

Exploring Machine Learning Techniques for Accurate Parkinsons Disease Detection: A Comprehensive Study

Shashank Jaiswal
Department of Computer Science
and Engineering
Lovely Professional University
Jalandhar, India
sjthief.4297@gmail.com

Mahendra Yadav
Department of Computer Science
and Engineering
Lovely Professional University
Jalandhar,
mahendra.12013205@gmail.c
om

Vivek Singh
Department of Computer Science
and Engineering
Lovely Professional University
Jalandhar, India
vivek.12010634@gmail.com

Vivek Singh
Department of Computer Science
and Engineering
Lovely Professional University
Jalandhar, India
vs292713@gmail.com

Lakshya Pratap Singh
Department of Computer Science
and Engineering
Lovely Professional University
Jalandhar, India
lakshyapratapsingh2001@gma
il.com

Shubham Singh
Department of Computer Science
and Engineering
Lovely Professional University
Jalandhar, India
shubh4579@gmail.com

Abstract—Parkinson's disease (PD) is a chronic and intricate neurodegenerative disorder that has a substantial impact on the personal and social aspects of individuals affected. PD manifests with symptoms that not only impact motor skills, but also result in speech impairments and cognitive deficiencies. Currently, the diagnosis of PD relies on the use of the Unified Parkinson's Disease Rating Scale (UPDRS), a comprehensive evaluation tool. However, the UPDRS is subjective and reliant on the expertise of the clinician. There is no specific test exclusively for PD and distinguishing PD-like symptoms from those caused by other conditions is challenging. Due to the slow progression of PD, even individuals displaying symptoms do not receive a definite diagnosis. By the time clear PD symptoms emerge, a significant amount of neural damage has already occurred. Consequently, the ability to accurately diagnose PD holds immense implications for therapeutic treatment and future advancements in therapy development. Constructing a model to identify PD is a crucial aspect of this study.

Keywords—Parkinson's disease, Machine learning, Diagnosis, Biomarkers, Feature selection, Classification algorithms.

I. INTRODUCTION

Parkinson's disease is a prevalent, chronic, progressive, and intricate neurodegenerative condition that is commonly associated with aging. Its primary symptoms include tremors at rest, muscular stiffness, slow movement, and difficulty maintaining posture [1]. This disease has a significant impact on a person's mobility, flexibility, and overall bodily function. Additionally, it can lead to secondary motor-related

complications, such as freezing and shuffling gait, which are challenging to treat using current medical therapies [2]. Therefore, early detection of Parkinson's disease is crucial, as it allows for timely medical interventions that can greatly improve the patient's condition. Unfortunately, there is no definitive test available to confirm the presence of Parkinson's disease. Diagnosing a disorder that can only be identified through specific motor and non-motor symptoms poses a challenge, and no single test can reliably diagnose Parkinson's disease. In recent years, there has been extensive research on utilizing machine learning methods to detect Parkinson's disease based on factors like voice, handwriting, and walking patterns [3]. This current study aims to assess the capabilities of machine learning in detecting Parkinson's disease by analyzing and selecting specific voice features. The voice data will be obtained from the UCI machine learning repository, and the study will also evaluate previous methods of PD detection using the UCI database and analyze their comparative performance [4].

Arguably, the most pivotal factor in utilizing machine learning to predict PD is the extensive volume of publicly accessible data acquired through the research of Parkinson's disease. This wealth of data has paved the way for extensive data mining and enhanced comprehension of the pathology associated with Parkinson's disease [5]. However, the diverse nature of the data, stemming from a range of tests and heterogeneous databases, often poses a challenge for individuals who are not well-versed in biomedical analysis of large-scale datasets. Several researchers have utilized machine learning techniques to predict Parkinson's disease, drawing on methods employed for predicting diseases affecting other organs [6]. For instance, Schapire and Singer (2000) employed a

boosting algorithm that utilized support vector machine and logistic regression to predict PD. The outcome was patient-specific probability predictions that achieved an AUC of 0.7. Similarly, Shim et al. (2009) utilized an artificial neural network to predict the occurrence of dyskinesia, an involuntary movement disorder [7]. The model exhibited a sensitivity of 0.725 and a specificity of 93%. These findings indicate that employing machine learning in these predictions has the potential to provide significant benefits to individuals with Parkinson's disease [8]. In recent years, there has been an increase in the utilization of machine learning algorithms for the detection and evaluation of Parkinson's disease (PD). Machine learning, as an analytical tool, constructs logical models based on data to make predictions. The application of machine learning algorithms in biomedical and healthcare investigations has grown due to the availability of extensive data from diverse testing systems. These algorithms have been applied to forecast the onset of PD at its initial stages, anticipate the effects of levodopa medication, and assess motor fluctuation alterations [9].

Parkinson is a gradual nervous system disability that disturbs movement. Around the globe, around 6.3 million individuals are said to have Parkinson's disease. The quantity is predicted to double and triple by 2030. Parkinson's disease exists in many symptoms such as shivering, poor balance, or noticeable slowness of motion [10]. To diagnose Parkinson's disease, there are approximately 14 varied symptom tests done by neurologists. However, the diagnosis rate is no more than 75%. This can majorly distress the chemistry of the brain within patients. A high rate of precise diagnosis can develop a long-term for the symptom patients. There are several families of methods that have been used for accurate diagnosis of Parkinson's disease. These include techniques based on voice and speech analysis, using functional magnetic resonance imaging, and other methodologies [11].

