



IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: IV Month of publication: April 2023

DOI: https://doi.org/10.22214/ijraset.2023.51241

www.ijraset.com

Call: 🕥 08813907089 🔰 E-mail ID: ijraset@gmail.com



Parkinson's Disease Detection using Ensemble Learning

Sachchit Kolekar¹, Naman Jain², Amit Mete³, Prof. Nilesh Kulal⁴ Department of Computer Engineering, Terna Engineering College, Nerul, Maharashtra, India

Abstract: In this decade of rapid developments in medical science, most research fail to focus on age related disorders. These are illnesses that manifest their symptoms at a far later stage, making complete recovery practically impossible. Parkinson's disease (PD) is the brain's second most prevalent neurodegenerative condition. One may claim that it is nearly incurable and causes significant suffering to people. All of this indicates that there is an impending demand for accurate, trustworthy, and expandable Parkinson's disease diagnosis. A problem of this magnitude necessitates the automation of the diagnostic to lead accurate and reliable results. Most Parkinson's disease patients have some type of speech impairment or dysphonia, making speech measures and indicators one of the most essential parts in PD prediction. The Goal of this work is to compare various machine learning models in successfully predicting the severity of Parkinson's disease and develop an effective and accurate model to help diagnose the disease accurately at an earlier stage, which could help doctors assist in cure and recovery of PD patients. We want to use the Parkinson's Telemonitoring dataset obtained from the UCI ML repository for the aforementioned purpose. Five Different Classification algorithms, including decision tree, random forest, logistic regression, support vector machine, and k-nearest neighbors, were used to create individual models. The Ensemble learning method was then applied to combine the predictions of these individual.

I. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative movement disease where the symptoms gradually develop start with a slight tremor in one hand and a feeling of stiffness in the body and it became worse over time. It affects over 6 million people worldwide. At present there is no conclusive result for this disease by non-specialist clinicians, particularly in the early stage of the disease where identification of the symptoms is very difficult in its earlier stages. Diagnosis of Parkinson's disease (PD) is commonly based on medical observations and assessment of clinical signs, including the characterization of a variety of motor symptoms. However, traditional diagnostic approaches may suffer from subjectivity as they rely on the evaluation of movements that are sometimes subtle to human eyes and therefore difficult to classify, leading to possible misclassification. In the meantime, early non-motor symptoms of PD may be mild and can be caused by many other conditions. Therefore, these symptoms are often overlooked, making diagnosis of PD at an early stage challenging. To address these difficulties and to refine the diagnosis and assessment procedures of PD, machine learning methods have been implemented for the classification of PD. This program is for the people who wants to know if they have Parkinson disease or not. The scope of this project is to show high accuracy of detecting Parkinson's disease in early stage

II. RELATED WORK

Early Detection of Parkinson's disease using Deep Learning and Machine learning [2016] By Wu Wang, Junho Lee, Fouzi Harrou. This study proposed a deep learning model to automatically discriminate normal individuals and patients affected by PD based on premotor features (i.e., Rapid Eye Movement (REM) sleep Behaviour Disorder (RBD) and olfactory loss). The pro- posed deep learning model showed good detection capacity by reaching an accuracy of 96.45%.

III. METHODOLOGY

A. Description of Data

The first dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voices recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD. (a.)



International Journal for Research in Applied Science & Engineering Technology (IJRASET)

