
ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN2321-2152

www.ijmece .com

Vol 12, Issue.2, 2024

PARKINSON'S DISEASE PREDICTION

[1] Mihaj begum, [2]Sai priya mathangi, [3]2. Rathnam Shivaani Siri

[1]Associate professor in Department of Information Technology Bhoj Reddy Engineering College for Women

[2,3] UG Scholars in Department of Information Technology Bhoj Reddy Engineering College for Women

Abstract

Accurately detecting Parkinson's disease (PD) at an early stage is certainly indispensable for slowing down its progress and providing patients the possibility of accessing to disease- modifying therapy. Towards this end, the preemptor stage in PD should be carefully monitored. An innovative deep-learning technique is introduced to early uncover whether an individual is affected with PD or not based on pre motor features. Specifically, to uncover PD at an early stage, several indicators have been considered in this study, including Rapid Eye Movement and olfactory loss, Cerebrospinal fluid data, and dopaminergic imaging markers. A comparison between the proposed deep learning model and twelve machine learning and ensemble learning methods based on relatively small data including 183 healthy individuals and 401 early PD patients shows the superior detection performance of the designed model, which achieves the highest accuracy, 96.45% on average. Besides detecting the PD, we also provide the feature importance on the PD detection process based on the Boosting method.

Key words: Parkinson's disease, Machine learning, Classification, Random forest

I INTRODUCTION

Parkinson's disease (PD) is becoming an important degenerative disease of the central nervous system, affecting the quality of lives of millions of seniors worldwide. Symptoms of PD can progress differently from one person to another because of the variety of the disease. Patients with Parkinson may show symptoms including tremors mainly at rest. Different types of tremors are possible: tremors in hands, limb rigidity, and gait and balance problems. Generally, two types of symptoms of PD can be distinguished: movement-related (i.e., motor) and unrelated to movement (non-motor). In fact, patients showing non-motor symptoms are more affected than whose main symptoms are motor.

Non-motor symptoms may include depression, sleep behavior disorders, loss of sense of smell, and cognitive impairment. It has been reported by the Centers for Disease Control and Prevention (CDC) that PD complications are ranked as the 14th leading causes of death in the

United States. To date, the cause of PD rests principally unknown. Particularly, the economic burden due to direct and indirect cost of PD covering treatment, social security payments, and lost income is estimated to be approximately $52 billion per year in the United States alone. Actually, the number of people affected by PD has exceeded 10 million worldwide. It should be noted that the timely detection of the PD facilitates rapid treatment and alleviate symptoms significantly as reported in. Therefore, detection of PD at an earlier stage is certainly akey element to slowing down its progression and could give patients the possibility of accessing to disease-modifying therapy, when available. Till now, there is no way to diagnose Parkinson's disease (PD). However, there are various symptoms and diagnostic tests used in combination. Several biomarkers have been investigated by scientists to early identify PD to slow down the disease process. Currently, all therapies used for PD improve symptoms without slowing or halting the disease progression. Various methods have proposed to help detection PD based on different kinds of measurements including speech data, gait patterns, force tracking data, smell identification data and spontaneous cardiovascular oscillations.

II LITERATURE SURVEY

Voice features of the patients are assumed to be 90% helpful to identify the presence of Parkinson's disease. Different researchers applied different features of the dataset to predict the disease. Many of the authors used the voices of the patients to analyze Parkinson's. In general, the speech problems are two and are hypophonia and dysarthria. Hypophonia means a very weak and soft tone from a person. Dysarthria means a slow voice that is understandable at one time caused due to the central nervous system. Arvind Kumar Tiwari proposed a paper "Machine Learning-based Approaches for Prediction of Parkinson's Disease". The dataset used in this paper is voice recordings of the patients. And in this paper, the author chooses the most important features among all features to predict Parkinson's disease by using minimum redundancy maximum relevance feature selection algorithms. The author applied different machine learning algorithms and compared them. Random Forest provides the highest accuracy of 90.3%. Carlo Ricciardi, et al, proposed a paper "Using gait analysis' parameters to classify Parkinsonism: A data mining approach". In this paper, the author compares performance analysis using two algorithms. The algorithms used are Random Forest and Gradient Boosted Trees. In this paper, the PD patients at different stages were taken into consideration and are identified as typical and atypical based on gait analysis using a data mining approach.

