Prediction of Parkinson's Disease using Machine Learning

1.Mrs. B. Lakshmi Sowjanya 2. Mr. Chinthakindi Rajendra

sowji54321@gmail.com rajendrachinthakindi@gmail.com

3.Ms. Emmadishetty Rakshitha 4.Ms. Kothi RamyaSri 5.Ms. Elishala Renuka

patellrakshitha@gmail.com kothiramyasri@gmail.com elishalarenuka123@gmail.com

[1]Assistant Professor, Department of CSE

[2,3,4,5] UG Students, Department of CSE

Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India

Abstract: Parkinson's disease is a progressive neurodegenerative disorder that gradually impairs motor and non-motor functions. It results from the deterioration or death ofdopamine-producing neurons in specific regions of the brain, leading to symptoms such as vocal difficulties, handwriting changes, gait abnormalities, and challenges in performing everyday tasks. These symptoms intensify over time, increasing the severity of the condition. In this study, we propose a methodology for predicting the severity of Parkinson's disease using deep neural networks, leveraging the UCI Parkinson's Telemonitoring Vocal Data Set. We developed a neural network model to estimate disease severity and implemented a machine learning model to detect the presence of the disorder. Classification of Parkinson's disease was performed using both a deep neural network and a Random Forest classifier.

Keywords: Parkinson's Disease, Machine Learning (ML), Supervised Learning, Classification Algorithms, Biomedical Data, Voice Measurements, Feature Extraction, Data Preprocessing, Standardization, Normalization, Feature Selection, Support Vector Machine (SVM), Random Forest, Logistic Regression, XGBoost, Deep Learning, TensorFlow, Keras, PyTorch, Confusion Matrix, Accuracy, Precision, Recall, F1-Score, ROC-AUC Curve, Hyperparameter Tuning, Cross-Validation, Model Evaluation, Predictive Modeling, UCI Parkinson's Dataset, Feature Importance, SHAP, LIME, Model Deployment, Healthcare Analytics, Early Diagnosis, Voice-based Detection.

1. INTRODUCTION

Parkinson's disease (PD) is a neuromotor disorder that begins with a mild tremor in the hand and a feeling of stiffness in the body and progresses over time. Parkinson's disease is caused by the loss of

neurons in the brain that produce the chemical messenger dopamine. Abnormal brain activity occurs when levels of an amino acid known as dopamine decrease. The exact cause of Parkinson's disease remains a mystery, but multiple factors, including genes, environment and triggers, appear to be involved. Each person experiences the symptoms in the Parkinson's disease differently. Early symptoms are not important and are overlooked. In most cases, symptoms start on one side of the body and spread to the other side.

2. LITERATURE SURVEY

- **Resul Das –** "A comparison of multiple classification methods for diagnosis of Parkinson's disease"
- **Srishti Grover et al.** – "Predicting Severity of Parkinson's Disease Using Deep Learning"
- **Elmehdi Benmalek et al**. – "UPDRS tracking using linear regression and neural network"
- **Warwick R. Adams** – "High-accuracy detection of early Parkinson's Disease using finger movement while typing"

3.PROPOSED SYSTEM:

The proposed work, there are four different classification algorithms were selected along with the two feature compressing methods as CFS with best-first search and Gain ratio with ranker mechanism. As described in the literature survey each algorithm is designed with an obtainable process in an optimized form, such a selected process may not be utilized to build a more competent method. The proposed method investigate and analyze four chosen method such as Hidden Markov Model (HMM), Artificial Neural Network (ANN), Support Vector Machine (SVM) and Decision Tree (J48) along with two other feature compressing methods. [2, 17] After analyzing these feature compressing methods, combine them with the linear models. And if any data is mismatched then reexamine with the other employed technique to improve the QoS.

MODULES USED

User Workflow:

- Users register with email and mobile, and get activated by the admin.
- They can upload datasets (in float format) and run classification algorithms (like XGBoost, SVM, ANN).
- Users can write reviews (positive, negative, neutral) after predictions, influencing the results.

Admin Workflow:

- Admin activates users and manages overall data.

- Admin can view results and the calculated accuracy of algorithms.

Data Preprocessing:

- Cleansed data is split into 60% for training and 40% for testing.
- Techniques used include noise removal, handling missing data, and adjusting values.

Machine Learning:

- Various classifiers are used (XGBoost, Extra Trees, AdaBoost, RF, SVM, ANN, Decision Tree).
- The highest accuracy classifier is selected as the best.

