ParkinSense:

A machine learning approach to classify Parkinson's Disease

Industry / Inhouse Project

Review - I

Team Members :

Aaryan Mahadik, D12C, 42

Kunal Khubchandani, D12C, 37

Raj Tandon, D12C, 62

Bhavika Valecha, D12C, 69

Project Mentor : Mrs. Veena Trivedi

Group number : 50

Content

Introduction to Project

- Parkinson's Disease (PD) is a complex neurodegenerative disorder that affects millions of people worldwide.
- Parkinson's Disease leads to a range of motor and non-motor symptoms. The motor symptoms, including tremors, bradykinesia (slowness of movement), rigidity, and postural instability, significantly impact an individual's ability to perform daily activities and diminish their quality of life.
- Early diagnosis and intervention play a critical role in managing PD and improving the prognosis for affected individuals.
- In recent years, there has been growing interest in leveraging non-invasive and easily accessible data sources, such as speech signals, to aid in the early detection of Parkinson's Disease.
- Researchers have observed distinctive alterations in speech patterns and acoustic features among individuals with Parkinson's Disease,these changes include variations in pitch, intensity, speech rate, phonation time, and articulation, among others.

Lacuna in the existing system

- Research Gaps
- Patient Care Gaps
- Data and Technology Gaps
- Awareness and Education Gaps
- Funding and Resources Gaps
- Collaboration Gaps

Problem Definition

- More than 10 million people are struggling with Parkinson’s Disease.
- Most people don’t have any idea if they have Parkinson’s Disease, hence even if they have Parkinson’s Disease, they aren’t treated.
- The current systems have issues such as limited accuracy, variability in systems, etc.
- The main motive of this project is to build a system which is able to accurately identify whether a person has Parkinson’s Disease or not using the data in their speech sample.

Literature Survey

Sr. no	Title	Year of publication	Abstract
1.	A Deep Learning Based Method for Parkinson's Disease Detection Using Dynamic Features of Speech	2021	They propose to apply Bidirectional long-short term memory (LSTM) model to capture time-series dynamic features of a speech signal for detecting PD.
2.	Development of Intelligent Parkinson Disease Detection System Based on Machine Learning Techniques Using Speech Signal	2021	We proposed an enhanced methodology based on utilizing SMOTE to balance the dataset, due to the available dataset is imbalanced. then adopted extra tree classifier with k-fold technique after we balanced the dataset with SMOTE.

Literature Survey

Sr. no	Title	Year of publication	Abstract
3.	Speech-based solution to Parkinson's disease management	2021	This paper presents pilot experiment to detect presence of dysarthria in speech and detect level of severity based on deep learning approach. Automated feature extraction and classification using convolutional neural network shows 77.48% accuracy on test samples of TORGO database with five fold validation. Using transfer learning, system performance is further analyzed for gender specific performance as well as in detection of severity of disease.
4.	Parkinson's Disease Detection from Voice and Speech Data Using Machine Learning	2020	Their proposed model introduced different data preprocessing methods, such as data standardization, multicolinearity diagnosis, dimensionality reduction technique to improve the quality of data.

Literature Survey

Sr. no	Title	Year of publication	Abstract
5.	Gradient boosting for Parkinson's disease diagnosis from voice recordings	2020	Parkinson's Disease (PD) is a clinically diagnosed neurodegenerative disorder that affects both motor and non-motor neural circuits. Speech deterioration (hypokinetic dysarthria) is a common symptom, which often presents early in the disease course. Machine learning can help movement disorders specialists improve their diagnostic accuracy using non-invasive and inexpensive voice recordings.
6.	Parkinson's Disease Diagnosis Using Machine Learning and Voice	2019	This paper explores the effectiveness of using supervised classification algorithms, such as deep neural networks, to accurately diagnose individuals with the disease.

Methodology Employed

Feature Extraction: Various acoustic, prosodic, and phonemic features will be extracted from the preprocessed speech signals. These features will capture unique characteristics of speech patterns associated with PD.

Feature Selection: Feature selection techniques will be employed to identify the most relevant and discriminative features for PD detection. This step aims to enhance model efficiency and interpretability.

Machine Learning Model Development: State-of-the-art machine learning algorithms, such as support vector machines (SVMs), random forests, and deep neural networks, will be implemented and trained on the selected features. Ensemble methods may be employed to improve model performance.

