Journal of Medical and Health Studies (JMHS)
ISSN: 2710-1452
DOI: 10.32996/jmhs
Journal Homepage: www.al-kindipublisher.com/index.php/jmhs

JMHS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Predictive Modeling and Early Detection of Parkinson's Disease Using Machine Learning

Irin Akter Liza[1]✉, Ekramul Hasan[2], Md Musa Haque[3], Shah Foysal Hossain[4], Md Al Amin[5], Shahriar Ahmed[6]
[1]College of Graduate and Professional Studies (CGPS), Trine University, Detroit, Michigan, USA
[2]College of Engineering and Technology, Westcliff University, Irvine, California, USA
[356]School of Business, International American University, Los Angeles, California, USA.
[4]School of IT, Washington University of Science and Technology, Alexandria, Virginia, USA

Corresponding Author: Irin Akter Liza, **E-mail**: iliza22@my.trine.edu

| ABSTRACT
Parkinson's disease is a progressive neurodegenerative disorder that impacts millions of citizens in the USA, mainly targeting the motor system and causing debilitating symptoms such as rigidity, tremors, and bradykinesia. The diagnosis of Parkinson's disease is presently heavily dependent on clinical evaluations and neurological examinations, targeting the detection of motor dysfunction. The principal aim of this study was to use machine learning as a means for early detection and prediction of Parkinson's disease. The dataset utilized for this study was the Parkinson's Disease Dataset, retrieved from the UCI Machine Learning Repository, which included comprehensive biomedical voice measurements from a cohort of individuals, 23 of whom are diagnosed with Parkinson's disease and 8 who are healthy controls. This dataset included a set of features extracted from voice recordings. These include parameters like fundamental frequency (pitch), amplitude variation, jitter, shimmer, and several phonation-related measures known to reflect early vocal impairments associated with the disease in question. This research project deployed three credible and proven algorithms: logistic regression, random forest, and the Support Vector Machines. Besides, this study employed a combination of metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, which are more holistic toward the model performance. According to the metric performance results of the three models, several key insights were drawn into the early detection of Parkinson's Disease. Particularly, it was clear that the Random Forest model had superior accuracy and was the most reliable in classifying positive cases and healthy patients, which could be that this model turned out to be most reliable in an early detection setting. In that respect, predictive modeling in Parkinson's disease is a capability frontier that has the potential to make a significant difference in clinical decision-making. Supported by Machine Learning algorithms, clinicians can trace the minute patterns in patient data, which are not easily visible through any other diagnostic means. In such cases, early detection of Parkinson's Disorder with vocal biomarkers or motor assessment may afford healthcare professionals opportunities for early interventions that might retard the disease process.

| KEYWORDS

Parkinson's Disease; Early detection; Predictive Modelling; Machine Learning; Random Forest; Logistic Regression

| ARTICLE INFORMATION

ACCEPTED: 10 October 2024 **PUBLISHED:** 12 November 2024 **DOI:** 10.32996/jmhs.2024.5.4.12

Introduction

Background and Motivation

According to Govindu & Palwe (2023), Parkinson's disease (PD) is a progressive neurodegenerative disorder that impacts millions of citizens in the USA, mainly targeting the motor system and causing debilitating symptoms such as rigidity, tremors, and

bradykinesia (slowness of movement). The condition is chronic and currently considered incurable; thus, it greatly impacts the quality of life for those affected over time. The rationale for early detection of Parkinson's disease is that it offers a window of opportunities for effective management through interventions that can delay the progression of the disease and improve outcomes, enhancing quality of life. Nasiruddin et al. (2024), argue that unfortunately, early diagnosis of Parkinson's remains very difficult. This comes mainly from subtle and variable early symptoms that are normally mistaken for normal signs of aging or symptoms of some neurological conditions. Indeed, by the time clinical diagnosis of Parkinson's disease can be made through evident motor impairments, significant neuronal loss has already occurred, particularly in the dopaminergic neurons of the substantia nigra. This delay in diagnosis underlines the urgent need for more sensitive and specific diagnostic techniques that disclose Parkinson's disease at the earliest stage before overt symptoms appear.

Junaid et al. (2023), contended that the diagnosis of Parkinson's disease is presently heavily dependent on clinical evaluations and neurological examinations, targeting the detection of motor dysfunction. Traditional diagnostic approaches to the disease have several serious drawbacks: there are no definitive biomarkers for Parkinson's; this hinders the process of diagnosing the neurodegenerative condition at early stages. Besides, there is considerable variability in the expression of the disease from one individual to another, which may be another factor leading to misdiagnosis or delayed treatment. Al Amin et al. (2024), posited that with the progressive course of the disease, there is an increasing interest in predictive models that use the medical history of patients to identify reproducible patterns that are associated with a pre-motor diagnosis of Parkinson's disease. It is here that ML comes into play: a strong ally, offering new visions for early detection by using advanced data analysis techniques.

Bhowmik et al. (2024), asserted that machine learning is a subfield of artificial intelligence that provides a structured approach to constructing algorithms capable of identifying complex patterns in big datasets. The Parkinson's models can also be trained, in the context of this disease, on a wide range of data sources-from voice recordings, handwriting samples, and brain imaging to all genetic data, for early detection. Bortty et al. (2024), articulated that an important motivation for the application of machine learning techniques within healthcare, especially for neurodegenerative disorders, has been their capabilities concerning high-dimensional data and finding hidden relationships that are not so easy to find via more traditional statistical methods. This will hopefully provide increased diagnostic accuracy and may also help further our knowledge of disease progression, potentiating the discovery of new biomarkers.