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder and is renowned for its symptoms of bradykinesia, rigidity, and resting tremor [12]. Although not fatal, it is a progressive condition that significantly impacts the quality of life for individuals affected by it. The etiology of PD remains largely unidentified, and unfortunately, a cure for the disease is nonexistent. However, there are treatment options available to manage its symptoms [13]. Consequently, exploring effective methods for early detection of PD is imperative. Early detection plays a crucial role in slowing disease progression, gauging treatment efficacy, and advancing our comprehension of its causes by studying at-risk populations. Above all, a successful detection method is vital towards achieving a cure for PD [14].

The objective of this project is to employ machine learning algorithms in the detection of PD in patients.

This endeavor encompasses the evaluation of various algorithms, feature selection techniques, and utilization of diverse datasets. The efficacy of these methods will be assessed through cross-validation, measuring accuracy, sensitivity, specificity, and the area under the ROC curve. By leveraging machine learning, we anticipate the development of a highly accurate PD detection method. Such a method can be instrumental in supporting research efforts, facilitating the discovery of a cure for PD, and enabling early detection and monitoring of the disease within a clinical setting [15].

The study aims to provide an accurate assessment of the most effective contemporary machine learning techniques for analyzing clinical data and optimizing their outcomes. A comprehensive analysis of various machine learning methods will be conducted to categorize individuals as either healthy or diseased, taking into consideration the levels of severity [16]. To determine the most precise classification, feature selection and different approaches to data transformation will be compared. This analysis will offer valuable guidance to individuals looking to utilize similar types of data. There is a considerable issue in that the data pertaining to Parkinson's Disease commonly possesses a low number of dimensions. This is since clinical data primarily consists of a limited range of value types, the majority of which are categorical and ordinal [18]. Given that machine learning techniques generally perform more effectively with continuous data of a greater number of dimensions, our objective will be to enhance the dimensionality and usefulness of the data. This can be accomplished by identifying sections of the data that can be linked to various tests and more elaborate patient records [19].

The aim of the study is to use machine learning techniques to distinguish between individuals with good health and those diagnosed with Parkinson's disease (PD), as well as identify the different stages of PD severity. It is of utmost importance to develop a reliable method to accurately assess the severity of PD, enabling appropriate treatment selection and effective drug trials for patients. Currently, there is no definitive measurement available. Although the widely used Unified Parkinson's Disease Rating Scale (UPDRS) is subjective, researchers often make individual adjustments. Quantitatively assessing the impact of the disease on patients' quality of life holds significant value.

II. LITERATURE REVIEW

A published article authored by Allison W. Willis, M.D., M.S.C.E., and Anne-Marie A. Wills, M.D., M.P.H., affiliated with "The Parkinson's Disease Foundation," underscores the pervasive impact of Parkinson's Disease (PD), which currently afflicts approximately one million Americans and an estimated four million individuals globally. PD, characterized by

its gradual onset and progression, predominantly affects individuals aged 50 and above, although instances of early-onset PD have been documented in individuals aged 20 to 40. Despite extensive research efforts, the precise etiology of PD remains elusive, and preventative measures have yet to be identified. As a chronic neurodegenerative disorder of the central nervous system, PD primarily impairs motor function. However, its effects extend beyond motor symptoms, encompassing a spectrum of non-motor manifestations, notably psychiatric and cognitive disturbances, thereby amplifying the intricacy of the disease's clinical presentation [20].

Machine learning, as expounded by Ganapathy et al. (2014), represents a captivating facet of artificial intelligence (AI) that endows computers with the capacity to learn from input data iteratively, minimizing the need for explicit programming by human operators. Leveraging mathematical models derived from training datasets, machine learning algorithms adeptly process input information to generate output, thereby facilitating predictive analytics and decision-making with remarkable efficiency. The ascendancy of machine learning has been particularly pronounced in diverse medical disciplines, spanning neurology, radiology, and genomics, with a view towards refining prognostic models, augmenting clinical decision-making, and enhancing procedural outcomes [21].

In recent years, a burgeoning body of research has harnessed machine learning techniques to discern patterns within digital datasets obtained from cohorts of PD patients, yielding notable strides in diagnostic accuracy vis-à-vis distinguishing PD patients from their healthy counterparts. Moreover, computational analyses facilitated by machine learning methodologies have proven instrumental in furnishing differential diagnoses for parkinsonian syndromes and evaluating the efficacy of therapeutic interventions [22]. Evidently, the burgeoning integration of machine learning augurs well for conferring substantive clinical utility within the realm of Parkinson's disease. As this interdisciplinary frontier advances, machine learning algorithms are poised to refine their predictive prowess, engendering more nuanced and tailored therapeutic modalities tailored to the distinctive needs of individual PD patients. Harnessing the vast troves of big data and sophisticated computational methodologies empowers researchers and clinicians alike to glean invaluable insights into the multifaceted pathophysiological underpinnings of PD, thereby facilitating early detection, precise diagnostic delineation, and the formulation of targeted therapeutic regimens [23].

