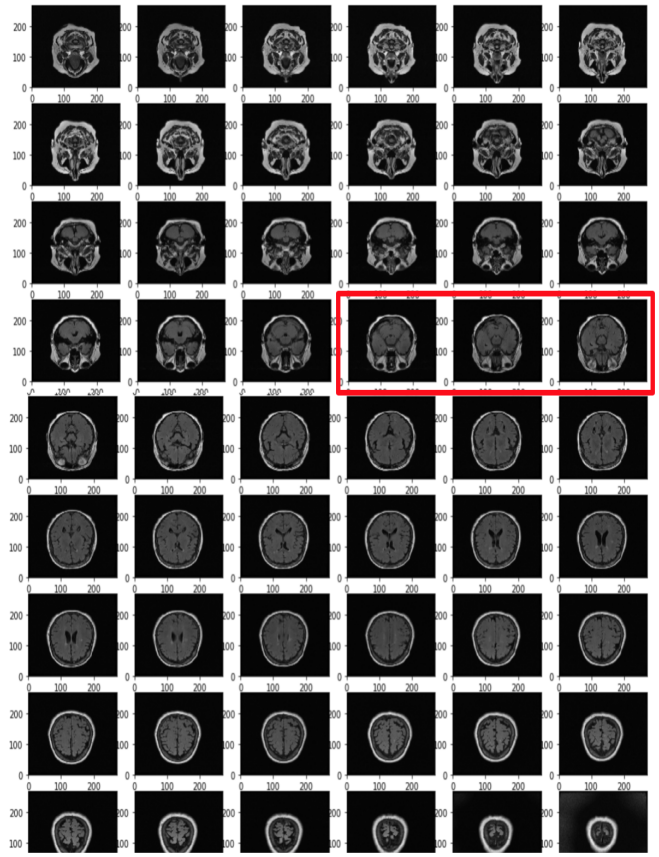


Fig. 2. T2 Axial MRI scan of Parkinson's Disease patient, with slices containing substantia nigra highlighted in red. Inset in Figure 3

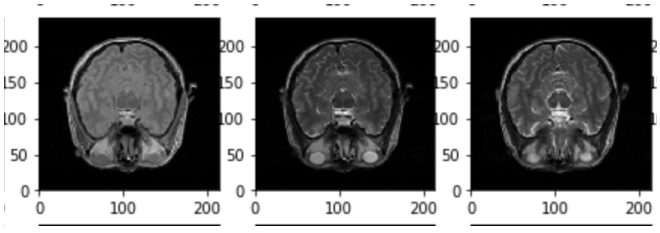


Fig. 3. Inset of section of MRI with slices containing substantia nigra.

MRI scans were filtered down to include only T2 Axial images for heterogeneity, and only scans with greater than 50 horizontal slices were included to guarantee a minimal resolution. From these filtered images, a region of about 10 slices around 40% from the bottom of the MRI scan was collected from each scan. The goal in this was to guarantee selection of the slices containing the substantia nigra, but also to minimize the collection of irrelevant slices to have homogeneity in the data. Each slice is also repeated 3 times along the z-axis so that it can be fed into the CNN model.

In order to have these slices be useful for a model, linear normalization was performed to get a consistent color gradient across all slices. We also perform the specific pre-processing technique required by each convolutional network by using the technique provided by the model itself. After normalization, images were ready to be fed into a CNN to extract features.

In order to extract a more useful representation of MRI scans, a brain segmentation technique was also attempted to isolate the brain tissue from the rest of the skull and other tissues in the head. Python's deepbrain extractor successfully drew out brain segmentations from NIFTI files, a different type of MRI scan, as seen in figure 4.

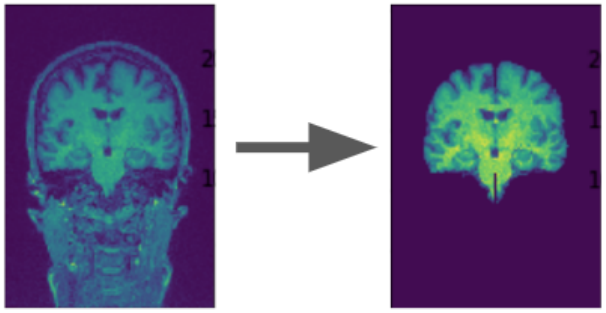


Fig. 4. Brain segmentation as performed by Python deepbrain extractor. Draws out and highlights brain tissue from the rest of the head.