The comparative analysis gives the result as Random Forest obtained the highest accuracy

of 86.4%. This model also helped clinicians to distinguish PD patients at an early stage. Dr.Anupam Bhatia and Raunak Sulekh proposed a paper “Predictive Model for Parkinson’s Disease through Naive Bayes Classification”. This paper summarizes that the Naïve Bayes classifier is used to analyse the performance of the dataset. The dataset used in this paper is recorded speech signals. The features in this dataset are voice measures and the aim is to predict the PD 0 for healthy and 1 for a person having the disease. In this paper, the author has used the Rapid miner tool to analyse the data. Naïve Bayes classifier produces 98.5% accuracy. This model helps doctors and patients to detect the disease and take preventive.

Dragana Miljkovic, et al, proposed a paper “Machine Learning and Data Mining Methods for Managing Parkinson’s Disease”. In this paper, the author concluded that based on the medical tests taken by the patients the Predictor part was able to predict the 15 different Parkinson’s symptoms separately. The machine learning and data mining techniques are applied on different symptoms separately and gives an accuracy range between 57.1% and 77.4% where tremor detection has the highest accuracy. M. Abdar and M. Zomorodi-Moghadam proposed a paper “Impact of Patients’ Gender on Parkinson’s disease using Classification Algorithms”. In this paper, the author chooses the UCI

PD dataset for finding the accuracy of Parkinson’s using SVM and Bayesian Network algorithms. The author chooses the most ten important features in the dataset to predict PD. The output variable is Sex and other factors are input, the author provides an approach for finding relationships between genders. The result obtained is SVM algorithm gives better performance than Bayesian Network with 90.98% accuracy. Md. Redone Hassan, et al, proposed a paper “A Knowledge Base Data Mining based on Parkinson ’s disease”. In this paper, different classification algorithms are used to predict Parkinson’s disease such as SVM, KNN, and Decision Tree. These algorithms are applied to the training dataset and provides different accuracies. The paper summarizes that the Decision tree algorithm provides 78.2% precision compared to the remaining algorithms.

Mehrbakhsh Nilashi, et al, proposed a paper “A hybrid intelligent system for the prediction of Parkinson’s Disease progression using Machine Learning techniques”. In this paper, Unified Parkinson's Disease Rating Scale (UPDRS) is mostly used to assess Parkinsonism. The author described the relationship between speech signals and UPDRS is important to find Parkinson’s.In this paper, the Incremental Support Vector Machine is used to predict Total-

UPDRS and Motor UPDRS. The author concluded that a combination of ISVR, SOM, and NIPALS are used to get effective results to predict the UPDRS .

III PROBLEM STATEMENT

Many of the people aged 65 or more do have a neurodegenerative disease, which has no cure. If we detect the disease in the early stages, then we can control it. Almost 30% of the patients are facing this incurable disease. Current treatment is available for patients who have minor symptoms. If these symptoms cannot be found at the early stages, it leads to death. The main cause for Parkinson's disease is the accumulation of protein molecules in the neuron which gets misfolded and hence causing Parkinson's disease. So till now, researchers got the symptoms and the root causes i.e. from where this disease had evolved. But very few symptoms have come to their cure and there are many symptoms that have no solution. So in this era where Parkinson's disease is increasing, it is very important to find the solution which can predict it in its early stages.

IV EXISTING SYSTEM

PD is detected at the secondary stage only (Dopamine deficiency) which leads to medical challenges. Also doctor has to manually examine and suggest medical diagnosis in which the symptoms might vary from person to person so suggesting medicine is also a challenge. Thus, the mental disorders are been poorly characterized and have many health complications. Existing system is not effective in early prediction and accurate medicinal diagnosis to the affected people.

Disadvantages:

➢ MRI and CT scan- conventional MRI cannot detect early signs of parkinsom's disease.