Objectives

The principal aim of this study is to use machine learning as a means for early detection and prediction of Parkinson's disease. In this research project, we build predictive models that identify subtle changes in the physiological and neurological profiles of patients with the possibility of developing into an affliction of Parkinson's disease, before the appearance of typical symptoms that may be identified with classic diagnostic methods. This research work, therefore, investigates a variety of features and biomarkers that could become early signs of disease, namely changes in voice, changes in motor function, and even molecular biomarkers. We will discuss how best to optimize different machine learning algorithms, such as the techniques of supervised learning-based decision trees, support vector machines, and deep neural networks, which are capable of providing better diagnostic accuracy and reliability.

Literature Review

Overview of Parkinson's Disease

As per Fadavi & Fadavi (2024), Parkinson's Disease is the second most common neurodegenerative disorder in the world, after Alzheimer's disease, and mainly impacts the motor system. It was first described by James Parkinson in 1817, and since then much has been learned about its clinical and pathological features. Parkinson's disease is a neurodegenerative disorder that affects the dopaminergic neurons in the substantia nigra, a part of the brain with large roles in motor control. Lin et al. (2022), upholds that the degeneration causes the significant loss of dopamine, an important neurotransmitter involved in the execution of smooth and controlled muscle movements. Characteristic symptoms include tremors, bradykinesia or slowness of movement, muscle rigidity, and postural instability. Other nonmotor symptoms that may appear in the course of this disease include cognitive impairment, mood disorders, sleep disturbances, and autonomic dysfunction. The fact that many of these nonmotor symptoms can antedate the motor signs of Parkinson's disease by several years suggests a long preclinical phase for the illness.

Dutta et al. (2024), asserted that the pathophysiology of Parkinson's disease is nary and complex, due to an interaction between genetic and environmental causes. Although the cause is still unknown, oxidative stress, mitochondrial dysfunction, and accumulation of misfolded alpha-synuclein proteins are also believed to play a major role in neuronal deaths. This is not just motoric neurodegeneration but also includes other neural systems, implying a wide range of symptoms (Raval et al.,2020). The clinical course of the disease is generally progressive, with the advancement of motor symptoms leading to severe disability later in the disease course. Early diagnosis is important because neuronal loss is irreversible, initiating therapeutic intervention early in the disease course can slow the disease course and improve outcomes.

Senturk (2020), indicated that early diagnosis is particularly important for PD because it may provide the potential to institute neuroprotective treatments that might delay the onset of severe motor symptoms. Thereby, symptomatic pharmacological and non-pharmacological therapies, which include replacement therapy with dopamine, deep brain stimulation, and physical exercises, reduce symptoms and improve quality of life. These treatments are effective, but most efficacious when given early in the course of the disease, and there is, therefore, an increasing need for robust diagnostic tools to identify PD at its earliest stage. Despite this need, early diagnosis remains difficult due to the subtlety and variability of initial symptoms, which can usually be readily explained by normal aging or other medical conditions.

Current Diagnostic Techniques

According to Zhang (2022), for a long period, the diagnosis is traditionally based on clinical examination, with an emphasis on motor symptoms such as tremors, bradykinesia, and rigidity. Neurologists diagnose, among others, according to the United Kingdom Parkinson's Disease Society Brain Bank criteria, but these are largely based on the observation of motor symptoms and the response to dopaminergic therapy. Clinical diagnoses are subjective, however, based on the interpretations of the clinician and thus often variable. Furthermore, by the time motor symptoms become apparent, there has already been a significant loss of dopaminergic neurons, indicating patients are diagnosed relatively late in the course of the disease.

Lipnick (2021), posits that imaging techniques such as MRI, PET, and SPECT are used in addition to clinical evaluations in assessing brain structural and functional changes. These neuroimaging techniques may provide important information on the degree of neurodegeneration but are expensive, not widely available, and may not detect early-stage disease. Moreover, imaging findings are not absolute for Parkinson's disease since they can present themselves in other neurological disorders. There is also a rising interest in the application of biomarkers for early detection through the analysis of cerebrospinal fluid and genetic testing. None of these approaches is, however, considered yet part of routine clinical practice because of problems with sensitivity, specificity, and cost.

The newest medical research has identified early biomarkers constellation of features that precede motor symptoms of Parkinson's disease. For instance, several studies have explored olfactory dysfunction, REM sleep behaviour disorder, and constipation as early forerunners to the development of PD. Other researchers are focused on alpha-synuclein aggregates, supposedly having a central role in the pathology of the disease and therefore can also serve as a biomarker (Yuan et al. 2021). Despite all these developments, no single test or biomarker would diagnose the presence of PD at an early stage. This need has, consequently, been the interest in developing this capability for early diagnosis, especially in the development of innovative approaches using machine learning models that can help in the improvement of early detection and prediction of these conditions.

The Role of Machine Learning in Medical Diagnosis

Wang et al. (2020), asserted that the advent of machine learning in healthcare has transformed the space of medical diagnostics and allowed the study of complex data sets for patterns that could not be easily seen by human clinicians. Machine learning is part of artificial intelligence, which involves the development of algorithms that become capable of learning from and making predictions or decisions from data. These models, when applied to disease detection, for instance, may consider several sources of data input: clinical records, imaging data, genetic information, and even patient-reported outcomes. Some advantages accruing from the use of ML in medical diagnosis revolve around handling large volumes of data in minimal time as well as offering improved diagnostic accuracy and personalized treatment based on a patient's unique profile.