Moreover, the assimilation of machine learning paradigms into the clinical management milieu of PD holds transformative potential. Machine learning algorithms, adept at parsing copious volumes of heterogeneous data from disparate sources, excel in uncovering latent patterns and trends that might elude human cognition. This enables clinicians to craft

bespoke treatment strategies that are attuned to the idiosyncratic needs and prognostic nuances of individual patients. By harnessing the prowess of machine learning, healthcare practitioners stand poised to optimize therapeutic interventions, ameliorate patient outcomes, and engender a tangible enhancement in the overall quality of life for PD patients [24].

In summation, the convergence of machine learning and Parkinson's disease heralds a new epoch of innovation across the realms of research, diagnosis, and therapeutics. By harnessing the transformative potential of artificial intelligence, investigators and clinicians alike are endowed with a potent toolkit to unravel the enigmas of this debilitating neurological malady [25]. With continued strides in technological sophistication and data analytics, the horizon of PD management is suffused with promise, auguring a paradigm shift in our understanding and management of this affliction. Machine learning, as a vanguard of innovation, holds sway over the trajectory of Parkinson's disease research and clinical practice, offering the prospect of improved prognostication, personalized therapeutic interventions, and enhanced outcomes for patients worldwide [26].

The tools and techniques of machine learning ranging from principal component analysis to neural networks have also played a part in making this data manageable, and have had an impact on the studies of PD. More recent machine learning techniques, such as support vector machines, have proved to be the most successful and promise viable predictive models gets tighter problem as those support vector machines can produce an accurate diagnosis in which is a category problem in it which attempt to distinguish between set of things [16]. These categorization support vector machines are proved to be useful when attempting to classify different stages of PD severity and mixed groups of control and those with various other types of parkinsonism. Having identified the effectiveness of support vector machines and the necessity of a diagnosis from its mixed definite and possible PD groups, there is an opportunity for prediction and diagnosis. This was proved by a recent study that built a model to make a predictive diagnosis of probable PD and assessing its accuracy compared to a confirmatory diagnostic using the unified Parkinson's disease rating scale [27].

To the best of our knowledge, there have been four primary studies conducted to systematically analyze the performance of motor disability patterns in predicting future Parkinson's disease (PD) in untreated individuals with early parkinsonian signs. The longest study, which lasted for 8 years, involved 184 patients from central Scotland who were referred by primary care physicians to a single neurologist (D.G.) for parkinsonism assessment. This study included both new cases and previously diagnosed patients. However, it should be noted that this study did not encompass a wide range of patients as only 53 of them underwent a single assessment [28].

The second study was carried out in France, the third in Norway, and the fourth in the United Kingdom. Although the specific criteria for parkinsonism varied slightly across these studies, they generally required patients to exhibit at least two cardinal signs of PD while excluding secondary causes of parkinsonism. Three of these studies had a relatively small number of patients and lasted between one to three years [29]. It is important to acknowledge that these studies have certain limitations when it comes to developing predictive models using motor disability patterns. The duration of follow-up in these studies, which determines the occurrence of PD or other parkinsonian syndromes, has been relatively short. Additionally, many patients did not receive a definitive diagnostic outcome, which further hinders the ability to differentiate short-term predictions [30].

III. EXISTING RESEARCH GAP

Kelly et al. (2015) proposed a Parkinson detection system using acoustic analysis that is processed and analyzed using data mining techniques. Rajpara et al. (2013) proposed Parkinson's Disease Diagnostic Assistant for android mobile devices. This application acts as an assistant to differentiate and classify walking patterns between normal human beings and Parkinson's disease (PD) patients using artificial neural network. A wireless shoe-based insole is used as an input device to record and measure the pressure of the foot for differentiating the patterns [31].

The classification schema is replicating an analysis conducted by a group of European researchers with our database. The specific goal is to classify PD patients and elderly healthy controls, while avoiding misclassification between the PD patients and neurologically impaired elderly. The findings suggest that the ability to distinguish between PD patients and elderly healthy controls is very robust, though statistically significant feature distinctions between certain impaired PD patients and their neurologically impaired classmates can be difficult to ascertain. There are still a lot of gaps in the existing Parkinson's disease detection research. Most of the previous studies only focus on one type of algorithm. In Rishikesh et al. (2013), the main focus of the research is to compare the classification results in a newly postprocessed gait database of Parkinson's disease (PD) patients against the results of similar classification analyses in the original database [31].

IV. DATA SET

The classification schema is replicating an analysis conducted by a group of European researchers with our database. The specific goal is to classify PD patients and elderly healthy controls, while avoiding misclassification between the PD patients and neurologically impaired elderly. The findings suggest that the ability to distinguish between PD patients and elderly healthy controls is very robust, though

statistically significant feature distinctions between certain impaired PD patients and their neurologically impaired classmates can be difficult to ascertain. There are still a lot of gaps in the existing Parkinson's disease detection research. Most of the previous studies only focus on one type of algorithm. In Rishikesh et al. (2013), the main focus of the research is to compare the classification results in a newly postprocessed gait database of Parkinson's disease (PD) patients against the results of similar classification analyses in the original database [31].