➢ PET scan - is used to assess activity and function of brain region involved in movements.

➢ SPECT scan – can reveal changes in brain chemistry, such as a decrease in dopamine.

➢ Current treatments for parkinson's disease focus on managing symptoms rather than preventing or slowing the progression of the disease. There is a need for more targeted therapies that can address the underlying causes halt or slow the degenerative process.

V PROPOSED SYSTEM

by using machine learning techniques, the problem can be solved with minimal error rate. The voice dataset of Parkinson's disease from the Kaggle website is used as input and it helps in performing the training and testing of the model. We propose a hybrid and accurate results analyzing patient voice. Thus, combining the results, the doctor can conclude normality or abnormality and prescribe the medicine based on the affected stage

ISSN2321-2152
www.ijmece .com
Vol 12, Issue.2, 2024

Advantages:

 A well-designed system could incorporate early detection methods, allowing for the identification of individuals at risk before the onset of symptoms. Early intervention and treatment may help slow down or prevent the progression of the disease.

 A proposed system could take into account individual genetic, environmental, and lifestyle factors to provide personalized risk assessments.

 The system could offer evidence-based lifestyle recommendations, including diet, exercise, and other health-promoting behaviors that may help reduce the risk of Parkinson's disease.

 Continuous monitoring of relevant health indicators and feedback mechanisms could be integrated into the system. This would allow individuals to track their progress in adopting preventive measures and make necessary adjustments.

VI IMPLEMENTATION

Programs are composed of one or more independently developed modules. A module description provides detailed information about a module and its supported components. The included description is available directly by making environment check for supported components. The modules are

 Pre-Processing

 Build Model

 Train Model

 Creating UI

 Integration of UI with model

 Test Model

Pre-Processing:

This is the module that is implemented in Project Stage 1. The pre-processing done on the statistics that have been obtained from previous patients voice data. The raw statistics are too broad and hence not useful for our model. The actual statistics that will be useful are aggregate statistics which will help us predict the disease accurately. Hence, we go through an entire report of patient's vocal data in Parkinson disease prediction and calculate the aggregate statistics for every patient

Build Model:

In this section, we will explore data and find patterns, which may be useful for feature engineering later. Since the goal of the project is to build the car price prediction model, the main purpose of this chapter is to find relations between features and the target column (selling_price_inr)

Train Model:

In this module, we must train the machine learning model for it to be fully used for prediction

of Parkinson disease. We have used SK-Learn for training the model. The data split for training80% and testing 20% of the total data. Training a model simply means learning (determining) good values for all the weights and the bias from labeled examples. In supervised learning, a machine learning algorithm builds a model by examining many examples and attempting.

Creating UI:

Frontend: The front-end of this project is created using Stream lit.The purpose of creating a front-end is for an easy and clean UI. This UI consists of 4 pages:

- Home page: This page gives us in detail description about the Parkinson disease detection.

- Data Info Page: This page describes the data which is used to train the model is shown in this page.

- Predict Page : In this page we will be giving the values as input and gets the response in form of prediction

- Visualization Page : In this page different graphs are shown for the data which weused to train the model.

Backend: We have used various python libraries for backend functionality. It is a lightweight and fast Python-based framework. In our project it is used as an interface

between the user queries and Machine learning model i.e., it fetches the prediction results, stats, etc from our ML model. Routing to different pages is done here.

Integration of UI with model:

In this approach, you describe the UI and related aspects (e.g., task, domain, context of use) in models. These models are then transformed into the final UI automatically. It involves methods, models, processes, and algorithms to achieve this transformation. The idea is to reduce the effort needed for UI creation by leveraging model transformations. Transformation rules define how to query data from the source model and convert it into

Packing your ML Model:

- Securing your packaged ML model

Planning of how to serve/expose your ML model to the consumers (as a REST service, etc.) If you have planned for a REST service, then, creating a REST API for your ML model

- Securing your REST API

Deploying your REST service in production (using Docker and Kubernetes)

ISSN2321-2152
www.ijmece .com
Vol 12, Issue.2, 2024

 Packaging your ML model

Instead of saving the ML model as it is, you can package your ML model (say mnist), and create a .pkl file for it, using Python's joblib library.

