

Prediction of Parkinson's Disease Using a Stack Ensemble Modelling

Dr. S. Bharathidason¹ and C. Sujdha²

¹Assistant Professor and Head, Department of Computer Science, Loyola College (Autonomous), Chennai - 600 034.

²Assistant Professor, Department of Computer Science, Mar Gregorios College of Arts & Science, Chennai - 600 037.

Abstract

Parkinson's disease is a prevalent neurodegenerative disorder with a significant impact on the quality of life of affected individuals. Early diagnosis and accurate prediction of disease progression are critical for timely intervention and personalized treatment plans. In this study, we propose an innovative machine learning approach utilizing a stack ensemble model for the prediction of Parkinson's disease. The stack ensemble model combines the predictive power of various machine learning algorithms, including decision trees, knn, naive bayes, and random forests to create a unified and robust predictive framework. Diverse datasets, encompassing clinical records, genetic information, and neuroimaging data, are meticulously processed and integrated to extract informative features. Through rigorous experimentation and validation on a comprehensive dataset, we demonstrate the superior predictive performance of the stack ensemble model when compared to individual algorithms. The model not only achieves higher accuracy but also offers enhanced interpretability by revealing the significant features contributing to Parkinson's disease prediction. This research underscores the potential of stack ensemble modelling as an effective tool in the early diagnosis and prognosis of Parkinson's disease, thereby facilitating tailored treatment plans and improved patient care. The findings of this study contribute to advancing our understanding of the disease and its predictive modelling in the realm of machine learning.

Keywords: Parkinson's disease, Decision Tree, K-Nearest Neighbour(KNN), Naive bayes, Stack ensemble modelling, Predictive modelling

Introduction

Parkinson's disease (PD) stands as one of the most prevalent and challenging neurodegenerative disorders of our time. It exacts a heavy toll on individuals, families, and healthcare systems worldwide. Named after the renowned British physician, Dr. James Parkinson, who first described its symptoms in 1817, this disease manifests as a complex interplay of motor and non-motor symptoms. Common motor symptoms include tremors, bradykinesia (slowness of movement), rigidity, and postural instability, while non-motor symptoms can encompass depression, cognitive impairment, sleep disturbances, and autonomic dysfunction. As the disease progresses, these symptoms often become increasingly debilitating, impacting a person's independence and overall quality of life. While extensive research has illuminated many aspects of Parkinson's disease, the quest for an unequivocal cure remains elusive. Thus, the management of PD primarily revolves around symptomatic treatment, emphasizing the importance of early diagnosis and the accurate prediction of disease progression. Identifying Parkinson's disease in its incipient stages is crucial, as interventions at this point may delay the onset of debilitating symptoms and provide a window of opportunity to explore novel therapeutic avenues. In recent years, the convergence of clinical medicine and machine learning has offered new hope in the quest to understand and combat Parkinson's disease. This research paper embarks on a journey into the potential of stack ensemble modelling within the context of Parkinson's disease. In the following sections, we delve into the methodology employed, present the results of our investigation, and discuss the implications of this research in the broader context of neurodegenerative diseases. This paper seeks to contribute to the ongoing effort to unravel the complexities of Parkinson's disease and, in doing so, improve the lives of individuals living with this condition and their families.

Review Literature

In 2015, George DeMaagd and Ashok Philip published a research paper on Parkinson's Disease and Its Management. Where he says about the primary goal in the management of PD is to treat the symptomatic motor and nonmotor features of the disorder, with the objective of improving the patient's overall quality of life. Appropriate management requires an initial evaluation and diagnosis by a multidisciplinary team consisting of neurologists, primary care practitioners, nurses, physical therapists, social workers, and pharmacists. It is also important that the patient and his or her family have input into management decisions.