However, these were not able to be implemented in the neural network due to a constraint on processing power and time.

C. Biomarker Data

The set of biomarkers selected for the model, termed the Curated Data Cut, were specifically developed by the PPMI organization as the most previously established significant markers that relate to the development and progression of Parkinsons disease. The markers include demographic information, lab test results, symptom scores, and genotype information.

One of the most significant biomarkers included in this set was Hoehn & Yahr stage. The Hoehn & Yahr scale is a five point staging mechanism that medical professionals use to evaluate the severity of the progression of a patients disease. Stage one expresses mild symptoms of Parkinsons and symptoms on only one side of the body. Stage two expresses early stage symptoms of Parkinsons on both sides of the body. Stage three expresses mid-stage symptoms characterized by loss of balance and slowness of movement. Stage four expresses severely disabling symptoms including loss of the ability to walk and the requirement of outside

assistance. Stage five expresses the most advanced symptoms of the disease and characterized by complete confinement of movement. This Hoehn & Yahr scale, as the most popular industry standard for staging the disease, is the primary marker that our model used to track the disease progression. [10]

Of note is the fact that Hoehn & Yahr stages 3, 4, and 5 were categorized together in the dataset and marked as later stage disease with the number 3.

1) Preprocessing of the Biomarker Data: There were originally 683 patients in the curated biomarker dataset. After filtering the dataset to include the patients that also had MRI data, there were 419 patients in the dataset.

The dataset was originally split into two separate sets, a baseline dataset with measurements for patients at the time of enrollment in the study and a progression dataset with measurements for patients up to three years after enrollment. These datasets were joined on patient ID such that any baseline measurements such as demographic information were included with other progressive biomarker entry measurements. So each entry for each patient was considered one cross-sectional data point in the model. This created 1603 visit entries in the dataset.

There were originally 128 biomarkers included in this dataset, however many of these biomarkers were not recorded for all patients. First biomarkers missing from more than 20% of data points were excluded from the final model. This reduced the amount of biomarkers to 81. After this, patient entries missing any data points were excluded from the dataset. This left 906 visit entries in the dataset.

IV. Approach

Originally, the goal of the project was to utilize MRI scans and biomarker data to predict the progression of Parkinson's disease. However, upon reassessment of the heterogeneity of the data, the project shifted towards classifying patients by stage. With this new goal in mind, a CNN was utilized to generate an image model, and logistic regression was used to build a model to predict PD stage from biomarker data.

A. Marking Progression in Subjects

For each entry, we generated a feature to mark whether the patient progressed to the next Hoehn & Yahr stage in the next twelve months. We defined progression as a step to the next stage or any further stage at the next annual visit. Remaining at the same stage or regressing in stage was marked as lack of progression. This resulted in 12.1% of entries marked as progression and 87.9% marked as not progression.

With these labels, modeling progression proved to be fairly difficult because of the large amount of heterogeneity. Parkinson's symptoms progress at different rates for different patients, and in this short 4-year period of data from the PPMI database, pulling specific features that correlated was progression was difficult. Additionally, defining progression from stage 0 to 1 as similar to progression from stage 2 to 3 may have led to some biasing as well due to different features marking each of these individual transitions.

B. Marking Staging in Subjects

Upon reevaluation, we changed our goal to trying to determine a way to more objectively subtype Parkinson's patients since even the clinically-used Hoehn & Yahr scale appeared to be subjective with respect to symptom severity. Accordingly, we wanted to use an unsupervised approach to determine staging or some other type of subtyping based on more objective features as opposed to the more subjective and aggregated Hoehn & Yahr score.

With this goal in mind, we first tested our pipelines by creating a CNN to separate healthy controls from PD patients. Based on features extracted from the neural network, we wanted to cluster together patients with similar features and see whether the model picked specific subtypes or other commonalities between specific PD patients.

C. Biomarker Model Construction

1) Progression Modeling: To predict whether a patient would progress in Parkinsons stage based on visit entries with biomarker information in the next 12 months, we constructed a Logistic Regression model. The model was constructed with the sklearn Python module. The feature inputs were the preprocessed biomarkers as discussed in the Dataset section and the labels were the binary classification of whether the patient had progressed at the next visit. The dataset was randomly divided into train, validation, and test sets on a 60/20/20 ratio split. The model was fit to the train set. The best parameters of the model were chosen based on the AUC on the validation set. The model was then evaluated on the test set.

