Data Imputation on Multimodal Datasets for Parkinson's Disease Classification

Hetvi Jethwani
Mathematics Department
IIT Delhi
mt1180754@iitd.ac.in

Bhumika Chopra
Mathematics Department
IIT Delhi
mt1180748@iitd.ac.in

***Abstract*—This report details the experiments conducted by us for term project of MTL782.**

I. Introduction

Parkinson's Disease (PD) is a chronic neuro-degenerative disease which affects the motor system in early stages, and non-motor symptoms manifest as it progresses. It is characterized by tremors, rigidity of muscles, difficulty in walking, and bradykinesia. Additional cognitive impairment in the form of apathy and depression are common, and in advanced stages patients often develop dementia. The goal of our project is to learn how to classify if a patient has PD or not, given speech and hand-drawn image data specific to the patient.

Past work in the domain of PD classification includes the utilization of a wide variety of data- wearable-sensor data ([6],[1],[11], [15]), data related to genetic factors ([5]), brain MRI scans ([10]), data about gut microbiota ([7]), speech data ([13],[2],[4],[9]), and data obtained from drawings ([14]). The trade-off however is data collection, and collecting large amounts of data is quite difficult. During our literature survey of the problem we observed that there are a lot of datasets, but all of them are small ($<$ 200 subjects), and often have multiple samples from the same subject. This motivated us to ask if information from these small, multi-modal datasets be merged together for better generalizability on unseen data. Due to unavailability of many of the datasets in public domain, we ultimately decided to proceed using 2 speech datasets ([9],[12]) and 1 hand-drawn image dataset ([14]). The aim of this work is to assess if merging of different datasets using techniques like data imputation can help us train models with less bias which are more generalizable.

II. Dataset Details

A. Speech Data

We utilize information from 2 datasets throughout our report- Multiple Sound Recordings (MSR) Dataset [9], and Sri Lanka (SL) Dataset [12]. We first detail the common speech features, then talk about preprocessing done on individual data sets, and then mention how they're utilized together.

1) MSR Dataset: The training data was obtained from 20 Parkinson's Disease (PD) patients and 20 healthy people. From each subject, 26 sound recordings were taken with different types of sounds- sustained vowels, numbers from 1 to 10, some short sentences and words. 26 features are extracted from every individual recording (namely the features mentioned in Table 1). In addition to this, we also have the Unified Parkinson's Disease Rating Scale (UPDRS) score of each subject in the training set with Parkinson's disease which tells us about the severity of the disease. Since we aim to solve a binary classification problem, we don't use these values. A test data file is also provided- it contains recordings of sustained vowels 'a' and 'o' three times each per subject for 28 subjects, and each subject is a PD patient. In addition to the features, each dataset also has information about unique subject ID and a binary variable called status which is 1 if subject has PD.

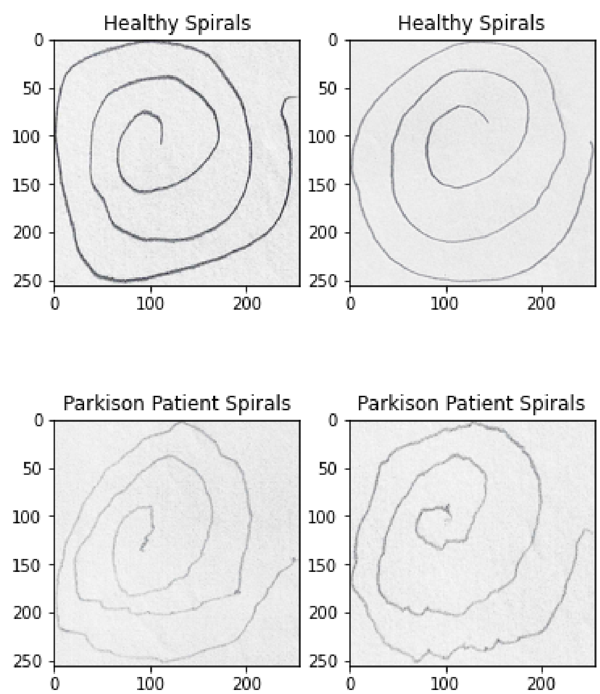
Feature	**Description of speech feature**
Jitter Values	Maximum perturbation of fundamental frequency as a percentage of pitch period duration
Local	Jitter ratio/1000
Local, Absolute	Jitter ratio*average period duration
rap	Relative average perturbation with 3-point estimate
ppq5	Relative average perturbation with 5-point estimate
ddp	3*rap
Shimmer Values	Maximum variation in peak amplitudes of successive pitch periods
Local	Avg. absolute difference between amplitudes of consecutive periods divided by avg. amplitude
Local, dB	Avg. absolute log (base 10) of difference between amplitudes of consecutive periods
apq3	3-point amplitude perturbation quotient
apq5	5-point amplitude perturbation quotient
apq11	11-point amplitude perturbation quotient
dda	Average absolute difference between the amplitudes of consecutive periods
AC	Autocorrelation
NTH	Noise to Harmonic ratio
HTN	Harmonic to Noise ratio, highly sensitive
Pitch	Median, mean, min, max, and standard deviation of pitch are measured
Pulses	No. of pulses are measured
Period	Number of periods, mean period and standard deviation of period are measured
Unvoiced Frames	Fraction of pitch frames that are unvoiced
Voice Breaks	Number of voice breaks, degree of voice breaks are measured

2) SL Dataset: This data set was obtained from 15 PD patients in Sri Lanka, and consists of 6 sound recordings per subject taken with sustained vowel sounds. 26 features are extracted from every recording, and they are similar to the MSR dataset.

B. Hand Drawing Data

The data set [14] was obtained from 55 volunteers- 28 healthy and 27 Parkinson's affected, ranging from healthy (UPDRS = 0) to severely affected patients (UPDRS >24). The participants were shown an archimedean spiral guided by bright spots and asked to draw the spiral by connecting all the dots. A similar process was done for drawing waves. The sketching was recorded using commercially available A3 size tablet (Wacom Intuos Pro Large). The A3 paper was placed on the tablet, and Ink pen (Wacom Intuos ink pen) was used to sketch the spiral. This pen senses the location of contact, x and y, and the pressure, pr, between the tip and the paper. The data was then segmented between each pen-down and corresponding pen-up time instances; pen-down identified based on pr > 0, and given an index label, i with m_i being the total number of samples of the segment. Features such as total length of each segment, mean pr per segment, total time duration of each segment, etc were computed. We were unable to get access to the extracted feature data and therefore have used the images available in the data set for further experimentation.

From the HD data set we obtain 102 healthy patients spiral and wave drawings and 72 PD patients spiral and wave drawings.