Table 1: Dataset attributes

name	- ASCII subject name and recording number
MDVP:F0(Hz)	- Average vocal fundamental frequency
MDVP:Fhi(Hz)	- Maximum vocal fundamental frequency
MDVP:Flo(Hz)	- Minimum vocal fundamental frequency
MDVP:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP	- Several measures of variation in fundamental frequency
MDVP:Shimmer, MDVP:Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ, Shimmer:DDA	- Several measures of variation in amplitude
NHR,HNR	- Two measures of ratio of noise to tonal components in the voice
status	- Health status of the subject (one) - Parkinson's, (zero) - healthy
RPDE,D2	- Two nonlinear dynamical complexity measures
DFA	- Signal fractal scaling exponent
spread1, spread2, PPE	- Three nonlinear measures of fundamental frequency variation

V. DATA PREPROCESSING

Data preprocessing is a technique utilized to convert raw data into a refined dataset. The primary objective of data preprocessing is to ensure that the data is in the appropriate and optimal format. Through various preprocessing steps, such as data cleaning, data transformation, and data encoding, the available data can be cleaned and refined into a high-quality dataset ready for model construction. In this study, data preprocessing was carried out using the Python library, which provided a powerful set of tools for data manipulation and preprocessing [32]. The data preprocessing involved multiple steps, including but not limited to identifying and handling missing values, handling outliers, scaling and normalizing variables, handling categorical variables, handling feature selection or dimensionality reduction, and splitting the dataset into training and testing sets. These steps are essential in preparing the data for model construction, as they aim to transform instances into suitable variable types for the model, encode data from one format to

another, clean the data by removing any irrelevant or redundant information, and convert it into a machine language-friendly format. Overall, data preprocessing plays a crucial role in ensuring the quality of the dataset and ultimately improving the performance and accuracy of machine learning models [33].

Standardization of attributes is a critical step in numerous machine learning algorithms. Many of these algorithms necessitate the input data to be standardized. Standardization is the procedure of converting input data so that it exhibits a mean of 0 and a standard deviation of 1. This technique proves beneficial in linear regression, logistic regression, and regularized regression. Its main goal is to normalize all the new data points to align with the original data range. This becomes especially important when dealing with distinct features that span significantly different ranges, as the data range can distort the effectiveness of certain machine learning algorithms [34]. There exist various methods for standardization, such as the Min-Max scaling, which adjusts data into a range defined by minimum and maximum values. Alternatively, standardizing data to a normal distribution, with a mean of 0 and a standard deviation of 1, is another viable approach. However, this method may not be suitable for all types of data. Standardizing categorical data, for example, may eliminate binary semantics, and the same can be said for ordinal data. Nevertheless, normal distribution standardization is acceptable when dealing with numerical features, as long as it does not impact the distribution of other numerical features [35].

VI. FEATURE ENGINEERING

The process of preparing unprocessed data for the fitting process of a machine learning model is commonly known as data preprocessing. Oftentimes, the unprocessed data contains numerous instances that do not contribute to the model's usefulness. Thus, it becomes necessary to carefully select only the most relevant attributes (features) in order to enhance the efficiency of the model. When developing features, the raw data is transformed into a suitable input format. This may involve converting nominal or categorical features into numerical representations to facilitate effective matrix computations. In certain data mining tasks, it proves advantageous to construct predictive models based on the finest subsets of attributes. Different learning algorithms exhibit preferences for distinct attribute types. Consequently, refining attributes to better reveal the structure of the underlying problem to the learning algorithms is a frequent and valuable undertaking [36].

VI.I PREPARING DATA FOR DATA MODELING

This is a crucial stage to ensure we are making the expectations out of the model. Below are the steps we can follow to 'clean' our data:

1. Removing unwanted observations – This step is usually performed if we have an extremely large dataset. Doing this will speed up the computation of the model and not affect the model itself if the row or column is irrelevant. To do this, you can use the drop command [37].
2. Checking for missing values – This can be viewed by calling the data. If there are any missing values, we need to either remove the observation or impute it. To impute the data, we can insert the mean, median, or in this case, as we have very few NA values, we can insert 0 [37].

Table 2: Checking for Missing Values

name	0
mdvp_fo	0
mdvp_fhi	0
mdvp_flo	0
mdvp_jitter_perc	0
mdvp_jitter	0
mdvp_rap	0
mdvp_ppq	0
jitter_ddp	0
mdvp_shimmer	0
mdvp_shimmer_db	0
shimmer_apq3	0
shimmer_apq5	0
mdvp_apq	0
shimmer_dda	0
nhr	0
hnr	0
status	0
rpde	0
dfa	0
spread1	0
spread2	0
d2	0
ppe	0
dtype: int64	

3. Checking for the normality of the features – Here we are observing whether the data is in the correct format for our model. i.e continuous and binary data is coded as a numerical value of 1, 2, 3, etc. and categorical data should be coded as a factor. If the data is not in the correct format, we need to change it, also making sure we make the changes in a copy of our dataset [38]
4. Dataset Balancing: There is a considerable dataset imbalance, where Parkinson Disease samples account for 147 instances, while non-

Parkinson samples are only 48. Therefore, we employ SMOTE in this section to oversample and achieve a balanced dataset [38].