2) Classification & Staging: To evaluate the pipeline, we tested the models performance on a more basic classification task, determining healthy control subjects from enrolled subjects with Parkinson's disease. The same feature inputs were used in both classification tasks. Additionally, a multi-class logistic regression model was utilized to stage patients via a "one-versus-all" algorithm. For each stage, a separate LR model was run where only the patients of that stage were classified as positive, training the model to recognize specific features that correspond to specific stages of Parkinson's.

D. CNN Model Construction

We use a Convolutional Neural Network to process the MRI images. We use a Convolutional Neural Network to process the MRI images to stage the patients. We label MRI scans of the patients as Healthy (0) and PD (1). We also separately label the scans in the four stages $0, 1, 2, 3$. Patients separated into stages are marked whether they progress or not not based on patient visit information in the next 12 months.

1) Progression Modeling: Initially, we based our model on a paper by Esmaeilzadeh et Al. [11] in which they created a 3-D Convolutional Neural Network to classify patients as PD or Healthy Controls to predict whether a patient was going to progress from their current stage given the MRI image in the next 12 months. The data was split into train and test sets in a 80/20 ratio. The model was fit on the MRI images and trained to determine progression. However,

we realized that our progression task was difficult given the immense variation of the scans as well as the small amount of Axial samples available for each stage. We faced the same challenges with the progression task related to MR-Images as we did regarding the Biomarkers.

2) Classification & Staging: We then decided to use a CNN for classificaiton and staging. For our network, we use pre-trained models provided by `Keras`. We use ResNet50 pre-trained with ImageNet weights so that the basic understanding of edges and curves are preserved. We then add an additional GlobalAveragePooling2D layer and a Dense layer of 1024 units. Each layer uses Rectified Linear-Unit (ReLU) as the activation function. For output, we use the Softmax output with the corresponding number of classes (2 for classification, 4 for stages. We use Categorical Crossentropy as our loss function with the Adam optimizer.

The data is split into train and test sets in a 80/20 ratio using `sklearn train_test_split` function. We first train the model on the basic task of classifying patients to differentiate between healthy control subjects and subjects with Parkinson's disease. We then trained a separate model to stage the patients into the four stages. The model is evaluated on the test set to performance

E. Clustering

After attempting to stage the patients classify and stage patients using the MRI images, we cluster the MRI images to determine what features the CNN is learning from each image. For each model, we take the output of the last Dense layer. The output of this layer is the 1024-dimensional vector representation of each MRI image as learned by the CNN. These representations are passed into a kMeans clustering algorithm with $n = 2$ or 4 clusters depending on the task. For each cluster, we look at how many patients of each label exist in each cluster to learn whether the representations provide any useful information regarding the clinical condition of each patient.

V. Challenges

A few major challenges arose in the completion of this project from the perspectives of infrastructure, data cleanliness and processing, and overall availability of data.

On the infrastructure side, we ran into some issues hosting our images and having the processing power to run specific algorithms on the data. Initially, we tried putting the images on Google Cloud, which provides the ability to collaborate and run separate instances; however, we were unable to get the images into Google Cloud, and we had trouble utilizing programs such as Jupyter, so we switched to an AWS instance to host the data instead. On this platform, we were able to run analyses on the image data, but we were only able to use one instance, which hurt our ability to work in parallel and run more analyses due to limitations in kernel and memory usage of the instance.

With regards to the imaging data, MRI images are notoriously heterogeneous, especially in the case of a disease with as much heterogeneity in itself as Parkinson's. MRI images, being first and foremost for clinical use as opposed to computational analysis, came in a bunch of different sizes and resolutions, which made the preprocessing step fairly difficult. Additionally, because of the different sizes, we had to estimate where our desired brain slices were for finding out useful information regarding PD. From viewing a few sample images, we estimated the area containing the substantia nigra to be around 40% of the way up the brain, but it is possible that there are some slices that were incorrectly included in the dataset.