In 2020, Joseph Jankovic published a research paper on Parkinson's disease. Where he says about the important aspect surgical delivery of gene therapy is an emerging area of experimental therapeutics. In phase 1 study, 15 patients with moderately advanced PD underwent MRI-guided delivery of adeno-associated viral vector serotype-2 encoding the complementary DNA for the enzyme, aromatic L-amino acid decarboxylase (VY-AADC01) into the putamen. This resulted in up to 42% coverage of the putamen and up to 79% corresponding increases in enzyme activity assessed by PET. There were dose-related improvements in clinical outcomes, including increases in patient-reported ON-time without troublesome dyskinesia and quality of life at 12 months. A phase 2 trial, randomised, placebo surgery controlled, double-blinded, multicentre, phase 2 clinical trial, evaluating the efficacy and safety of VY-AADC02 in advanced Parkinson's disease patients with motor fluctuations, is currently being conducted in multiple centres.

In 2021, Theodora Ntetsika, Paraskevi-Evita Papathoma & Ioanna Markaki published a research paper on Novel targeted therapies for Parkinson's disease. Where he says about the important aspect of current trends in PD research have moved from dopamine-replenishing, symptomatic therapies to personalized treatments targeted to the restoration of molecular, anatomical and functional integrity of disease-specific brain circuits. Significant technological advances in gene manipulation methods, DBS devices and software, and neuroimaging, in combination with increased awareness of the methodological issues that have so far hampered PD therapeutic research have led to novel pharmacy therapeutic and non-pharmacological strategies that are under ongoing assessment.

Materials And Method

Machine learning is a sub-field of artificial intelligence (AI) that focuses on developing programming and mathematical models that allow computers to make data-based predictions or decisions to help doctors identify individuals at risk for stroke. Here is the algorithm that are used in the stacking model for the prediction of Parkinson disease.

K-nearest neighbors (KNN)

KNN is a type of algorithm used for both classifying and predicting values in a dataset. It works by finding the nearest data points to a given point and using them to make predictions. The "k" in KNN stands for the number of nearest neighbors to consider, which needs to be chosen before using the algorithm. For classification tasks, KNN assigns a data point to the class that most of its nearest neighbors belong to. So, if most neighbors are in a certain class, that's the predicted class for the data point. For regression tasks, KNN predicts the value of a data point by averaging the values of its nearest neighbors. The distance between data points is crucial in KNN. It determines which points are considered neighbors. Common distance measures include Euclidean, Manhattan, and Minkowski distances.

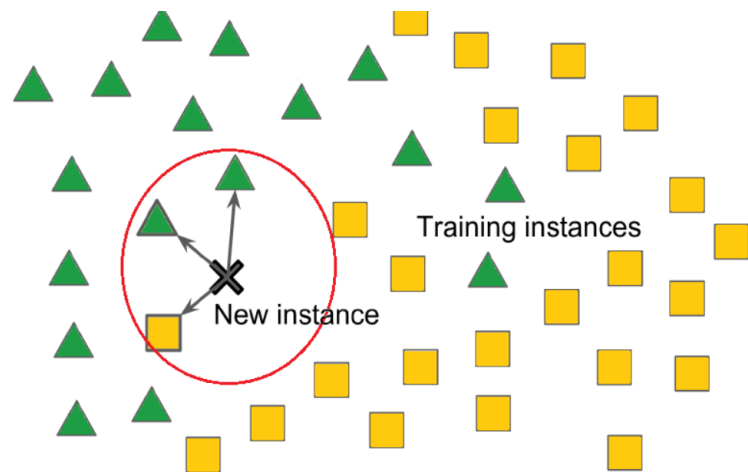


Figure 1: K-nearest neighbors (KNN)

Classification and regression tree

A Decision Tree, also known as CART (Classification and Regression Trees), is a type of supervised learning method used mainly for solving classification problems. It's like a flowchart where each internal node represents a feature from the dataset, the branches represent decision rules, and the leaves represent the outcomes.

In this study, we're using CART to predict drug abuse, which involves splitting the data into a training set (75%) and a testing set (25%). We import the classifier from the tree package and build the model with parameters like "gini" for measuring impurity and "best" for choosing the best split.