Visualization Module:

- Displays input datasets, training progress, and prediction charts
- Helps users analyse results through graphs (e.g., line charts, bar graphs)

TECHNOLOGIES USED

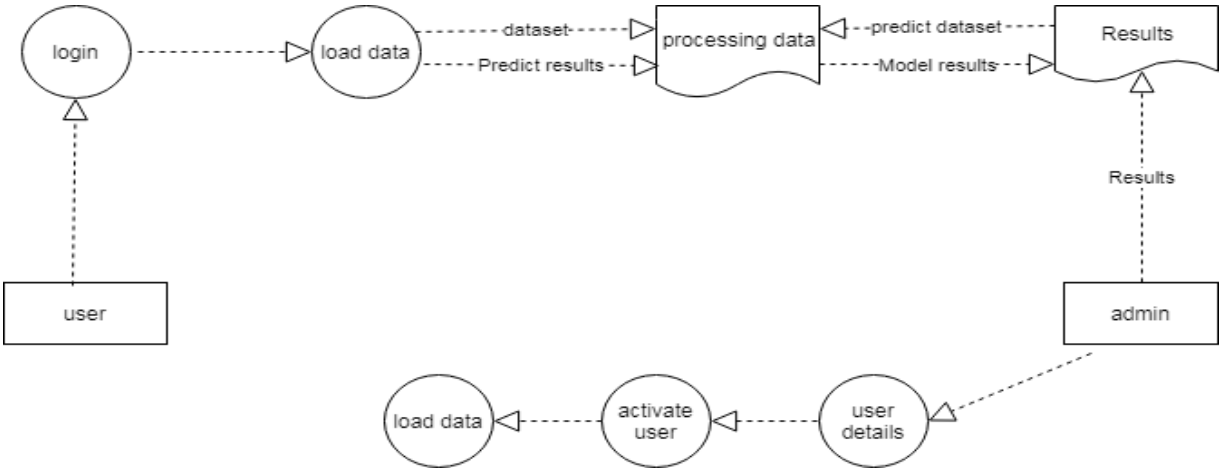
Programming Language: Python
Framework: Django
Tools: PyCharm, Visual Studio Code
Database: SQLite
Operating System: Windows 10
Frontend: HTML, CSS, JavaScript

ADVANTAGES OF SYSTEM:

- The comparative values show that the proposed method obtains higher accuracy when compared with other existing methods.
- The efficiency and suitability of the proposed approach are compared with other suggested methods.

4. ARCHITECTURE

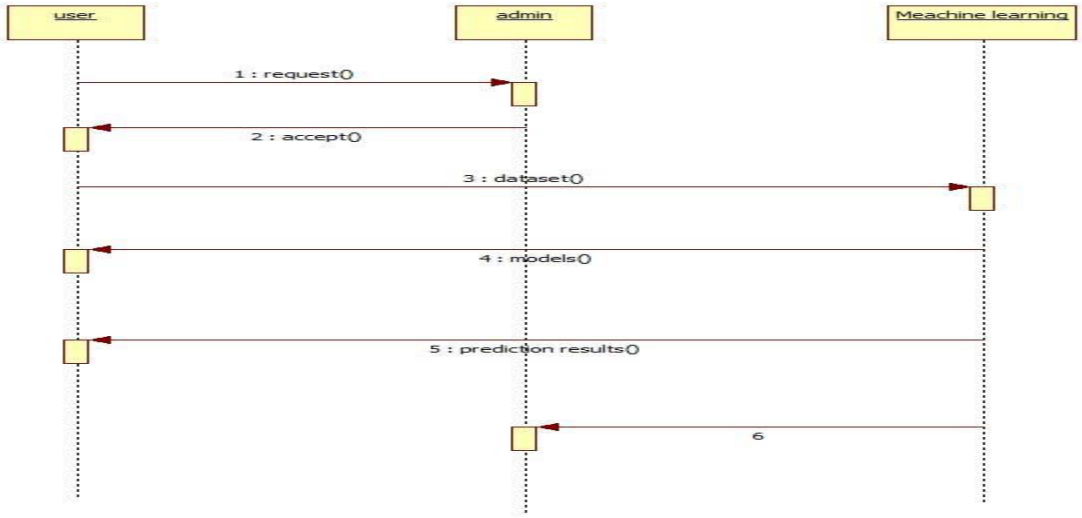
A **Data Flow Diagram (DFD)**, also known as a **bubble chart**, is a graphical tool used to represent a system by illustrating the **input data**, the **processes** that modify this data, and the **output** generated. It models system components like processes, data, external entities, and information flows. DFDs show how information moves and is transformed throughout the system. They can be represented at multiple levels of abstraction, with each level adding more detail about the system's functions and data flow.