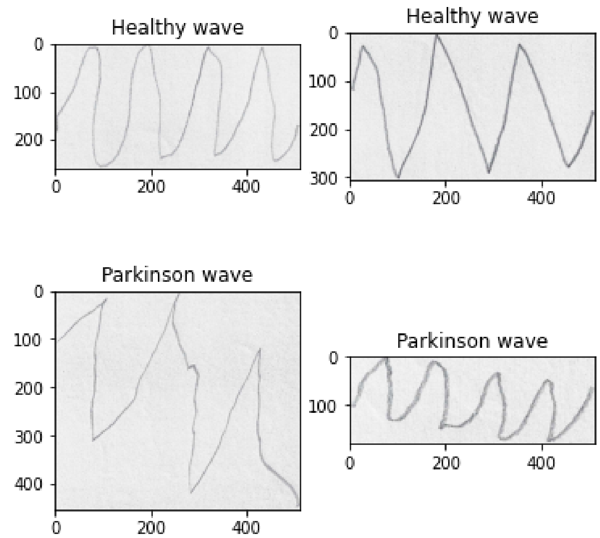


III. Data Preprocessing

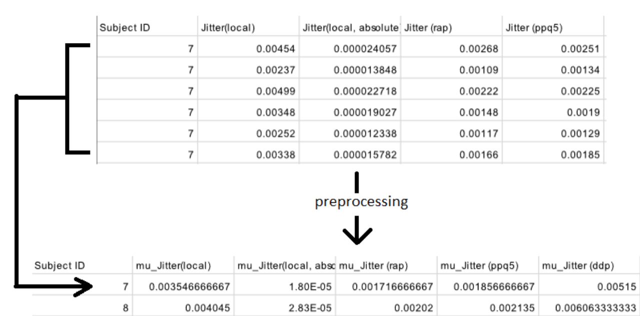
A. Speech data

Ensuring uniformity- Dataset provided as .txt file was first written down to a .csv file and organized for convenience in usage.

Ensuring i.i.d.- Note that we have multiple recordings for each individual, which implies the i.i.d assumption won't hold



over our data. To make sure the plausibility of the independence assumption holds, we use a series of statistical measures to capture the distribution of an individual's recordings of different sound types. Namely, if we have f features per recording, x recordings per individual and n individuals, we have a total of $x * n$ data-points. We transform this to a table with n data-points by performing the following operations- for a given individual, we find 7 statistical measures, namely the mean, median, 10% trimmed mean, 25% trimmed mean, standard deviation, inter-quartile range, and mean absolute deviation per feature. Hence, every individual now has $7 * f$ features associated with it instead of f but we have only one data-point per individual and this ensures that we can assume our data to be i.i.d.



Subject ID	Jitter(local)	Jitter(local, absolute	Jitter (rap)	Jitter (ppq5)
7	0.00454	0.000024057	0.00268	0.00251
7	0.00237	0.000013848	0.00109	0.00134
7	0.00499	0.000022718	0.00222	0.00225
7	0.00348	0.000019027	0.00148	0.0019
7	0.00252	0.000012338	0.00117	0.00129
7	0.00338	0.000015782	0.00166	0.00185

Subject ID	mu_Jitter(local)	mu_Jitter(local, absc	mu_Jitter (rap)	mu_Jitter (ppq5)	mu_Jitter (ddp)
7	0.003546666667	1.80E-05	0.001716666667	0.001856666667	0.00515
8	0.004045	2.83E-05	0.00202	0.002135	0.006063333333

Fig. 1. Reduction of multiple recordings for an individual to 1 datapoint

Merging of datasets- After making all data uniform, and making sure that the number of features in both are same via cleaning and removing extra features, we merge the SL dataset with the MSR dataset. It shall be referred to as MSR+SL dataset from this point on wards.

B. Image data

The images available are blurry and in RGB format. First we convert the images to grayscale (lower dimension), and then do filtering to increase contrast and extract the essential features only for further experimentation.

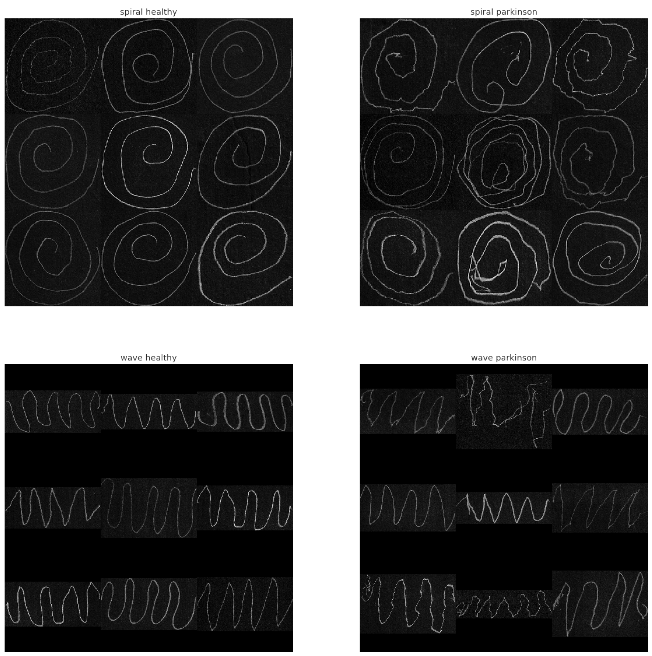


Fig. 2. Original raw images

After filtering and segmenting the images we obtain clearer and contrasting images.

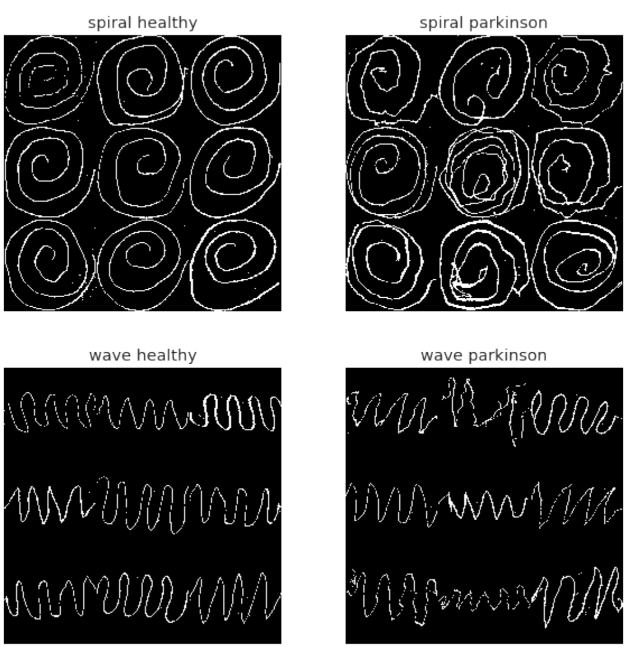


Fig. 3. Filtered images

We use this filtered data for further experiments.

