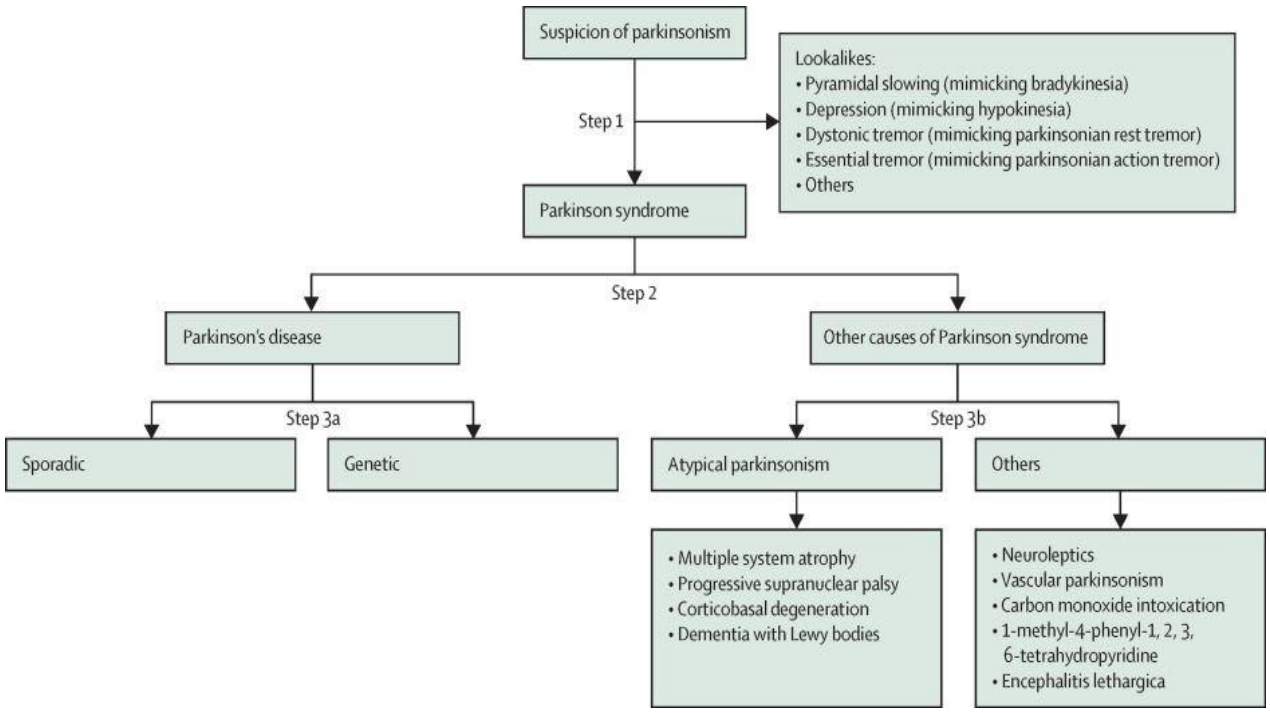
Sreevarshan S (TL)
Vishwa S (M1 lead)
Sreehari Pranesh K (M2 Lead)
Vinith Kumar S (M3 Lead)

Project Design Phase -I
Solution architecture

Parkinson's Disease Detection using ML

General Detection of PD through medical process:



System Architecture:

Machine learning has given computer systems the ability to automatically learn without being explicitly programmed. In this, the author has used three machine learning algorithms (Logistic Regression, KNN, and Naïve Bayes). The architecture diagram describes the high-level overview of major system components and important working relationships. It represents the flow of execution and it involves the following five major steps:

The architecture diagram is defined with the flow of the process which is used to refine the raw data and used for predicting the Parkinson's data.

The next step is preprocessing the collected raw data into an understandable format.

Then we have to train the data by splitting the dataset into train data and test data.