During the past decade, different machine learning methodologies have been adopted in the diagnosis of neurodegenerative diseases, including Alzheimer's and Parkinson's disease. This model may be of great value during early diagnostic scenarios due to its capability of assessing very minute clinical changes across patients' data before any symptom development. Several machine learning algorithms have been applied to voice recordings, gait patterns, handwriting samples, and neuroimaging data to conduct early detection of PD. For instance, the early identified features of Parkinson's disease include reduced vocal volume and speech rate variation. ML models run on these speech features also tend to classify with high accuracy between Parkinson's and healthy subjects (Yuan et al. 2021). Handwriting analysis through ML techniques can also recognize micrographia, small cramped handwriting being one of the most well-acknowledged early symptoms of Parkinson's disease. Previous studies have revealed the potential of ML models for improving the early diagnosis of PD. For example, studies have identified that support vector machines, random forests, and artificial neural networks classify patients with Parkinson's disease from healthy individuals with a high degree of accuracy. Most of the models are trained on the combination of clinical features, imaging data, and other biomarkers to improve their predictive power.

Lin et al. (2022) report that more than 90% of classification accuracy in the diagnosis of Parkinson's disease was obtained using MRI by an SVM-based ML model. Their study used a dataset of voice recordings and built an ML algorithm to predict

symptom severity in Parkinson's with high accuracy. Despite such developments, the application of ML for diagnosing PD remains in its infancy, as several challenges stand in the way. In the first place, the main issue is heterogeneity: symptoms of this disease can be variable, and the speed of development may be different for different patients. That is why it's hard to develop only one model that would fit all. Moreover, Junaid et al. (2021), argued that most ML models need huge and well-annotated datasets to show high performance; this factor might be very hard to get in a clinical setting. Also, there are concerns about the interpretability of the ML model: many of the algorithms that produce the most accurate results, such as deep learning models, act like "black boxes" and cannot give clear explanations for the predictions they make. This lack of transparency acts as a barrier to their translation into clinical applications, as interpretability forms the bedrock on which confidence from practitioners is obtained.

Method

Data Collection and Preprocessing

The dataset utilized for this study was the Parkinson's Disease Dataset, retrieved from the UCI Machine Learning Repository, which included comprehensive biomedical voice measurements from a cohort of 31 individuals, 23 of whom are diagnosed with Parkinson's disease and 8 who are healthy controls. This dataset included a set of features extracted from voice recordings. These include parameters like fundamental frequency (pitch), amplitude variation, jitter, shimmer, and several phonation-related measures known to reflect early vocal impairments associated with the disease in question (Pro-AI-Robikul, 2024). The target attribute is the status variable, which originally included a binary value of 1 for the presence of Parkinson's disease and 0 for its absence. Besides the fact that the main data set constitutes the ground, supplementary data considered included patient clinical records, demographic information, and medical history, which could improve model accuracy where applicable. This component further included the age profiles of the patients, the onset of the symptoms, and even the status of their medication, hence shedding a better light on disease progression and further improving the robustness of the machine learning applied in early detection efforts.

Data Preprocessing

Data preprocessing is an essential step in ensuring that the dataset used is clean, reliable, and ready for machine learning analysis. First of all, data cleaning included an overview of the detected missing values, duplicates, and inconsistencies in the dataset. There was no need for imputation since the Parkinson's Disease Dataset had no missing values. Any missing values in some other datasets could have been imputed using a mean, median, or mode imputation technique or anything as complex as KNN imputation. Secondly, other than checking the dataset for gaps, we also looked at the outliers that might affect the skewed performance of the model. Outliers in the data were identified by statistical methods, which include IQR and Z-score analyses. If there were any outliers, they were treated by either capping transformation or removal processes to retain the integrity of the dataset (Pro-AI-Robikul, 2024). This protocol would also involve encoding categorical variables into numerical variables using techniques such as one-hot encoding or label encoding, although the current dataset is essentially composed of numerical features.

Thirdly, the next preprocessing step involved feature scaling due to the very different ranges among the values of the attributes of the dataset. To scale the features to identical range values, techniques such as Min-Max normalization or Standardization, also known as Z-score normalization, can be used; this is necessary in those algorithms sensitive to the magnitude of features, for example, SVM and K-NN. Normalization made the features such as pitch measurements and amplitude variations come out on a similar scale; in other words, no single feature because it has a high range dominated the model's result (Pro-AI-Robikul, 2024). Besides, correlation matrices were drawn to check on multicollinearity in the dataset, hence making sure highly correlated features were combined or removed to avoid redundancy and give a better performance towards the model. Finally, the data was then divided into training and testing sets, which are usually 80% to learn the model and the rest for testing, to evaluate the performance of the model on unseen data and test the robustness and generalizability of the predictive model developed to detect Parkinson's disease early.

Feature Engineering and Selection

Essentially, feature engineering and selection are crucial steps in the development of efficient machine learning models. This is especially true in medical diagnostics, where the presence of relevant biomarkers increases the predictive accuracy manifold. In the case of Parkinson's disease detection, there has been a focus on feature transformation and selection that captures the subtleties in vocal impairments forming an early manifestation of the disease itself. The original biomedical speech characteristics were Fo, Shimmer, Jitter, and noise-to-harmonic ratio. These feature extraction algorithms, PCA and LDA, can be conducted to reduce the dimensions in the dataset but preserve the maximum variance that is relevant for classification. This will help in the summarization of information from various correlated features into fewer uncorrelated components. Model interpretability will increase and the risk of overfitting is reduced.

