iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

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

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

Parkinson's Disease Detection using Ensemble Learning

Sachchit Kolekar[1] , Naman Jain[2], Amit Mete[3], Prof. Nilesh Kulal[4]
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.)

©IJRASET: All Rights are Reserved

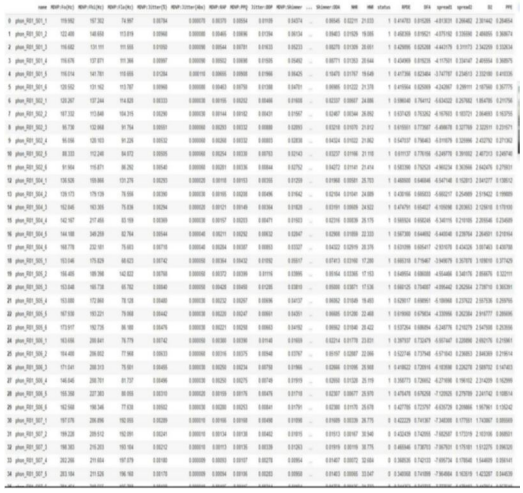
There are no missing values after Data cleaning (b.) No duplicate rows in the given dataset Attribute Information: name - ASCII subject name and recording number • mdvp_fo_hz: Average vocal fundamental frequency (Actual column name MDVP:Fo(Hz) • mdvp_fhi_hz :Maximum vocal fundamental frequency (Actual column name MDVP:F hi(Hz)) • mdvp_flo_hz :Minimum vocal fundamental frequency (Actual column name MDVP:F lo(Hz)) • mdvp_jitter_in_percent, 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 colu mn 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

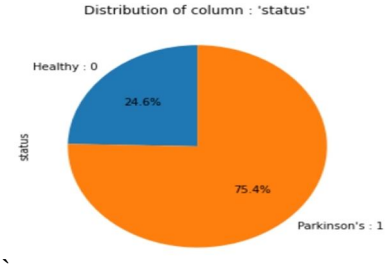
The process of analyzing and building a model from a dataset involves several crucial steps. The first step is to load the dataset, which involves importing the data from a file or a database. Once the dataset is loaded, the next step is to perform univariate and bivariate analysis, which includes analyzing 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



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.

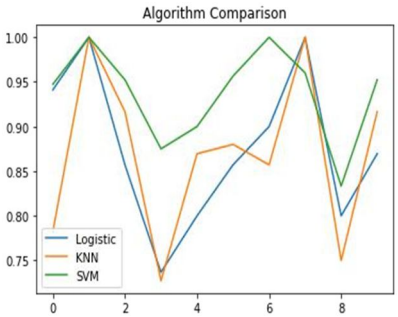


From above we can see out of 195 patients, 48 patients (24.6 %) are healthy and 147 patients (75.4%) patients are having Parkinson's disease.

B. Model Building Part

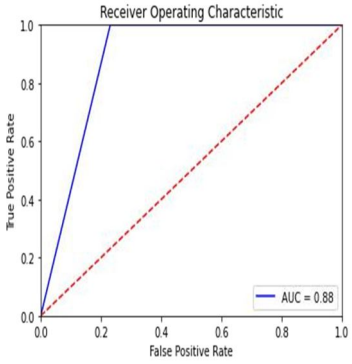
Model building is a critical part of machine learning that involves creating an algorithm that can make predictions based on input data. The goal of model building is to develop a model that can accurately predict outcomes for new data inputs. In this essay, we will discuss the various stages involved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