For creating and restoring the .pkl file, you can either use joblib library or pickle library of Python. You normally use joblib to save an object having large data, else, you use pickle library. Here, in this case, we have used joblib library..pkl file is nothing but a serialized pickle file, which if you want, you can compress it further, to save storage space, using say Python's gzip library. After you apply compression, your ML model file name will look like – mnist.pkl.gz

VII RESULTS

	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	...	Shim
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.00007	0.00370	0.00554	0.01109	0.04374	...	
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	...	
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.00009	0.00544	0.00781	0.01633	0.05233	...	
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	...	
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	...	
5	phon_R01_S01_6	120.552	131.162	113.787	0.00968	0.00008	0.00463	0.00750	0.01388	0.04701	...	
6	phon_R01_S02_1	120.267	137.244	114.820	0.00333	0.00003	0.00155	0.00202	0.00466	0.01608	...	
7	phon_R01_S02_2	107.332	113.840	104.315	0.00290	0.00003	0.00144	0.00182	0.00431	0.01567	...	
8	phon_R01_S02_3	95.730	132.068	91.754	0.00551	0.00006	0.00293	0.00332	0.00880	0.02093	...	
9	phon_R01_S02_4	95.056	120.103	91.226	0.00532	0.00006	0.00268	0.00332	0.00803	0.02838	...	

10 rows × 24 columns

Data Pre-processing

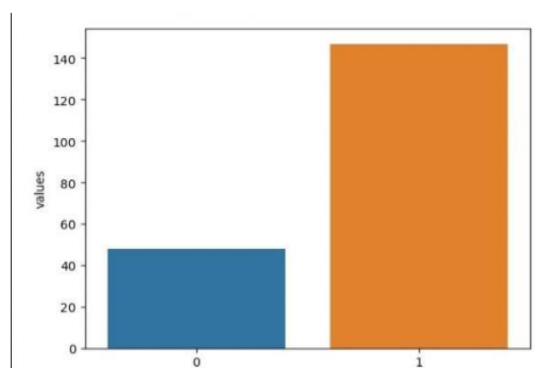
```
name                 object
MDVP:Fo(Hz)         float64
MDVP:Fhi(Hz)        float64
MDVP:Flo(Hz)        float64
MDVP:Jitter(%)      float64
MDVP:Jitter(Abs)    float64
MDVP:RAP            float64
MDVP:PPQ            float64
Jitter:DDP          float64
MDVP:Shimmer        float64
MDVP:Shimmer(dB)    float64
Shimmer:APQ3        float64
Shimmer:APQ5        float64
MDVP:APQ            float64
Shimmer:DDA         float64
NHR                 float64
HNR                 float64
status                int64
RPDE                float64
DFA                 float64
spread1             float64
spread2             float64
D2                  float64
PPE                 float64
dtype: object
```

