DETECTING PARKINSON'S DISEASE USING MACHINE LEARNING

NALAIYA THIRAN PROJECT REPORT

2022

Submitted by

VASUNDARA T	312419205112
VISHNUPRIYA N	312419205119
MUKESH S	312419205064
SWETHA K	312419205105

Team ID: PNT2022TMID28464

Content

TEAM ID: PNT2022TMID28464

DETECTING PARKINSON'S DISEASE USING MACHINE LEARNING

1. Introduction

1.1 Project Overview

More than 10 million people are living with Parkinson's Disease worldwide, according to the Parkinson's Foundation. While Parkinson's cannot be cured, early detection along with proper medication can significantly improve symptoms and quality of life. Parkinson's disease is a brain disorder that causes unintended or uncontrollable movements, such as shaking, stiffness, and difficulty with balance and coordination. The researchers found that the drawing speed was slower and the pen pressure is lower among Parkinson's patients. One of the indications of Parkinson's is tremors and rigidity in the muscles, making it difficult to draw smooth spirals and waves. It is possible to detect Parkinson's disease using the drawings alone instead of measuring the speed and pressure of the pen on paper.

1.2 Purpose

Symptoms usually begin gradually and worsen over time. As the disease progresses, people may have difficulty walking and talking. They may also have mental and behavioral changes, sleep problems, depression, memory difficulties, and fatigue. Early detection helps patients. Although there is no cure for Parkinson's disease, medicines, surgical treatment, and other therapies can often relieve some symptoms.

2. Literature Survey

2.1 Existing Problem

Speech or voice data is assumed to be 90% helpful to diagnose a person for identifying presence of disease. In general, Person with PD suffer from speech problems, which can be categorized into two: hypophonia and dysarthria. Hypophonia indicates very soft and weak voice from a person and dysarthria indicate slow speech or voice, that can hardly be understood at one time and this causes because of damage to central nervous system.

So, most of the clinicians who treat PD patients observe dysarthria and try to rehabilitate with specific treatments to improvise vocal intensity

2.2 References

Author Name: DAVID GIL A, MAGNUS JOHNSON B

Title of the Paper: Diagnosing Parkinson by using Artificial Neural Networks and Support Vector Machines

Description: He found that with a smaller number of neurons at hidden layer both training set and test sets performed poorly. With higher number of neurons, the training set performed well with high risk of over fitting. The ideal solution for this layer was found to be 13 neurons.

Author Name: Mohammad S Islam

Title of the Paper: The Mechanistic Role of Thymoquinone in Parkinson's Disease: Focus on Neuroprotection in Pre-Clinical Studies

Description: He compared various ML techniques based on their performance accuracies in determining whether person is having PD or not and mentioned that new classifier may be built to get better accuracies.

Author Name: Kazi Amit Hasan

Title of the Paper: Classification of Parkinson's Disease by Analyzing Multiple Vocal Features Sets

Description: He used different classification methods RF, KNN, Decision Tree, Logistic Regression (LR), SVM, and Naïve Bayes for detection of PD. The best result achieved by Decision Tree and Random Forest (RF) classification methods. The data mining techniques may be a more popular in many field of medical, business, railway, education etc. They are most commonly used for medical diagnosis and disease prediction at the early stage. The data mining is employed for healthcare sector in industrial societies.

Author Name: Shail Raval

Title of the Paper: A Comparative Study of Early Detection of Parkinson's Disease using Machine Learning Techniques

Description: For the detection of PD they include all the aspects such as biological data, chemical data and genetic data. In this paper they mainly focused on the symptoms like rigidity, Tremor at rest, changing voice etc. The secure data transmission is proposed through authentication check, duplication check and faulty node detection. The proposed method is applicable to long ranges of transmission. It is also supporting a retransmission concept.

Author Name: Rajalakshmi Shenbaga Moorthy

Title of the Paper: Freezing of Gait Prediction in Parkinsons Patients Using Neural Network

Description: Novel analytic system for Parkinson’s disease Prediction mechanism using Improved Radial Basis Function Neural Network (IRBFNN). RNNs is during a one

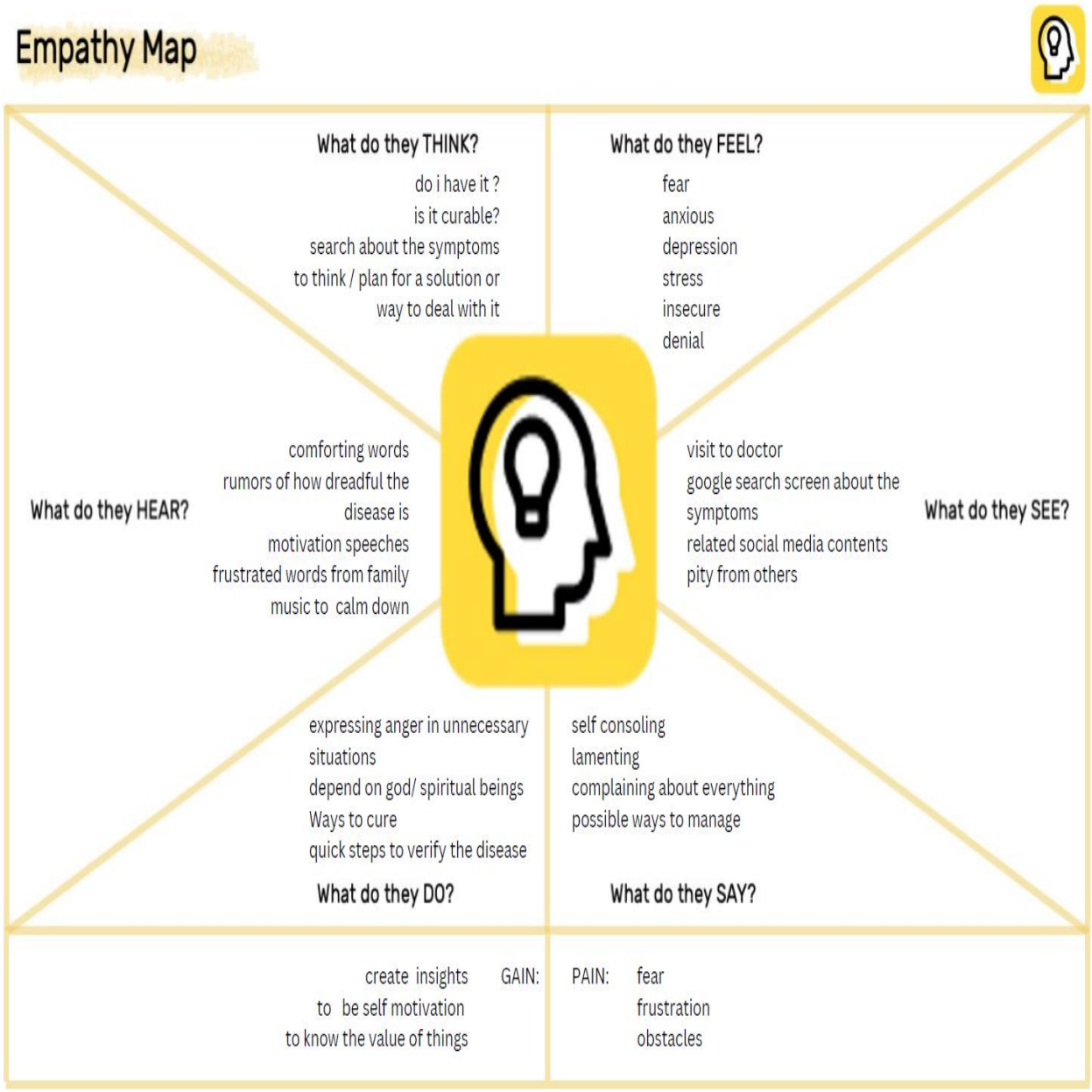
among the deep learning models that are used for modeling the arbitrary length sequences by applying a transition function to all or any it’s hidden states during a recursive manner.