IV. Descriptive Data Analysis

Throughout this section, we will use statistical techniques to analyse and draw inferences about the data.

A. Speech Data

We have 2 different speech datasets - MSR and SL. We have done statistical analysis on the MSR data set and the merged MSR+SL data sets. The preprocessed MSR dataset consists of 40 subjects (rows) and 184 features (columns) while the merged MSR+SL dataset consists of 55 subjects (rows) and 184 features (columns).

We first construct a heatmap selecting the top 20 features which have highest correlation with Parkinson detection for both the datasets. We observe that for the MSR dataset, Jitter and Shimmer measures are most correlated with the status of PD. Jitter in healthy subjects is generally less than one percent of the pitch period, but large values for it are often observed in PD patients. Large values for shimmer variation are also observed in PD patients. In the merged dataset, we observe that features related to the Harmonic to Noise ratio are most correlated with the status of PD. This is a measure of the periodic to non-periodic components of speech, and PD patients are usually observed to have a low value for it, due to hoarseness and incomplete closure of the middle-part of the larynx [8].

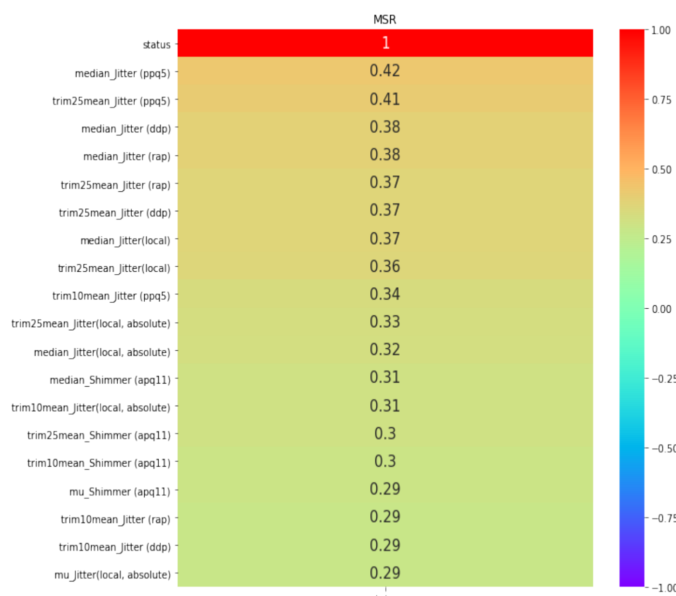


Fig. 4. Heatmap MSR

We make scatter plots between some of the top extracted features to see if they're correlated amongst themselves, and get a better insight of the data.

From the plots we draw the inference that -

- Amplitude variation measures are linearly correlated with frequency variation measures.
- Mean period is directly correlated with the mean pitch and therefore we can use 1 and do away with the other for further analysis.

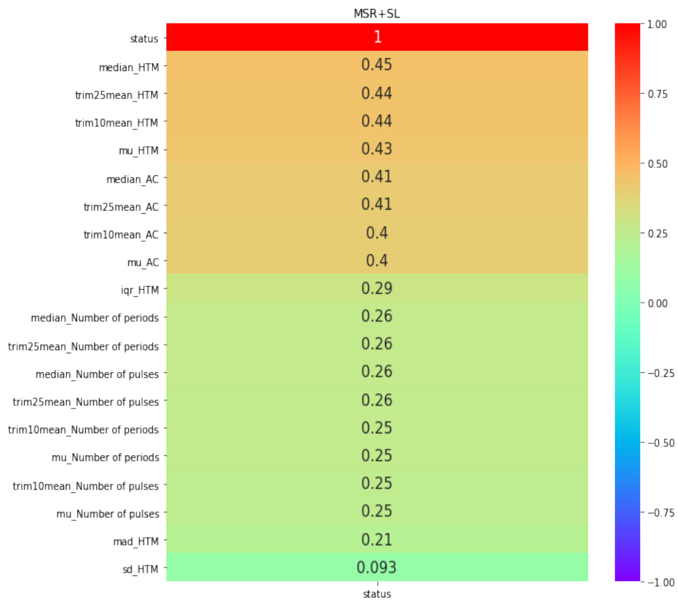


Fig. 5. Heatmap MSR+SL

- Most of the data outliers are from PD positive patients.
- Frequency variation measures (jitter, RAP, PPQ, DAP) and amplitude. variation measures (Shimmer APQ3 , shimmer APQ5, DDA) are highly correlated among themselves.
- HNR is higher for people not having PD whereas the range for people with PD is large.
- Spread 1 has outlier for PD people and Spread 2 has outlier for healthy people.

Finally, we construct the density plots to observe the changes in distribution of fraction of locally unvoiced time frames and the mean sample period before and after merging the datasets. (8)

As can be seen from the plots the density distribution for mean sample period remains the same while the distribution of fraction of locally unvoiced time frames gets skewed towards the right.

From the above analysis we get an understanding of the features in our dataset, the correlations between the features which helps in identifying the relatively more important and distinguishing features and finally the statistical measures of the merged dataset and how it is different from the original ones.

B. Image Data

We have done most of the image preprocessing and feature extraction in the above sections. Here we have tried to visualize the dataset and see if we can directly observe any outliers or distinguishing features just by observing the data. First, we extract the skeletonized images 11.

By plotting the extracted skeleton pixels as points and rescaling we can overlay all of the images on top of each other for better visualization 12.

