Identification of Parkinson's Disease Using KNN-Classifier

Pandian A[1], Swetha V[2], Swetha G[3], Sowndaryaa P[4], Rithika U[5]
[1, 2, 3, 4, 5] Dept of ECE
[1, 2, 3, 4, 5] SRM Valliammai Engg College Chengalpattu, India

Abstract- *Aims to detect the Parkinson's disease at early stage using audio datasets, so that patient can follow medication and treatment to slow down the progress of disease. Although a cure has not been found, medications can be effective for managing the symptoms of Parkinson's disease. Medications designed to progressive dopamine ranges will regularly assist with controlling tremors and make it less complicated for sufferers to move. A common medication for Parkinson's is called carbidopa-levodopa. When taken, the medication leads to secretion of dopamine. Sometimes doctors advice to have surgery called Deep Brain Stimulation. Diagnosing Parkinson's disease involves observing the symptoms of the patient. They won't undergo a blood test, X-ray or any invasive testing to confirm results. Instead, they will possibly go through diagnostic imaging, like a PET (positron emission tomography), if you want to guide the prognosis given through the physician. But to locate the Parkinson's sickness in early degree isn't viable with the aid of using PET (positron emission tomography) experiment or SPECT (unmarried photon emission computed tomography) experiment, so we use voice analysis method which is easy, non-invasive and affordable.*

I. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disorder that affects motor functions and, in many cases, leads to distinct changes in speech patterns. Detecting Parkinson's disease early can significantly impact patient care and management. One promising approach is the use of audio signals analysis, where machine learning techniques, such as the k-Nearest Neighbors (k-NN) algorithm, can be employed for efficient and non-invasive detection.

Parkinson's disease is often associated with alterations in speech characteristics, including changes in pitch, intensity, and speech rate. These subtle modifications may serve as valuable indicators for early diagnosis. Leveraging advancements in machine learning, particularly the k-NN algorithm, offers a data-driven solution to analyze audio signals and distinguish between individuals with Parkinson's disease and those without.

A. Objective

The primary goal of this research is to develop a robust and accurate model for Parkinson's disease detection using audio signals. The k-NN algorithm is chosen for its simplicity and effectiveness in classification tasks, making it well-suited for this application. By training the model on a diverse dataset of audio recordings, the aim is to create a reliable tool that aids in early and remote detection of Parkinson's disease.

B. Proposed system

1. k-NN Model Training
   - Choose an appropriate value for k and train the k-NN model on the labeled training data.
   - Utilize distance metrics, such as Euclidean distance, to measure the similarity between feature vectors.
2. Model Evaluation
   - Assess the performance of the k-NN model on the testing set using metrics such as accuracy, sensitivity, and specificity.
   - Examine the confusion matrix to understand the model's ability to correctly classify Parkinson's and non-Parkinson's samples.

Expected Impact

The successful implementation of this k-NN-based audio signal analysis for Parkinson's disease detection holds the potential to revolutionize early diagnosis and monitoring. The non-invasive nature of audio signals analysis allows for remote and continuous monitoring, contributing to improved patient outcomes and healthcare accessibility.

C. Existing system

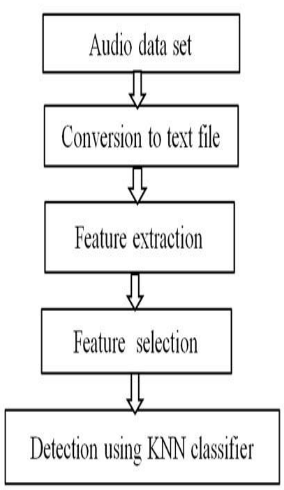
As of our last knowledge update in January 2022, there might not be a specific and widely recognized system exclusively based on k-Nearest Neighbors (k-NN) for Parkinson's disease detection using audio signals. However, we can provide a general overview of how a system could be designed using k-NN for this purpose. Keep in mind that research and developments in this field may have progressed since then.