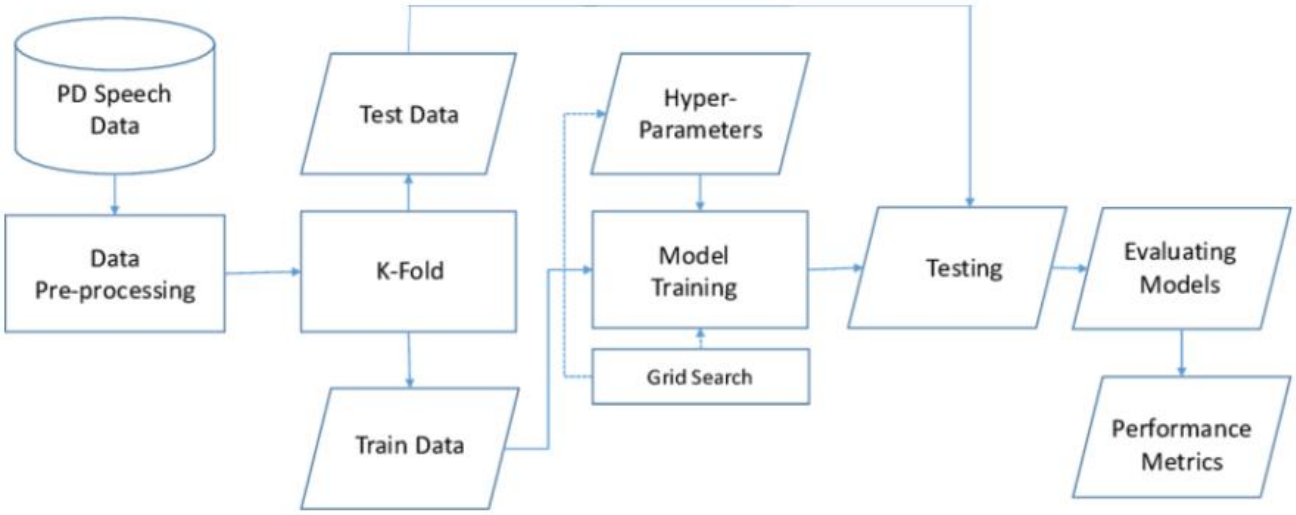
The Parkinson's data is evaluated with the application of a machine learning algorithm that is Logistic Regression, KNN, and Naïve Bayes algorithm, and the classification accuracy of this model is found.

After training the data with these algorithms we have to test on the same algorithms.

Finally, the result of these three algorithms is compared on the basis of classification accuracy.

Sreevarshan S (TL)
Vishwa S (M1 lead)
Sreehari Pranesh K (M2 Lead)
Vinith Kumar S (M3 Lead)

Block Diagram for Parkinson's disease detection:



3.2.1 Speech Dataset
3.2.2 Pre-processing data
3.2.3 Training data
3.2.4 Apply Machine Learning Algorithms
    3.2.4.1 KNN
    3.2.4.2 Naive Bayes
    3.2.4.3 Logistic Regression
3.2.5 Testing Data