Finally, the dataset, although the biggest available dataset specifically aimed towards looking at Parkinson's progression, was still fairly sparse, especially for later stages, which made it difficult to learn a good representation without heavily overfitting to the few points that existed for each stage. Additionally, a lot less patients were included in the imaging data as opposed to the biomarker data, likely as a result of the cost and unproved efficacy of using MRI to model Parkinson's progression. Because of this, linking was necessarily between the biomarker data and the images, which also dropped the number of available patients from the biomarker dataset. In the end, we had only 4 patients with imaging and biomarker data for PD stages 3-5, which severely limited our ability to gain insights about this specific group.

VI. Results & Evaluation

A. Biomarker Model Results

1) Classifying Healthy versus Parkinson's Subjects: First, the efficacy of the Biomarker Logistic Regression model was tested with the task of classifying subjects as healthy controls or those with Parkinsons disease. The parameters for the model were an inverse regularization constant of 0.1 and a penalty function of L1 regularization. These parameters were chosen as the best parameters for the model based on results from the progression prediction task, as discussed in the section below. The AUC score of the model on the train set was 0.992. The AUC score of the model on the validation set was 0.98. The AUC score of the model on the test set was 0.986 as shown in Figure 5. This illustrates that the model could effectively classify patients into control and those affected by the disease based on patient biomarkers at different time points.

The features ordered by how strongly they contributed positively or negatively to the predictive model are listed in Table .

2) Classifying Patients Based on Stage: The efficacy of the biomarkers on evaluating Hoehn and Yahr stage was tested with the multi-class task of classifying visit entries for different patients by categorical stage number. A One-vs-the-rest (OvR) multi class model was fit to the biomarker data, fitting one classifier per categorical stage. This resulted in an AUC score of 0.992 on stage 0 prediction, AUC score of 0.636 on stage 1 prediction, AUC score of 0.899 on stage 2 prediction, and AUC score of 0.673 on stage 3 prediction. These results are illustrated in Figure 6. It is notable that

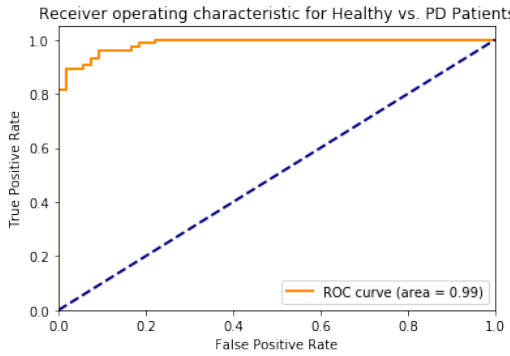


Fig. 5. Shows the performance of presence of Parkinson disease prediction model.

TABLE I

FEATURES CONTRIBUTING TO PRESENCE OF PD PREDICTIONMODEL

Biomarker	Model Contribution
MDS-UPDRS Part III Score	+
MDS-UPDRS Total Score	+
Epworth Sleepiness Scale Score	-
STAI Trait Sub-score	-
Age at Baseline	-
MDS-UPDRS Part II Score	+
MDS-UPDRS Part I Score	-
Semantic Fluency Total Score	+
Symbol Digit Modalities Score	+
CSF p-tau (2016 assay)	-

stages 0 and 2 were by far the most represented in the dataset, and performed significantly better with the model.

3) Classifying Patients Based on Progression: The Biomarker Logistic Regression model was then performed on the classification task of whether a patient would progress to a further Hoehn and Yahr stage within twelve months. Five options were tested for inverse regularization constant, 0.01, 0.1, 0.25, 0.5, and 1. Two options were tested for the penalty function, L1 and L2 regularization. All possible combination options of these two parameters were tested. The parameter combination that had the best performance on the validation set was an inverse regularization constant of 0.1 and a penalty function of L1 regularization. This model had an AUC score of 0.697 on the test set as shown in Figure 7.

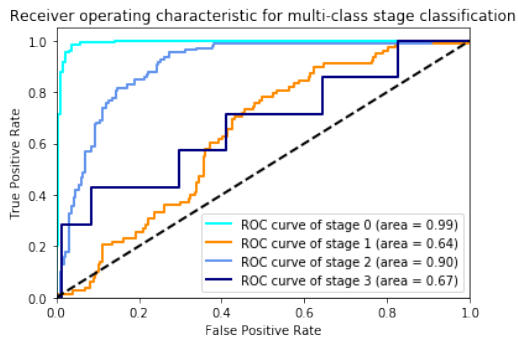


Fig. 6. Shows the performance of Parkinson disease stage prediction models.