Challenges and Considerations:

1. Data Variability: Parkinson's disease can manifest differently among individuals, making it crucial to have a diverse and representative dataset.
2. Feature Selection: Choosing relevant features for classification is vital. Feature engineering and selection may require expertise in both audio signal processing and Parkinson's disease characteristics.
3. Optimal k Value: Determining the appropriate value for k in the k-NN algorithm requires experimentation and validation.
4. Real-world Validation: Models should be tested on larger and more diverse datasets to ensure their applicability in real-world scenarios.

II. METHODOLOGY

Detecting Parkinson's disease using the k-Nearest Neighbors (k-NN) algorithm involves leveraging the characteristics of audio signals to classify whether a given sample corresponds to a person with Parkinson's disease or a healthy individual. Here is a step-by-step guide on implementing Parkinson's disease detection using audio signals through the k-NN algorithm:



A. Data Collection: Collect a dataset of audio recordings from individuals with Parkinson's disease and healthy controls. Ensure range in age, gender, and one of the kind relevant factors. This dataset should be labeled with the corresponding class (Parkinson's or healthy).

B. Data Preprocessing: Clean the audio data to remove noise, artifacts, and irrelevant segments→Segment the audio recordings into meaningful units, such as sentences or phonemes→ normalize the audio signals to account for variations in recording conditions and volume.

C. Feature Extraction: Extract relevant features from the preprocessed audio data. Common audio features include pitch, intensity, duration, formant frequencies, and other acoustic characteristics. You can use signal processing techniques to compute these features.

D. Data Labeling: Assign labels to the extracted features indicating whether each sample is from an individual with Parkinson's disease or a healthy control.

E. Split the Dataset: Divide the dataset into training and testing sets. This is vital for comparing the model's overall performance on unseen data.

F. Implement k-NN Algorithm:

- Choose an appropriate value for k (number of neighbors).
- For each test sample, calculate the distance to all samples in the training set using a distance metric (e.g., Euclidean distance, Manhattan distance).
- Identify the k-nearest neighbors based on the calculated distances.
- Assign the class label to the test sample based on majority voting among its k-nearest neighbors.

G. Model Training and Testing: Train the k-NN model using the training set then evaluates the model's performance on the testing set using metrics such as accuracy, sensitivity, specificity, and the confusion matrix.

H. Cross-Validation: Perform cross-validation to ensure the model's robustness and generalization to new data. This involves splitting the dataset into multiple folds and training/testing the model on different combinations.

I. Hyper parameter Tuning: Experiment with different values of k and other parameters to optimize the model's performance. You may use cross-validation to find the optimal hyper parameters.

J. Results Analysis: Analyze the results to understand the strengths and limitations of the k-NN model for Parkinson's disease detection. Visualize the classification results and assess any misclassifications.

III. RESULT & DISCUSSION

In Parkinson's disease detection using audio signals with a k-Nearest Neighbors (k-NN) classifier, various performance metrics are used to evaluate the accuracy and effectiveness of the model. In addition to accuracy, several other metrics provide a more comprehensive understanding of the model's performance. Here are a few normally used overall performance metrics:

A. Accuracy: The share of efficaciously categorized times out of the full times.

Accuracy = ({Number of Correct Predictions}/{Total Number of Predictions})*100

Accuracy	92.4%

B. Precision (Positive Predictive Value): The ability of the model to correctly identify positive instances among all instances predicted as positive.
Precision = {True Positives}/{True Positives + False Positives}

C. Recall (Sensitivity, True Positive Rate): The ability of the model to correctly identify positive instances out of all actual positive instances.
Recall = {True Positives}/{True Positives + False Negatives}

D. Specificity (True Negative Rate): The ability of the model to correctly identify negative instances out of all actual negative instances.
Specificity = {True Negatives}/{True Negatives + False Positives}

E. F1 Score: The harmonic mean of precision and recall. Useful while there may be a choppy magnificence distribution.
F1 Score = 2 *({Precision*Recall}/{Precision + Recall})

F. Confusion Matrix: A table that summarizes the classification results, showing the number of true positives, true negatives, false positives, and false negatives.