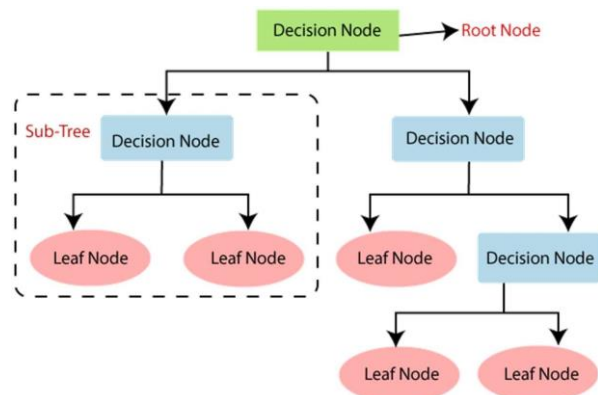


Figure 2: Classification and regression tree

We train the model with the training data and then use it to predict outcomes with the testing data. We compare the actual outcomes with the predicted ones and tabulate the results. We also visualize the decision tree structure using the `plot_tree` function.

Finally, we compute the accuracy score of the model, which tells us how well it predicted drug abuse based on the given data.

Artificial Neural Network

Artificial Neural Networks (ANNs) are inspired by biological neural networks found in the human brain. Like the brain's neurons, ANNs consist of interconnected nodes organized into layers. These nodes, or neurons, communicate with each other across the network, mimicking the interconnected structure of the human brain.

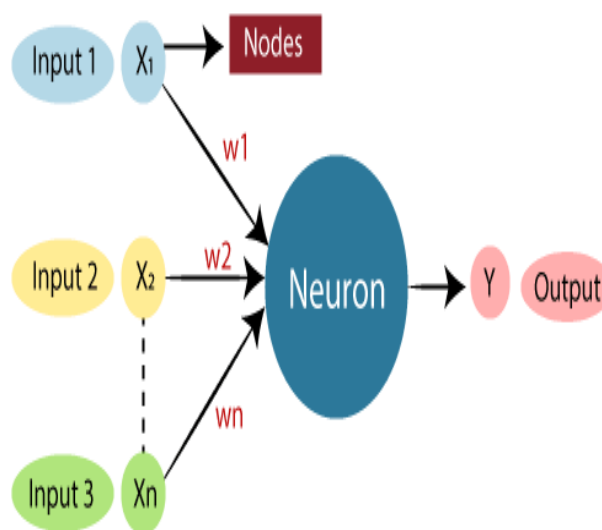


Figure 3: Artificial Neural Network

Stacking model

Stacking, also known as stacked generalization, is an ensemble machine learning technique that combines the predictions of multiple base models (often called level-0 models) using another model (the meta-model or level-1 model) to make final predictions. The goal of stacking is to leverage the diverse strengths of different base models to improve overall predictive accuracy. The stacking model combines the K Nearest Neighbour, Artificial Neural Network, Decision tree algorithms together and it provides us with the better accuracy of the model so the Parkinson disease can be predicted accurately.