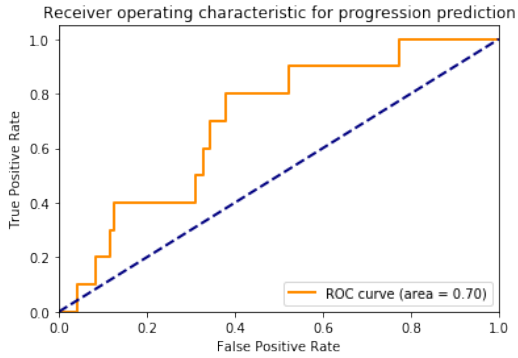


Fig. 7. Shows the performance of PD progression prediction model.

The features that contributed positively or negatively to the predictive model are listed in Table .

TABLE II

FEATURES CONTRIBUTING TO PD PROGRESSION PREDICTION MODEL.

Biomarker	Model Contribution
TD/PIGD classification	-
Total Rigidity Score	-
Epworth Sleepiness Scale Score	-
SCOPA-AUT Gastrointestinal (GI) Score	+
Tremor Score	
HVLT Immediate/Total Recall	+
REM Sleep Behavior Disorder Questionnaire Score	-
MDS-UPDRS Part II Score	+
Categorical SNCA rs356181 Genotype	-
MOCA Score	+
SCOPA-AUT Sexual Dysfunction Score	+
Semantic Fluency Score	-
Animal subscore	+
Age at Baseline	-
Symbol Digit Modalities Score	-
CSF p-tau (2016 assay)	-

B. CNN Model Results

1) Classifying Healthy vs Parkinon's Subjects: We evaluated the trained ResNet50 Model on the test set of MRI images. We can see the results of this model in Figure 8. This shows that the model is able to perform slightly better than average in classifying PD patients versus Healthy Controls. The model performed with an error rate of 0.45 on the test set in classifying healthy and control patients. We speculate that the poor performance of this model might be attributed to the high variance in our original Axial MRI scans that were used for this task. The performance might also be attributed to the specific slices of the MRI images that we are obtaining. We attempted to train the model on the top 20 slices of the MRI Axial scan instead of the middle 10 and were able to find an error rate of 0.2. This result was surprising as the top 20 slices of the brain clinically correspond to the lower brain which does not provide any useful information regarding Parkinson's disease. However, when looking at the

middle 10 slices containing the substantia nigra, we find poor performance in our model.

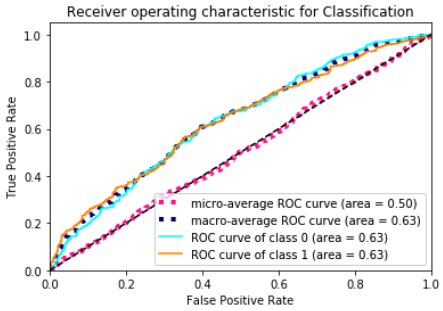


Fig. 8. Receiver Operating Curve for the Classificaiton CNN Model using ResNet50. We find that the model has an AUC of 0.63 for each class, indicating that it is able to perform slightly better than average in classifying PD vs Control Patients

2) Staging of Parkinson's Subjects: We evaluated the trained ResNet50 Model on the test set of MRI images, and the results can be seen in Figure 9. We see that the ROC for stage 0 and stage 2 are the highest (0.59 and 0.65 respectively). This can be attributed to the fact that there are more patients in these two stages in the dataset. The trends in this model closely follow that of the Biomarker data, which was also interesting. Using this model, we get an error rate of 0.33 for classifying stages of Parkinson's Patients given MRI scans. This leads us to believe that MRI scans do provide some useful information regarding staging Parkinsons' patients, but it is possible that our method of approach for preprocessing data needs to be more rigorous to achieve better results on our model.