Data cleaning and data visualization

################################ name ############################

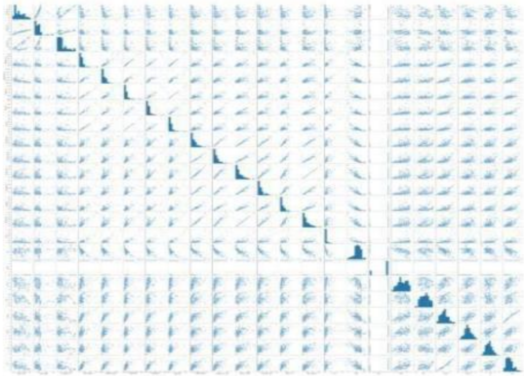
{'phon_R01_S16_3', 'phon_R01_S31_4', 'phon_R01_S43_5', 'phon_R01_S49_6', 'phon_R01_S10_2', 'phon_R01_S21_5', 'phon_R01_S35_6', 'phon_R01_S20_6', 'phon_R01_S32_5', 'phon_R01_S39_5', 'phon_R01_S19_2', 'phon_R01_S24_6', 'phon_R01_S49_5', 'phon_R01_S19_4', 'phon_R01_S06_6', 'phon_R01_S50_6', 'phon_R01_S06_1', 'phon_R01_S25_2', 'phon_R01_S27_1', 'phon_R01_S43_4', 'phon_R01_S16_2', 'phon_R01_S16_6', 'phon_R01_S13_1', 'phon_R01_S16_4', 'phon_R01_S31_6', 'phon_R01_S27_5', 'phon_R01_S07_6', 'phon_R01_S10_1', 'phon_R01_S24_4', 'phon_R01_S33_1', 'phon_R01_S18_2', 'phon_R01_S44_2', 'phon_R01_S08_6', 'phon_R01_S26_5', 'phon_R01_S35_1', 'phon_R01_S44_5', 'phon_R01_S43_2', 'phon_R01_S42_6', 'phon_R01_S16_5', 'phon_R01_S34_3', 'phon_R01_S05_6', 'phon_R01_S50_4', 'phon_R01_S04_4', 'phon_R01_S26_3', 'phon_R01_S10_3', 'phon_R01_S34_1', 'phon_R01_S08_5', 'phon_R01_S02_3', 'phon_R01_S13_6', 'phon_R01_S39_4', 'phon_R01_S21_1', 'phon_R01_S32_2', 'phon_R01_S42_4', 'phon_R01_S04_6', 'phon_R01_S20_5', 'phon_R01_S13_2', 'phon_R01_S22_2', 'phon_R01_S43_1', 'phon_R01_S27_3', 'phon_R01_S13_4', 'phon_R01_S32_4', 'phon_R01_S19_6', 'phon_R01_S17_2', 'phon_R01_S01_2', 'phon_R01_S17_5', 'phon_R01_S05_3', 'phon_R01_S19_3', 'phon_R01_S05_2', 'phon_R01_S25_4', 'phon_R01_S27_2', 'phon_R01_S37_2', 'phon_R01_S06_4', 'phon_R01_S01_5', 'phon_R01_S07_4', 'phon_R01_S34_2', 'phon_R01_S19_1', 'phon_R01_S25_6', 'phon_R01_S33_6', 'phon_R01_S37_5', 'phon_R01_S31_1', 'phon_R01_S01_4', 'phon_R01_S49_1', 'phon_R01_S44_1', 'phon_R01_S02_6', 'phon_R01_S08_2', 'phon_R01_S37_6', 'phon_R01_S19_5', 'phon_R01_S50_5', 'phon_R01_S21_3', 'phon_R01_S35_3', 'phon_R01_S04_1', 'phon_R01_S02_5', 'phon_R01_S18_5', 'phon_R01_S24_1', 'phon_R01_S02_2', 'phon_R01_S27_4', 'phon_R01_S39_1', 'phon_R01_S20_3', 'phon_R01_S31_3', 'phon_R01_S34_4', 'phon_R01_S17_6', 'phon_R01_S21_7', 'phon_R01_S35_5', 'phon_R01_S26_2', 'phon_R01_S02_1', 'phon_R01_S44_4', 'phon_R01_S20_2', 'phon_R01_S25_1', 'phon_R01_S07_2', 'phon_R01_S26_4', 'phon_R01_S31_5', 'phon_R01_S18_3', 'phon_R01_S22_4', 'phon_R01_S10_6', 'phon_R01_S18_6', 'phon_R01_S22_1', 'phon_R01_S22_6', 'phon_R01_S07_5',