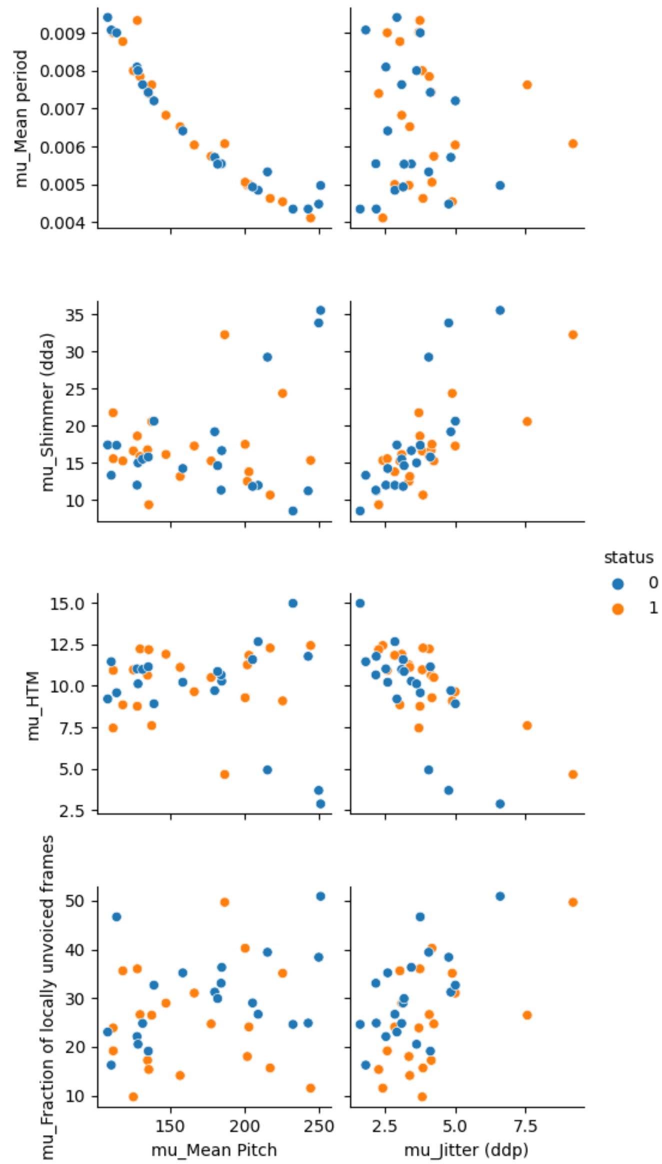


Fig. 6. Scatter plots MSR

As observed the healthy patients are significantly more consistent than the Parkinson's.

V. Baseline Classifiers

We apply a range of baseline classifiers, document our results, and interpret them in this section.

A. Speech Data

For each hypothesis class, we used 10-fold cross validation to validate our results. The class with highest validation accuracy was considered to be most generalizable, and test accuracy is reported on that.

Multi-Layer Perceptron (MLP) performs best for both MSR and MSR+SL. We use a MLP with 5 layers, and hidden layer sizes as (100,50,25) respectively, and we use the 'lbfgs' optimizer instead of gradient descent which uses a quasi-Newton