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

There are no missing values after Data cleaning (b.) No duplicate rows in the given dataset Attribute Information: name - ASCII subject name and recording number • mdvp_fo_hz: Average vocal fundamental frequency (Actual column name MDVP:Fo(Hz) • mdvp_fhi_hz :Maximum vocal fundamental frequency (Actual column name MDVP:F hi(Hz)) • mdvp_flo_hz :Minimum vocal fundamental frequency (Actual column name MDVP:F lo(Hz)) • mdvp_jitter_abs, mdvp_rap, mdvp_ppq, jitter_ddp :Several m easures of variation in fundamental frequency (Actual column names MDVP:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP respectively) • mdvp_shimmer, mdvp_shimmer_db, shimmer_apq3, shimmer_apq5, mdvp_apq, shim mer_dda :Several measures of variation in amplitude (Actual column names MDVP:S himmer, MDVP:Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ, Shim mer:DDA • nhr, hnr :Two measures of ratio of noise to tonal components in the voice (Actual column names NHR, HNR respectively) • rpde, d2 :Two nonlinear dynamical complexity measures • dfa - Signal fractal scaling exponent (Actual column name DFA) • spread1, spread2, ppe :Three nonlinear measures of fundamental frequency variation (Actualy column names spread1, spread2, PPE respectively) 20 • status - Health status of the subject (1): Parkinson's ; (0) :healthy

B. Methodology Used To Perform Experiment

The process of analyzing and building a model from a dataset involves several crucial steps. The first step is to load the dataset, which involves importing the data from a file or a database. Once the dataset is loaded, the next step is to perform univariate and bivariate analysis, which includes analyzing each variable separately and then analyzing the relationship between two variables. This analysis helps in determining basic statistics such as central values, spread, tails, and relationships between variables. After analyzing the data, the next step is to split the dataset into training and test sets in the ratio of 70:30 for training and testing the model, respectively. The data must be prepared for training by scaling the data and getting rid of any missing values. Once the data is prepared, standard classification algorithms such as Logistic Regression, Naive Bayes', SVM, k-NN, etc., are trained on the data. The accuracy of the trained models is tested using a meta-classifier, and the accuracy of the test data is noted. Furthermore, standard ensemble models such as Random Forest, Bagging, Boosting, etc., are trained on the data, and their accuracy is noted. Finally, all the models are compared, and the best model is selected for deployment. The model deployment process involves creating a user API, which allows users to interact with the pre-built model. The front-end web page is connected to the pre-built model, which allows users to input data and receive predictions from the model. With these steps, a complete data analysis and model building process can be carried out, resulting in a reliable and accurate model that can be deployed for various applications. The applications of the model built using the steps mentioned in the previous paragraph are vast and varied. The model can be used in several fields, including but not limited to: 1.Healthcare: The model can be used to predict the likelihood of a patient developing a certain disease based on their medical history and other relevant factors. 2. Finance: The model can be used to predict stock prices, identify fraudulent transactions, and determine creditworthiness. 3. Marketing: The model can be used to identify customer behavior patterns and create targeted marketing campaigns.