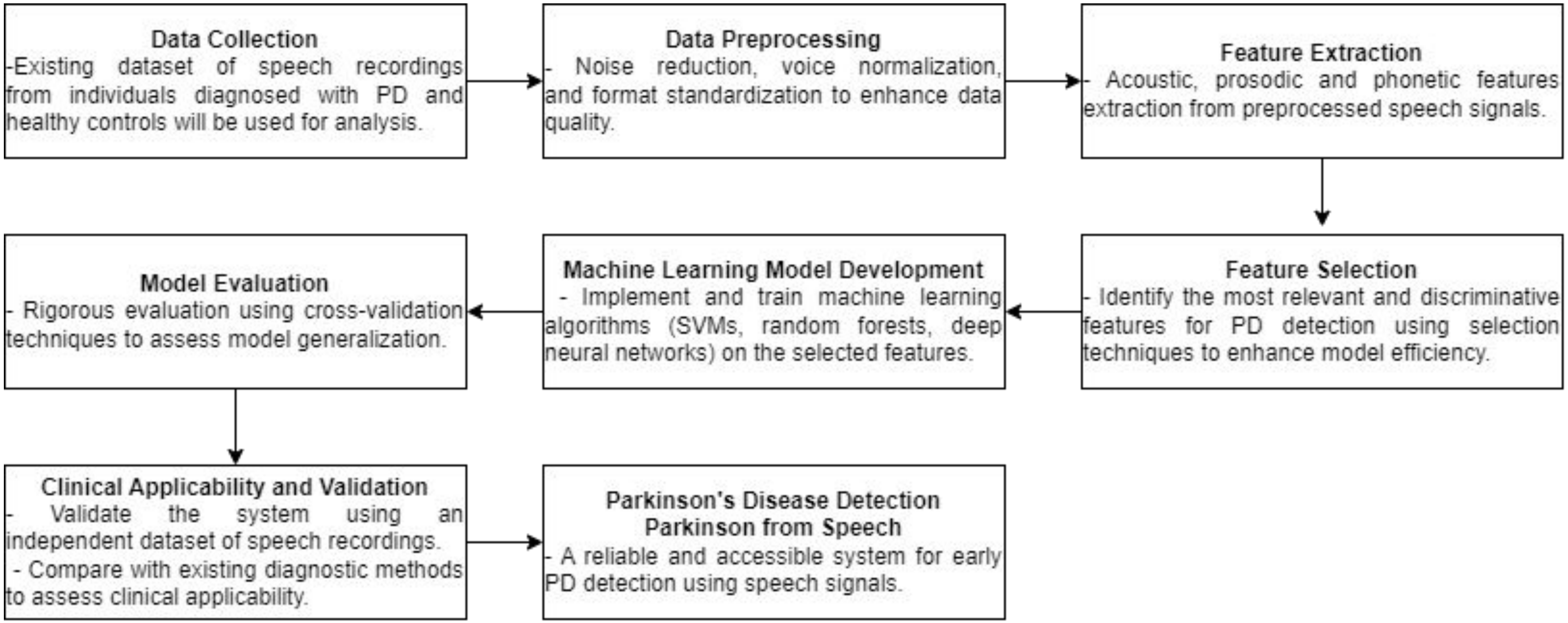
Model Evaluation: The developed models will be rigorously evaluated using cross-validation techniques to assess their generalization capabilities. Stratified sampling will ensure balanced representation of PD and healthy cases in each fold.

Clinical Applicability and Validation: The proposed system will be validated using an independent dataset of speech recordings. The system's performance will be compared with existing diagnostic methods, including clinical assessments and neuroimaging techniques, to assess its clinical applicability.