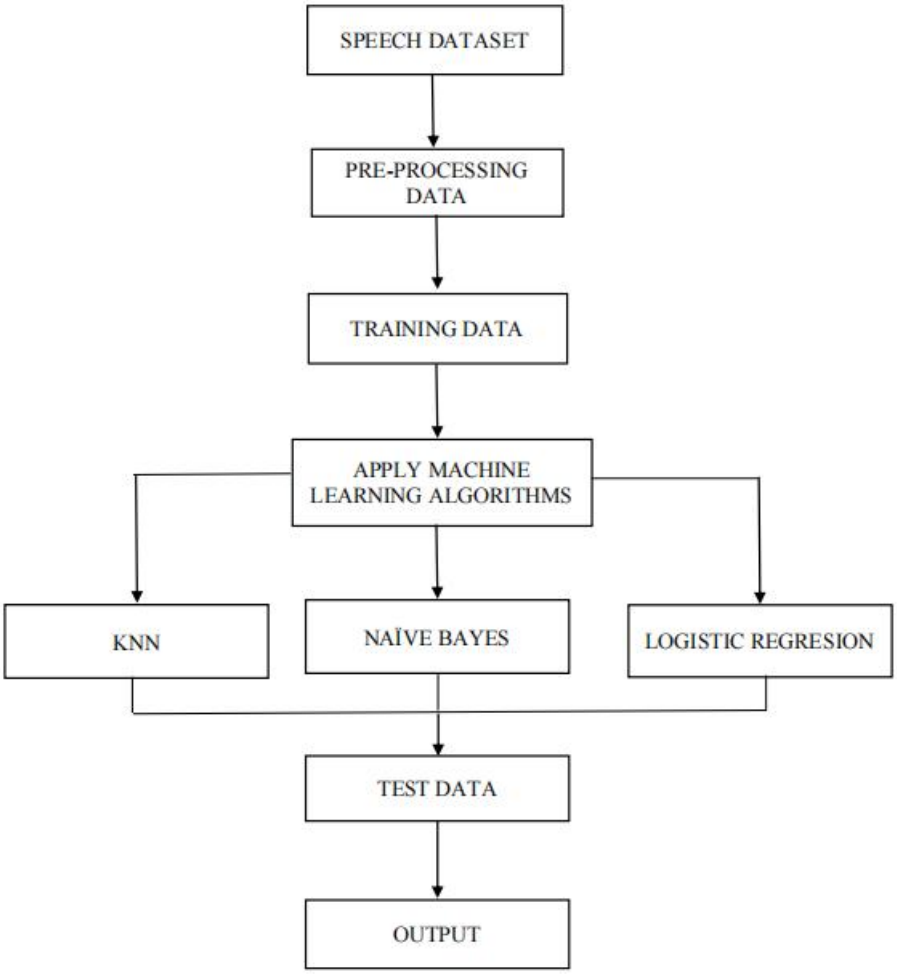


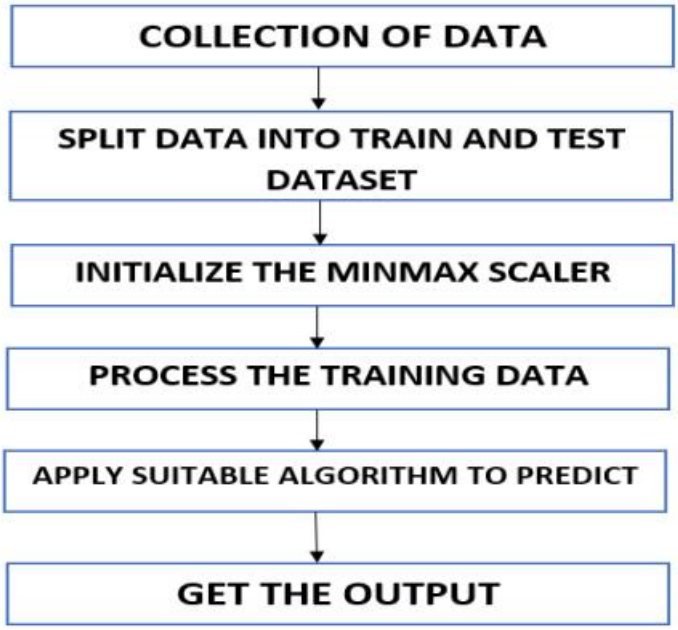
Fig. System Architecture

DATAFLOW DIAGRAM :

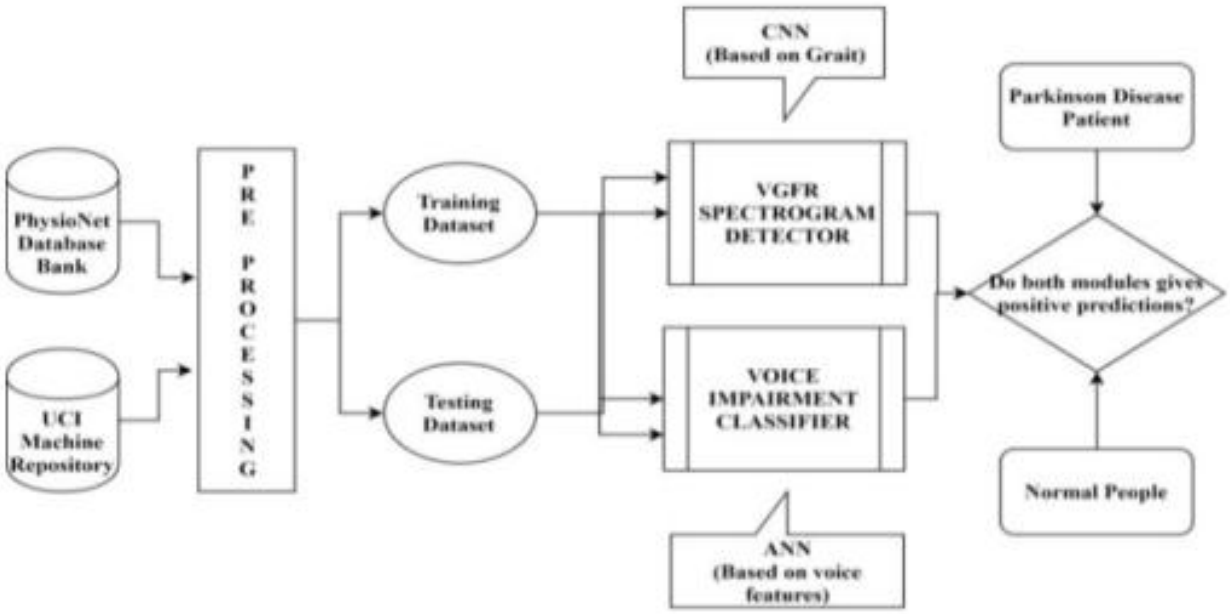
The Data flow diagram explains the basic flow data through various steps of Parkinson's disease Detection.