Moreover, domain-specific knowledge plays a fundamental role in feature engineering. Composite features, such as the ratio of harmonic components, known as HNR, and aggregated measures of jitter and shimmer, were devised to inherently improve the model with the capability of capturing subtle voice irregularities. Advanced RFE and Regularization-L1/ Lasso technique systematically remove the less impactful features. It allows ranking features according to their predictive power and is useful when there are large counts of variables. Other statistical methods of evaluating the relevance of each feature regarding the target variable included mutual information gain and Chi-square tests. In this regard, the aim is to choose a subset of features that maximizes the model's accuracy and lowers computation complexity to enhance generalizability.

Model Selection

The choice of a proper machine learning model is very crucial for reliable predictions in early detection. Characteristics such as features being numerical and categorical in nature and possible class imbalances drove the choice of the models. In that respect, this research project deployed three credible and proven algorithms, namely, logistic regression, random forest and the Support Vector Machines. Normally, Logistics Regression is the baseline since it is so simple and interpretable. It works well for cases where the relationship between features and the target variable is approximately linear. It has serious limitations in terms of capturing the complex nonlinear patterns present within biomedical data, especially datasets in which subtle variations in voice features are indicative of the presence of Parkinson's disease.

The quest to overcome non-linearity requires the application of models such as Support Vector Machines and Random Forests. Because SVM works effectively in high-dimensional spaces, utilizing kernel functions (such as a Radial Basis Function) to manage nonlinear decision boundaries, it will be a good fit for classification in nuanced voice characteristics for healthy vs. Parkinson's patients. SVMs are fairly resistant to overfitting in general and particularly in those cases where the dataset may not be too large. On the other hand, Random Forest is an ensemble technique that builds many trees and averages their predictions to improve accuracy by reducing variance. This is useful in noisy data and provides some insight into feature importance that can be used in further refinement of feature selection.

Training and Testing Framework

Designing an efficient predictive algorithm for early Parkinson's disease detection utilizing machine learning requires a robust training and testing framework. That is, it should be able to perform well not only on the data it has been trained on but also generalize well when new and unseen data are presented. An essential first step involves dataset division into training and testing sets. Then, typically, an 80-20 or 70-30 split was used whereby 80% or 70% of the data was devoted to training, and 20% or 30% was reserved for testing so that a balance is struck between providing enough examples for the model to learn from and testing it upon a different subset objectively. While the actual fitting is done with the help of a training set, where it learns the pattern and relationship between the data values, it is the testing set that provides an unbiased estimate as to the accuracy of the model and its predictive capability.

Moreover, a simple split into data and cross-validation techniques were used to increase the robustness of the model. The most applied so far is the so-called K-fold cross-validation, where the dataset is divided into k equal-sized folds (commonly, k = 5 or k = 10). The model is trained on k-1 folds and tested on the one that remains. This process repeats k times; each fold once serves as the test set, after which the final performance metric is averaged out of the results. This technique prevented overfitting since the model is validated, not on some test set, but on multiple subsets, hence it gives a more reliable estimate of the performance of the model. Another useful advanced technique is stratified cross-validation. It is very often used when one is dealing with an imbalanced dataset, such as medical diagnostics, where the number of Parkinson's disease cases is significantly lower than healthy cases. Stratified cross-validation makes sure that the same proportional balance of positive-labeled Parkinson's cases and negative-labeled healthy cases are held in each fold as it is in the original dataset, allowing for a more balanced evaluation and reducing bias.

Hyperparameter Tuning

Hyperparameter tuning is an essential procedure to maximize the performance of machine learning models. Contrasting with model parameters, which have been learned from training data, the hyperparameters are chosen before the actual training starts and decide on the characteristics of the learning algorithm. The ultimate goal here is to find a combination of hyperparameters that maximizes performance on the validation set. The common techniques for hyperparameter tuning are grid search and random search.

Grid search is a hyperparameter-tuning method that involves the pre-specification of a set of possible values for each hyperparameter and then an exhaustive search over the space of possible combinations. Examples of hyperparameters for the Support Vector Machines model are the regularization parameter and the type of kernel, which can be linear, polynomial, or radial basis function. If we choose three possible values for 'C' (e.g., 0.1, 1, 10) and two possible kernels, grid search would evaluate all

combinations total of six different models in this case find the best one. The advantages of this method are clear: it is exhaustive, but very expensive to compute when we have many hyperparameters with a large number of possible values.

On the other hand, random search randomly samples within a certain range of hyperparameter combinations to be evaluated. In many cases, this will be far more efficient than a grid search for which only a few of the hyperparameters matter to performance. The advantage of random search over grid search is that in much less time, it can explore more of the space of the hyperparameters, which often yields better results. Besides this, more sophisticated methods like Bayesian optimization, Tree-structured Parzen Estimator (TPE), and Hyperband probabilistically model performance predictions of previous results to make the hyperparameter tuning far more efficient. In practice, both grid search and random search evaluate the performance of each hyperparameter configuration with cross-validation to ensure those are not due to chance. Then, taking as the best set of hyperparameters, the final model is trained on the whole training set and subsequently evaluated on the independent testing set to estimate its generalization performance.

Evaluation Metrics

The performance evaluation of machine learning models, especially in a healthcare setting, requires a comprehensive set of metrics to capture various aspects of predictive accuracy. It is just a single metric in this case: accuracy. It may prove misleading for real datasets where classes are always imbalanced. Therefore, this study employed a combination of metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, which are more holistic toward the model performance. Subsequently, we compared the results with both baseline models and previous studies and ensured that the proposed approach contributes to early diagnosis and better management of the disease, thus leading to better patient outcomes of Parkinson's Disorder.