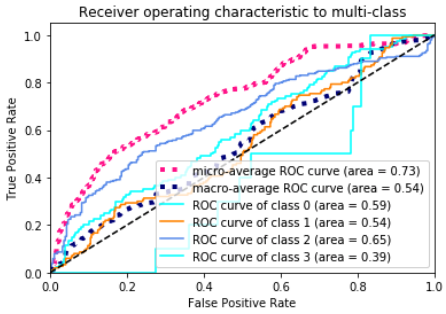


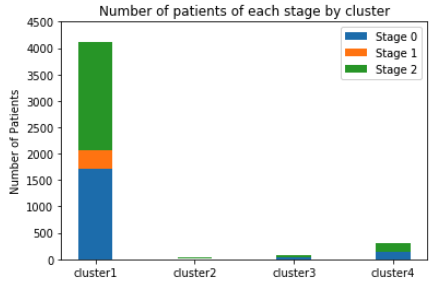
Fig. 9. Receiver Operating Curve for the Classification CNN Model using ResNet50. We find that the model has an AUC of 0.65 for for stage 2 and 0.59 for stage 0, which makes since there are more patient images of those two stages. The curve also looks very similar to the one using the Biomarker data, with the trends for each stages being relatively the same.

C. Clustering

After running the MRI images through a CNN, we took the features extracted by the last layer of the network as the feature representation of each image. We then ran a kMeans clustering algorithm on the image feature maps from each classification task (Healthy vs PD and the Staging Task) to determine whether any features are represented by each cluster.

For the classification task, we cluster the extracted features into two clusters. We then find the number of patients labelled 0 or 1 (Control vs PD) in each cluster. We find that each cluster has approximately the same ratio of patients of each label: each of the two clusters contains 70% Parkinson's patients and 30% control patients.

We find more interesting results when running the clustering algorithm on the extracted features from the stage classification task in Figure 10. We see that most of the patients fall in the first cluster, which has an even distribution of each stage. Cluster 2 seems to contain entirely stage 2 patients, and the two stages still dominate entirely in the last two clusters. This can probably once again be attributed to the large number of stage 0 and stage 2 patients available in the dataset. Given these clusters of patients, we find that our model was not able to find significant differences in the patient feature vectors to be able to cluster them efficiently into various stages.




```
Cluster 1: [0.303, 0.332, 0.364]
Cluster 2: [0.0, 1.0, 0.0]
Cluster 3: [0.443, 0.114, 0.443]
Cluster 4: [0.397, 0.162, 0.441]
```

Fig. 10. (Top) Clusters and labelled patients in each cluster. We can see that most patients fall in cluster 1 which has an equal distribution of stage 0 and stage 1 patients. We also see that Cluster 2 is composed of entirely stage 2 patients, and stages 0 and 2 dominate in two clusters. (Bottom) Percentage of each stage $(0, 1, 2$ respectively) in each cluster.

VII. Discussion & Future Steps

Overall, our model did successfully replicate what others have done regarding diagnosis of Parkinson's disease via MRI imaging, suggesting that the pipeline does work for at least the diagnosis of PD. However, it struggled with modeling the progression and staging of Parkinson's due to a lack of data as well as heterogeneous data for the patients that did exist. Thus, although MRI imaging has proven potential in PD diagnosis, we are still struggling to find an equivalent in either the specific biomarkers or MRI data to determine PD stage.

There are a number of potential factors that could have contributed to the low accuracy of the CNN models. One obvious factor is the low availability of data. Focusing on only Axial T2 slices significantly cut down our dataset to only a few thousand images, which is not significantly enough to perform a classification task. Another potential problem could have been the pre-processing of the data. As this was the first time our team members worked with

MRI images, understanding the data and the ways to preprocess the data took a significant amount of time. A third potential problem is that the clinical parts of our approach might be incorrect: Parkinson's disease manifests itself in certain regions of the brain, which we attempted to capture by only looking at the relevant middle slice locations of the brain. However, due to our limited prior clinical knowledge of both the disease and MRI scans, it is possible that we are looking in the incorrect location for any indication of the disease.

In terms of future steps, further analysis could be done on the segmented brain images (extracted via Deepbrain or another extractor). Segmentation could help highlight the specific brain features that we are interested in while removing features that may unnecessarily distract the model. Additionally, the use of different image types, such as DaTscans, which are generally used for early-stage detection, may be more telling or provide more specific information about later staging. The addition of more data, especially for later-stage patients, could also provide the model with a lot more samples to go off of, which would greatly increase the chance of finding unique and common features of these stages of Parkinson's. Finally, a more rigorous analysis of the selected slices may improve data quality, and more rigorous and planned training, or a more in-depth examination of the pooled layers, could allow us to apply more domain knowledge to improve model performance.