Data flow diagram

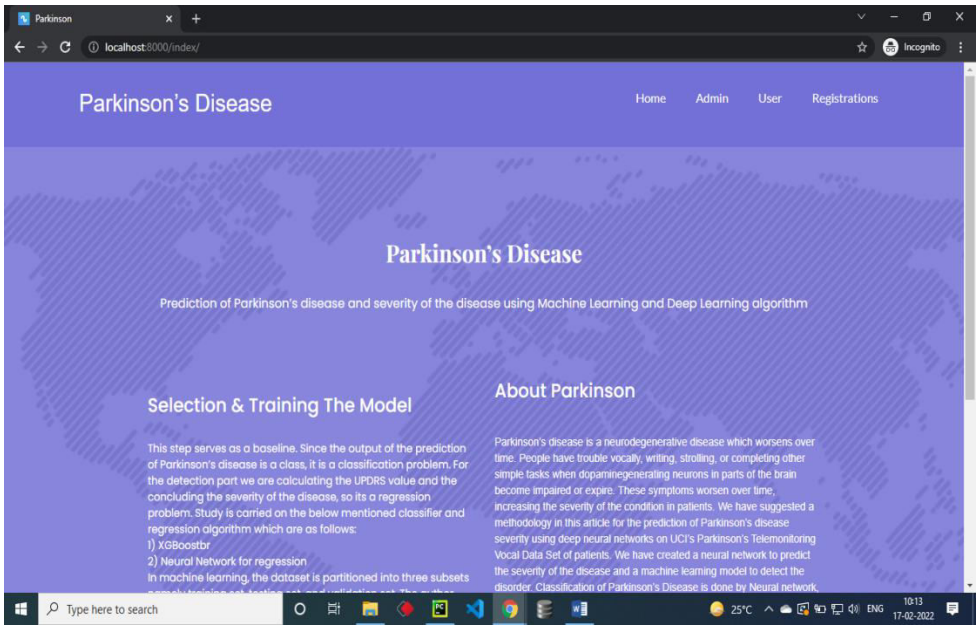
SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams

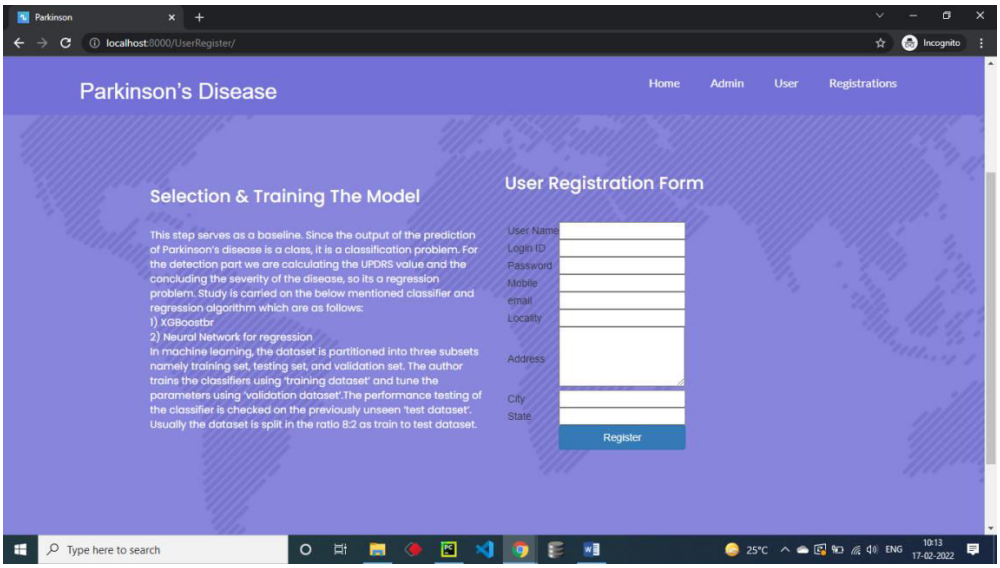


5.OUTPUT SCREENS

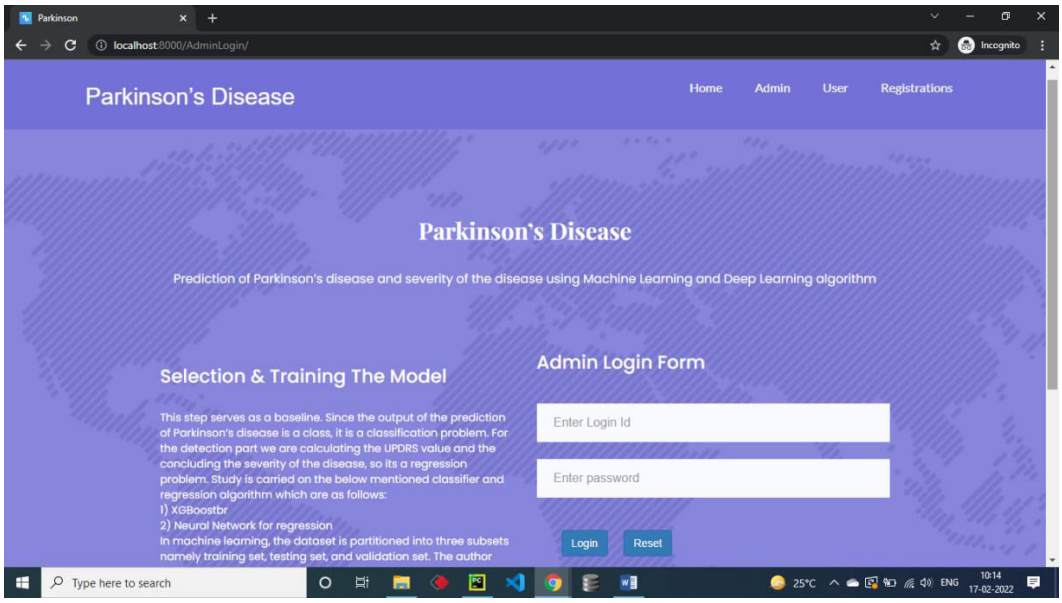
Home Page



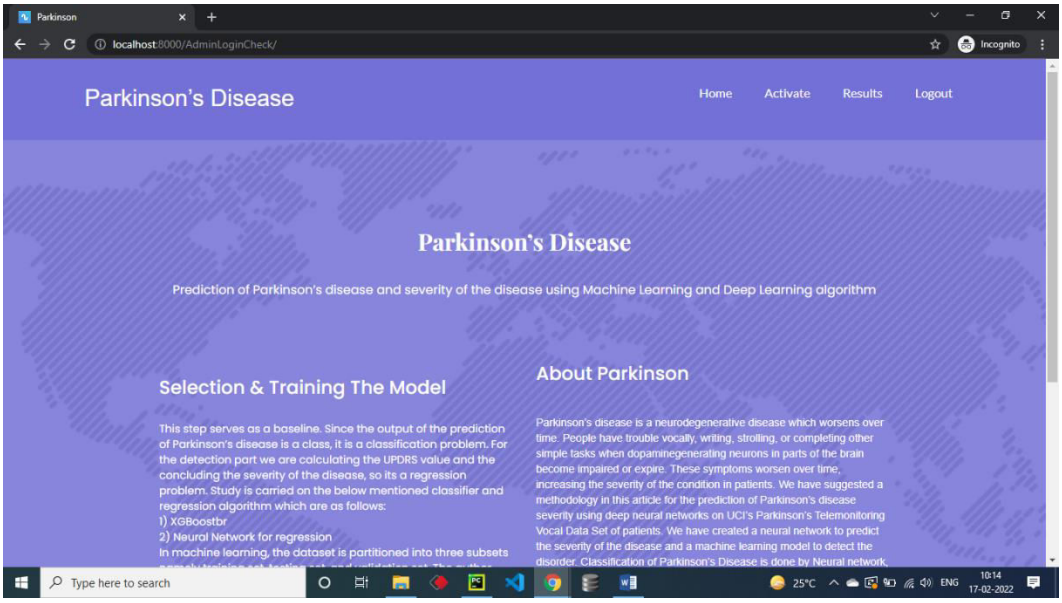
User Register Form



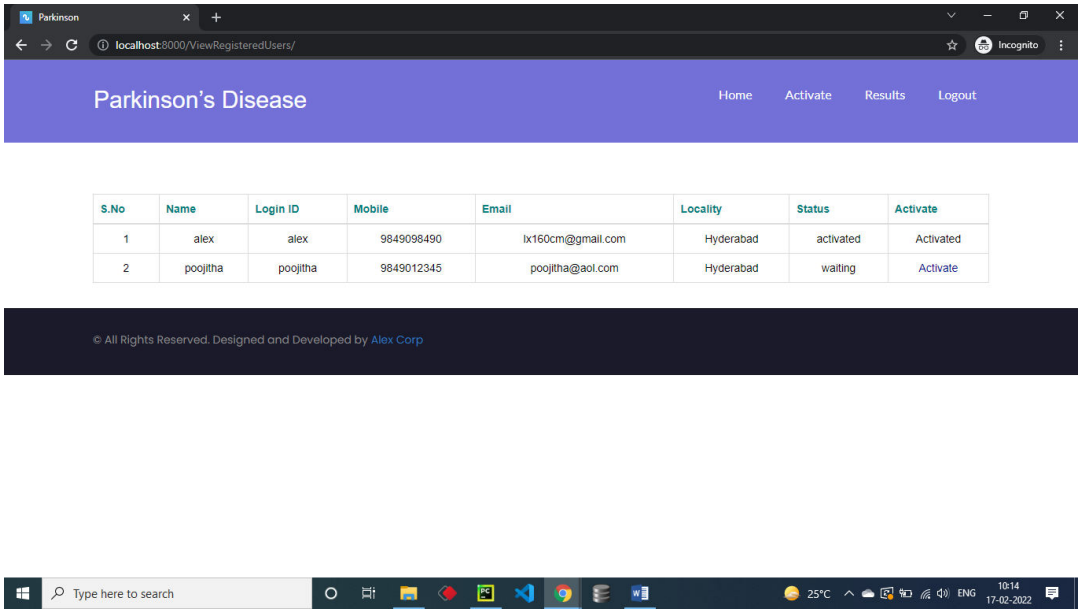
Admin Login Form



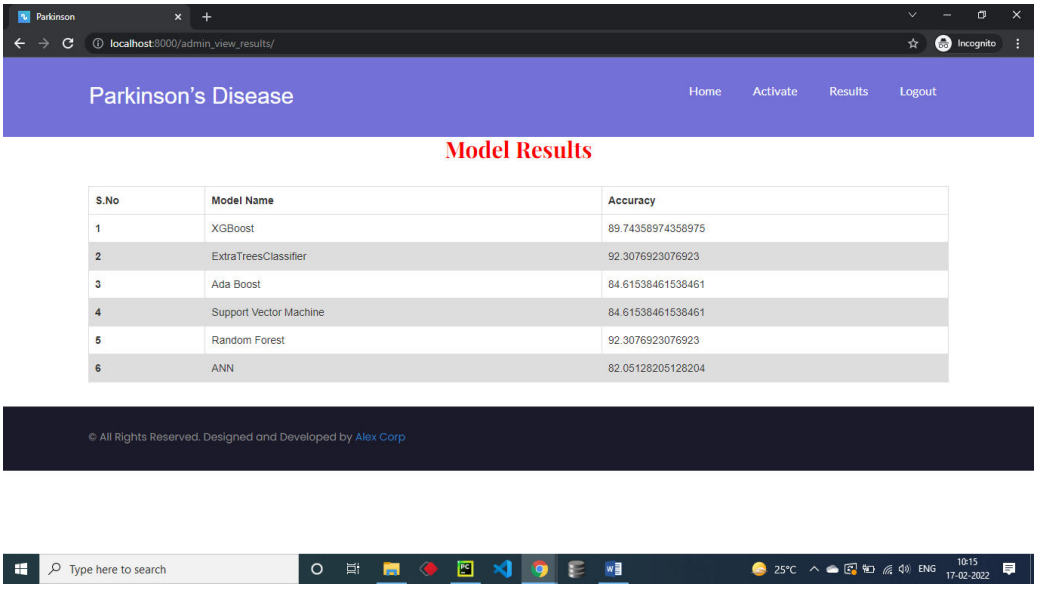
Admin Home Page



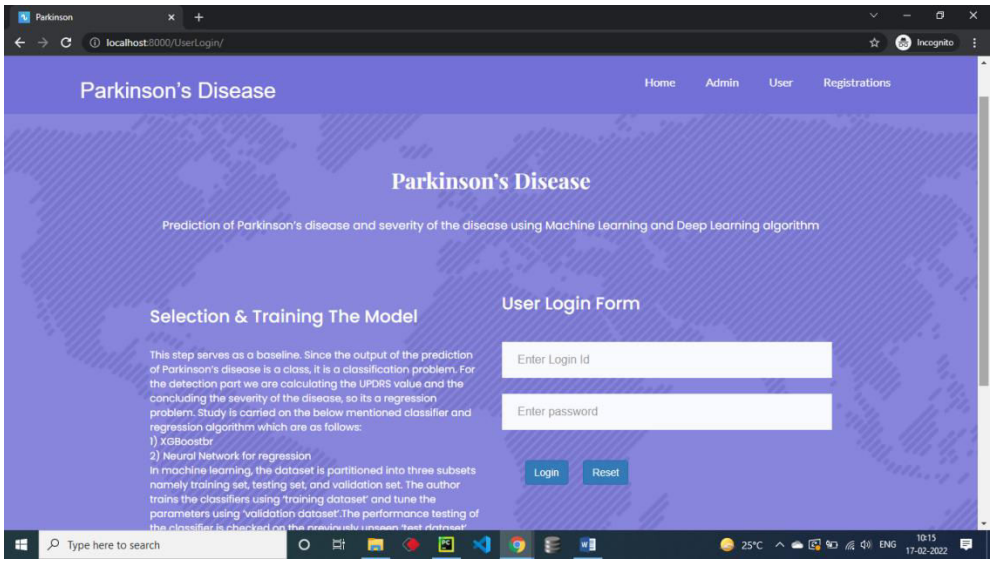
Admin view user and activate



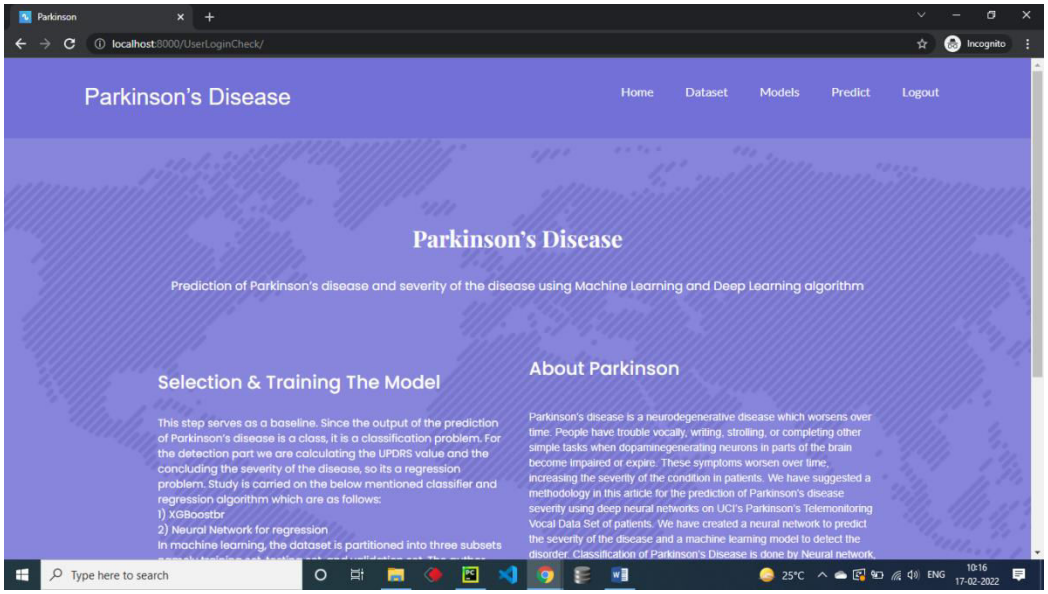
Admin view Results



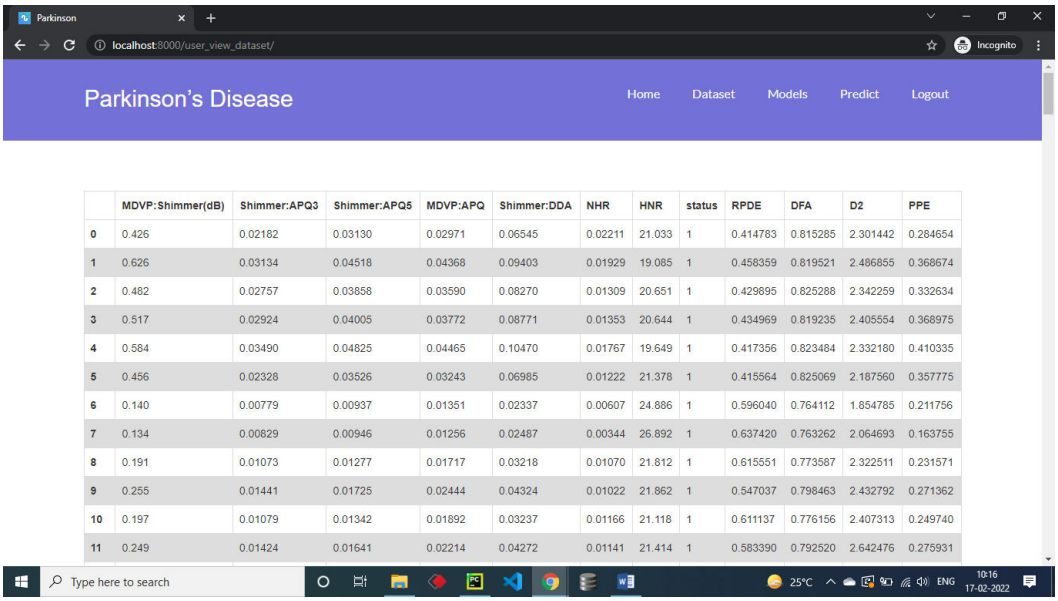