Results and Discussion

Exploratory Data Analysis (EDA)

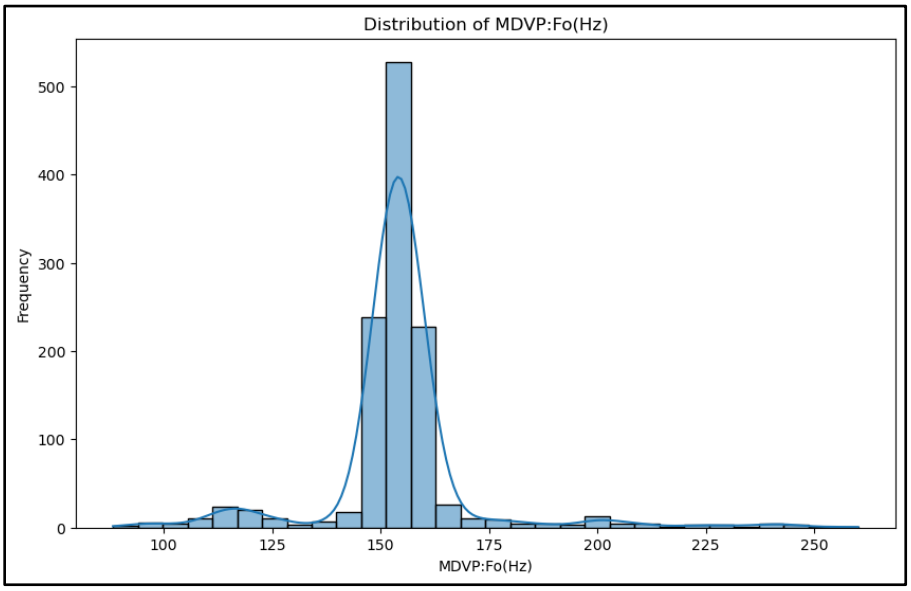

Figure 1: Portrays the Distribution of MDVP: Fo (Hz)

Hz feature, which portrays the instrumental frequency (pitch) from voice recordings in the Parkinson's disease dataset. The histogram above has a superimposed KDE curve, which reflects right-skewness for this variable. Therefore, most of the frequency values fall around 140-160 Hz, making the pitch frequency of most of the patients in this dataset almost similar, suggesting a central tendency. However, noticeable tails are extending towards the lower and higher ends, at frequencies below 100 Hz and above 200 Hz, respectively, possibly indicating the presence of outliers or patients with atypical pitch characteristics. The skewed nature of the distribution indicates that a small number of persons have much higher or lower values of fundamental frequency, which may reflect variations in the severity of impairment in Parkinson's patients. These are minor deviations that might become critical in distinguishing between persons affected with Parkinson's and healthy controls in predictive modelling.

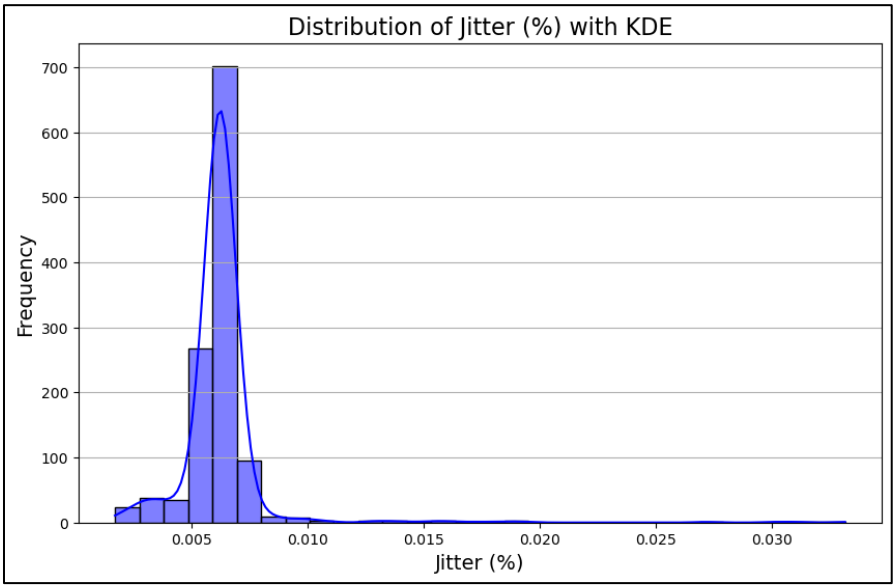

Figure 2: Displays the Distribution of Jitter (%) with KDE

The above chart distribution of Jitter (%), a measure of variation in voice recording frequency, along with its Kernel Density Estimate (KDE) curve to visualize the underlying distribution. The distribution looks highly right-skewed, and the highest density of data is between 0.004% to 0.006% Jitter. The peak frequency count is around 0.005%, meaning most of the people in this dataset have a voice frequency variation around this value, which could be a typical feature in Parkinson's patients. The long tail extending toward the right shows there are a few people with larger jitter values, possibly outliers in the dataset, and showing more marked irregularities in voices. Skewness to the right suggests that part of the patients have higher jitter values and hence greater voice instability, which could become an important discriminator for Parkinson's disease severity or even progression in predictive models. Some such outliers may cause problems during preprocessing to improve model performance.