2.3 Problem Statement Definition

Parkinson's disease is a brain disorder that causes unintended or uncontrollable movements, such as shaking, stiffness, and difficulty with balance and coordination. Although most people with Parkinson's first develop the disease after age 60, about 5% to 10% experience onset before the age of 50. Early-onset forms of Parkinson's are often, but not always, inherited, and some forms have been linked to specific gene mutations. The most prominent signs and symptoms of Parkinson's disease occur when nerve cells in the basal ganglia, an area of the brain that controls movement, become impaired and or die.

3. Ideation and Proposed Solution

3.1 Empathy Map



3.2 Ideation and Brainstorming

1. Approach by the user point of view:

When a user is undergoing some similar symptoms related to the Parkinson's disease, the user is in the need to detect whether it is Parkinson's. Parkinson's disease is a progressive disorder that affects the nervous system and the parts of the body controlled by the brain and the nerves. They are mostly slow progressing but there are some symptoms can be noticed like the tremor in one hand, muscle stiffness, slowing of movement, stooped posture, lack of facial expression, decreased arm swing and soft or low voice.

2. Creating a point of view template:

USER	NEEDS	INSIGHT
A person having similar symptoms to the parkinson's disease.	It can help to reduce the efforts of the disease.	It can help in the early detection of the disease. It can help to ease the feelings felt by the disease.
A person assumes to have the disease	To avoid a misunderstanding	It can help to reduce the stress and other factors affecting the patient.

3. Anticipation of the solution:

When the user experiences symptoms similar to the symptoms related to parkinson's, they are in a need to detect whether they have the disease or not. The most efficient method of detection is by using the detection by the machine learning algorithm. Consultation with a doctor can be done after the initial detection by the algorithm. It can help with a faster detection and a more accurate method than guessing and other incorrect methods.

4. Ideate:

The ways in detecting the Parkinson's disease can be done by various methods. The most common is by doing a laboratory tests by the suggestion of a doctor. It can be the best method but it can be time consuming and of high cost. As there is a possibility of not having the disease an earlier detection can help in reducing the anxiety felt by the patient. A random guess can be done by an experienced professional but it is not accurate to depend on and it may lead to wrong diagnosis and treatment for a healthy person which may cause other diseases. With the help of the external sources, a wrong assumption can be made by themselves. An initial method to detect the disease before a doctor consultation can have a greater advantage. It can be efficiently provided by the machine learning algorithm. During the evaluation, the severity of the disease may also be validated. An earlier detection

caN prevent the symptoms to worsen further. By training the data set to get an accurate result can help with a correct prediction with the test data.

5. Success of the project:
This project of detecting the parkinson's disease is mainly based on the machine learning algorithm which are more accurate in the prediction model. Computer make less error compared to the humans. So, the success of the project is mostly higher than other methods

3.3 Proposed Solution

SNO.	PARAMETER	DESCRIPTION
1	PROBLEM STATEMENT	Parkinson's disease disorder is a brain disorder that causes unintended or uncontrollable movements, such as shaking, stiffness, and difficulty with balance and coordination
2	IDEA/SOLUTION DESCRIPTION	Detection of Parkinson's disease using the drawings alone instead of measuring the speed and pressure of the pen on paper. Histogram of Oriented Gradients (HOG) image descriptor along with a Random Forest classifier to automatically detect Parkinson's disease.
3	NOVELTY/ UNIQUENESS	A test that includes sketching a spiral on a piece of paper could be used to identify people who are at risk of getting Parkinson's disease, according to Australian researchers. All three measurements were used to determine whether a participant had Parkinson's disease or not can be used for accurate measures.
4	SOCIAL IMPACT/CUSTOMER SATISFACTION	Less error. Easily implemented. Convenient to use. Cost-free. Avoids to travel. Time can be saved.
5	BUSINESS MODEL (REVENUE MODEL)	Welfare based project. Free source. Anyone can use and suggest it to others.
6	SCALABILITY OF SOLUTION	It is a SAS where adjustment can be made anywhere flexible to change, accessed by any number of users.

3.3 Problem Solution Fit

4.RequirementAnalysis

4.1 Functional Requirement

Following are the functional requirements of the proposed solution

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Image upload	The drawings drawn by the user are uploaded here for analyses.
FR-2	Parkinson prediction	User drawings are used to detect it by the model
FR-3	Result Generation	With the given data the result can be generated as whether the person is healthy or affected.

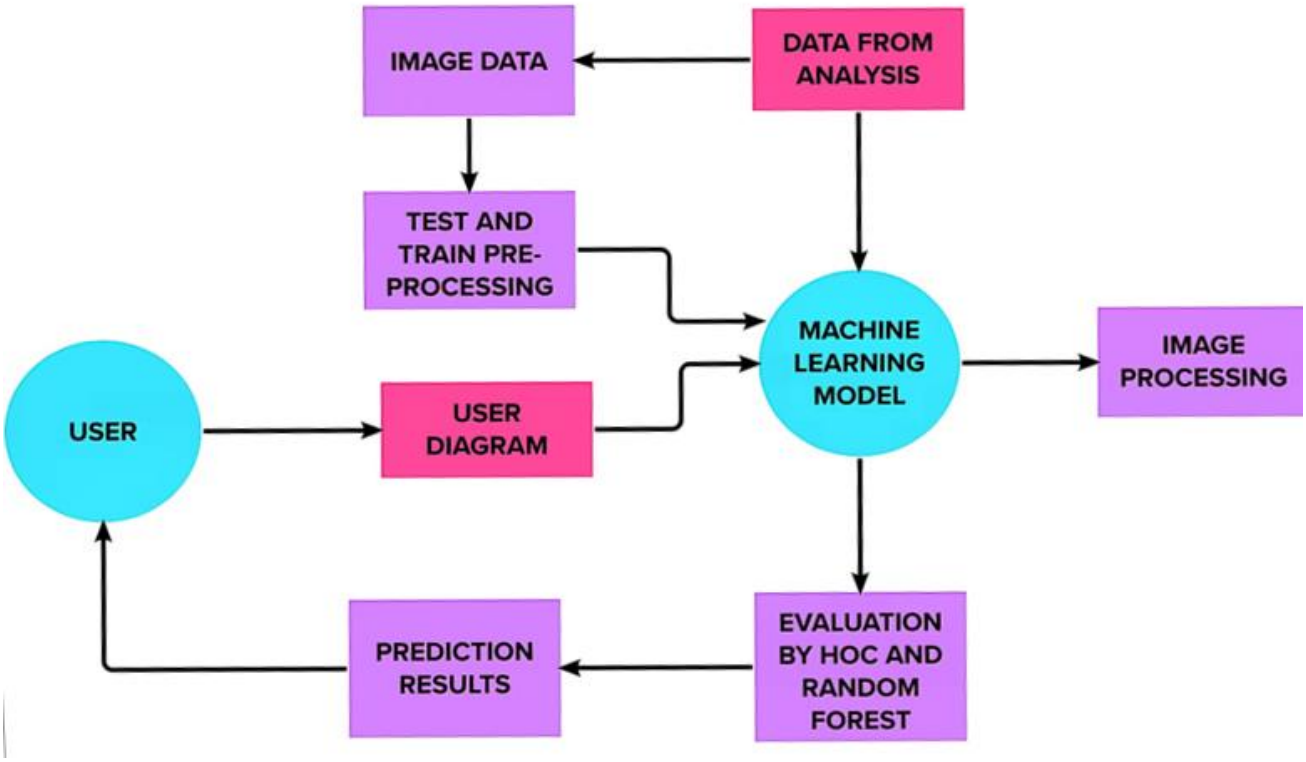
Non-Functional Requirements:

Following are the functional requirements of the proposed solution.