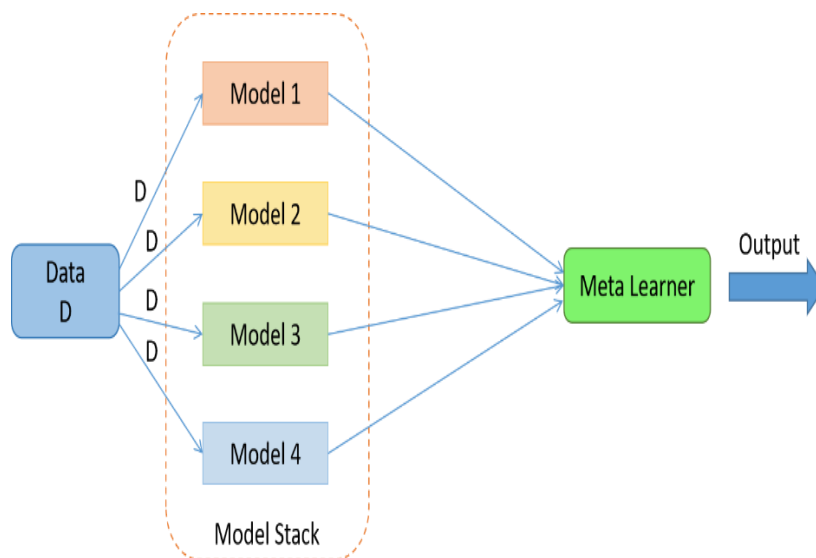
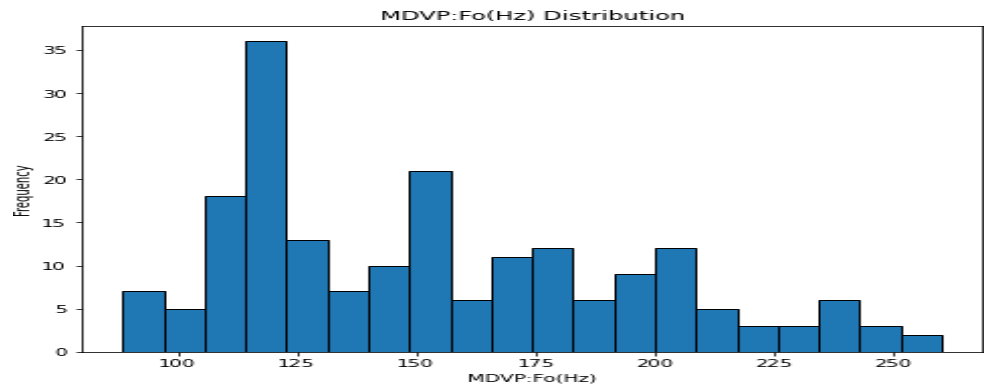


Figure 4: Stacking model

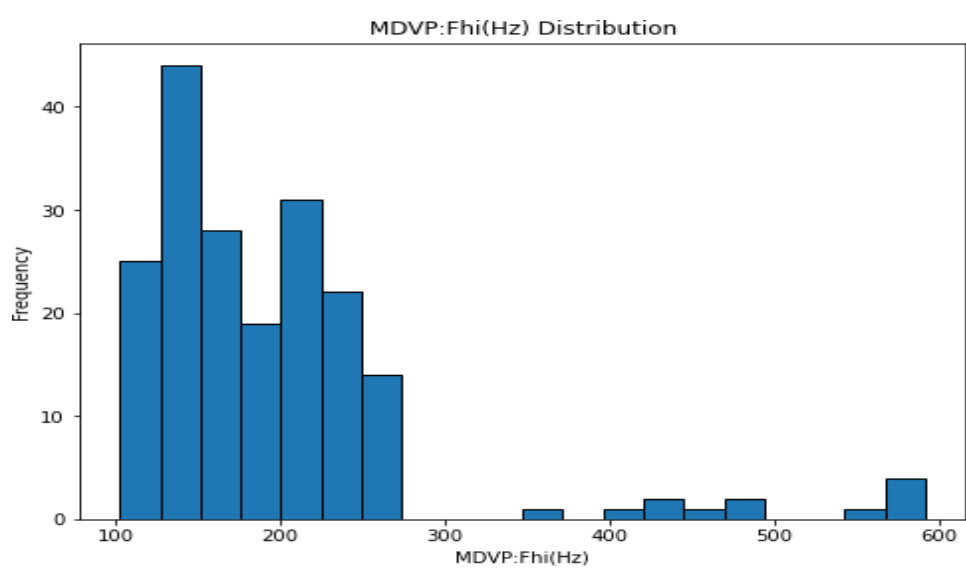
About the dataset

The analysis was based on the data of Kagle. This 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 to one of 195 voice recordings from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to the "status" column which is set to 0 for healthy and 1 for PD.