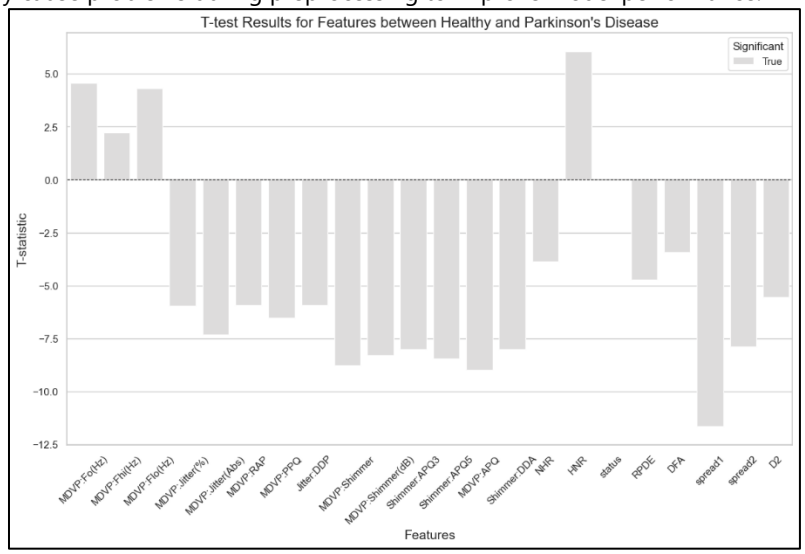

Figure 3: Exhibits the T-test Results for Features between Healthy and Parkinson's Disease

The above chart shows **T-test results** for various features that distinguish between healthy individuals and those who have Parkinson's disease. The y-axis depicts the T-statistic values, which give the extent of each feature in taking a significant difference in the two groups. Positive T-statistics are the features where healthy individuals have higher values, whereas negative T-statistics are the features where higher values apply to Parkinson's patients. Grey bars correspond to features that are statistically significant to separate these two groups. Notice that the features **MDVP: Fo(Hz)** and **MDVP: Fhi(Hz)** presented positive T-statistics, which means their values were much higher in healthy controls, while features related to voice irregularities (**Jitter**, **Shimmer**, and **NHR- Noise-to-Harmonic Ratio**) showed very negative T-statistics, thus supporting higher values in patients with Parkinson's. These results put into evidence the potential of some of the voice metrics, above all frequency and amplitude variations, to identify healthy individuals from subjects with Parkinson's disease as a fact, thus underlying their possible use in predictive modelling for early diagnosis.

Descriptive Analysis

Performance Metric	Random Forest	Support Vector Machine	Logistic Regression
Accuracy	94.87%	87.18%	89.74%
Precision [class 0]	1.00	0.67	1.00
Precision [class 1]	0.94	0.91	0.89
Recall [class 0]	0.71	0.57	0.43
Recall [class 1]	1.00	0.94	1.00
F1-Score [Class 0]	0.83	0.62	0.60
F1-Score [Class 1]	0.97	0.92	0.94

Table 1: Visualizes the Models Performance Summary

Model Performance

A. Logistic Regression

Classification Report:

	precision	recall	f1-score	support
0.0	1.00	0.43	0.60	7
1.0	0.89	1.00	0.94	32
accuracy			0.90	39
macro avg	0.94	0.71	0.77	39
weighted avg	0.91	0.90	0.88	39

Table 2: Depicts Logistic Regression Classification Report

This classification report above represents very good overall performance by a Logistic Regression model, with an accuracy of about 89.7%. In class 1.0, the model achieves perfect recall at 1.00 and high precision at 0.89, bringing about an f1-score of 0.94 with 32 samples. Class 0.0 yields perfect precision at 1.00 but lower recall at 0.43. Hence, it yields an f1-score of 0.60 with 7 samples. The macro average treats both classes equally, irrespective of sample size. Here, the metrics are balanced, with precision at 0.94, recall at 0.71, and an f1-score of 0.77. The weighted average takes into consideration the class imbalance and shows very strong overall performance with precision at 0.91, recall at 0.90, and a f1-score at 0.88 across all 39 samples.

B. Random Forest

Classification Report:

	precision	recall	f1-score	support
0.0	1.00	0.71	0.83	7
1.0	0.94	1.00	0.97	32
accuracy			0.95	39
macro avg	0.97	0.86	0.90	39
weighted avg	0.95	0.95	0.95	39

Table 3: Showcases Random Forest Classification Report

Excellent performance was also demonstrated by this Random Forest model; overall, the accuracy is 94.9%. For class 1.0, with 32 samples, this model achieves perfect recall, giving an impressive 0.97 f1-score. For class 0.0, with 7 samples, perfect precision is achieved at 1.00, with the recall improved from the logistic regression model to 0.7143, giving a strongly performing 0.83 f1-score. The macro average metrics are very high; with a precision of 0.97, recall of 0.86, and an f1-score of 0.90. The weighted average has its metrics all at 0.95 for precision, recall, and f1-score, which shows what type of balance there is in robust classification performance across the 39 samples, even with class imbalance.

C. Support Vector Machines

Classification Report:

	precision	recall	f1-score	support
0.0	0.67	0.57	0.62	7
1.0	0.91	0.94	0.92	32
accuracy			0.87	39
macro avg	0.79	0.75	0.77	39
weighted avg	0.87	0.87	0.87	39

Table 4: Displays Support Vector Machines Classification Report

The support vector machine model performs well but is relatively lower compared to the previous models, considering an overall accuracy of 87.2%. Class 1.0 has 32 samples, the model is performed very well, having high precision at 0.91 with a great recall of 0.94, thereby giving an f1-score of 0.92. Class 0.0, with 7 samples, does considerably worse, though, with a precision of 0.67 and a recall of 0.57, hence an f1-score of 0.62. Consequently, this imbalance in performance between classes is reflected by macro averages of precision, recall, and f1-score at 0.79, 0.75, and 0.77, respectively. The weighted averages are more consistent metrics, at 0.87 across precision, recall, and f1-score, although these are lower than those decided using the weighted averages for both the Random Forest and Logistic Regression models.