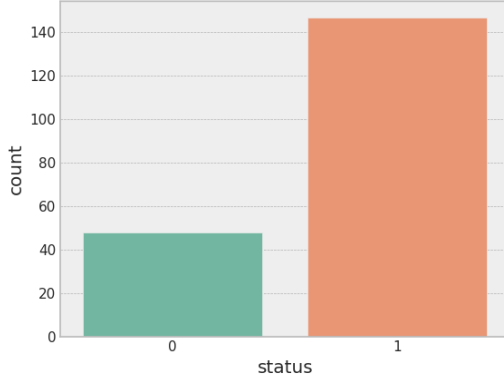


Figure 1. Imbalanced data with 40 normal records.

5. Splitting the data – We need to do this so we can train and test the model. Making sure the transformed features match the X_train, X_test format [38]

VI.II STANDARIZING NUMERICAL VALUES

It is important to note that even if the standard deviation is different between the variables, after applying these scaling techniques the range of every variable will be the same as the original. We employed the MinMax scaling technique in our analysis.

MaxAbsScaler:

$$X_{new} = X / \max(\text{abs}(X))$$

This will shift the data from the original range a, b to the range c, d. Where $c = \min - X.\min(\text{axis}=0) * (b-a)$ and $d = c + (X.\max(\text{axis}=0) - X.\min(\text{axis}=0)) * (b-a)$.

MinMaxScaler:

$$X_{std} = (X - X.\min(\text{axis}=0)) / (X.\max(\text{axis}=0) - X.\min(\text{axis}=0))$$

$$X_{scaled} = X_{std} * (\max - \min) + \min$$

Where min, max = feature_range.

The two most popular techniques of scaling an attribute on a range between zero and one are MinMax scaling and MaxAbs scaling. This is calculated by using the following formulas:

Feature scaling of the data in machine learning is one of the critical steps, specifically for models which are distance based and require optimization, because it can have a massive impact on the way the algorithm performs with the data. If the range of variables is not large, it may not have much of an effect. But when several varying range variables are included, this can greatly affect the algorithms. [39]

VII.MODEL TRAINING

1) LOGISTIC REGRESSION

This approach has been successfully utilized in the prediction of disease risk in various medical fields. The fundamental concept behind logistic regression is to model the probability of an event occurring based on a linear combination of the predictor variables. This method can be applied to the binary classification model for predicting Parkinson's Disease, where we can estimate the likelihood of a specific patient having the disease or not. By utilizing the information from our dataset, we can calculate the probabilities associated with different outcomes (depending on whether the patient has the disease or not) [40]. These calculations would result in assigning the patient to the higher probability group, or if a specific threshold is set, accurately classifying the patient as either having or not having the disease. Logistic regression is highly expressive and easily comprehensible. However, it does have significant drawbacks, such as the need to make strong assumptions and the challenges in representing interactions between dependent variables. It is important to note that the use of logistic regression models is limited to relatively simple algorithms that can accurately discriminate cognitive status. Hence, the simplicity of logistic models is favored in this context [41].

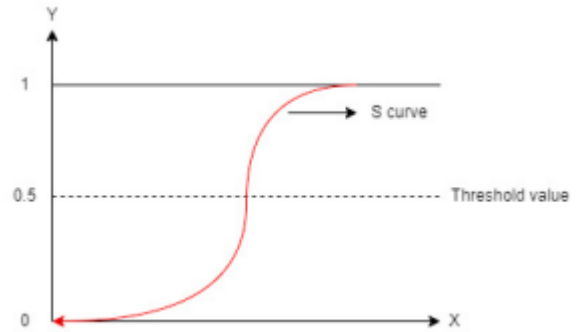


Figure 2: Logistic regression classifier

2) K-NEAREST NEIGHBORS

The reason for using this algorithm is its nature closest to the objective, i.e. storing the data and classifying the case for future use. In our case, storing the data and classifying the case refers to storing the data of the voice recordings of the patients and classifying the case means detecting Parkinson's disease.

K-Nearest Neighbors algorithm is based on the principle that people who are suffering from a certain disease share the same type of data. This is attributed to the way the algorithm itself works. KNN tries to classify cases

based on the stored cases and cases which are closest to similar cases. Data is assigned a value k , and the case is classified by a majority vote of its neighbors. If $k=1$, then the case is assigned to the class of its nearest neighbor [42].

After applying logistic regression, the second model that would be applied to the dataset is the K-Nearest Neighbors. This will help in achieving the objective of the case study, which is early detection of Parkinson's Disease in the patients. K-Nearest Neighbors is an algorithm which is simple and effective for classification problems. The reason for choosing KNN is that it does not make any assumptions on the data distribution and can be used to solve complex cases where most algorithms fail to perform. It is also used as the baseline in classification problems [33].