User login form



User Home Page

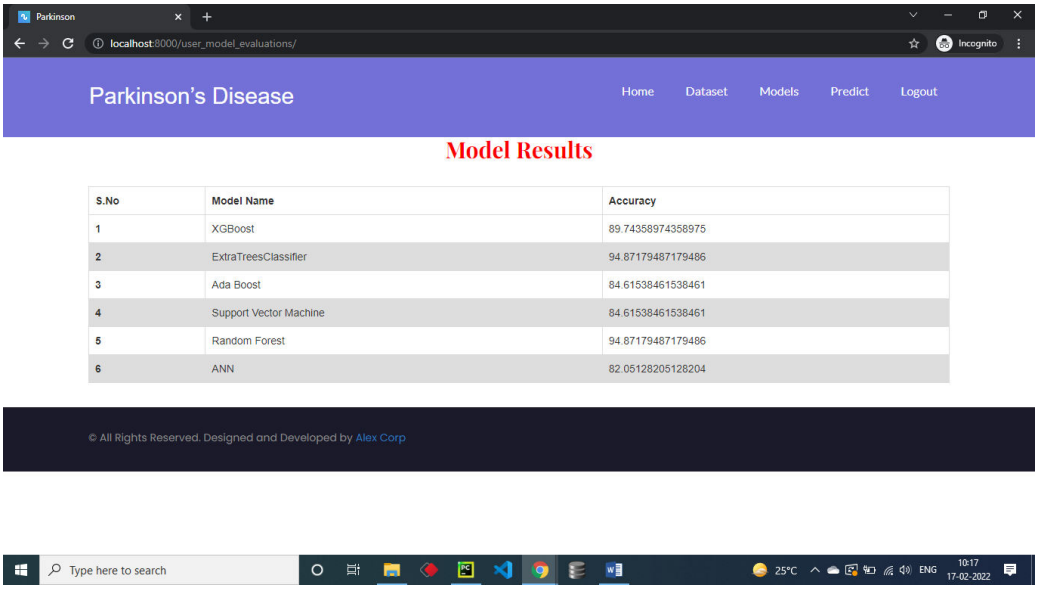


User view Dataset

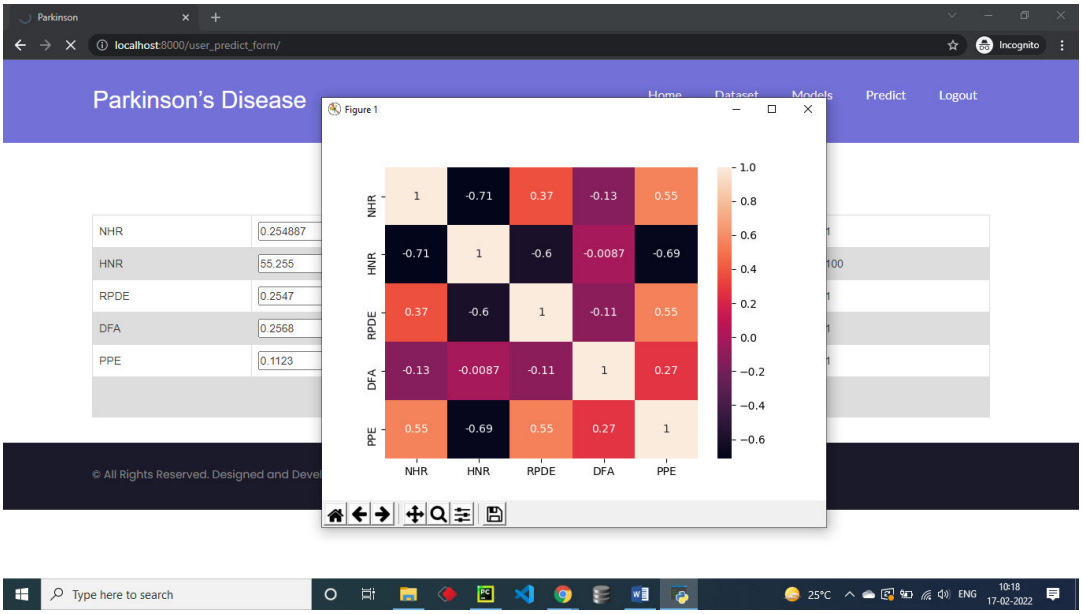


	MDVP:Shimmer(dB)	Shimmer:APQ3	Shimmer:APQ5	MDVP:APQ	Shimmer:DDA	NHR	HNR	status	RPDE	DFA	D2	PPE
0	0.426	0.02182	0.03130	0.02971	0.06545	0.02211	21.033	1	0.414783	0.815285	2.301442	0.284654
1	0.626	0.03134	0.04518	0.04368	0.09403	0.01929	19.085	1	0.458359	0.819521	2.486855	0.368674