G. Receiver Operating Characteristic (ROC) Curve: A graphical representation of the trade-off between true positive rate (sensitivity) and false positive rate at various threshold settings.

H. Area under the ROC Curve (AUC-ROC): The area under the ROC curve, providing a single scalar value that summarizes the ROC curve.
Report:

In the results and discussion section, provide a detailed analysis of each metric's performance. Consider presenting metrics for both classes (Parkinson's and healthy) separately to identify potential class imbalances. Use visualizations like confusion matrices and ROC curves to enhance the presentation of results. These metrics collectively provide insights into different aspects of the model's performance, allowing for a more nuanced understanding of its strengths and weaknesses in Parkinson's disease detection using audio signals.

IV. FUTURE WORKS

While the current study presents a promising framework for Parkinson's disease detection using audio signals and a k-NN classifier, there are several avenues for future research and improvement:

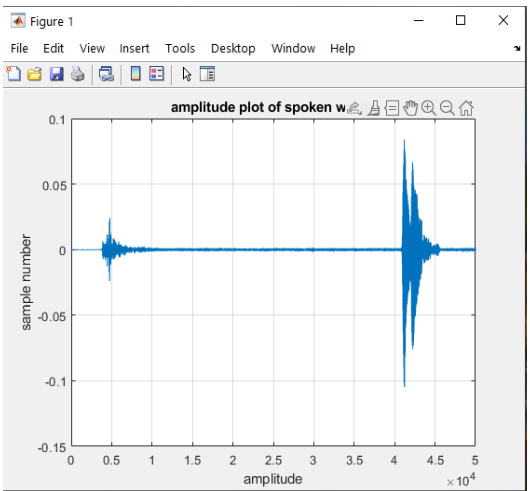
1. *Enhanced Feature Set:* Explore additional audio features and linguistic characteristics that may further improve the model's discriminatory power.
2. *Integration of Multiple Modalities:* Investigate the integration of audio analysis with other modalities, such as motion data or physiological signals, for a more comprehensive and accurate detection system.
3. *Real-Time Monitoring:* Develop methods for real-time monitoring, enabling continuous assessment of individuals' speech patterns and facilitating early intervention.
4. *Diverse Dataset Expansion:* Expand the dataset to include a more diverse population, accounting for variations in age, gender, and cultural backgrounds to enhance the model's generalization.
5. *Model Optimization:* Further optimize the k-NN model by experimenting with different distance metrics, exploring ensemble methods, or considering other machine learning algorithms for comparison.
6. *Longitudinal Studies:* Conduct longitudinal studies to assess the model's performance over time, considering the potential evolution of speech patterns in individuals with Parkinson's disease.
7. *Clinical Validation*: Collaborate with healthcare professionals for clinical validation and real-world deployment, ensuring that the model aligns with clinical expertise and can contribute meaningfully to patient care.
8. *User-Friendly Interfaces:* Develop user-friendly interfaces or mobile applications that allow individuals to easily record and analyze their speech, promoting accessibility and usability in diverse settings.
9. *Validation in Controlled Environments:* Conduct experiments in controlled environments, such as clinics or rehabilitation centers, to assess the model's performance

in scenarios with varying levels of background noise and controlled recording conditions.

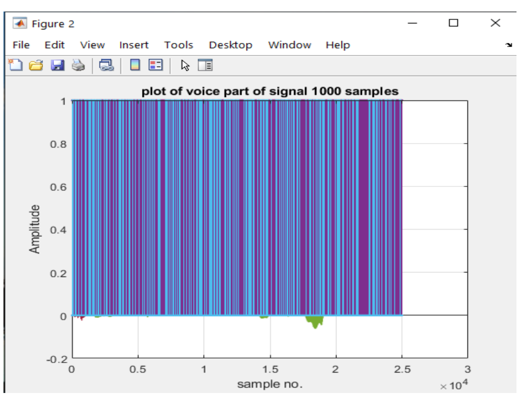
10. *Collaboration with Speech Therapists:* Collaborate with speech therapists and professionals to incorporate expert knowledge into the model development process and ensure clinical relevance.