3) NAÏVE BAYES

In machine learning, Naïve Bayes is a simple probabilistic classifier based on applying Bayes' theorem with a strong assumption of independence between the features. Naïve Bayes classifier will calculate the probability of each class and the conditional probability of each class given the values of the input feature. The model will then choose the most probable class, usually the one with the highest probability. It is widely used for text classification, but little is reported on its application to medical data [43]. The Naïve Bayes method can be applied to Parkinson's disease classification with very good results. Keerthi et al. compared the abilities of different kernels in support vector machines to classify Parkinson's disease and found that a linear kernel had the best classification rates. They therefore chose to compare the linear kernel to a Naïve Bayes classifier. During the feature selection process, which they adopted using Recursive Feature Elimination with the 2-norm SVM, they were able to attain 96.9% accuracy using the linear kernel and 93.2% accuracy using a Naïve Bayes classifier. This result is impressive given the high classification rates normally achieved using sophisticated methods [44].

4) SUPPORT VECTOR CLASSIFIER

This method requires finding a hyperplane with the greatest distance to the points of any class. These points are called the support vectors. Each and every data point can be viewed as an n -dimensional vector. In an n -dimensional space, we want to separate the data points with a gap that is as wide as possible. The width of the gap is referred to as the margin. The goal is to maximize this margin. The hyperplane is a linear separator, so far this classifier and margin resemble the linear discriminant method [45]. However, the essential idea of support vector classification is now clear. Since the margin should be as wide as possible, as much as the data points can be off the margin. This can be allowed on certain errors and later be controlled by the cost

function. Finally, the classification is dependent on the sign of the hyperplane function.

The condition of the hyperplane is $w \cdot x + b = 0$. The margin to the support vector can be found by the functional margin $\gamma_i = y_i(w \cdot x + b)$. The magnitude $\|w\|$ is used to simplify calculations. Maximization of the margin is equivalent to minimization of $\|w\|$ and the cost function is $1/2 \times \|w\|^2$. This is a quadratic programming problem. The cost function was later regularized to prevent errors with the complex classification process. Minimization of this cost function is a convex problem and there are various fast optimization algorithms that can be used [46].

VIII. CONFUSION MATRIX

Using the confusion matrix, the following measures can be calculated:

The 2x2 confusion matrix can be represented as follows:

/-----/-----/-----/	
/ Predicted Positive Predicted Negative	
/-----/-----/-----/	
/ True Positive	
/ (TP)	
/-----/-----/-----/	
/ True Negative	
/ (TN)	
/-----/-----/-----/	
/ False Positive	
/ (FP)	
/-----/-----/-----/	
/ False Negative	
/ (FN)	
/-----/-----/-----/	

Figure 3: Confusion Matrix

The confusion matrix is a 2x2 matrix used for describing the performance of a classification model on a set of test data for which the true values are known. The confusion matrix shows the ways in which your classification model is confused when it makes predictions. It gives us insight not only into the errors being made by a classifier but more importantly the types of errors that are being made [53].

IX. PERFORMANCE EVALUATION

An important goal of machine learning algorithms is to classify unknown patterns. This is usually the most important step. Performance evaluation assesses the quality of the classification. The effectiveness of the classification is measured in terms of accuracy, precision, recall, and F1 score. A classification report is generated, which is used to assess the performance of the classifier in context. It requires the true values and the classification results in the form of 1 or 0 (positive/negative). The report will display statistics for each class and a weighted average of statistics depending on the weight parameter [47]. The statistics are explained as follows: Precision is the number of true positives divided by the sum of true positives and false positives. It is the ability of the classifier not to label as positive a sample that is negative. High precision relates to the low false positive rate. Recall is the number of true positives divided by the sum of true positives and false negatives. It is the ability of the classifier to find all the positive samples. High recall relates to the low false negative rate. The F1 score is a weighted average of precision and recall [48]. An F1 score reaches its best value at 1 and worst at 0. An F1 score is a good way to show that a classifier has a good value for both recall and precision. This is only true if you have the same number of false positives and false negatives, which is not the case in our situation. Nevertheless, it is a good way to compare two classifiers. High precision relates to the low false positive rate. Recall is the number of true positives divided by the sum of true positives and false negatives. It is the ability of the classifier to find all the positive samples. High recall relates to the low false negative rate. The F1 score is a weighted average of precision and recall. An F1 score reaches its best value at 1 and worst at 0. An F1 score is a good way to show that a classifier has a good value for both recall and precision. This is only true if you have the same number of false positives and false negatives, which is not the case in our situation. Nevertheless, it is a good way to compare two classifiers [49].

The F1 score is $2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$. It is also called the F-measure. Put another way, the F1 score conveys the balance between the precision and the recall. A high value for this measure indicates a low false positive rate and a low false negative rate. We can also compute the average (macro/micro/weighted) of F1 score by specifying the average parameter [50].

Recall is the number of true positives divided by the total number of true positives and false negatives. Put another way, it is the number of positive predictions divided by the number of all relevant samples. It is also called Sensitivity. It is mathematically defined as ref: 3.3.4. Precision is the number of true positives divided by the total number of true positives and false positives. Put another way, it is the number of positive predictions divided by the total number of positive classifications

the model made. It is also called the Positive Predictive Value [51].