2	0.482	0.02757	0.03858	0.03590	0.08270	0.01309	20.651	1	0.429895	0.825288	2.342259	0.332634
3	0.517	0.02924	0.04005	0.03772	0.08771	0.01353	20.644	1	0.434969	0.819235	2.405554	0.368975
4	0.584	0.03490	0.04825	0.04465	0.10470	0.01767	19.649	1	0.417356	0.823484	2.332180	0.410335
5	0.456	0.02328	0.03526	0.03243	0.06985	0.01222	21.378	1	0.415564	0.825069	2.187560	0.357775
6	0.140	0.00779	0.00937	0.01351	0.02337	0.00607	24.886	1	0.596040	0.764112	1.854785	0.211756
7	0.134	0.00829	0.00946	0.01256	0.02487	0.00344	26.892	1	0.637420	0.763262	2.064693	0.163755
8	0.191	0.01073	0.01277	0.01717	0.03218	0.01070	21.812	1	0.615551	0.773587	2.322511	0.231571
9	0.255	0.01441	0.01725	0.02444	0.04324	0.01022	21.862	1	0.547037	0.798463	2.432792	0.271362
10	0.197	0.01079	0.01342	0.01892	0.03237	0.01166	21.118	1	0.611137	0.776156	2.407313	0.249740
11	0.249	0.01424	0.01641	0.02214	0.04272	0.01141	21.414	1	0.583390	0.792520	2.642476	0.275931