IV. IMPLEMENTATION

A. Data Analysis Part

1000	$\mathcal{R}(\theta,h_0(t_0))$	89.01(1)	RMP:/la(k)	RNP(litter(3)	4949:32tter(48x)	100.60	100.00	litter 89	RP.Sime	Sime 04		-	status	8956	95	and a	qual		m	ł.
# pectrum	115.952	197.302	74.997	1076	1,00075	11075	1 11554	1.01125	114574	0.0546	1.021	21,033	. 1	0.414703	11128	4100	1299402	2301442	121454	ł.
1 (44,80,80,20)	122 408	148.658	15.019	1 10940	1 300000	12345	11116	10094	1.00134	11940	11155	19.05	1	145258	129521	4375182	133686	2.488355	13874	ć.
2 (HH,RUSU)	116.682	131.111	11.555	10190	1,000,00	0354	10701	10103	1.0233	0.05275	12:35	21.01	1	0.425076	112528	440175	\$31973	2.342299	0.332834	ł.
3 photo80,80,4	1645	107.871	11.36	0.00997	100000	0.0952	11000	1055	11542	0.0071	1100	2164	1	14009	11525	41750	1384	2 49554	13655	ł,
4 (44,50,50)	16014	141.701	10.65	1124	0.00110	0.0065	11008	11196	13425	1141	11:10	1964	1	14736	132344	-1147107	125613	233248	14103	ł.
5 percentation	121.52	101.92	10.767	1.096	1,00000	1040	110750	1008	1,14701	1.0365	11/22	21378	1	0.415564	1.12510	42097	1299111	210756	4.36775	i
4 photology	121.267	137.244	14.525	10033	100030	1015	1822	1.0466	11935	14257	12807	24386	1	0.596340	1764112	-6434022	125762	184%	121750	ł
7 phot, 801, 502, 2	107.332	10.548	194.315	0.00290	0.000130	0.00144	10102	1 (042)	11157	1.047	0.00344	26.852	1	113745	175362	416760	1 10721	206483	1.1575	1
1 photology 1	96.730	132.048	91.754	1.000	1 300 500	0.0020	0.0032	1000	0.02055	1121	12107	2182	1	14/501	0.77367	4.007	1.327768	23251	1221571	1
9 phon, 761, 552, 4	95.054	128 102	91.228	0.00532	1 200060	11038	14032	110003	0 12830	0.94324	611022	21.842	1	1547137	17940	41107	1.325866	2.402762	12700	1
10 photo \$10,503	#30	1224	84.172	1 01575	1 202068	0.0254	0.00230	1.0753	11240	1.027	10196	21.118	1	641137	1770158	-6.246775	1.3/02	2.4720	12674	1
11 phot, 501, 502, 5	91.904	16471	86.252	1054	1,000082	0.00201	10038	1.0344	1.072	11472	121142	21.04	1	0.582390	6752526	49039	13056	2.642475	127901	č
12 photo20,000	136.535	151.004	101.278	1020	1,00020	10110	1.01153	10026	11/29	0.01968	1050	25.703	1	0.455520	15404	454714	11203	254/077	113812	i.
13 plus, R01, 504,2	105.173	175 125	78.556	0.00390	1 2001230	6.00105	1 11215	1066	1160	112164	12104	24.035	1	14016	198503	48021	125485	25/542	1 199825	i.
N photology M	1216	163.305	5.0	1029	1 300125	6.89121	1014	1004	11121	109	11988	24 322	1	1449	115427	410508	12083	2.0591	11700	ł.
15. photo301,001,0	142.157	217.458	15 159	1008	0.000138	0.0157	11000	110471	11150	0.02216	1003	25175	1	0.585524	149246	43415	421018	2 2155-46	12568	ł.
N phylicity	144,100	345,258	12.74	1054	12004	0.05211	111252	11052	112147	1.1290	12188	22.335	1	155730	154612	1400	123764	126621	121014	ł.