Prediction Insight

According to the performance results of the three models, several key insights are drawn into the early detection of PD. It is clear that the Random Forest model, with an accuracy of 94.9%, is the most reliable in classifying positive cases with an f1-score of 0.97 and healthy patients with an f1-score of 0.83, which could imply that this model may turn out to be most reliable in an early detection setting. The fact that precision is constantly high for all models, especially class 1.0, a determination based on positive predictions is highly reliable with few false positives, which is an important feature of patient screening. However, the different class recall rate, especially in class 0.0 from 43% Logistic Regression model to 71% Random Forest, indicates some early staged cases or subtle biomarkers may be missed, especially among healthy people. The biomarker performance patterns add that some biomarkers are more distinctly identifiable, as evidenced in the high class 1.0 metrics across all the models, which may be important for establishing a reliable early warning and recognition of high-risk patients. Thus, the Random Forest is particularly valuable in general for screening programs because with this algorithm, both the classes have balanced performance, and thus false positive and false negative cases are minimized.

Discussion

Clinical Implications

Predictive modelling in Parkinson's Disease is a capability frontier that has the potential to make much difference in clinical decision-making. Supported by Machine Learning algorithms, clinicians can trace the minute patterns in patient data, which are not easily visible through any other diagnostic means. In such cases, early detection of Parkinson's Disorder with vocal biomarkers or motor assessment may afford healthcare professionals opportunities for early interventions that might retard the disease process. These predictive models can provide clinically relevant decision support through various means of giving the clinicians a second opinion, either supportive or challenging, toward an accurate diagnosis and treatment for better patient outcomes. Additionally, the integration of these models into clinical workflows can be used to optimize resource utilization by prioritizing high-risk patients further for testing or specialist consultation. Since healthcare is moving toward personalized medicine, ML models can provide specific treatment recommendations based on an individual's patient profile, hence maximizing the efficacy of therapeutic strategies.

To successfully consolidate these predictive algorithms into clinical practice, a few recommendations can be made. First, health institutions must consider the adoption of an EHR system with a compatible base with Machine Learning algorithms so that predictive analytics can be dictated easily. Second, training healthcare professionals on the potentials and limitations of these models is important for their effective and responsible use. Third, collaboration between data scientists and clinicians is equally important for models that address specific clinical needs while maintaining high standards of care. Predictive models can eventually alter the management of PD by shifting treatment from a reactive to a proactive role, which can improve patient's quality of life and optimize healthcare system efficiency.

Limitations and Challenges

Despite the countless benefits, there are several limitations and challenges related to using Machine Learning algorithms in the clinical context, specifically concerning ethical and practical issues. Primary among the concerns is the ethics of sensitive patient data usage. Innumerable datasets used for training these models contain personal health information (PHI), the disclosure of which could be highly prone to breaches if not protected well. Ensuring patient privacy and data protection, according to the law-HIPAA, for example, an important thing. Guarantees on robust anonymization of data, and the use of secure data storage solutions for the protection of patients' information, should be ensured. Getting consent from the patients, especially when real-time applications are a concern, is another ethical challenge that needs to be met with appropriate transparency in communication and policy.

Another substantial issue is algorithm interpretability and transparency. Since many of those advanced ML models, including deep learning networks, are very complex and sometimes described as "black boxes"-it may be hard to understand how in the world the clinicians arrived at a particular prediction. This lack of interpretability may undermine trust in the model outputs and potentially result in resistance from healthcare professionals accustomed to evidence-based and consort consultation decision-making processes. Given this, there is also an increasing interest in interpretable ML models or the use of techniques like SHAP values that explain in as much detail as possible the contribution of each feature toward the model's prediction. Lastly, limitations of the present study have to be mentioned, such as dependence on particular datasets that may not reflect completely the diversity in a patient population. As such, vocal biomarker data can be limited to language, age, and even regional accents which will reduce the generalizability of the model.

Future Research Directions

Forging ahead, there are distinct avenues for future research that could elevate the predictive accuracy of ML models for Parkinson’s disease. First, the incorporation of more data sources into future research. Currently, present models use structured datasets such as vocal recordings, motor assessments, or demographic information. On the other hand, some unstructured data integration could be allowed to be made, such as MRI scans, genetic information, or even patient lifestyle data, including exercise and diet regimens. This would allow the training of more comprehensive models using multimodal datasets for improved diagnostic precision. The inclusion of longitudinal data on patients about the progression of the disease over time could further allow models to predict not only the onset but also the various stages of progression of Parkinson's, informing the choice of long-term treatment options.

Another promising direction forward is the inclusion of real-time patient data from wearable technology and devices of the IoT. Continuous, real-time data for gait, tremors, or speech can be obtained through wearable sensors of these features. This would enable dynamic adjustments in patient care. For example, such streams of data can feed predictive models, enabling early warnings of disease exacerbation. Therefore, this may trigger appropriate medical interventions. Real-time data can allow remote monitoring, reducing frequent hospital visits, a particular benefit to patients with impaired mobility. Nevertheless, this will involve further development of data processing and real-time analytics for correct and timely predictions.