©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

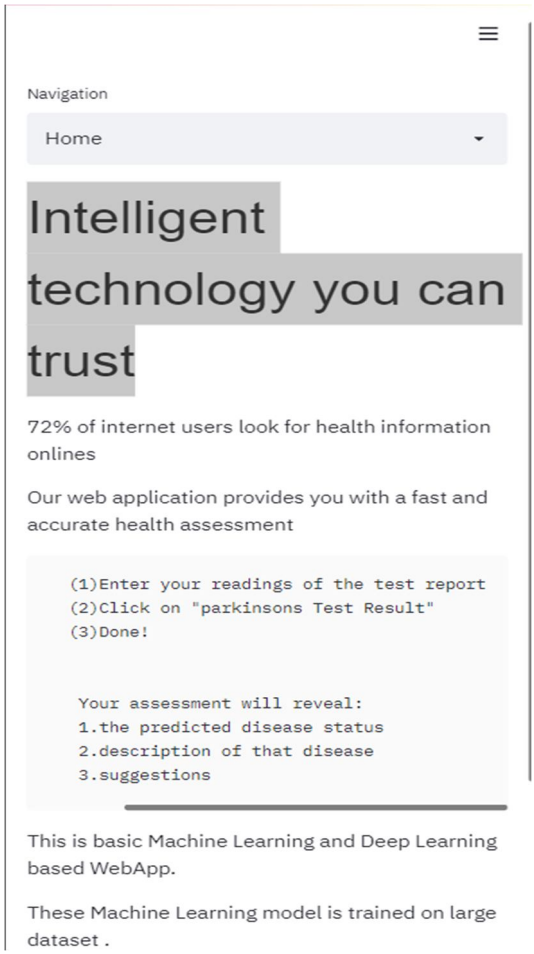
From the above comparison of different algorithms (Logistic Regression, K-nearest neighbors and Support Vector Machine) we can conclude that SVM (Support Vector Machine) performed slightly better than other algorithms.



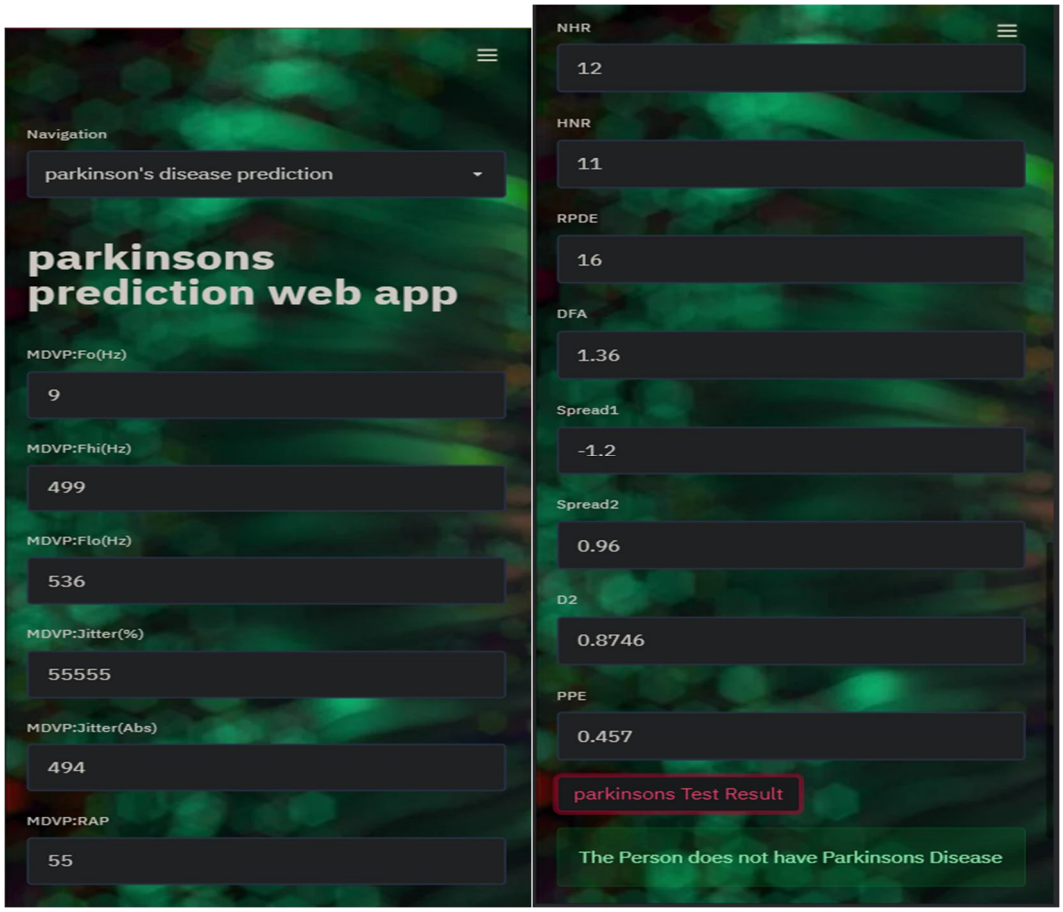
From the Stacked meta classifier Model, we can find out the following details:
Accuracy of the model: - 95%
Re-call of the model: - 100%
Precision of the model: - 94%
F1-Score of the model: - 97%
ROC-AUC: 88%