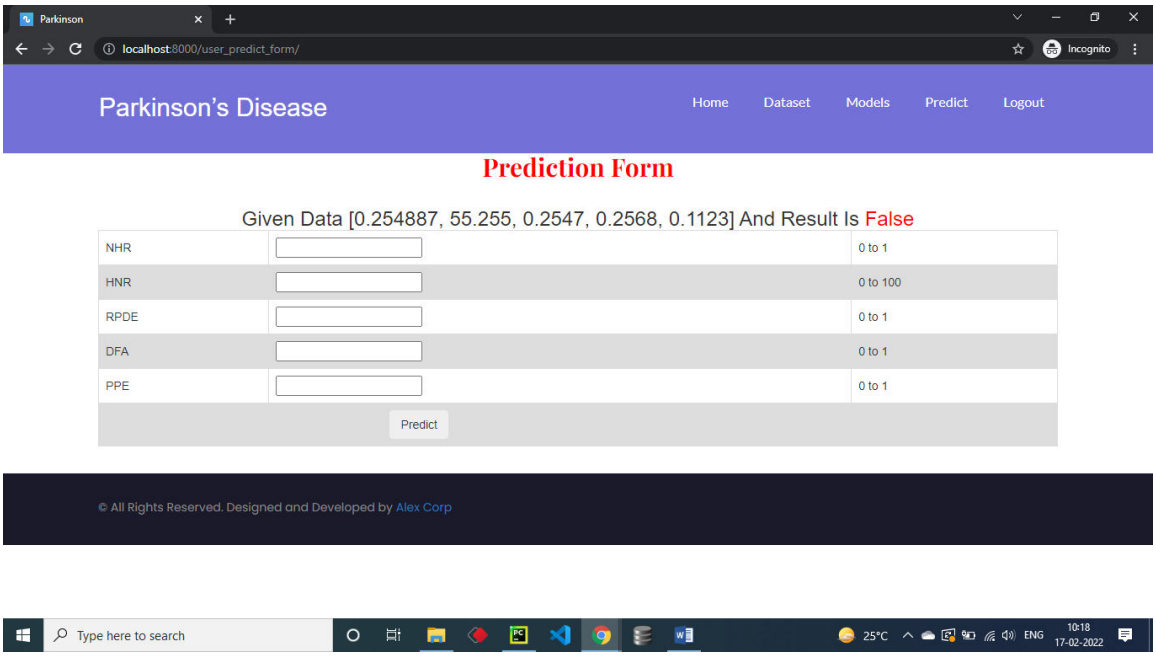
User view Models Results



Correlation Graph



Prediction Results



6.CONCLUSION

Parkinson's Disease (PD) is a significant research area due to its progressive natureand the importance of early-stage detection, which can greatly enhance patient health and quality of life. In this study, a proposed solution was able to effectively differentiate between early-stage Parkinson's patients and healthy individuals, achieving a sensitivity between 92% and 100%, a specificity of 95% to 100%, and an Area Under the Curve (AUC) ranging from 0.97 to 1.00. These performance metrics indicate a highly reliable diagnostic capability. It was also observed that Parkinson's Disease is more commonly detected in individuals over the age of 55, suggesting that age is a major risk factor. Additionally, the study found that females are more likely to develop Parkinson's than males, indicating a possible gender-based predisposition that warrants further investigation.

7.FURTHER ENHANCEMMENT

it can be concluded that the prognostication of Parkinson's Disease is very complex and is dependent on many variable factors which keeps on changing. If the features are properly selected we can get an optimized and efficient model which can get proper severity and extent of the spread of disease in the patientThe future scope for predicting Parkinson's disease using machine learning (ML) is quite promising and rapidly evolving. Here are some key areas where ML is making significant strides:

1. **Early Detection**: ML models are being developed to analyze various types of data, such as genetic information, clinical assessments, and lifestyle factors, to predict the onset of Parkinson's disease years before symptoms appear. This early detection can lead to timely interventions and better management of the disease1.
2. **Wearable Technology**: Wearable devices that monitor movement, sleep patterns, and other physiological parameters are being integrated with ML algorithms to detect early signs of Parkinson's disease. These devices provide continuous data, allowing for real-time monitoring and early intervention1.
3. **Voice and Speech Analysis**: ML models are being used to analyze changes in voice and speech patterns, which are often early indicators of Parkinson's disease. These models can detect subtle changes that may not be easily noticeable to human ears2.
4. **Imaging Techniques**: Advanced imaging techniques, such as MRI and CT scans, combined with ML algorithms, are being used to identify early structural and functional changes in the brain associated with Parkinson's disease.
5. **Polygenic Risk Scores**: ML models are incorporating polygenic risk scores, which take into account multiple genetic factors, to improve the accuracy of Parkinson's disease predictions.
6. **Ensemble Learning**: Techniques like Random Forest and XGBoost are being used to combine multiple ML models, increasing the accuracy and reliability of predictions.

These advancements in ML are paving the way for more accurate and early predictions of Parkinson's disease, ultimately leading to better patient outcomes and quality of life.

8.REFERENCES

[1] Das R. “Comparison of multiple classification methods for diagnosis of Parkinson disease”. *Expert Systems With Applications*, 37:1568–1572, 2010.

[2] Genain N, Huberth M, and Vidyashankar R. “Predicting Parkinson’s Disease Severity from Patient Voice Features.” 2014.

[3] Benmalek E, Elmhamdi J, and Jilbab A. “UPDRS tracking using linear regression and neural network for Parkinson’s disease prediction.” *International Journal Of Emerging Trends and Technology In Computer Science*, 4:189–193, 2015.

[4] Seeja K.R., Srishti Grover, Saloni Bhartia, Akshama, and Abhilasha Yadav. “Predicting Severity Of Parkinson’s Disease Using Deep Learning”. *Procedia Computer Science*, 132:1788–1794.

[5] Warwick R. Adams. “High-accuracy detection of early Parkinson’s Disease using multiple characteristics of finger movement while typing”. 2017.

[6] Cole B, Roy S, De Luca C, and Nawab S. “Dynamical Learning and Tracking of Tremor and Dyskinesia From Wearable Sensors.” *IEEE Transactions On Neural Systems And Rehabilitation Engineering*, 22:982–991, 2018.

[7] Tappy keystroke dataset.

[8] Parkinson’s telemonitoring dataset.

[9] Google collaborator.

[10] Keras api reference.

[11] Python machine learning course.

[12] Deep learning course.

[13] Multiple output neural networks.