- In the first step we will collect data like from the patient with different medical equipment.
- The collected data is now sent process for training the data which will be classified using different ML algorithms.
- After classification the data into Train and Test datasets, the datasets will be sent to predict whether the patient has the disease or not.

Sreevarshan S (TL)
Vishwa S (M1 lead)
Sreehari Pranesh K (M2 Lead)
Vinith Kumar S (M3 Lead)



CLASS DIAGRAM:



Modules Division :

Let us discuss about the various modules in our proposed system and what each module contributes in achieving our goal.

3.2.1 Speech Dataset:

The main aim of this step is to spot and acquire all data-related problems. during this step, we'd like to spot the various data sources, as data are often collected from various sources like files and databases. The number and quality of the collected data will determine the efficiency of the output. The more are going to be the info, the more accurate are going to be the prediction.

3.2.2 Data Pre-Processing:

The main aim of this step is to study and understand the nature of data that was acquired in the previous step and also to know the quality of data. A real-world data generally contains noises, missing values, and maybe in an unusable format that cannot be directly used for machine learning models. Data pre-processing is a required task for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model. Identifying duplicates in the dataset and removing them is also done in this step.

3.2.3 Training data:

Splitting the dataset into Training set and testing set:

In machine learning data preprocessing, we have to break our dataset into both training set and test set. This is often one among the crucial steps of knowledge preprocessing as by doing this, we will enhance the performance of our machine learning model.

Suppose, if we've given training to our machine learning model by a dataset and that we test it by a totally different dataset. Then, it'll create difficulties for our model to know the correlations between the models.

If we train our model alright and its training accuracy is additionally very high, but we offer a replacement dataset there to, then it'll decrease the performance. So we always attempt to make a machine learning model which performs well with the training set and also with the test dataset

3.2.4 Apply Machine Learning Algorithms:

Now, we've both the train and test data. The subsequent step is to spot the possible training methods and train our models. As this is often a classification problem, we've used three different classification methods KNN, Naive Bayes, and Logistic Regression. Each algorithm has been run over the Training dataset and their performance in terms of accuracy is evaluated alongside the prediction wiped out the testing data set.

3.2.5 Testing Data:

Once Parkinson’s disease Prediction model has been trained on the pre-processed dataset, then the model is tested using different data points. In this testing step, the model is checked for correctness and accuracy by providing a test dataset to it. All the training methods need to be verified for finding out the best model to be used. after fitting our model with training data, we used this model to predict values for the test dataset. These predicted values on testing data are used for model comparison and accurate calculation.

3.3 User Interface:

Our Front-End implementation is completed using HTML, CSS, JavaScript and Flask

Framework in Scientific Python Development Environment (Spyder). The user interface is extremely essential for any project because everyone who tries to utilize the system for a purpose will attempt to access it using an interface. Indeed, our system also features a user interface built to facilitate users to utilize the services we provide. Where we have used HTML a terminology, utilized for creating web sites. One among the useful aspects of HTML is, it can embed programs written during a scripting language like JavaScript, which is liable for affecting the behavior and content of web pages. CSS inclusion would affect the layout and appearance of the content.

System Configuration

Software Requirements:

1. Software:
- Spyder
- Google Colab

2. Operating System: Windows 10
3. Tools: Web Browser
4. Python Libraries: numpy, pandas, matplotlab, seaborn, sklearn, pickle.

Hardware Requirements:

1. RAM: 4 GB or above
2. Storage: 30 to 50 GB
3. Processor: Any Processor above 500MHz

CONCLUSION:

Parkinson's is the second most neuro-degenerative disease which has no cure. It results in difficulty of body movements, anxiety, breathing problems, loss of smell, depression, and speech. In this paper, the three different machine learning algorithms used to measure the performance are KNN, Naive Bayes, and Logistic Regression applied on the dataset. The author chose the voice features of patients as the dataset contains more than 700 features and finally took the ten important features that are useful to evaluate the system. The author compared all the three machine learning methods accuracies and based on this one prediction model is generated.The Solution Architecture for the application is done.