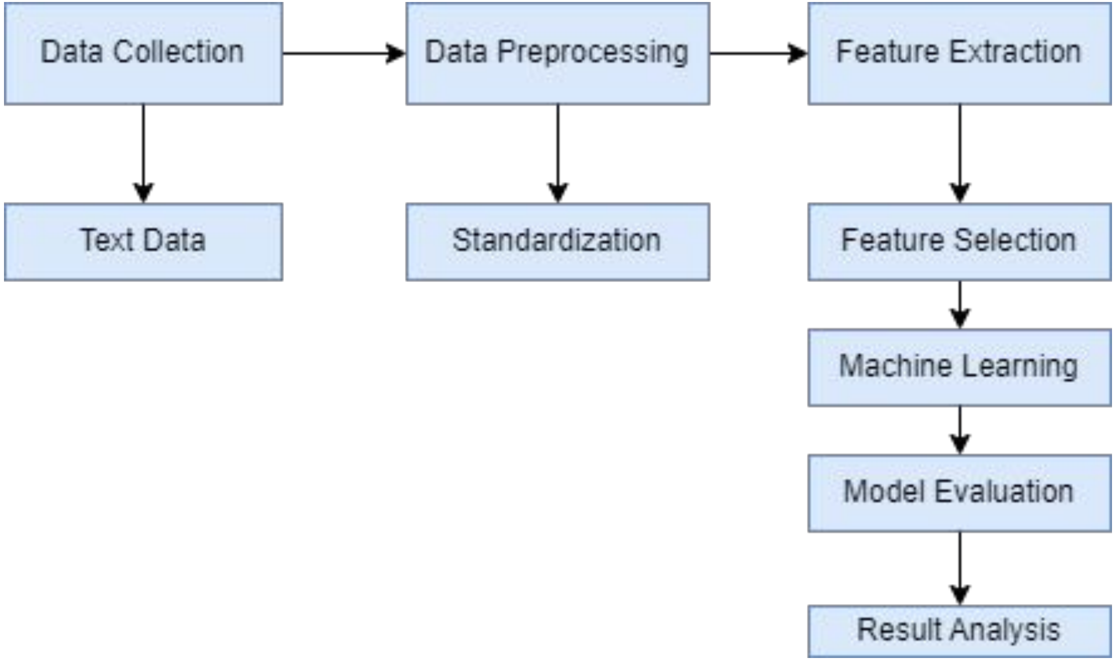
Requirements

SOFTWARE	HARDWARE	TOOLS	CONSTRAINTS
MATLAB	Standard computer with sufficient computational power to run machine learning algorithms.	Tableau	Constraints related to MATLAB installation in a machine

Block Diagram



Modular Diagram



Dataset

Attribute Information

The attribute information of the data set is briefly described below.

Attributes in the matrix columns:

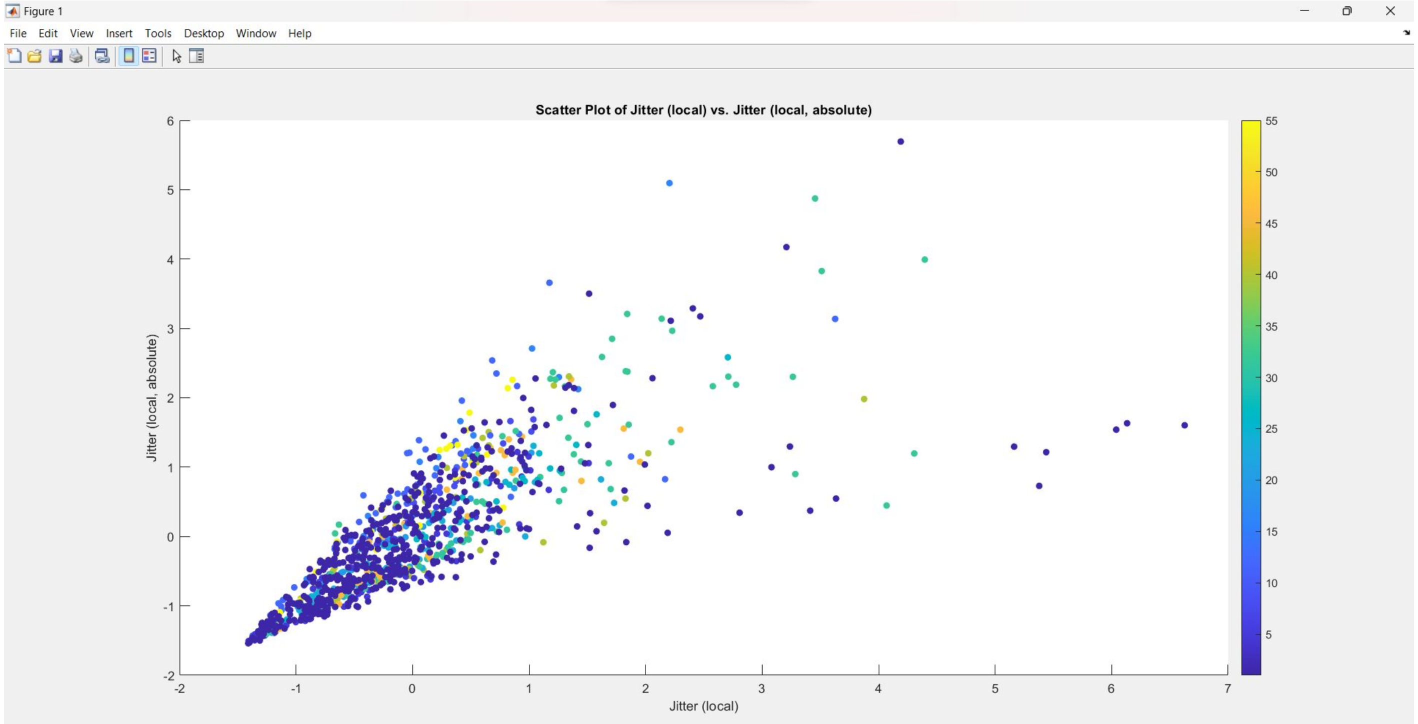
- name: ASCII subject name and recording number
- MDVP:Fo(Hz): Average vocal fundamental frequency
- MDVP:Fhi(Hz): Maximum vocal fundamental frequency
- MDVP:Flo(Hz): Minimum vocal fundamental frequency
- MDVP:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP: Various measures indicating fundamental frequency variation
- MDVP:Shimmer, MDVP:Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ, Shimmer:DDA: Various measures indicating amplitude variation
- NHR, HNR: Two measures representing the ratio of noise to tonal components in the voice
- status: Health status of the subject - (1) Parkinson's, (0) healthy
- RPDE, D2: Two nonlinear measures of dynamical complexity
- DFA: Signal fractal scaling exponent
- spread1, spread2, PPE: Three nonlinear measures indicating fundamental frequency variation