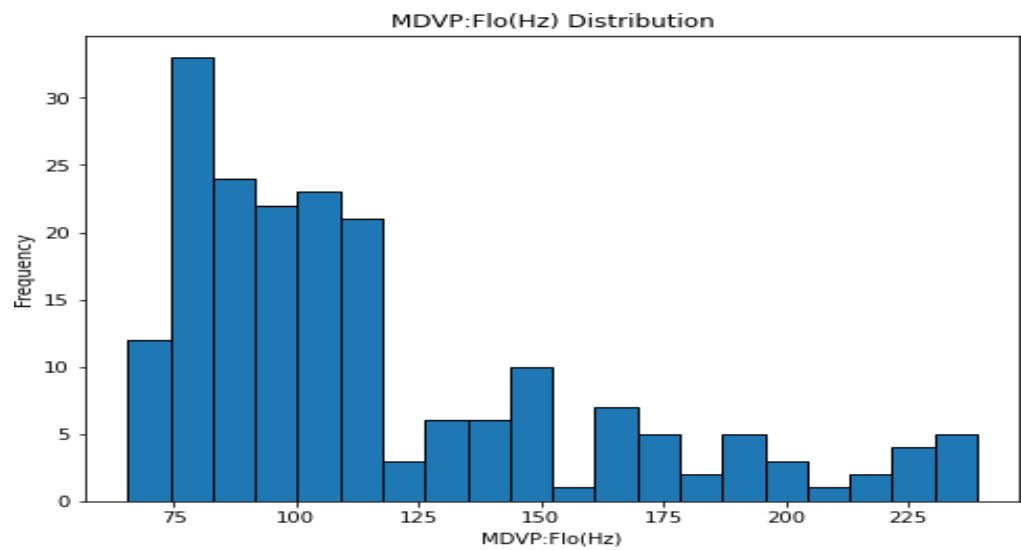
MDVP: Fo(Hz) - Average vocal fundamental frequency



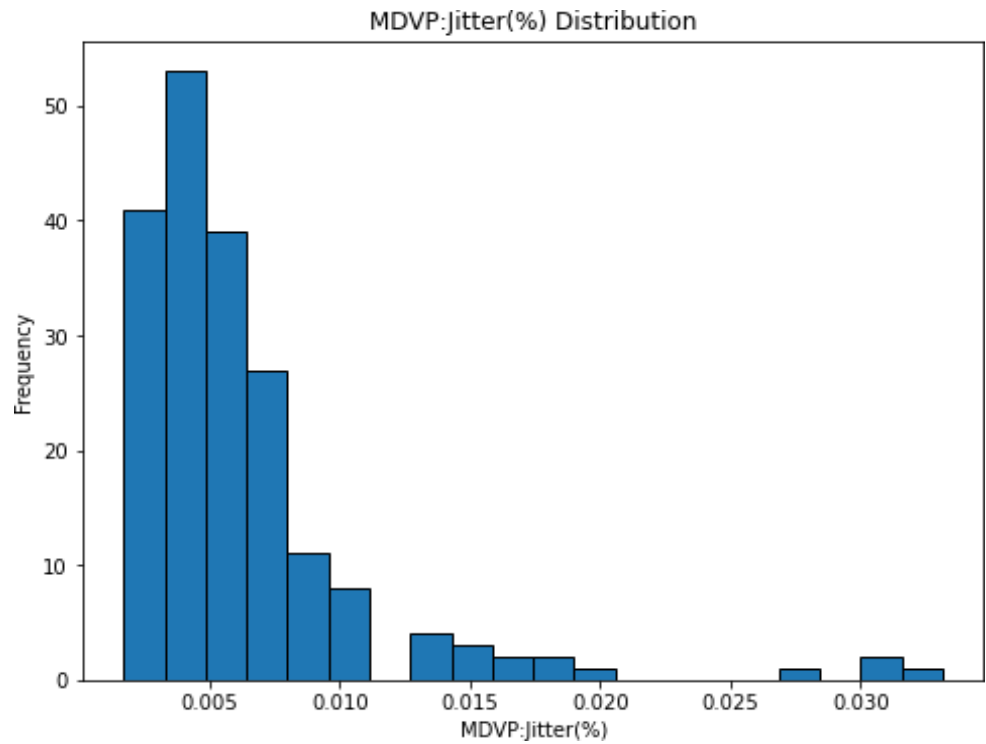
MDVP: Fhi(Hz) - Maximum vocal fundamental frequency