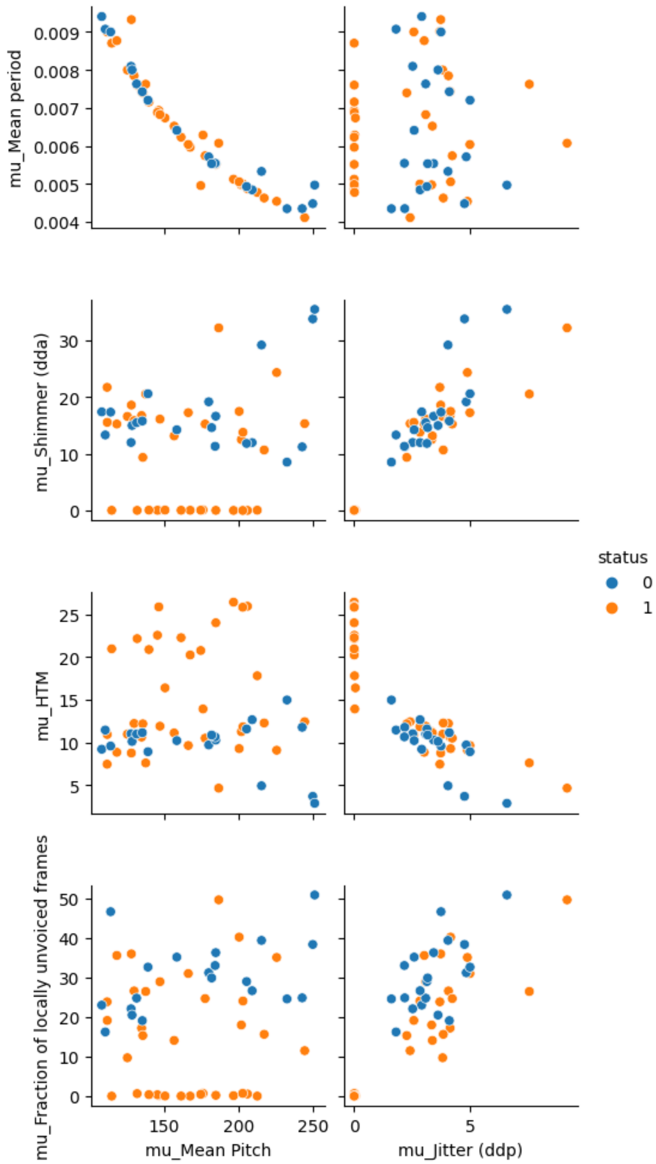


Fig. 7. Scatter plots MSR+SL

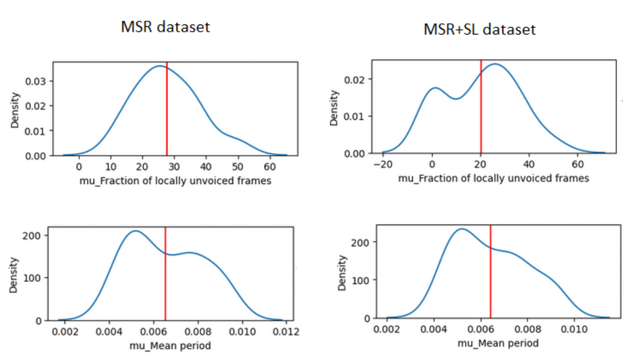


Fig. 8. Density plots

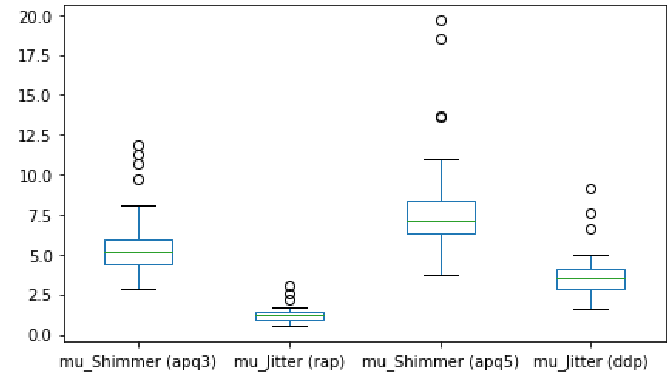


Fig. 9. Box plots MSR

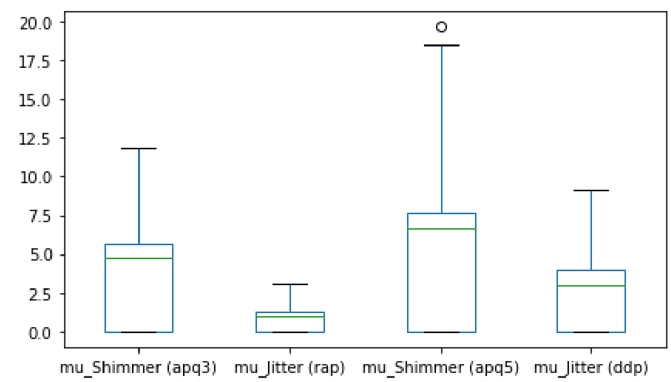


Fig. 10. Box plots MSR+SL

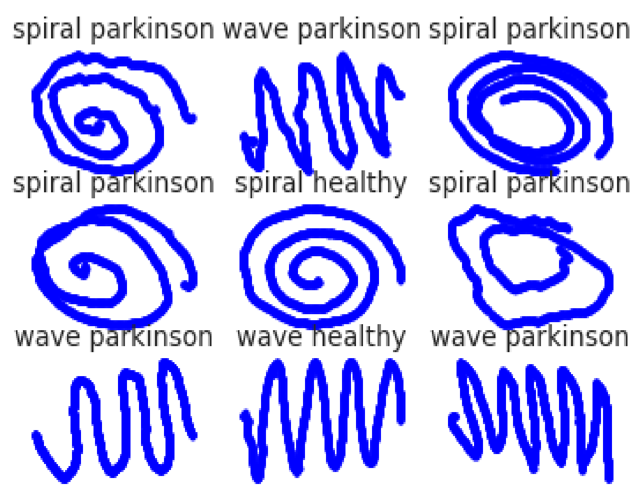


Fig. 11. Extracted skeleton from image

method of optimization. It was shown to be better on smaller datasets [3]. After choosing to model the problem using the MLP as defined, we find the test accuracies to be 0.75 and 0.83 respectively, both better than the validation accuracies. The area under ROC curve for the same approximately comes out to be 0.7 and 0.88 respectively. Note that here several

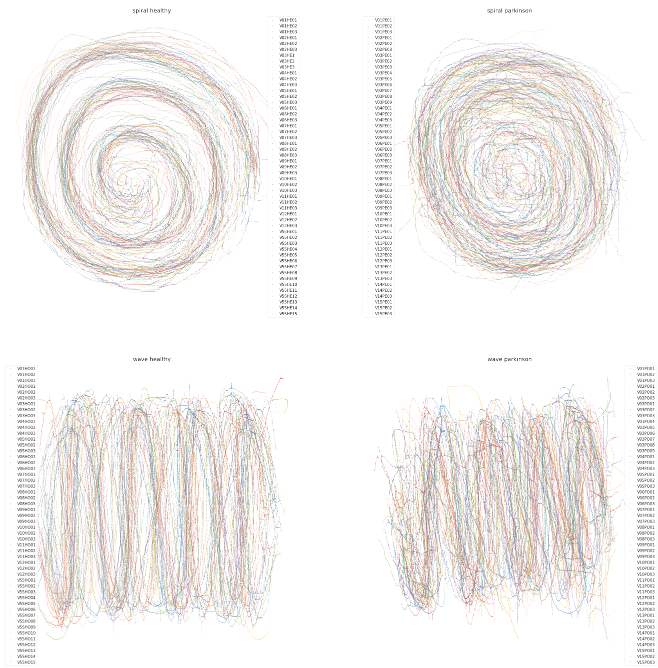


Fig. 12. Overlay images

10 fold cross validation, mean validation accuracy		
Model	MSR dataset	MSR+SL dataset
Logistic Regression	0.667	0.715
Gaussian Naive Bayes	0.558	0.715
k-NN, k=6	0.492	0.695
SVM with rbf kernel	0.45	0.625
Decision Tree	0.5	0.74
Random Forest	0.592	0.695
AdaBoost	0.517	0.735
GradientBoost	0.5	0.74
xgboost	0.667	0.775
MLP	0.7	0.75

Test Accuracy		
Model	**MSR**	**MSR+SL**
Logistic Regression	0.75	0.667
Gaussian Naive Bayes	0.75	0.833
k-NN, k=6	0.75	0.667
SVM with rbf kernel	0.5	0.667
Decision Tree	0.5	0.667
Random Forest	0.75	0.667
AdaBoost	0.75	0.667
GradientBoost	0.5	0.667
xgboost	0.5	0.667
Multi Layer Perceptron	0.75	0.833

models tend to perform as "good" as MLP but that is only specific to the common test-train split. To assess which model would perform well without being affected by the specificity of data, we must factor in the validation accuracy too. Cross validation helps us in understanding how well our model will generalize to the data collected in the future.

Note that consistently merged MSR+SL dataset gives a better validation accuracy and test accuracy than MSR dataset, thus we show that even merging of existing datasets can considerably improve model generalizability for PD classification.

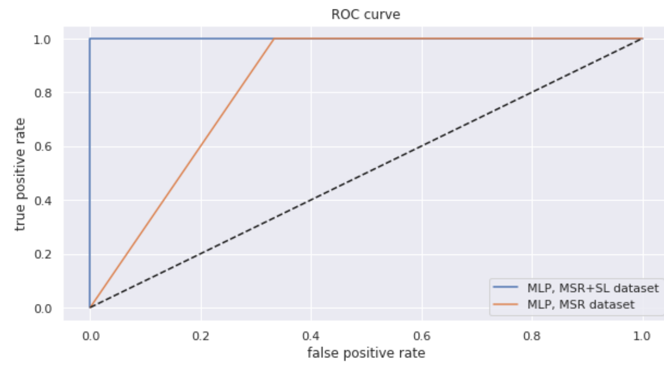


Fig. 13. Same classifier trained on merged dataset performs better

B. Image Data

We have trained a simple CNN and ResNet model on the image data set separately for spirals and wave. The CNN has 4 layers - 2 convolutional and 2 dense. Trained using Adam optimizer with MSE loss and ReLU and Sigmoid activation functions. We have used early stopping criteria to give the best fit model as soon as possible.

ResNet performs better but the difference between the 2 models is not significant. In the later sections we have merged the speech and image data sets to get better accuracy and cross validation scores.

As in case of speech data we have use 10-fold cross validation to validate our results and the class with highest validation score is considered for reporting test accuracy.