Dataset

Data Collection

The source for numeric data is from the Kaggle dataset, encompassing key attributes like MDVP:Fo(Hz), MDVP:Fhi(Hz), MDVP:Flo(Hz), jitter and shimmer measures, NHR, HNR, and various nonlinear complexity measures. The dataset includes voice modulations from individuals with Parkinson's disease and a control group. The control group consists of individuals who do not have Parkinson's disease and serves as a reference for comparative analysis. This inclusion of a control group allows for a thorough examination of the distinctive speech patterns associated with Parkinson's disease, enabling the development of a robust and reliable detection model. The dataset's diversity, covering both affected individuals and the control group, contributes to the generalizability and validity of our proposed system for Parkinson's disease detection using speech signals.

Implementation
Support Vector Machines (Accuracy = 84.52%)



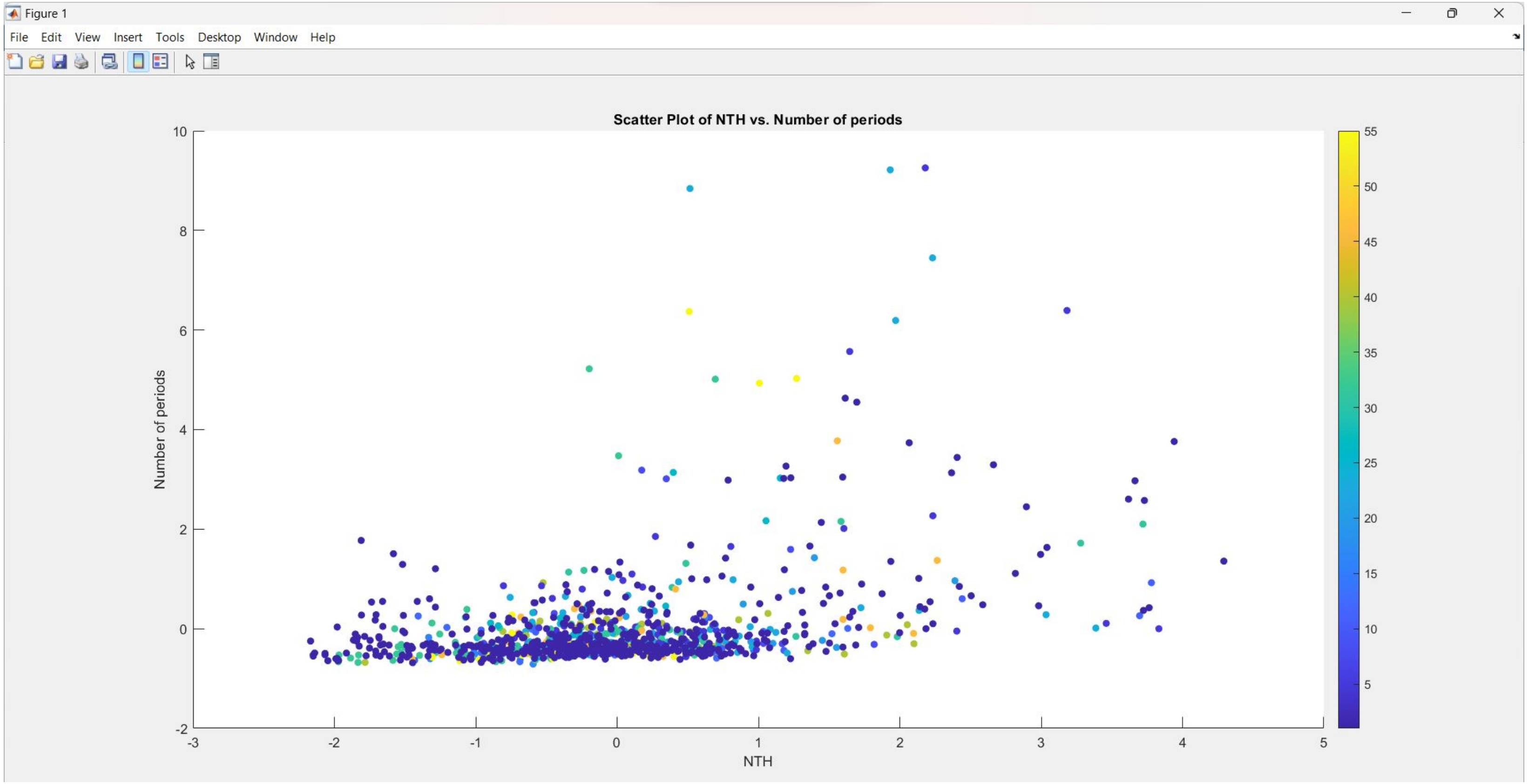


Figure 1
File Edit View Insert Tools Desktop Window Help
Scatter Plot of NTH vs. Number of periods
Number of periods
NTH
10
8
6
4
2
0
-2
-3
-2
-1
0
1
2
3
4
5
55
50
45
40
35
30
25
20
15
10
5

Implementation
K-Nearest Neighbors (Accuracy = 50.00%)

Command Window

New to MATLAB? See resources for Getting Started.

```
>> test = load("C:\Users\dell\OneDrive\Documents\MATLAB\test_data.txt");
train = load("C:\Users\dell\OneDrive\Documents\MATLAB\train_data.txt");
train_features = train(:, 1:end-1);
trainLabels = train(:,end);
K=5;
knn_model = fitcknn(train_features,trainLabels,'NumNeighbors',K);
predictions=predict(knn_model,test);
testLabels = test(:,end);
confusionMatrix=confusionmat(testLabels,predictions);
>> accuracy = sum(diag(confusionMatrix))/sum(confusionMatrix(:));
>> disp(['Accuracy: ',num2str(accuracy)]);
Accuracy: 0.5
>>
```

Implementation
XGBoost (Accuracy = 87.17%)

```
Accuracy score on test data:  0.8717948717948718
```

Conclusion

The proposed project aims to develop a reliable and accessible system for the early detection of Parkinson's Disease using speech signals. By leveraging advanced signal processing techniques and machine learning algorithms, the system will be capable of extracting relevant features from speech recordings and accurately discriminating between individuals with PD and healthy controls. This research has significant implications for healthcare diagnostics and could lead to early intervention, improved disease management, and enhanced quality of life for individuals living with PD.

References

- https://en.wikipedia.org/wiki/Parkinson%27s_disease
- https://en.wikipedia.org/wiki/Artificial_intelligence
- https://en.wikipedia.org/wiki/Machine_learning
- https://en.wikipedia.org/wiki/Deep_learning
- https://www.google.com/
- https://archive.ics.uci.edu/dataset/301/parkinson+speech+dataset+with+multiple+types+of+sound+recordings