17 phot,521,524,3	162.778	222.101	75.60	1.00710	1,00040	0.00234	1007	1000	6.0327	0.04322	11299	21.5%	1	142/098	1.635417	-2101070	140428	3.007403	14070	ł.
10 photo \$10,500	152146	175.83	60.623	101742	1 300 55	1004	0.0432	10952	12517	(1743)	113%	17.288	1	144575	17547	-196675	13707	1109010	137145	ł.
19 photo (10), (10)	155.455	105 294	10.02	1.0710	1 000158	6.98372	10038	1.0116	0.1386	0.0054	11316	17.153	1	14854	0.636(200	455466	13075	2 85475	0.322111	ł.
2 per/00,003	153.04	16.73	65.782	1004	1 000150	1545	1040	11/28	1000	0.05200	10071	17.536	1	160-25	175457	42540	122564	2.728710	130291	ł
21 photo10,000,000	153.88	172.068	11.53	1040	0,00030	6.0222	11057	1.0886	114137	0.0002	11166	19.40	1	162817	1498961	41008	1237622	2,557536	1,255%	ł.
22 phot,821,525,5	107.538	193.221	75.00	10140	1,000138	6 86228	1804	1000	0.04251	0.0685	1128	22.468	1	11/90	0.679534	433998	120254	2 9 16 7 77	125506	ł.
23 phot,901,555,5	173.917	192.728	8.10	11147	0.000130	0.86221	111251	0.00663	11452	0.0992	12194	21.42	1	1537264	110034	434778	0.210279	2.547500	12359	ċ.
34 phot,521,531,3	162458	20.64	31.75	11174	100050	11030	11036	12114	12163	6.62214	6.01778	23.821	1	1.397937	172479	45040	1 225656	2.02/5	12:591	ł.
25 phot, 801, 516, 2	154.405	296.012	17.96	0.00833	0.000068	0.0216	11075	1154	1.0767	1657	11287	22.06	1	15274	173754	45780	0.25683	2,84593	121974	ł.
A photology at	171241	28.313	75.921	1000	1,000130	1129	0.0234	1.0750	1296	0.0266	12105	25.94	1	1419522	1.725914	41030	120270	2 585752	114/43	ł.
8 photo30,500,4	16.545	28.70	81.737	101496	1,000130	0.0250	1.025	10749	1.01919	0.0293	11028	25.77	1	0.356773	0.729852	427986	119912	2314299	11299	ł.
2 pho.30,303	195.398	227 363	8.05	1.0210	1 3 3 3 5 1	1105	10175	1.045	03/76	0.02307	13977	25370	1	147547	16723	1.0825	1,279785	224/742	11051	ł.
2 photo101,001,001	162.568	196346	17.638	1,0552	100030	1120	1 11253	0.00041	1.11791	0.02300	11171	201	1	142776	6.725797	44372	125866	1,95790	11920	i.
3 photo (10,00)	197.078	25.55	152.055	1028	1.00010	1016	1010	1048	11/290	1198	1.013	28.775	1	0.422229	174387	-734200	1.17757	17087	1,0000	ł
21 photo 501,507,3	199.228	28512	152.091	1004	0.000010	0.00134	100	0.00402	1105	0.01513	13117	3134	1	14049	17496	-140307	1173218	2102108	1980	ł.
12 photo \$10,000	198,383	2620	193.194	1.020	1,00010	8.00113	1003	10133	0.01263	0.01919	0.02116	31.7%	1	14554	1736703	-110702	1175101	1.912275	119622	ł.
33 photo,501,507,4	282.288	211.604	197.079	0.00100	1 200009	0.00000	110107	1.0270	0.02854	0.01407	0.0072	12 584		13453	17433	-186734	11784	154625	125614	ł.
34 photo 301,507,5	201104	211.525	16.10	1007	100008	0.00054	10106	1.020	10093	0.01403	1006	31147	1	13408	0.741095	-736464	110293	1423257	11463	ł.