Cross checking the column values

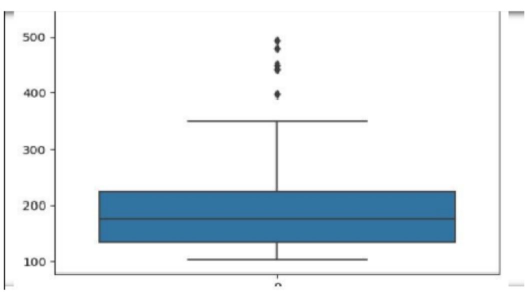


Checking imbalance between categorization

ISSN2321-2152
www.ijmece .com
Vol 12, Issue.2, 2024

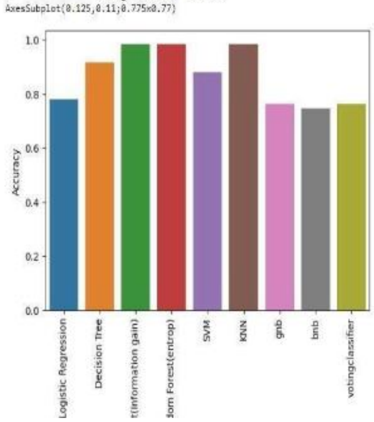


Visualization of correlation between columns



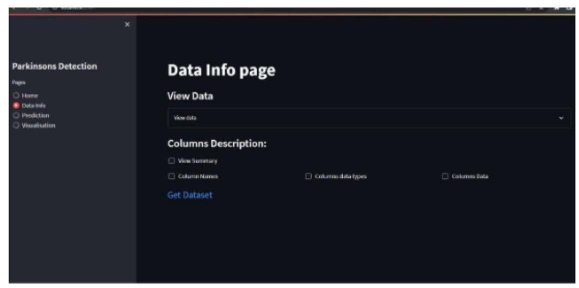
Finding outliers

```
0.9830508474576272
                          Method Used  Accuracy
0              Logistic Regression  0.779661
1                    Decision Tree  0.915254
2  Random Forest(Information gain)  0.983051
3           Random Forest(entrop)  0.983051
4                              SVM  0.881356
5                              KNN  0.983051
6                              gnb  0.762712
7                              bnb  0.745763
8                 votingclassifier  0.762712
AxesSubplot(0.125,0.11;0.775x0.77)
```

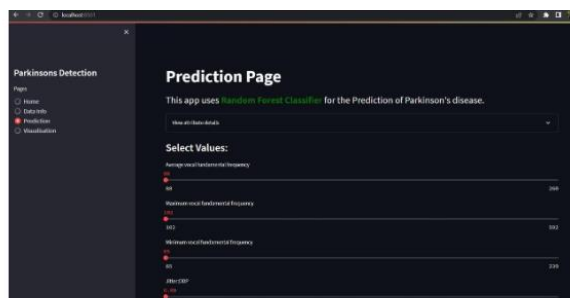


Testing accuracy

Data info page



Visualize page



Predict page

VIII CONCLUSION

Its aimed to predict Parkinson's disease using machine learning (ML) techniques. Parkinson's disease prediction project demonstrated the potential of ML algorithms to accurately predict the presence of the disease. This project has implications for early detection, personalized treatment, and improved healthcare outcomes for individuals with Parkinson's disease. We have used various algorithms for testing whether person have Parkinson's disease. Among all the algorithms Random Forest gave the highest accuracy 98%

REFERENCES

[1] Tasanas ,M. A., McSharry, P. E., Spielman, J., & Ramig, L. O. (2012). Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Transactions on Biomedical Engineering, 59(5), 1264-1271.

[2] Tsanas, A., & Tsipouras, M. G. (2014). Machine learning for the automatic assessment of Parkinson's disease. IEEE Transactions on Biomedical Engineering, 61(1), 88-97.

[3] Ossig, C., Antonini, A., Buhmann, C., Classen, J., Csoti, I., Falkenburger, B., ... & Reichmann, H. (2013). Deep brain stimulation in Parkinson's disease: a German multicentre study with 111 patients. Journal of Neurology, Neurosurgery & Psychiatry, 84(6), 670-675.

[4] Tsanas, A., & Little, M. A. (2014). Detecting Parkinson's disease from smartphones'movement data. In Proceedings of the 16th international conference on multimodal interaction (pp. 73-76).

[5] Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., ... & Apaydin, H. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics, 17(4), 828-834.

[6] Arlot, S., Celisse, A., & Hocquet, C. (2010). A survey of cross-validation procedures for model selection. Statistics Surveys, 4, 40-79