V. RESULT

A. *User Interface*



B. O/P



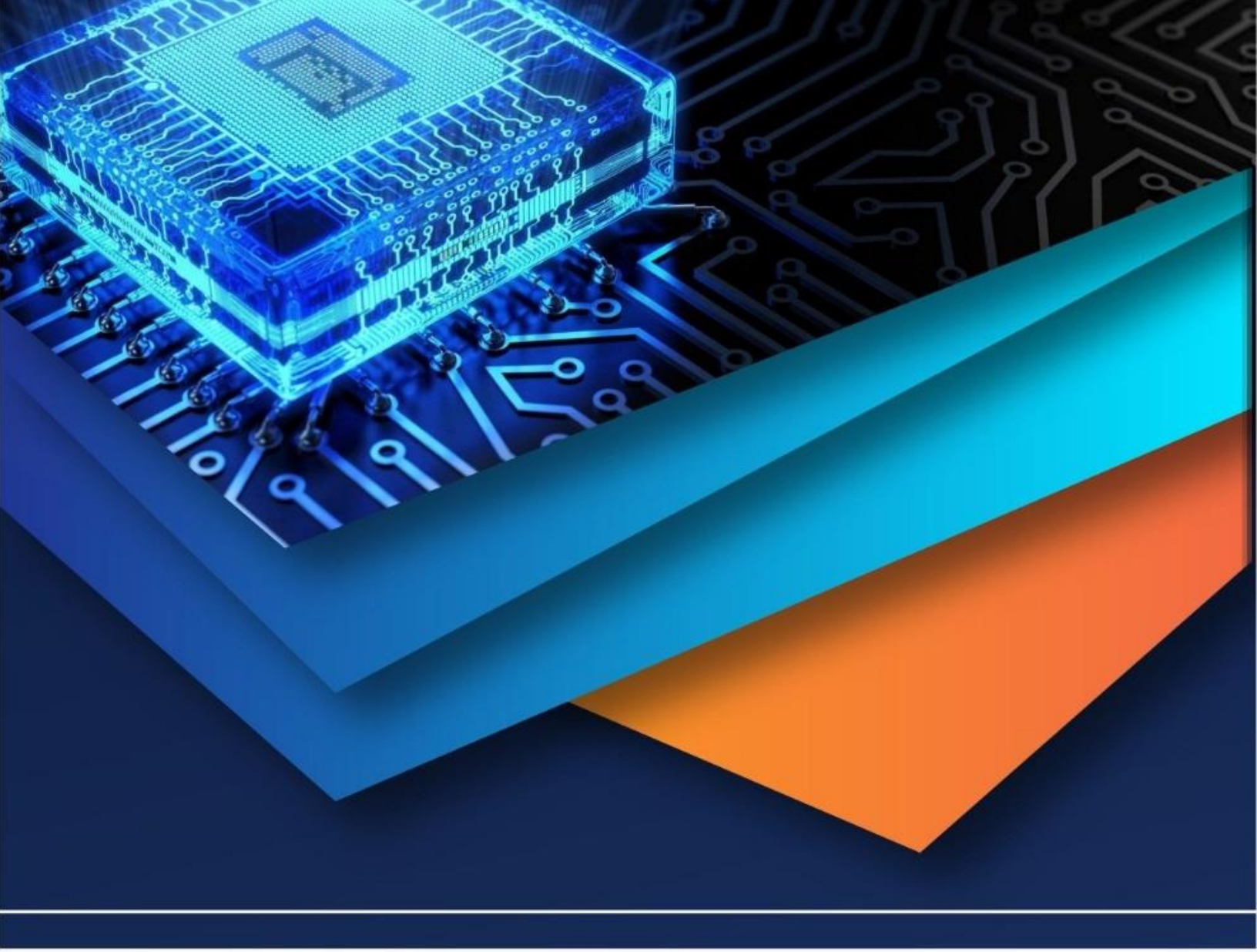
VI. CONCLUSION

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-algorithms-comparatively"
[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"



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)