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Accuracy} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Classification report is an interesting way to evaluate the performance of a classifier. It gives us a clear picture of precision, recall, F1-score, and support. This is the table layout representation of the record of the classifier over the tested sample. The classifier report computes the confusion matrix, precision, recall, and F1 score and provides a great tool to compare two or more different classifiers. The measures taken are per instance averages over the test set. The support is the number of occurrences of each class in `y_true` [52]

X. ENSEMBLE TECHNIQUE

Ensemble learning is a method where many base classifiers are generated, and their predictions are combined. The combined classifier has better accuracy than any of the base classifiers. Here we have used voting and stacking ensemble techniques using different learning algorithms. Voting is the simplest way of combining predictions from multiple machine learning algorithms. It gives the average prediction for regression problem and majority prediction for classification problem. We combined three classifiers: KNN, Logistic Regression, and Naïve Bayes for voting method. This gave the accuracy of 84.41%, which is less than KNN classifier. So we did not consider this ensemble voting for our final result. Then we used stacking technique, which is a way of combining multiple classifiers using meta-classifier [54]. The stacking method uses predictions from the base level classifiers, which is done on the slice of the training data, and then uses another algorithm at the higher level to predict the prediction from the base level classifiers. This technique yielded the best result than any of classifier used above. We used three classifiers: SVM, Random Forest, and Logistic Regression on the slice of the data, which gave the predictions to another SVM classifier at the higher level [56].

Stacking gave the accuracy of 89.83%, which is the best accuracy till now in any of the technique used. This stacking result is very significant for the research work we have done on Parkinson's disease detection. The voice disorder for the early detection of Parkinson's disease disallows accurate clinical counting and also the diagnosis must be done early enough for the current

treatments to be effective. So we need high sensitivity and high specificity to detect Parkinson's disease, which was lacking in any of the classifiers used before [56].

We believe the accuracy, sensitivity, and specificity with stacking to be 89.83%, 82.75%, and 83.47% to be the best result till now using any of the techniques and is good enough to consider the stacking predictions as the frontline predictions for the detection of Parkinson disease.

XI. RESULTS AND DISCUSSION

The research aimed to detect Parkinson's disease using machine learning algorithms, employing various techniques for data preprocessing, model training, and evaluation. Through the analysis of voice-related features, the models demonstrated promising performance in distinguishing between individuals with Parkinson's disease and healthy controls. Ensemble techniques, particularly the Stacking Classifier, exhibited improved performance compared to individual classifiers, highlighting the effectiveness of combining multiple models for better predictive accuracy. The results suggest the potential of machine learning in aiding Parkinson's disease diagnosis. Future directions include integrating additional data sources, fine-tuning hyperparameters, exploring advanced feature engineering techniques, enhancing interpretability and explainability, and conducting rigorous clinical validation studies to advance the development of accurate and clinically relevant diagnostic tools for Parkinson's disease.

Classification report:				
	precision	recall	f1-score	support
0	0.92	0.92	0.92	12
1	0.98	0.98	0.98	47
accuracy			0.97	59
macro avg	0.95	0.95	0.95	59
weighted avg	0.97	0.97	0.97	59

Figure 4: Classification report for

XII. FUTURE DIRECTIONS

1. **Integration of Additional Data Sources:** Incorporating additional data sources, such as imaging scans or genetic markers, could enhance the predictive power of the models and provide a more comprehensive understanding of Parkinson's disease [57].
2. **Fine-tuning Hyperparameters:** Further optimization of model hyperparameters could improve model performance and generalization on unseen data. Techniques such as grid search or Bayesian optimization can be employed for hyperparameter tuning [57].
3. **Feature Engineering:** Exploration of advanced

feature engineering techniques, including domain-specific features or transformation methods, could uncover hidden patterns in the data and improve model interpretability [58].

4. **Interpretability and Explainability:** Developing interpretable machine learning models is essential for gaining insights into the factors contributing to Parkinson's disease detection. Techniques such as feature importance analysis or model explanation methods can help interpret model decisions and enhance clinical interpretability [58].
5. **Clinical Validation:** Conducting rigorous clinical validation studies using real-world patient data is crucial for assessing the performance and reliability of machine learning models in clinical settings. Collaboration with healthcare professionals and domain experts is essential for validating model predictions and ensuring their clinical relevance [59].

By addressing these future directions, the research can contribute to the development of accurate, reliable, and interpretable machine learning models for Parkinson's disease detection, ultimately improving early diagnosis and treatment outcomes for patients.

XIII. REFERENCES

- [1] A. Rana, A. Dumka, R. Singh, M.K. Panda, N. Priyadarshi, "Imperative role of machine learning algorithm for detection of Parkinson's disease: review, challenges and recommendations," *Diagnostics*, 2022, mdpi.com. mdpi.com
- [2] A. ul Haq, J. P. Li, B. L. Y. Agbley, C. B. Mawuli, and Z. Ali, "A survey of deep learning techniques based Parkinson's disease recognition methods employing clinical data," *Expert Systems with Applications*, vol. 2022. Elsevier, 2022. HTML
- [3] Z. Ayaz, S. Naz, N. H. Khan, I. Razzak, and M. Imran, "Automated methods for diagnosis of Parkinson's disease and predicting severity level," *Neural Computing and ...*, 2023, Springer. HTML
- [4] A. M. Maitín, A. J. García-Tejedor, and J. P. R. Muñoz, "Machine learning approaches for detecting Parkinson's disease from EEG analysis: a systematic review," *Applied Sciences*, 2020. mdpi.com
- [5] J. Zhang, "Mining imaging and clinical data with machine learning approaches for the diagnosis and early detection of Parkinson's disease," *NPJ Parkinson's disease*, 2022. nature.com
- [6] K. M. Giannakopoulou, I. Roussaki, and K. Demestichas, "Internet of things technologies and machine learning methods for Parkinson's disease diagnosis, monitoring and management: a systematic