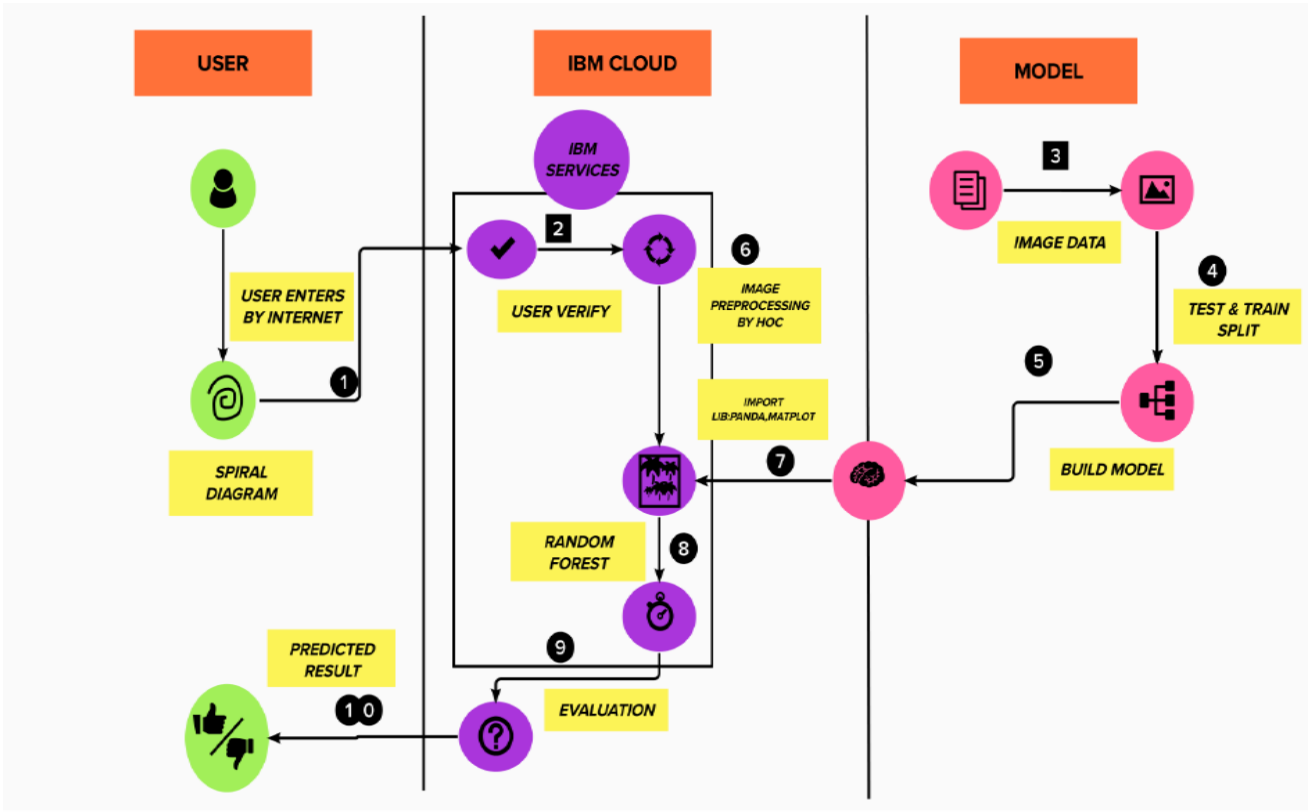
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	The user interface is easy to operate and user-friendly.
NFR-2	Security	The data given by the user will be very secure.
NFR-3	Reliability	Failure is not vital and user can access by refreshing.
NFR-4	Performance	Load time for the user interface screen will be not more than 2 seconds.
NFR-5	Availability	Maximum down time will be about 1 hour.
NFR-6	Scalability	System can handle about 1000 users at a given time.