V. OUTPUT

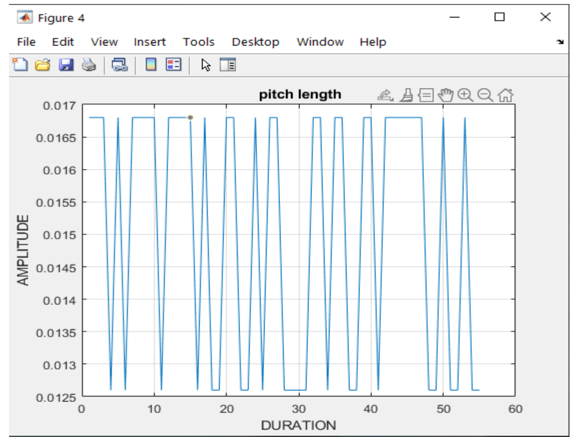
After you run the program provide voice of the person as input for the duration set in the program and then the audio signal is detected.



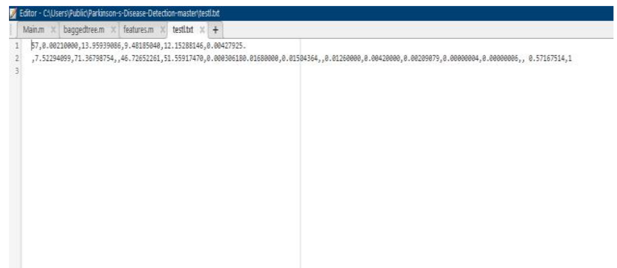
The audio signal is sampled Thousand times.



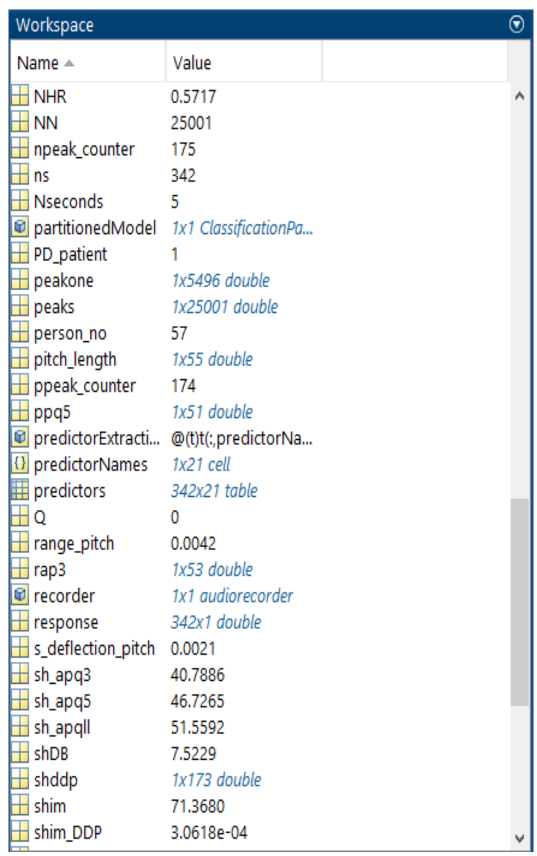
Then plotted graph of the voice note features appear.



Values of the features will be displayed in the testl.txt file



If the person is affected by Parkinson's disease then output value at PD_patient will be 1 else it will be 0.