10 fold cross validation, mean validation accuracy		
Model	Spiral	Wave
Simple CNN	0.7134	0.6829
ResNet	0.7172	0.7031

Test Accuracy		
Model	Spiral	Wave
Simple CNN	0.6128	0.5732
ResNet	0.6729	0.6325

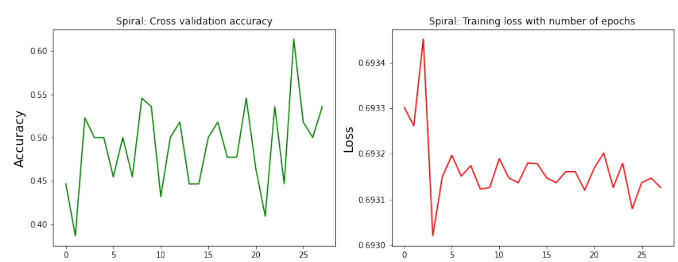


Fig. 14. Simple CNN trained on spiral data

As can be seen from the above plots, the spiral images give higher classification accuracy over the wave images.

VI. Novelty

In the above analysis we have used 3 different data sets, 2 of speech (SL and MSR) and 1 of hand drawings (HD). We

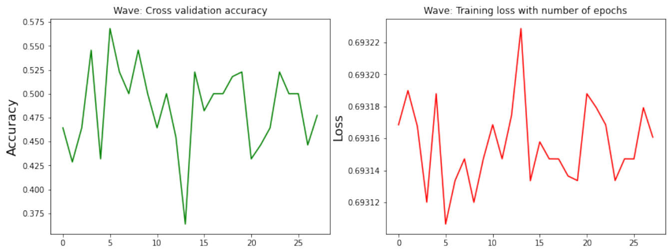


Fig. 15. Simple CNN trained on wave data

have trained different baseline models on the data sets and reported the accuracy and loss obtained.

Throughout our experiments we observed that the individual data sets were very small and could easily be overfitted upon by our models. In general the data for Parkinson's patients is available in small chunks in different formats. We propose a novel method to combine the different types of data sets available to generate a larger one and by doing so increase the accuracy of classification. We have used Generative Adversarial Networks for the same.

1) Generative Adversarial Networks (GANs): Generative modeling is an unsupervised learning task in machine learning that involves automatically discovering and learning the regularities or patterns in input data in such a way that the model can be used to generate or output new examples that plausibly could have been drawn from the original data set. GANs are a clever way of training a generative model by framing the problem as a supervised learning problem with two sub-models: the generator model that we train to generate new examples, and the discriminator model that tries to classify examples as either real or fake. The two models are then trained simultaneously. Figure 16 shows the working of a basic GAN.

2) Data Imputation: Our aim is to combine the SL, MSR and HD data sets. For this we try to first project the data available to a common latent space.

1) Merging MSR and HD data sets
2) Merging(MSR+SL) and HD data sets

In both the case we resize all the images available in the HD data set to 13x14 while maintaining the aspect ratios. Then we merge the 2 data sets. This will lead to many missing values in the data as the features of both the HD and MSR data set are different. To fill in the missing values we use a Generative Adversarial Imputation Network (GAIN), this method is known as data imputation.

The generator (G) observes some components of a real data vector, tries to impute the missing components conditioned on what is actually observed, and outputs a completed matrix. The discriminator (D) then takes a completed vector and attempts to determine which components are real and which were imputed. To ensure that D forces G to learn the desired distribution, D is provided some additional information in the form of a hint vector. The hint reveals to D partial information about the missingness of the original sample, which is used by D to focus its attention on the imputation quality of particular components. This hint ensures that G does in fact learn to generate according to the true data distribution. Figure 17 shows how the GAIN works with a toy example.

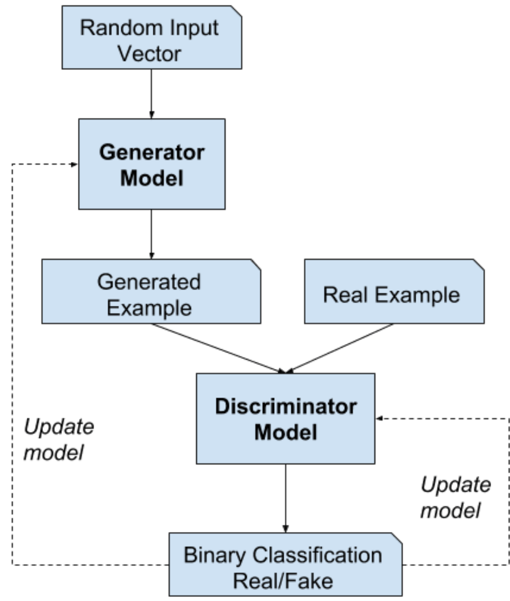


Fig. 16. GAN

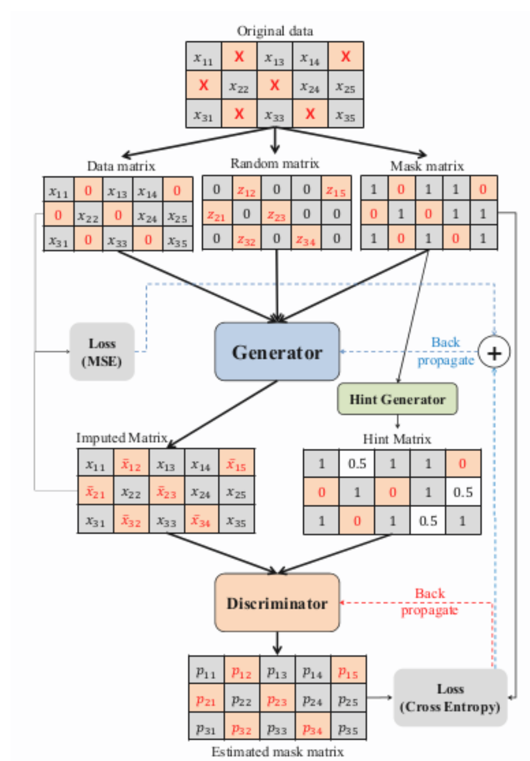


Fig. 17. Example working of GAIN

Using the above method we obtain the imputed data.

3) Baseline Classification on Imputed data: We first have a 20-80 split of test-train data. For each hypothesis class, we used 10-fold cross validation to validate our results. Along with this, we report the test accuracy for each model trained on the whole training dataset. Note that we have datasets of 2 kinds here- one is dataset where MSR dataset is imputed with HD data (referred to as MSR-HD), one is dataset where MSR+SL dataset is imputed with HD data (referred to as MSR+SL-HD). We first perform some baseline experiments on the resulting imputed tables.