Conclusion and Suggestions

The principal aim of this study was to use machine learning as a means for early detection and prediction of Parkinson's disease. The dataset utilized for this study was the Parkinson’s Disease Dataset, retrieved from the UCI Machine Learning Repository, which included comprehensive biomedical voice measurements from a cohort of 31 individuals, 23 of whom are diagnosed with Parkinson’s disease and 8 who are healthy controls. This dataset included a set of features extracted from voice recordings. These include parameters like fundamental frequency (pitch), amplitude variation, jitter, shimmer, and several phonation-related measures known to reflect early vocal impairments associated with the disease in question. This research project deployed three credible and proven algorithms, namely, logistic regression, random forest, and the Support Vector Machines. Besides, this study employed a combination of metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, which are more holistic toward the model performance. According to the metric performance results of the three models, several key insights were drawn into the early detection of Parkinson's Disease. Particularly, it was clear that the Random Forest model had superior accuracy and was the most reliable in classifying positive cases and healthy patients, which could be that this model turned out to be most reliable in an early detection setting. In that respect, predictive modeling in Parkinson's Disease is a capability frontier that has the potential to make much difference in clinical decision-making. Supported by Machine Learning algorithms, clinicians can trace the minute patterns in patient data, which are not easily visible through any other diagnostic means. In such cases, early detection of Parkinson's Disorder with vocal biomarkers or motor assessment may afford healthcare professionals opportunities for early interventions that might retard the disease process.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Al Amin, M., Liza, I. A., Hossain, S. Hasan, E., Haque, M. M., & Bortty, J. C. (2024). Predicting and Monitoring Anxiety and Depression: Advanced Machine Learning Techniques for Mental Health
[2] Bhowmik, P. K., Miah, M. N. 1., Uddin, M. K., Sizan, M. M. H., Pant, L., Islam, M. R., & Gurung, N. (2024). Advancing Heart Disease Prediction through Machine Learning: Techniques and Insights for Improved Cardiovascular Health. *British Journal of Nursing Studies*, 4(2), 35-50.
[3] Bortty, J. C., Bhowmik, P. K., Reza, S. A., Liza, I. A, Miah, M. N. 1., Chowdhury, M. S. R., & A1 Amin, M. (2024). Optimizing Lung Cancer Risk Prediction with Advanced Machine Learning Algorithms and Techniques. *Journal of Medical and Health Studies,* 5(4), 35-4R
[4] Dutta, S., Sikder, R., Islam, M. R., A1 Mukaddim, A., Hider, M. A., & Nasiruddin, M. (2024). Comparing the Effectiveness of Machine Learning Algorithms in Early Chronic Kidney Disease Detection. *Journal of Computer Science and Technology Studies*, 6(4), 77-91.
[5] Fadavi, N., & Fadavi, N. (2024). Early Recognition of Parkinson's Disease Through Acoustic Analysis and Machine Learning. *arXiv preprint arXiv:2407.16091*.
[6] Govindu, A., & Palwe, S. (2023). Early detection of Parkinson's disease using machine learning. *Procedia Computer Science*, *218*, 249-261.
[7] Hider, M. A., Nasiruddin, M., & Al Mukaddim, A. (2024). Early Disease Detection through Advanced Machine Learning Techniques: A Comprehensive Analysis and Implementation in Healthcare Systems. Revista de Inteligencia Artificial en Medicina, 15(1), 1010-1042.
[8] Junaid, M., Ali, S., Eid, F., El-Sappagh, S., & Abuhmed, T. (2023). Explainable machine learning models based on multimodal time-series data for the early detection of Parkinson's disease. *Computer Methods and Programs in Biomedicine*, *234*, 107495.
[9] Lin, C. H., Wang, F. C., Kuo, T. Y., Huang, P. W., Chen, S. F., & Fu, L. C. (2022). Early detection of Parkinson's disease by neural network models. *IEEE Access*, *10*, 19033-19044.
[10] Nasiruddin, Dutta, S., Sikder, R., Islam, M. R., Mukaddim, A. A., & Hider, M. A. (2024). Predicting Heart Failure Survival with Machine Learning: Assessing My Risk. *Journal of Computer Science and Technology Studies*, 6(3), 42-55.
[11] Pro-AI-Rokibul. (2024). *Predictive-Modeling-And-Early-Detection-Of-Parkinsons-Disease-Using-Machine-Learning/README.md at main · proAIrokibul/Predictive-Modeling-And-Early-Detection-Of-Parkinsons-Disease-Using-Machine-Learning*. GitHub. https://github.com/proAIrokibul/Predictive-Modeling-And-Early-Detection-Of-Parkinsons-Disease-Using-Machine-Learning/blob/main/README.md
[12] Raval, S., Balar, R., & Patel, V. (2020, June). A Comparative Study of Early Detection of Parkinson's Disease using Machine Learning Techniques. In *2020 4th International Conference on trends in Electronics and informatics (ICOEI)(48184)* (pp. 509-516). IEEE.
[13] Rahman, A., Karmakar, M., & Debnath, P. (2023). Predictive Analytics for Healthcare: Improving Patient Outcomes in the US through Machine Learning. Revista de Inteligencia Artificial en Medicina, 14(1), 595-624.
[14] Senturk, Z. K. (2020). Early diagnosis of Parkinson's disease using machine learning algorithms. *Medical hypotheses*, *138*, 109603.
[15] Wang, W., Lee, J., Harrou, F., & Sun, Y. (2020). Early detection of Parkinson's disease using deep learning and machine learning. *IEEE Access*, *8*, 147635-147646.
[16] Yuan, W., Beaulieu-Jones, B., Krolewski, R., Palmer, N., Veyrat-Follet, C., Frau, F., ... & Lipnick, S. L. (2021). Accelerating diagnosis of Parkinson's disease through risk prediction. *BMC neurology*, *21*(1), 201.
[17] Zhang, J. (2022). Mining imaging and clinical data with machine learning approaches for the diagnosis and early detection of Parkinson's disease. *npj Parkinson's Disease*, *8*(1), 1