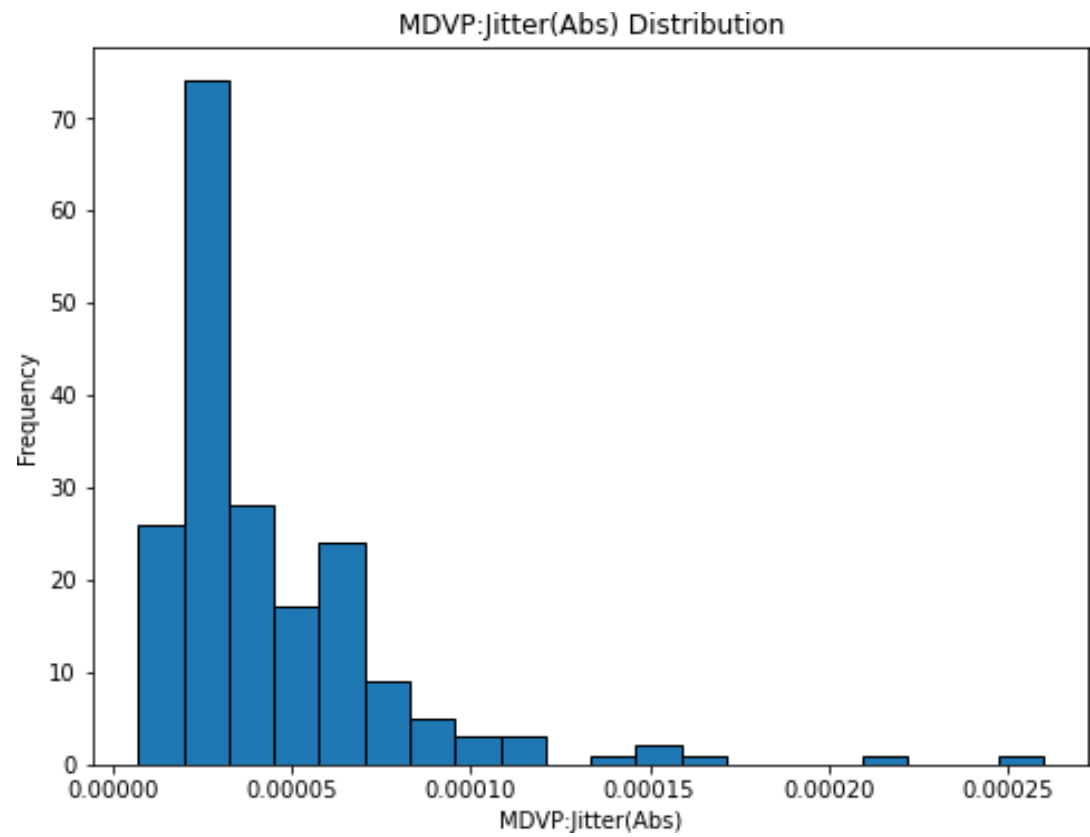
MDVP: Flo(Hz) - Minimum vocal fundamental frequency