10 fold cross validation, mean validation accuracy		
Model	**MSR-HD**	**MSR+SL-HD**
Logistic Regression	0.902	0.909
Gaussian Naive Bayes	0.68	0.548
k-NN, k=4	0.842	0.839
k-NN, k=6	0.843	0.849
SVM with poly kernel	0.847	0.88
SVM with rbf kernel	0.82	0.819
Decision Tree	0.874	0.784
Random Forest	0.908	0.884
AdaBoost	0.923	0.874
GradientBoost	0.901	0.894
xgboost	0.918	0.924
Multi Layer Perceptron	0.892	0.904

Test Accuracy		
Model	**MSR-Drawing**	**MSR+SL-Drawing**
Logistic Regression	0.524	0.713
Gaussian Naive Bayes	0.492	0.583
k-NN, k=4	0.524	0.739
k-NN, k=6	0.524	0.739
SVM with poly kernel	0.524	0.739
SVM with rbf kernel	0.524	0.739
Decision Tree	0.492	0.643
Random Forest	0.508	0.635
AdaBoost	0.492	0.6
GradientBoost	0.635	0.687
xgboost	0.635	0.661
Multi Layer Perceptron	0.524	0.739

Details of every model used- For Logistic Regression we use solver lbfgs; kNN is done with k=4, and 6 with distance measure as euclidean distance; degree 3 polynomial kernel is used for SVM, along with rbf kernel; Our decision tree uses gini criterion, and its depth is upper bound by 3; Random forest uses 50 estimators, with max features 12; Multi-Layer Perceptron is taken with 4 hidden layers of sizes (200,100,50,25) in order, and adam is used as the optimizer.

We observe that using the imputed data gives considerably better validation accuracy even with the baseline models, the validation accuracy is considerably better relative to the respective non-imputed speech and non-imputed image results. There is potential for using imputation to improve generalizability in domains where it is not possible to collect a lot of data.

4) A hybrid model: In the baseline case, we're losing valuable information provided to us by the structure of the image (which in itself is a huge motivation for using CNNs) so we propose a hybrid model as a workaround to this problem.

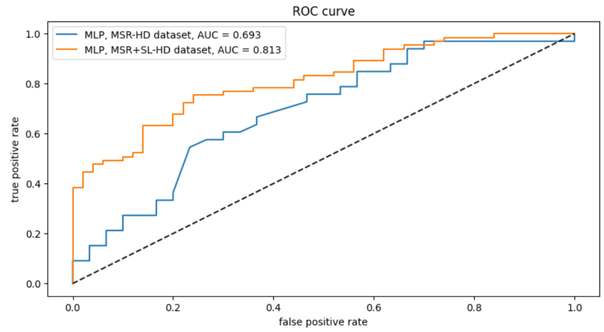


Fig. 18. Merged dataset gives considerably better classification results

We further test a hybrid model in which we use the imputed data set to obtain the latent space (common for speech and image) representations for the speech and image data. We train the obtained speech data using a MLP (classifier 1) and the images using a CNN (classifier 2). Finally we take a weighted average of the classification probabilities obtained from the 2 classifiers and make our final prediction. We set the threshold to 0.5, the class having probability > 0.5 is selected.

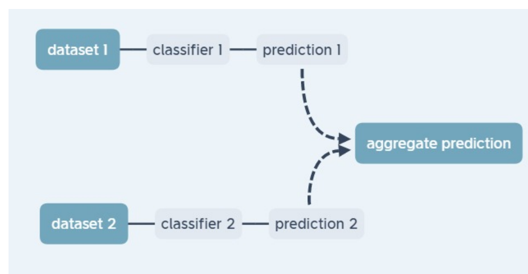


Fig. 19. Flow of the novel hybrid algorithm

$$Aggregate\,prediction = a*(prediction\,1) + b*(prediction\,2)$$

Where $(prediction\,1)$ is the probability that the MLP classifies the speech as belonging to a Parkinson patient and $(prediction\,2)$ is the probability that the CNN classifies the image to be drawn by a Parkinson patient. a, and b are the weights assigned to the prediction probabilities of the 2 models respectively. If $Aggregate\,prediction > 0.5$ then the overall prediction is that the person has Parkinson's else not.

Hybrid Model trained on MSR+HD dataset		
a	b	Test Accuracy
0.1	0.9	0.5238
0.6	0.4	0.6667
0.7	0.3	0.6508
0.8	0.2	0.5714
0.9	0.1	0.5396

VII. Entity Relationship Diagram [24]

Apart from classification of subjects as affected v/s healthy, from a diagnostic perspective our dataset can also be used

Fig. 20. MLP trained on (MSR+HD) imputed speech data

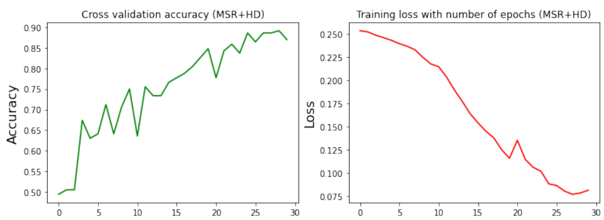


Fig. 21. CNN trained on (MSR+HD) imputed image data

Fig. 22. MLP trained on (MSR+SL+HD) imputed speech data

Fig. 23. CNN trained on (MSR+SL+HD) imputed image data

Hybrid Model trained on MSR+SL+HD dataset		
a	b	Test Accuracy
0.3	0.7	0.8412
0.4	0.6	0.8730
0.6	0.4	0.9206
0.7	0.3	0.9245
0.9	0.1	0.9365

for early diagnosis of PD and related disorders. To explore and detail other areas of application, we build an Entity-Relationship diagram of a toy database management system for companies to get various types of data (like speech, image, gait, etc.) from PD patients via research labs. Such data can be used to improve assistive products which make life easier for PD patients. For example, data related to speech in PD patients can help a company like Amazon improve their Amazon Echo product to understand PD patients more efficiently, data related to gait can help companies like Liftware to create better stabilizing spoons for PD patients, etc.

VIII. Conclusion and Future Work

. After all the experiments, we observed that using the imputed data gave better results and using the hybrid model, where we merged all 3 datasets (SL+MSR+HD), we were able to obtain an accuracy of 93.65%, which is higher than all the baseline models used before with any of the datasets. These experiments show that data imputation using GAIN is indeed a powerful tool as it allows merging of different types of data. The imputed data is more generalizable and gives better results. This technique will be essential for present day machine learning architectures, specially in fields like medical science where we do not have access to large amounts of data and the available data has been taken from different sources and therefore cannot be merged without introducing missing values.

In this report we have used a very novel method for imputation which gave surprisingly good results. In the future we could try to construct more complex GANs which would take in to account the different weightage of different features in the datasets, and impute the data using prior knowledge about the 2 datasets being merged to produce better results. There are also some video datasets for Parkinson's which we have not used in this project. In the future we could also try imputing 3 different modes of data (video+speech+images) and hope to obtain even better classification accuracy.