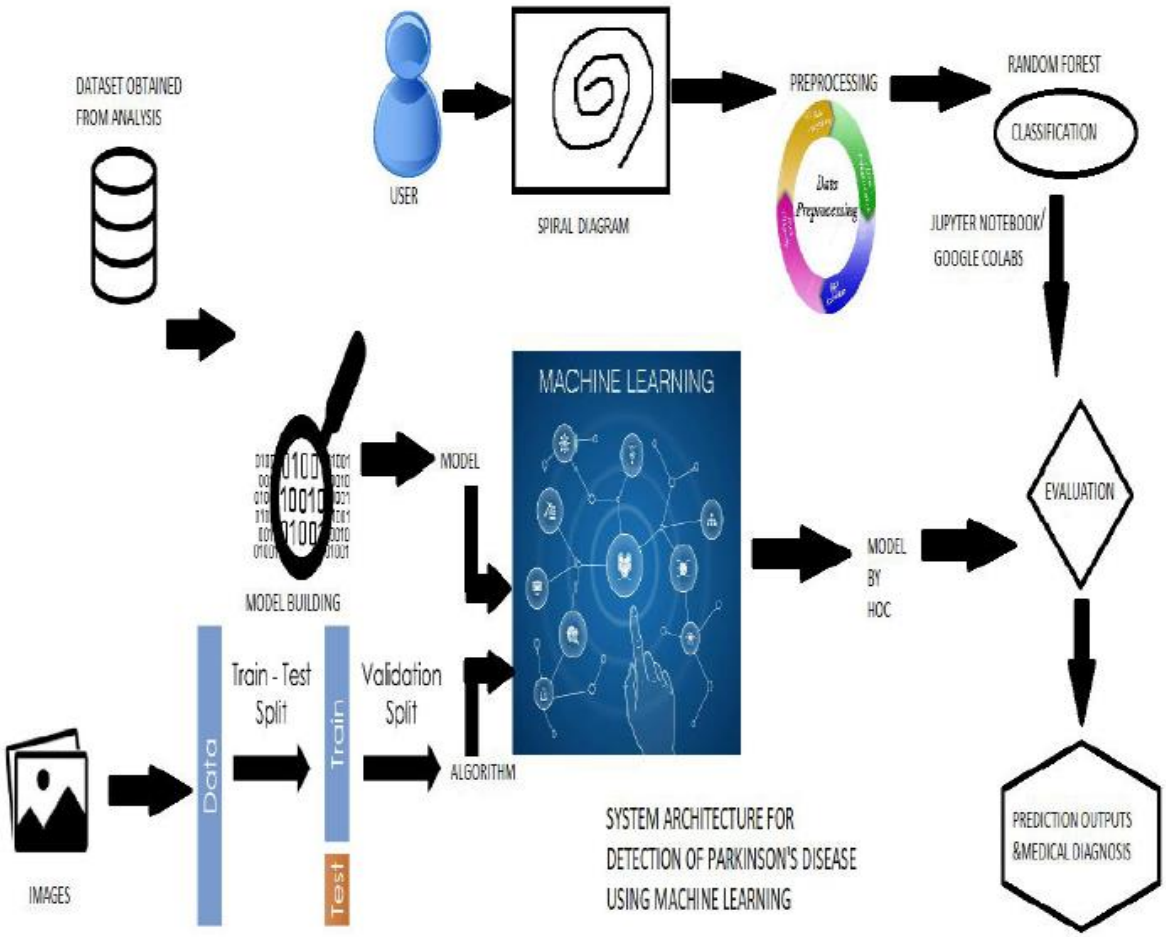
5 Project Design

5.1 Data Flow Diagram



5.2 Solution and Technical Architecture





6. Project Planning and Scheduling

6.1 Script Planning and Execution

SPRINT	FUNCTIONAL REQUIREMENTS	USER STORY NO.	USER STORY/ TASK	STORY POINTS	PRIORITY	TEAM MEMBERS
SPRINT-1	REGISTER	USN-1	As a user I can register by email, password and confirm password	2	HIGH	VASUNDARA T
SPRINT-1		USN-2	As a user get confirmation mail	2	HIGH	VISHNUPRIYA N
SPRINT-1		USN-3	As a user, I can register by other sources	1	LOW	SWETHA K
SPRINT-1	LOGIN	USN-4	As a user, I can login by gmail	2	HIGH	VASUNDARA T
SPRINT-1		USN-5	As a user, I can login by other sources	1	MEDIUM	SWETHA K
SPRINT-2	DASHBOARD	USN-6	As a user, I can view and access the dashboard	2	HIGH	SWETHA K
SPRINT-2	PREDICTION	USN-7	As a user, I can enter my details for prediction	2	HIGH	VISHNUPRIYA N
SPRINT-3		USN-8	As a user, I can view the accuracy of the model	1	MEDIUM	MUKESH S
SPRINT-3		USN-9	As a user, I can verify it with others	1	LOW	MUKESH S
SPRINT-4	RESULTS	USN-10	As a user, I can get output for parkinson	2	HIGH	VASUNDARA T
SPRINT-4		USN-11	As a user, I can decide to consult the doctor	1	MEDIUM	VISHNUPRIYA N
SPRINT-3	LOGOUT	USN-12	As a user, I can logout from it	1	HIGH	MUKESH S
SPRINT-4	LOGIN/LOGOUT	USN-1,2	As a admin, I can login or logout	2	HIGH	SWETHA K
SPRINT-4	ADMINISTRATION	USN-3	As a admin, I can oversee the model	2	HIGH	MUKESH S

6.2 Sprint Delivery Schedule

SPRINT	TOTAL STORY POINTS	DURATION	SPRINT START DATE	SPRINT END DATE	STORY POINTS COMPLETED (AS ON PLANNED END DATE)	SPRINT RELEASE DATE (ACTUAL)
SPRINT 1	20	6 DAYS	24 OCTOBER 2022	29 OCTOBER 2022	20	29 OCT 2022
SPRINT 2	20	6 DAYS	31 OCTOBER 2022	05 NOVEMBER 2022	20	05 NOV 2022
SPRINT 3	20	6 DAYS	07 NOVEMBER 2022	12 NOVEMBER 2022	20	12 NOV 2022
SPRINT 4	20	6 DAYS	14 NOVEMBER 2022	19 NOVEMBER 2022	20	19 NOV 2022