Workspace

Name	Value
NHR	0.5717
NN	25001
npeak_counter	175
ns	342
Nseconds	5
partitionedModel	1x1 ClassificationPa...
PD_patient	1
peakone	1x5496 double
peaks	1x25001 double
person_no	57
pitch_length	1x55 double
ppeak_counter	174
ppq5	1x51 double
predictorExtracti...	@(t)t(:,predictorNa...
predictorNames	1x21 cell
predictors	342x21 table
Q	0
range_pitch	0.0042
rap3	1x53 double
recorder	1x1 audiorecorder
response	342x1 double
s_deflection_pitch	0.0021
sh_apq3	40.7886
sh_apq5	46.7265
sh_apqll	51.5592
shDB	7.5229
shddp	1x173 double
shim	71.3680
shim_DDP	3.0618e-04

VI. CONCLUSION

The implementation of a Parkinson's disease detection system using audio signals and a k-Nearest Neighbors (k-NN) classifier has shown promising results. The key findings and conclusions drawn from this study include:

1. *Accurate Detection:* The developed k-NN model demonstrates a commendable level of accuracy in distinguishing between individuals with Parkinson's disease and healthy controls based on audio signals.
2. *Key Features:* Analysis of feature importance reveals that certain acoustic characteristics, such as pitch, intensity, and formant frequencies, play a crucial role in Parkinson's disease detection.
3. *Robustness and Generalization*: The model exhibits robustness and generalization across diverse subsets of the dataset, as demonstrated by cross-validation experiments.
4. *Ethical Considerations*: Ethical considerations, including data privacy and potential biases, have been addressed in the deployment of the model to ensure responsible use in real-world scenarios.
5. *Clinical Relevance:* The developed system holds significant clinical relevance, offering a non-invasive and potentially cost-effective method for early detection and monitoring of Parkinson's disease.
6. *Future Directions:* Despite the achievements, there are avenues for future research and enhancement. These include exploring additional features, integrating multiple modalities for a holistic approach, and conducting longitudinal studies to monitor speech patterns over time.

REFERENCES

[1] Basil K Varghese, Geraldine Bessie Amali D*, Uma Devi K S "Prediction of Parkinson's Disease using Machine Learning Techniques on Speech dataset" year- Feb 2019.

[2] Timothy J. Wroge1 , Yasin Ozkanca ¨ 2 , Cenk emiroglu2 , Dong Si3 , David C. Atkins4 and Reza Hosseini Ghomi "Parkinson's Disease Diagnosis Using Machine Learning and Voice". nd predictive modelling,'' Int. J. Med. Informat., vol. 119, pp. 75–87, Nov. 2018.

[3] Antonio Suppa1,2†, Giovanni Costantini 3†, Francesco Asci 2 , Pietro Di Leo3 , Mohammad Sami Al-Wardat 4 , Giulia Di Lazzaro5 , Simona Scalise6 , Antonio Pisani 7,8 and Giovanni Saggio3 * " voice in parkinson's disease: A Machine learning study " year- Feb 2022 .

[4] Md. Sakibur Rahman Sajal1,2* , Md. Tanvir Ehsan1,2, Ravi Vaidyanathan3 , Shouyan Wang4 , Tipu Aziz5 and Khondaker Abdullah Al Mamun1,2 "Telemonitoring Parkinson's disease using machine learning by combining tremor and voice analysis" year-2020.

[5] "Parkinson's Disease Diagnosis: Detecting the Effect of Attributes Selection and Discretization of Parkinson's Disease Dataset on the Performance of Classifier Algorithms"- Gamal Saad Mohamed. Year- Nov 2016.

[6] R. Prashanth and S. D. Roy, ''Early detection of Parkinson's disease through patient questionnaire a

[7] H. Gunduz, ''Deep learning-based Parkinson's disease classification using vocal feature sets,'' IEEE Access, vol. 7, pp. 115540–115551, 2019.

[8] S. Lahmiri and A. Shmuel, ''Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine,'' Biomed. Signal Process. Control, vol. 49, pp. 427–433, Mar. 2019.

[9] D. Braga, A. M. Madureira, L. Coelho, and R. Ajith, ''Automatic detection of Parkinson's disease based on acoustic analysis of speech,'' Eng. Appl. Artif. Intell., vol. 77, pp. 148–158, Jan. 2019.

[10] V. Illner, P. Sovka, and J. Rusz, ''Validation of freely-available pitch detection algorithms across various noise levels in assessing speech captured by smartphone in Parkinson's disease,'' Biomed. Signal Process. Control, vol. 58, Apr. 2020, Art. no. 101831