This github repository contains information about all the experiments we performed and the datasets used - Parkinson's Disease Classification

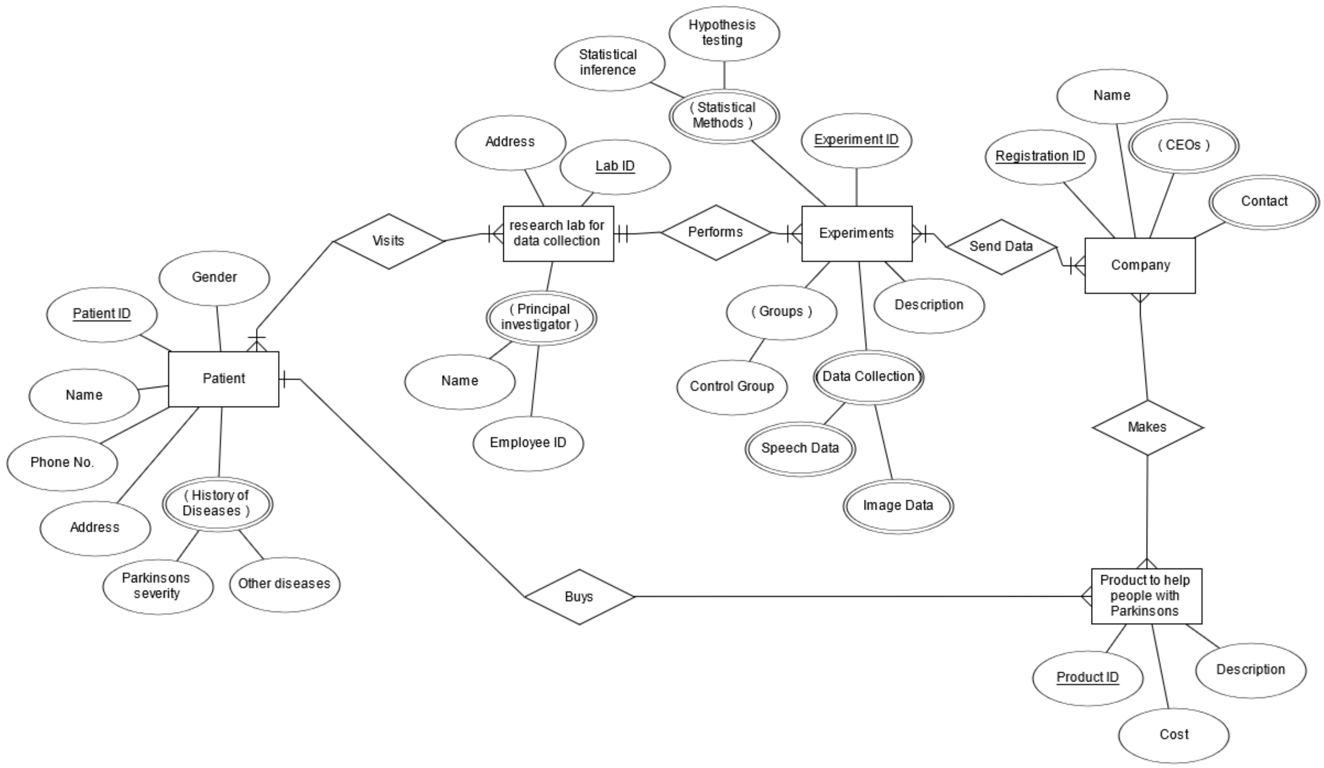


Fig. 24. ER diagram illustrating a toy DBMS for Parkinson's disease management

REFERENCES

[1] Bjoern M. Eskofier, Sunghoon I. Lee, Jean-Francois Daneault, Fatemeh N. Golabchi, Gabriela Ferreira-Carvalho, Gloria Vergara-Diaz, Stefano Sapienza, Gianluca Costante, Jochen Klucken, Thomas Kautz, and Paolo Bonato. Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson's disease assessment. pages 655–658, August 2016.

[2] A. Frid, E. J. Safra, H. Hazan, L. L. Lokey, D. Hilu, L. Manevitz, L. O. Ramig, and S. Sapir. Computational diagnosis of parkinson's disease directly from natural speech using machine learning techniques. In *2014 IEEE International Conference on Software Science, Technology and Engineering*, pages 50–53, 2014.

[3] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In Yee Whye Teh and Mike Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR.

[4] H. Hazan, D. Hilu, L. Manevitz, L. O. Ramig, and S. Sapir. Early diagnosis of parkinson's disease via machine learning on speech data. In *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*, pages 1–4, 2012.

[5] Sandhya Joshi, Deepa Shenoy, G.G. Vibhudendra Simha, P.L. Rrashmi, K.R. Venugopal, and L.M. Patnaik. Classification of Alzheimer's Disease and Parkinson's Disease by Using Machine Learning and Neural Network Methods. pages 218–222, February 2010.

[6] Ken J. Kubota, Jason A. Chen, and Max A. Little. Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures: Machine Learning for PD Wearables. *Movement Disorders*, (9):1314–1326, September 2016.

[7] Daniele Pietrucci, Adelaide Teofani, Valeria Unida, Rocco Cerroni, Silvia Biocca, Alessandro Stefani, and Alessandro Desideri. Can Gut Microbiota Be a Good Predictor for Parkinson's Disease? A Machine Learning Approach. *Brain Sciences*, 10(4):242, April 2020.

[8] M. Bonfiglio R. Chiaramonte. Acoustic analysis of voice in parkinson's disease: systematic review of vocal disability and meta-analysis of studies. *Neurologia*.

[9] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun. Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 17(4):828–834, 2013.

[10] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M.C. Gilardi, and A. Quattrone. Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy. *Journal of Neuroscience Methods*, 222:230–237, January 2014.

[11] Sachin Shetty and Y. S. Rao. SVM based machine learning approach to identify Parkinson's disease using gait analysis. pages 1–5, August 2016.

[12] Vaseekaran Varatharajah. Parkinson's Voice Data - Sri Lanka.

[13] T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins, and R. H. Ghomi. Parkinson's disease diagnosis using machine learning and voice. In *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–7, 2018.

[14] Poonam Zham, Dinesh K. Kumar, Peter Dabnichki, Sridhar Poosapadi Arjunan, and Sanjay Raghav. Distinguishing different stages of parkinson's disease using composite index of speed and pen-pressure of sketching a spiral. *Frontiers in Neurology*, 8:435, 2017.

[15] Andong Zhan, Srihari Mohan, Christopher Tarolli, Ruth B. Schneider, Jamie L. Adams, Saloni Sharma, Molly J. Elson, Kelsey L. Spear, Alistair M. Glidden, Max A. Little, Andreas Terzis, E. Ray Dorsey, and Suchi Saria. Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity: The Mobile Parkinson Disease Score. *JAMA Neurology*, 75(7):876–880, 07 2018.