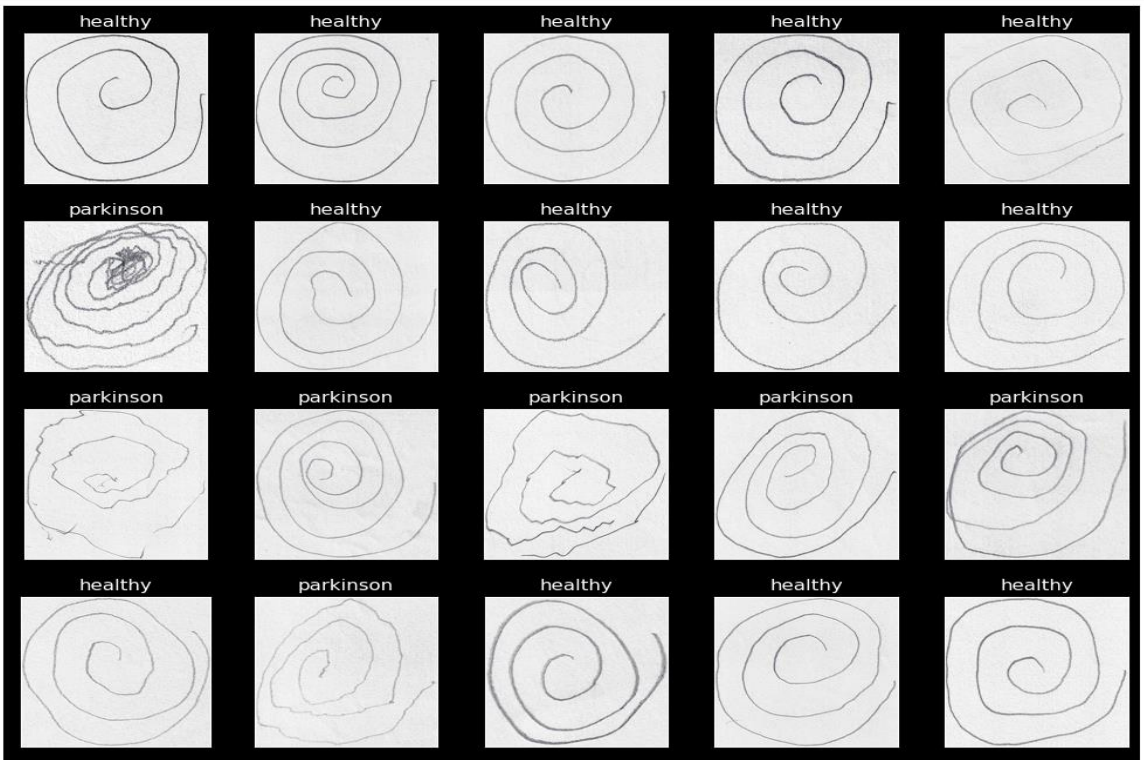
7 Coding And Solution

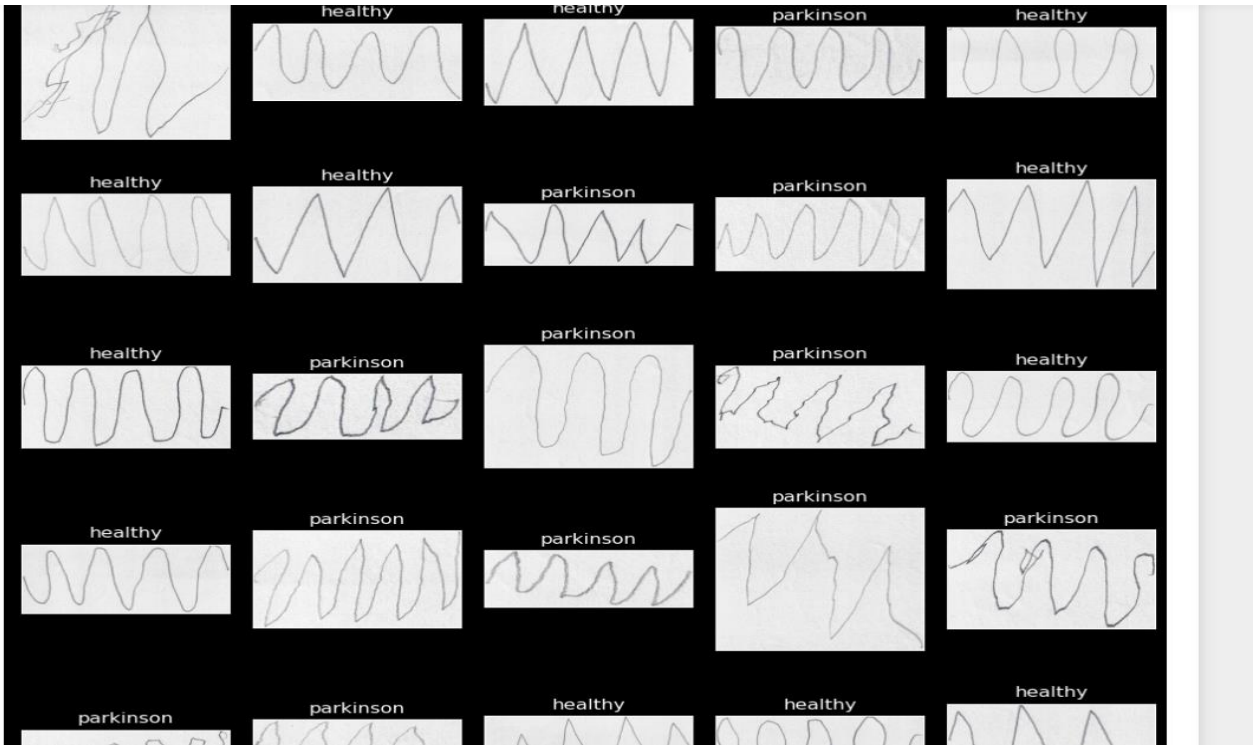
7.1 Machine Learning

Learning which model is best for the given Dataset

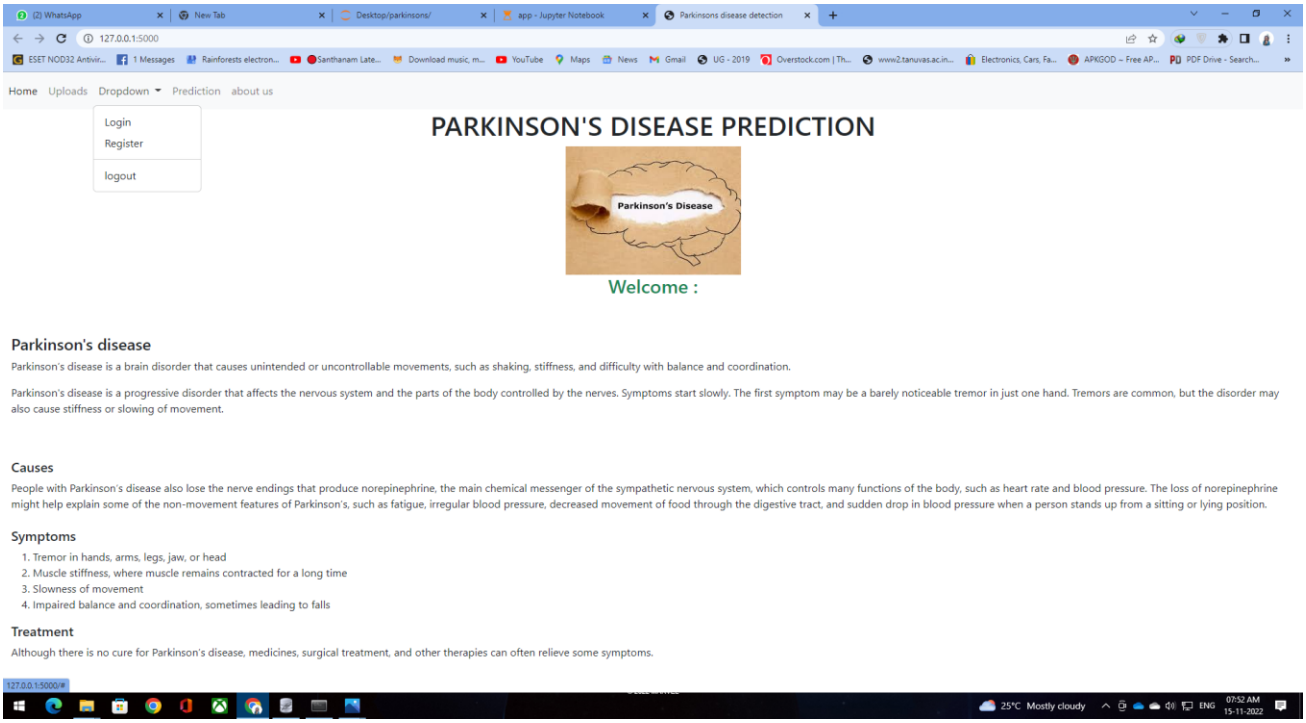
```
In [10]: Spiral_Train_Path = Path("C:/Users/Admin/Desktop/parkinsons/data/spiral/training")
         Spiral_Test_Path = Path("C:/Users/Admin/Desktop/parkinsons/data/spiral/testing")

In [11]: Spiral_Train_PNG_Path = list(Spiral_Train_Path.glob(r"*/*.png"))
         Spiral_Test_PNG_Path = list(Spiral_Test_Path.glob(r"*/*.png"))

In [12]: Spiral_Train_PNG_Labels = list(map(lambda x: os.path.split(os.path.split(x)[0])[1],Spiral_Train_PNG_Path))
         Spiral_Test_PNG_Labels = list(map(lambda x: os.path.split(os.path.split(x)[0])[1],Spiral_Test_PNG_Path))

In [13]: Spiral_Train_PNG_Path_Series = pd.Series(Spiral_Train_PNG_Path,name="PNG").astype(str)
         Spiral_Train_PNG_Labels_Series = pd.Series(Spiral_Train_PNG_Labels,name="CATEGORY")

In [14]: Spiral_Test_PNG_Path_Series = pd.Series(Spiral_Test_PNG_Path,name="PNG").astype(str)
         Spiral_Test_PNG_Labels_Series = pd.Series(Spiral_Test_PNG_Labels,name="CATEGORY")

In [15]: Main_Spiral_Train_Data = pd.concat([Spiral_Train_PNG_Path_Series,Spiral_Train_PNG_Labels_Series],axis=1)
```

```
plt.show()
```

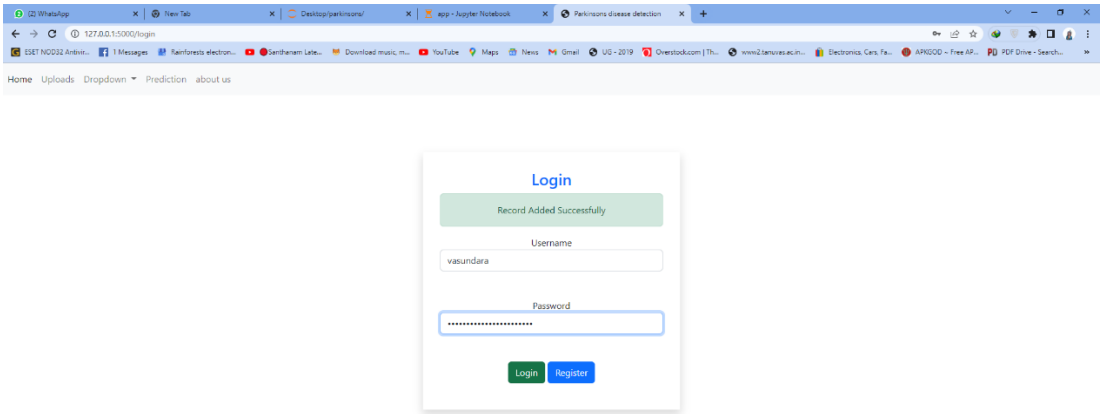

healthy
healthy
healthy
healthy
healthy
parkinson
healthy
healthy
healthy
healthy
parkinson
parkinson
parkinson
parkinson
parkinson
healthy
parkinson
healthy
healthy
healthy


healthy
healthy
parkinson
healthy
healthy
healthy
healthy
parkinson
parkinson
healthy
healthy
parkinson
parkinson
healthy
parkinson
parkinson
parkinson
parkinson
parkinson
healthy
healthy