International Journal for Research in Applied Science & Engineering Technology (IJRASET)

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Identifying the structure of the file, number of attributes, types of attributes and a general idea of likely challenges in the dataset Data analysis is a crucial part of machine learning, as it involves understanding the data that will be used to train a machine learning model. It includes a variety of techniques that help to identify patterns, trends, and relationships within the data, which can inform the selection of appropriate machine learning algorithms and preprocessing techniques.



From above we can see out of 195 patients, 48 patients (24.6 %) are healthy and 147 patients (75.4%) patients are having Parkinson's disease.

B. Model Building Part

Model building is a critical part of machine learning that involves creating an algorithm that can make predictions based on input data. The goal of model building is to develop a model that can accurately predict outcomes for new data inputs. In this essay, we will discuss the various stages involved in model building and some key considerations to keep in mind.

- 1) Data Preparation: The first step in model building is to prepare the data for use in training the model. This includes cleaning and transforming the data so that it can be used to create the model. The data should also be split into training, validation, and testing sets.
- 2) Choosing the Algorithm: The next step is to choose an algorithm that is appropriate for the data and the problem you are trying to solve. There are many algorithms available, including linear regression, logistic regression, decision trees, random forests, and neural networks. The choice of algorithm will depend on the type of data, the problem being addressed, and the desired outcome.
- 3) *Training the Model:* Once the algorithm has been selected, the next step is to train the model. This involves using the training set to adjust the parameters of the algorithm so that it can accurately predict outcomes for new data inputs. This is typically done using an iterative process, where the algorithm is adjusted based on the results of each iteration.
- 4) *Evaluating Model Performance:* Once the model has been trained, it is important to evaluate its performance. This is typically done using the validation set, which provides a measure of how well the model can generalize to new data inputs. The performance of the model is evaluated based on various metrics such as accuracy, precision, recall, and F1 score.
- 5) *Tuning the Model:* If the performance of the model is not satisfactory, it may be necessary to tune the model. This involves adjusting the parameters of the algorithm to improve its performance. This process may involve adjusting the learning rate, regularization parameters, or other hyperparameters.
- 6) *Testing the Model:* Once the model has been trained and tuned, it is important to test it on a separate testing set to evaluate its performance on new data inputs. This provides a final measure of how well the model can generalize to new data inputs.
- 7) *Deploying the Model:* Once the model has been developed and tested, it can be deployed for use in real-world applications. This may involve integrating the model into an existing software system, developing a new application that uses the model, or deploying the model as a service.





International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

From the above comparison of different algorithms (Logistic Regression, K-nearest neighbors and Support Vector Machine) we can conclude that SVM (Support Vector Machine) performed slightly better than other algorithms.



RESULT

From the Stacked meta classifier Model, we can find out the following details: Accuracy of the model: - 95% Re-call of the model: - 100% Precision of the model: - 94% F1-Score of the model: - 97% ROC-AUC: 88%

> = Navigation Home Intelligent technology you can trust 72% of internet users look for health information onlines Our web application provides you with a fast and accurate health assessment (1)Enter your readings of the test report (2)Click on "parkinsons Test Result" (3)Done! Your assessment will reveal: 1.the predicted disease status 2.description of that disease 3.suggestions This is basic Machine Learning and Deep Learning based WebApp.

V.

These Machine Learning model is trained on large dataset .



B. O/*P*



VI. CONCLUSION

For our project in this semester, we have studied and implemented basic and necessary requirements for our selected topic, Parkinson's Disease Detection System. We have studied related Research papers for Parkinson Disease Detection and designed the Flow and System Architecture for our project. We studied technologies and other necessary material for our project.

In conclusion, this study evaluated the performance of six different machine learning algorithms for a classification task. The accuracy and F1-score were used to compare the performance of these algorithms, including Logistic Regression, K-nearest neighbors, Support Vector Machine, Stacking, Random Forest, and Adaptive Boosting. Based on the results, the Stacking model outperformed the other models with a mean accuracy of 93% and a mean F1-score of 94%. It is important to note that the choice of the best model may vary depending on the problem, dataset, and evaluation metrics used. Nonetheless, this study provides valuable insights into the performance of various machine learning models and can be used as a reference for future studies in the field of classification tasks.

REFERENCES

- [1] Akkem Yaganteeswarudu, "Multi Disease Prediction Model by using Machine Learning and Flask API," Infoshare Systems, Pyramid Softsol Pvt Limited Hyderabad, India
- [2] Sharone Li," Streamlit Hands-On: From Zero to Your First Awesome Web App"
- [3] Mohamad Alissa," Parkinson's Disease Diagnosis Using Deep Learning", M.Sc. in Artificial Intelligence with Speech and Multimodal Interaction
- [4] Rudra A. Godse, , Smita S. Gunjal, Karan A. Jagtap & , Neha S. Mahamuni "Multiple-disease-prediction-using-different-machine-learning-algorithmscomparatively"
- [5] Geeta Rani," Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning"
- [6] C K Gomanthy,"The Parkinson's Disease Detection Using Machine Learning Techniques"











45.98



IMPACT FACTOR: 7.129







INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 🕓 (24*7 Support on Whatsapp)