VIII. Conclusion

Parkinson's disease is an important and prevalent neurodegenerative disorder, and it has been a challenge for researchers to work with due to its heterogeneity of symptoms and biomarkers. We present a staging methodology for Parkinson's disease using a convolutional neural network and clustering algorithm to define subtypes of patients based on MRI scans and biomarkers. Through this framework, we define some level of similarity between patients facing an otherwise fairly heterogeneous disease. Ideally, this type of subtyping will allow clinicians to separate different types of Parkinson's disease and specialize treatment accordingly, so that the disease can be most effectively managed for the many who face its symptoms every day.

IX. Contributions

Pramoda focused on infrastructure and image analysis. She developed the pipeline of data from the DICOM files to the final CNN features extracted, as well as the clustering algorithm to discover common features pulled out by the model. Ariel worked with the biomarker data and logistic regression modeling, as well as developing the relationship between patient ID numbers, labels of progression and stage, and MRI images. Erica looked at the biology of Parkinson's disease from the biomarker and MRI perspective and created the preliminary pipeline for running data through a CNN.

X. Acknowledgments

The authors would like to thank Professor David Sontag and our TA Irene Chen for their guidance and feedback through the project process. Additionally, we would like the entire HST.956/6.S897 staff, including Professor Peter Szolovits and TA Willie Boag, for their help, as well as Christina Ji for her help with obtaining the PPMI imaging data.

References

[1] Parkinson's Disease. https://www.aans.org/Patients/Neurosurgical-Conditions-and-Treatments/Parkinsons-Disease.

[2] Tuite, P. J., Mangia, S., & Michaeli, S. (2013). Magnetic Resonance Imaging (MRI) in Parkinson's Disease. Journal of Alzheimer's disease & Parkinsonism, Suppl 1, 001. doi:10.4172/2161-0460.S1-001.

[3] Shinde, S., Prasad, S., Saboo, Y., Kaushick, R., Saini, J., Pal, P. K., & Ingalhalikar, M. (2019). Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI. NeuroImage. Clinical, 22, 101748. Advance online publication. doi:10.1016/j.nicl.2019.101748

[4] JN Rutledge, SK Hilal, AJ Silver, R Defendini, and S Fahn. Study of movement disorders and brain iron by MR. American Journal of Roentgenology, 1987, 149:2, 365-379.

[5] Lehricy, Stphane, Sharman, Michael A., Santos, Clarisse Longo Dos, Paquin, Raphal & Gallea, Cecile (2012). Magnetic resonance imaging of the substantia nigra in Parkinson's disease. Movement Disorders, 27, 7. https://doi.org/10.1002/mds.25015.

[6] Schwarz, S. T., Rittman, T. , Gontu, V. , Morgan, P. S., Bajaj, N. and Auer, D. P. (2011), T1Weighted MRI shows stagedependent substantia nigra signal loss in Parkinson's disease. Mov. Disord., 26: 1633-1638. doi:10.1002/mds.23722

[7] R. Prashanth and Sumantra Dutta Roy (2018). Novel and improved stage estimation in Parkinson's disease using clinical scales and machine learning. Neurocomputing, 2018, 305:78-103. https://doi.org/10.1016/j.neucom.2018.04.049.

[8] Vasquez-Correa, J. C., Arias-Vergara, T., Orozco-Arroyave, J. R., Eskofier, B. M., Klucken, J., & Noth, E. (2018). Multimodal assessment of Parkinson's disease: a deep learning approach. IEEE journal of biomedical and health informatics. https://doi.org/10.1109/JBHI.2018.2866873.

[9] Faghri, F., Hashemi, S. H., Leonard, H., Scholz, S. W., Campbell, R. H., Nalls, M. A., & Singleton, A. B. (2018). Predicting onset, progression, and clinical subtypes of Parkinson disease using machine learning. bioRxiv, 338913. https://doi.org/10.1101/338913

[10] Mller, J., Wenning, G. K., Jellinger, K., McKee, A., Poewe, W., & Litvan, I. (2000). Progression of Hoehn and Yahr stages in Parkinsonian disorders: a clinicopathologic study. Neurology, 55(6), 888-891. https://doi.org/10.1002/mds.22875

[11] Esmaeilzadeh, Soheil Yang, Yao Adeli, Ehsan. (2018). End-to-End Parkinson Disease Diagnosis using Brain MR-Images by 3D-CNN.