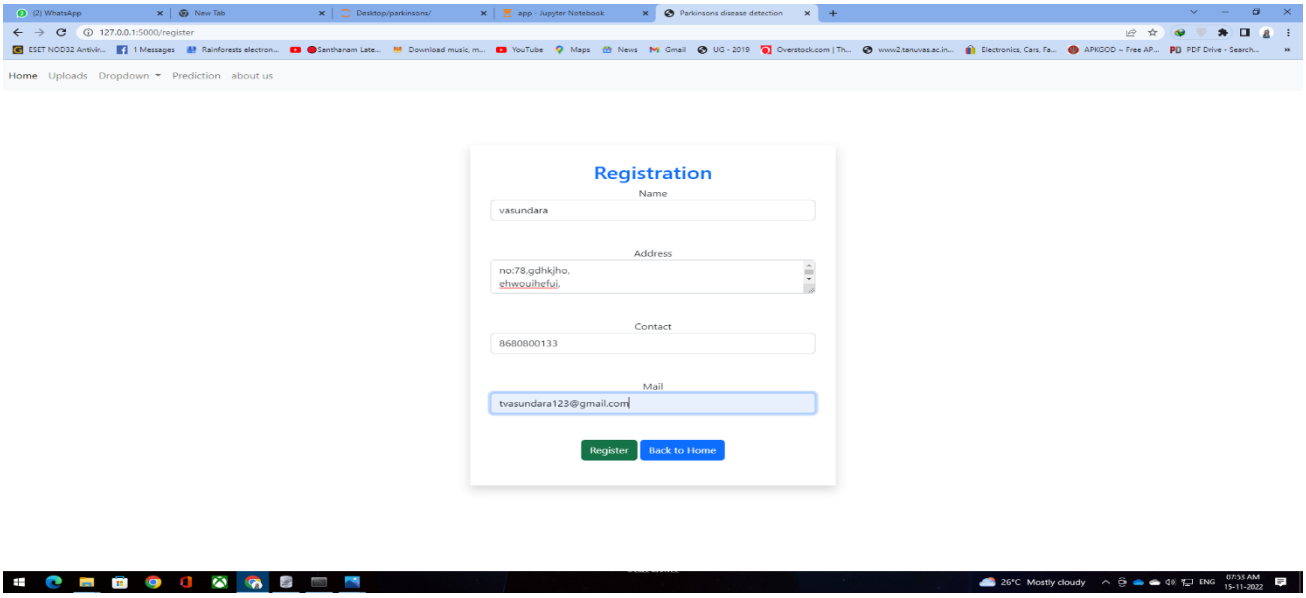
7.2 Dashboard



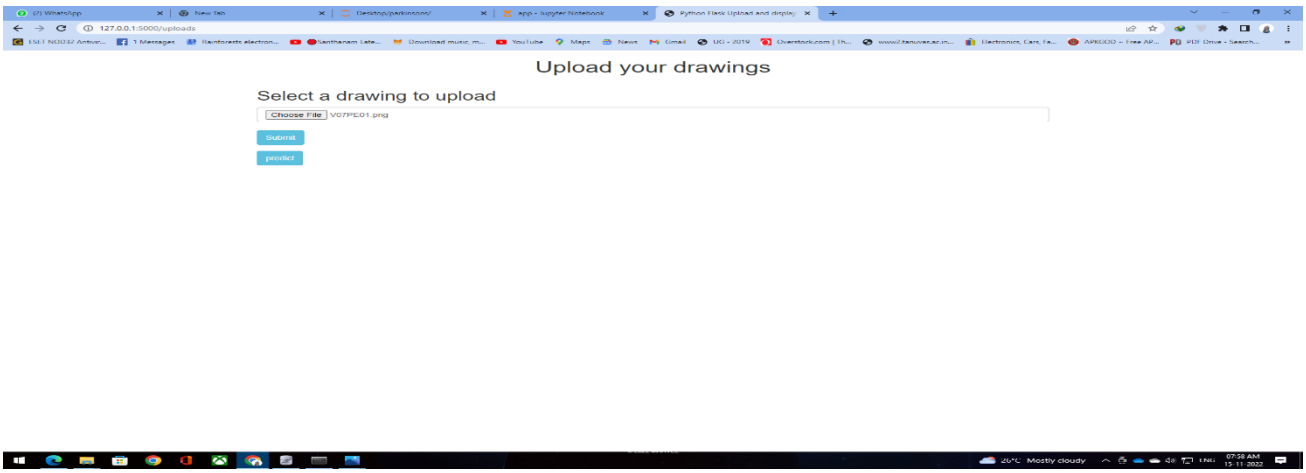
LOGIN



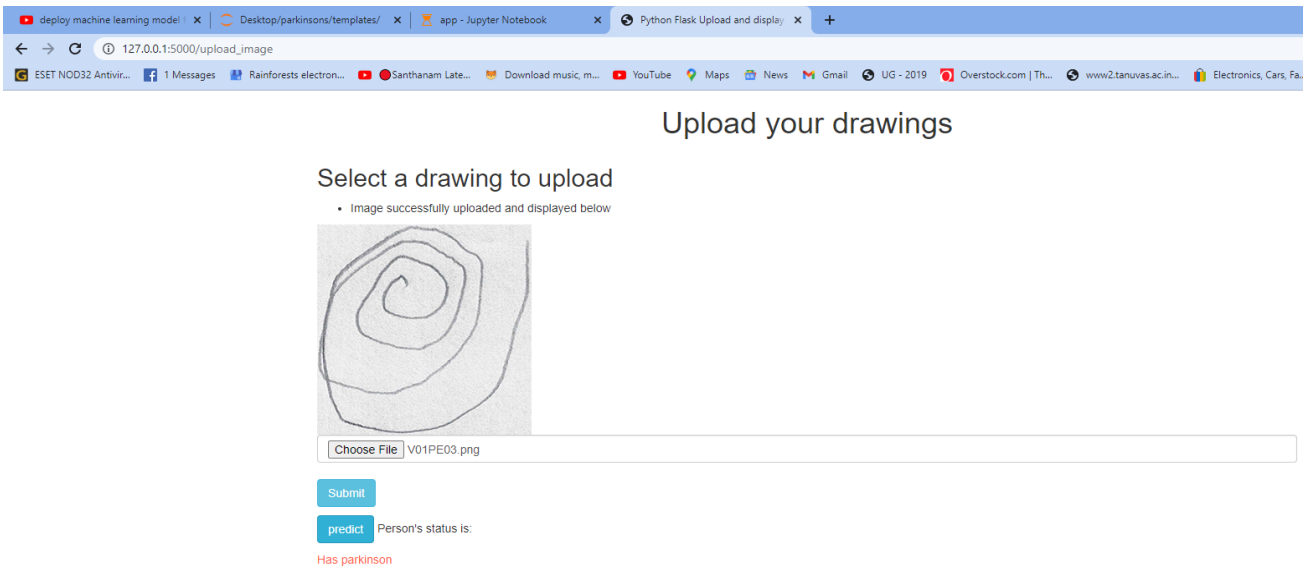
REGISTER



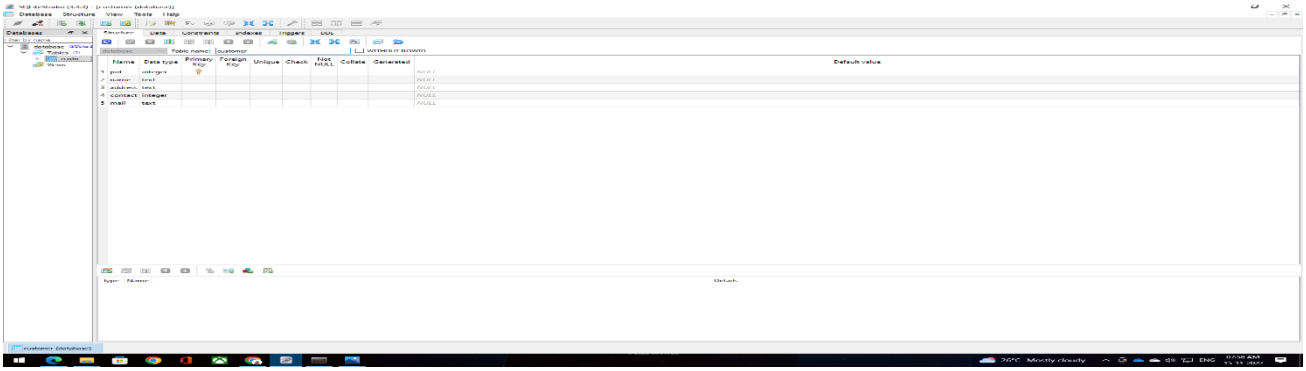
UPLOADS:

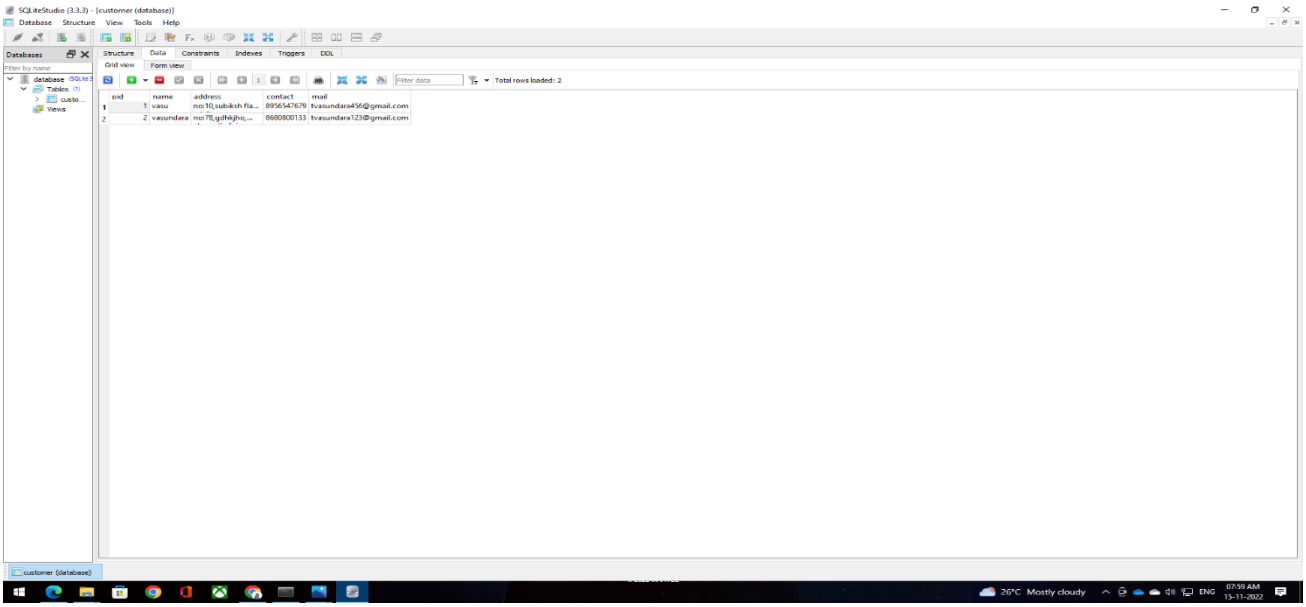


PREDICTION:



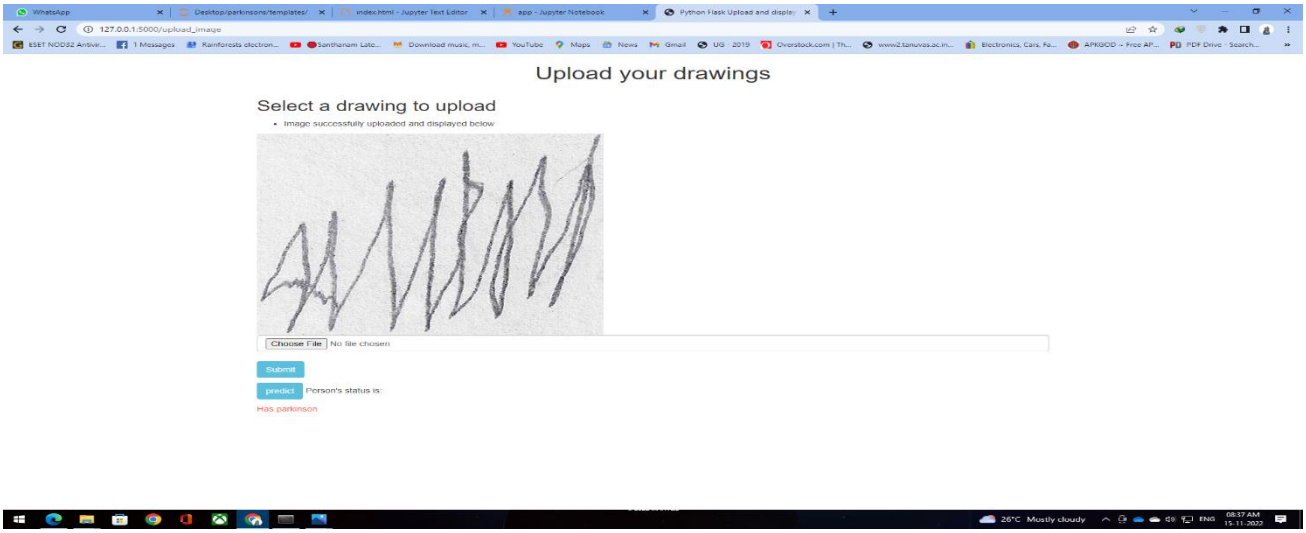
7.3 DATABASE SCHEMA:

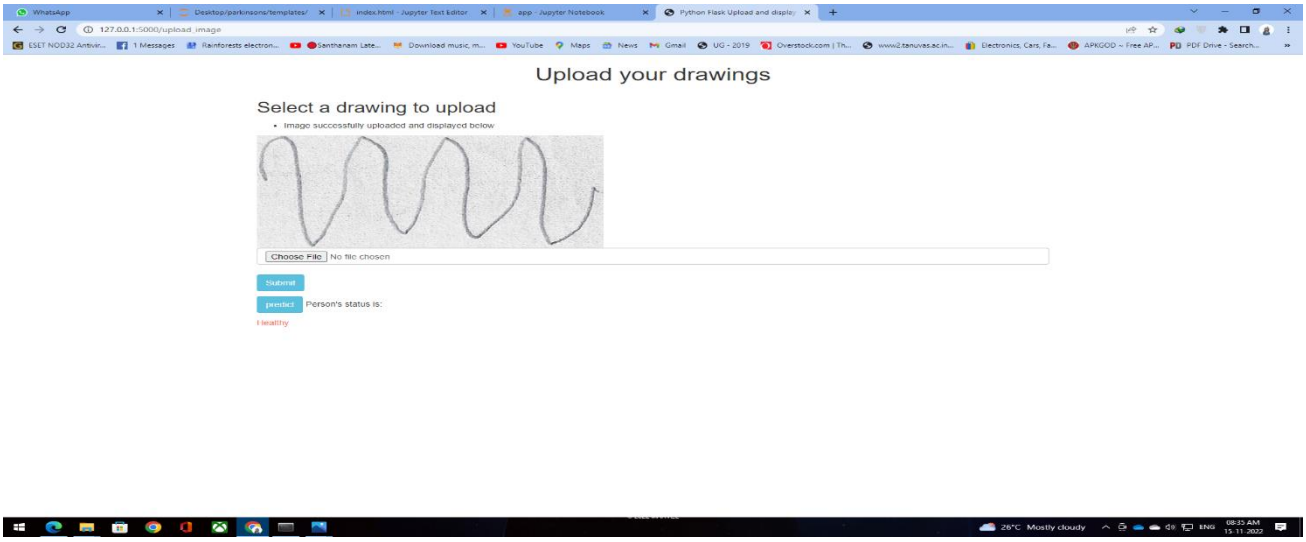




8. Testing

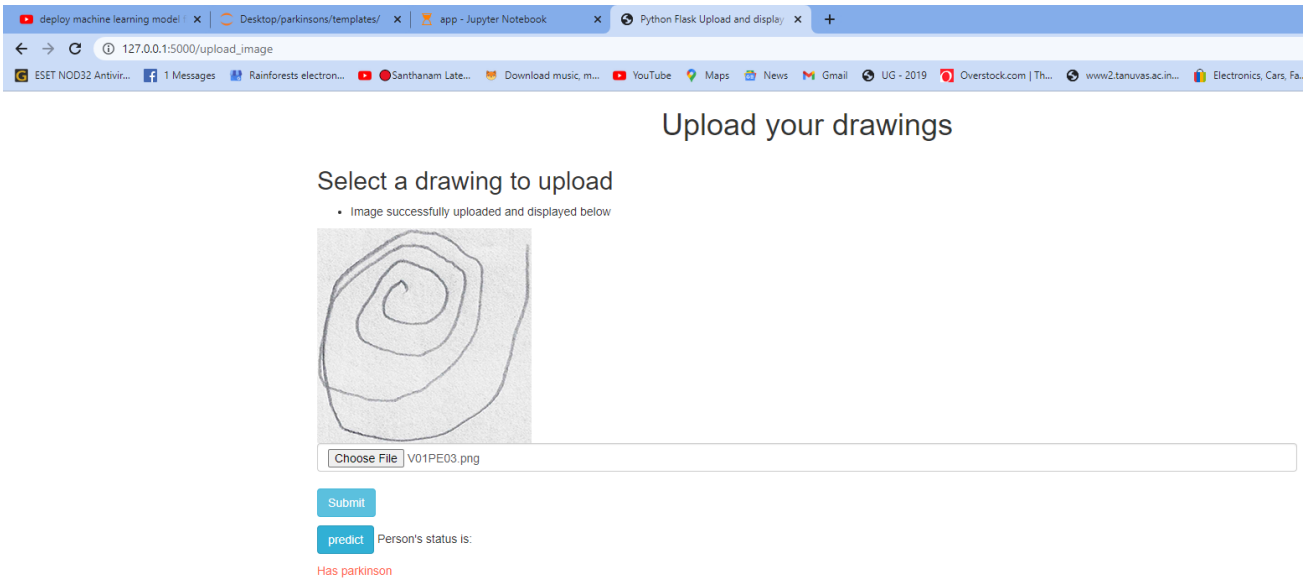
8.1 Test Cases



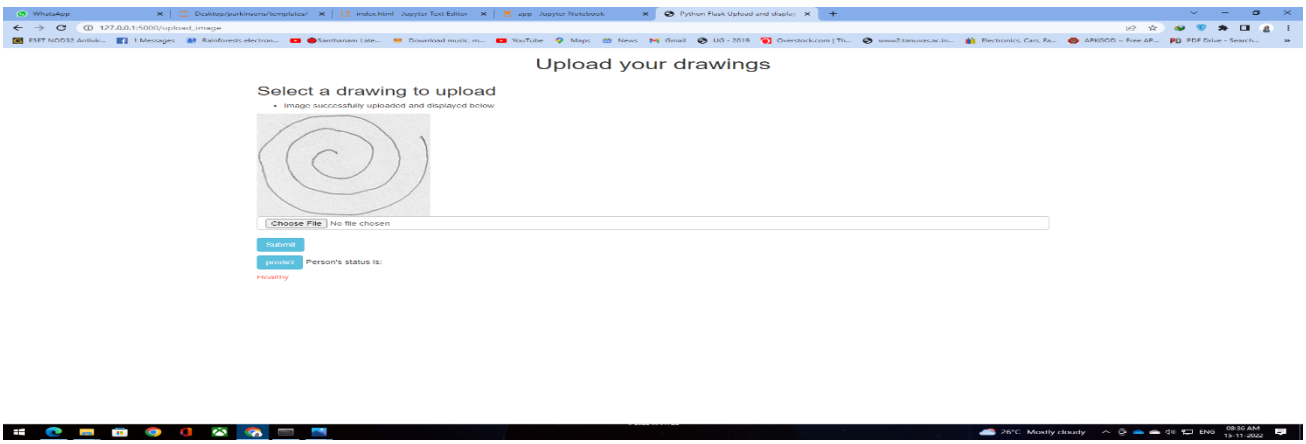


8.2 User acceptance Testing

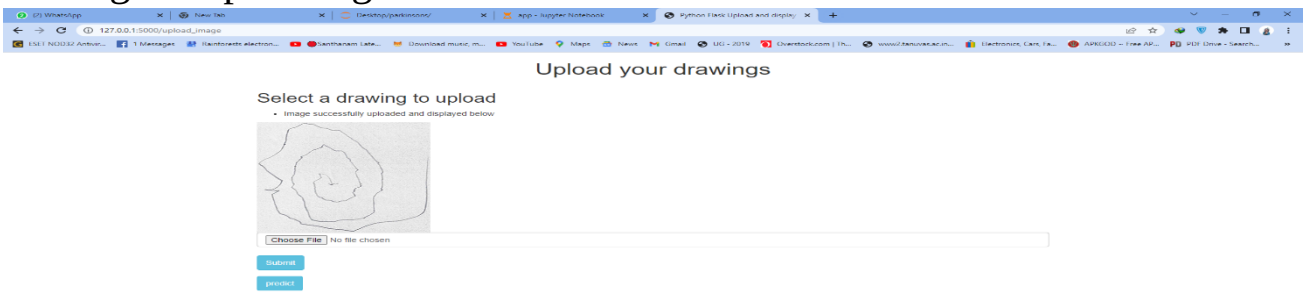
Testing a case where user has Parkinson



Testing a case where user has healthy



Testing for uploadings



Default test values for checking the prediction algorithm.

```
In [226]: print("TRAIN: ")
          print(Train_Wave_Set.class_indices)
          print(Train_Wave_Set.classes[0:5])
          print(Train_Wave_Set.image_shape)
          print("---"*20)
          print("VALIDATION: ")
          print(Validation_Wave_Set.class_indices)
          print(Validation_Wave_Set.classes[0:5])
          print(Validation_Wave_Set.image_shape)
          print("---"*20)
          print("TEST: ")
          print(Test_Wave_Set.class_indices)
          print(Test_Wave_Set.classes[0:5])
          print(Test_Wave_Set.image_shape)
```

```
TRAIN:
{'healthy': 0, 'parkinson': 1}
[1, 1, 0, 0, 1]
(256, 256, 1)
------------------------------------------------------------
VALIDATION:
{'healthy': 0, 'parkinson': 1}
[1, 0, 0, 1, 0]
(256, 256, 1)
------------------------------------------------------------
TEST:
{'healthy': 0, 'parkinson': 1}
[0, 0, 0, 1, 0]
(256, 256, 1)
```

9. Result

9.1 Performance Metrics

```
    print("-" * 30)
```

```
LinearRegression:
----------
nan
0.9335111546546421
------------------------------
PLSRegression:
----------
nan
0.9335098131842107
------------------------------
Ridge:
----------
nan
0.9185363197943562
------------------------------
Lasso:
----------
nan
0.5986568551269732
------------------------------
ElasticNet:
----------
nan
0.6081779237738685
------------------------------
KNeighborsRegressor:
----------
nan
0.6292853089020909
------------------------------
DecisionTreeRegressor:
----------
nan
0.8366600265340756
------------------------------
BaggingRegressor:
----------
nan
0.7563068160475616
------------------------------
RandomForestRegressor:
----------
nan
0.7853979882836472
```

10. Advantages Disadvantages

Advantages:

- One of the fastest and easiest way to determine if a person is likely to have Parkinson or not.
- Useful for predicting parkinson
- User Friendly
- early detection
- cost-free analysis

Disadvantages:

- Need a more acceptable prediction methods
- Users need to draw
- Does Not take null value as input
- Does not provide suggestions to the user.

11. Conclusion

There are currently no blood or laboratory tests to diagnose non-genetic cases of Parkinson's. The researchers found that the drawing speed was slower and the pen pressure is lower among Parkinson's patients. One of the indications of Parkinson's is tremors and rigidity in the muscles, making it difficult to draw smooth spirals and waves. It is possible to detect Parkinson's disease using the drawings alone instead of measuring the speed and pressure of the pen on paper.

12. Future Scope

There are many possible improvements that could be explored to improve the scalability and accuracy of this prediction system. As we have developed a generalized system, in future we can use this system for the analysis of different data sets. The performance of the health's diagnosis can be improved significantly by handling numerous class labels in the prediction process, and it can be another positive direction of research.

13. Appendix

Source Code:

IBM-Project-27762-1660065107/**Project Development Phase/**

Demo video:

IBM-Project-27762-1660065107/**Project report & demo video/**