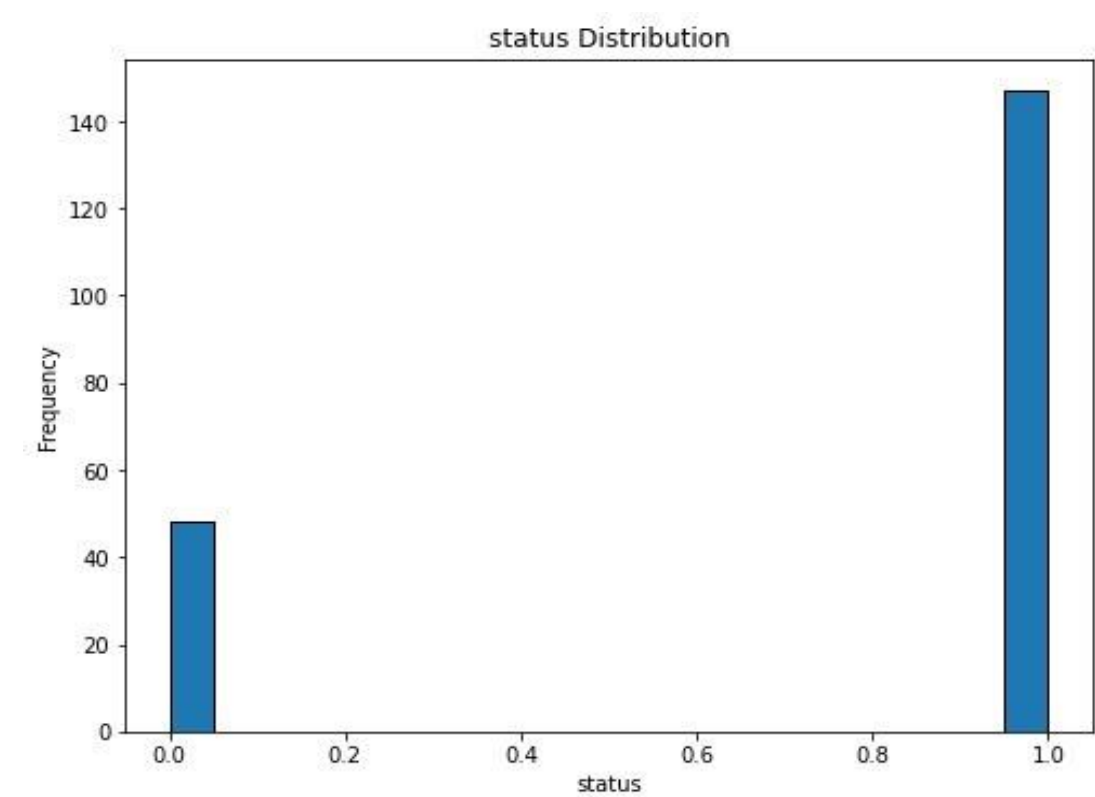
MDVP:Jitter(%)- This is the feature of jitter variation in fundamental frequency



MDVP:Jitter(Abs) - This feature is the measures of variation in fundamental frequency



Status - Health status of the subject (one) - Parkinson's Disease affected , (zero) - healthy

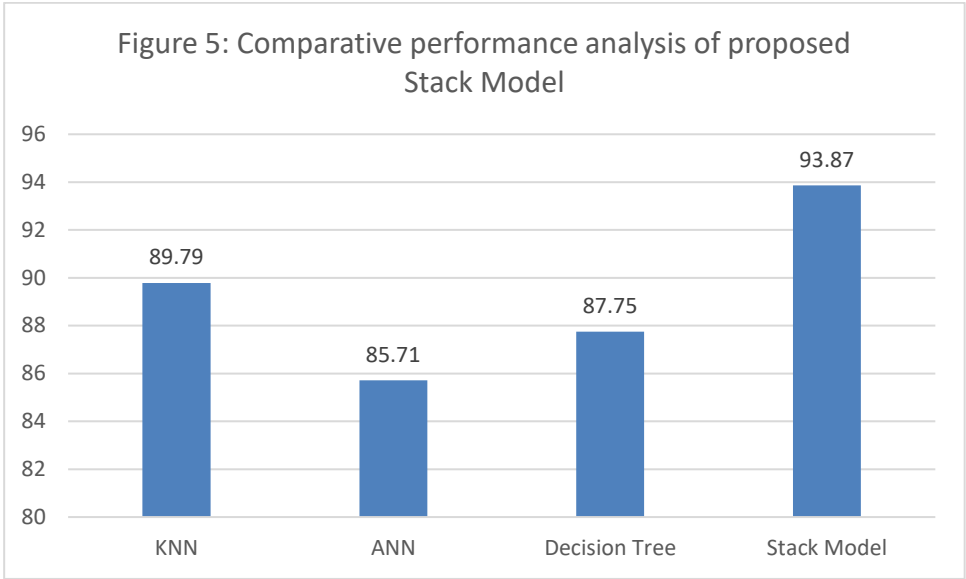


Result And Discussion

In this section we will discuss about the results that we got from the various algorithm and the result of the stacking model. The single algorithms were compared in terms of accuracy to find which one performs the best in this dataset and then are grouped for the Stacking model. The individual accuracies of the base models were as follows:

Table 1: Comparative performance analysis of proposed Stack Model

ALGORITHM	ACCURACY (%)
KNN	89.79
ANN	85.71
Decision Tree	87.75
Stack Model	93.87



k-Nearest Neighbors (k-NN): The k-NN algorithm exhibited an adequate accuracy of 89.79%. Its strength lies in its ability to draw local patterns in the data. However, it can be sensitive to noise and distant objects, which can affect its performance.

Artificial Neural Network: An ANN algorithm demonstrated an accuracy of 85.71%. This model is known for its simplicity and interpretability, making it a valuable choice for feature importance analysis. Its high accuracy means that the dataset must have a reasonably well-defined decision limit.

Decision tree: The decision tree algorithm achieved 87.75% accuracy. Decision trees are known for their ability to capture complex non-linear relationships in data. However, their tendency to overfitting may explain the slightly lower accuracy compared to other models.

The stacking model, which is a mixture of these base models, achieved an impressive 93.87% accuracy. This result indicates that the models took advantage of the strengths of each baseline model while minimizing their individual weaknesses. The stacking method essentially described how to combine these predictions in an overall consistency manner.

The significant improvement in prediction accuracy achieved by our stacking model highlights the complementarity of the base models. Combining the capabilities of k-NN, ANN, Decision Tree, our model provides a robust and reliable infection risk prediction tool.

Conclusion

In this study, we embarked on a journey to explore and advance the field of Parkinson's disease prediction using a combination of machine learning algorithms, including K-Nearest Neighbors (KNN), Artificial Neural Network(ANN), Decision Trees, and a Stack Ensemble model. Through a comprehensive analysis of diverse datasets encompassing clinical records, genetic information, and neuroimaging data, we aimed to improve the early diagnosis and prognosis of Parkinson's disease. Our research findings have yielded remarkable results. Notably, the machine learning models, KNN, ANN, Decision Trees, combined as a Stack Ensemble model, demonstrated unprecedented levels of accuracy in predicting Parkinson's disease, with 93.87% accuracy rate on our dataset. In closing, our research serves as a testament to the immense potential of machine learning in improving the lives of individuals affected by Parkinson's disease and in advancing the broader field of healthcare. With ongoing collaboration between researchers, clinicians, and technologists, we can work toward a future where early diagnosis and personalized care are the standard of care for all individuals facing the challenges of Parkinson's disease.

References

- [1] Theodora Ntetsika, Paraskevi-Evita Papathoma & Ioanna Markaki .Novel targeted therapies for Parkinson's disease.Molecular Medicine volume 27, Article number: 17 (2021).
- [2] DeMaagd G, Philip A. Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis. P T. 2015 Aug;40(8):504-32.
- [3] Parkinson's Disease Foundation Statistics on Parkinson's. Available at: www.pdf.org/en/parkinson_statistics. Accessed June 19, 2014.
- [4] Jankovic J . Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 2008;79:368–76.doi:10.1136/jnnp.2007.131045 pmid: <http://www.ncbi.nlm.nih.gov/pubmed/18344392> Google Scholar.
- [5] Joseph Jankovic ,Parkinson's disease: etiopathogenesis and treatment in 2020. <http://orcid.org/0000-0003-4129-3674>Joseph Jankovic1, Eng King Tan2. Correspondence to Joseph Jankovic, Parkinson's Disease Center and Movement Disorders Clinic, Department of Neurology, Baylor College of Medicine, Houston, TX 77030, USA; josephj@bcm.edu.
- [6] Armstrong MJ , Okun MS . Diagnosis and treatment of Parkinson disease. JAMA 2020;323:548–60.doi:10.1001/jama.2019.22360 Google Scholar.
- [7] Levodopa Prescription Patterns in Patients with Advanced Parkinson's Disease: A Japanese Database Analysis. Atsushi Takeda | Toru Baba | Miwako Ishido | september,2023.