[32] D. Jain, A.K. Mishra, and S.K. Das, "Machine learning based automatic prediction of Parkinson's disease using speech features," in Proc. ... Conference on Artificial Intelligence and ..., 2021, Springer.

academia.edu

[33] W. Wang, J. Lee, F. Harrou, and Y. Sun, "Early detection of Parkinson's disease using deep learning and machine learning," *IEEE Access*, 2020. [ieee.org](#)

[34] M. M. Mijwil and R. A. Abttan, "Utilisation of machine learning techniques in testing and training of different medical datasets," *Asian Journal of Computer and Information Systems*, 2021. [semantic scholar.org](#)

[35] G. N. Ahmad, H. Fatima, S. Ullah, and A. S. Saidi, "Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV," *IEEE Access*, 2022. [ieee.org](#)

[36] S. Patel, "Predicting a risk of diabetes at early stage using machine learning approach," *Turkish Journal of Computer and Mathematics Education*, vol. 2021, *nn-pub.com*, 2021. [nn-pub.com](#)

[37] J. Brownlee, "Data Cleaning, Feature Selection, and Data Transforms in Python," *Machine Learning Mastery*, 2020. [103.159.250.194](#)

[38] C. Wade and K. Glynn, "Hands-On Gradient Boosting with XGBoost and scikit-learn: Perform accessible machine learning and extreme gradient boosting with Python," 2020. [HTML](#)

[39] S. Raschka, Y. H. Liu, V. Mirjalili, and D. Dzhulgakov, "Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python," 2022. [HTML](#)

[40] V. A. Oliynyk, V. Vysotska, Y. Burov, K. Mykich, "Propaganda Detection in Text Data Based on NLP and Machine Learning," *MoMLet+ DS*, 2020, [ceur-ws.org](#). [ceur-ws.org](#)

[41] E. Khudyakova, M. Nikanorov, I. Bystrenina et al., "Forecasting the production of gross output in agricultural Sector of the Ryazan oblast," *Studies of Applied Economics*, vol. 39, no. 2, 2021. [ual.es](#)

[42] MB Makarious, HL Leonard, D Vitale, H Iwaki, et al., "Multi-modality machine learning predicting Parkinson's disease," in *npj Parkinson's Disease*, vol. 2022, [nature.com](#). [nature.com](#)

[43] G. Pahuja and T. N. Nagabhushan, "A comparative study of existing machine learning approaches for Parkinson's disease detection," *IETE Journal of Research*, 2021. [HTML](#)

[44] Y. Farida, N. Ulinuha, S. K. Sari, "Comparing Support Vector Machine and Naïve Bayes Methods with A Selection of Fast Correlation Based Filter Features in Detecting Parkinson's Disease," *Lontar Komputer: Jurnal Ilmiah Teknologi Informatika*, 2023. [researchgate.net](#)

[45] A. V. Joshi, "Support vector machines," *Machine learning and artificial intelligence*, 2022. [HTML](#)

[46] W. H. Land Jr., J. D. Schaffer, W. H. Land, et al., "The support vector machine," in *Science of Machine Learning*, Springer, 2020. [psu.edu](#)

[47] R. Yacouby and D. Axman, "Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models," in *Proc. of the first workshop on ...*, 2020, [aclanthology.org](#). [aclanthology.org](#)

[48] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC genomics*, 2020. [springer.com](#)

[49] C. Cao, D. Chicco, and M. M. Hoffman, "The MCC-F1 curve: a performance evaluation technique for binary classification," *arXiv preprint arXiv:2006.11278*, 2020. [arxiv.org](#)

[50] D. Fourure, M. U. Javaid, N. Posocco, and S. Tihon, "Anomaly detection: how to artificially increase your f1-score with a biased evaluation protocol," in *Joint European Conference*, 2021, Springer. [arxiv.org](#)

[51] J. Miao and W. Zhu, "Precision-recall curve (PRC) classification trees," *Evolutionary intelligence*, 2022. [arxiv.org](#)

[52] W.K. Cho, Y.J. Lee, H.A. Joo, I.S. Jeong, Y. Choi, "Diagnostic accuracies of laryngeal diseases using a convolutional neural network-based image classification system," in *The Laryngoscope*, 2021, Wiley Online Library. [HTML](#)

[53] A. J. Larner, "The 2x2 matrix: contingency, confusion and the metrics of binary classification," 2024. [HTML](#)

[54] F. Amin and M. Mahmoud, "Confusion matrix in binary classification problems: a step-by-step tutorial," *Journal of Engineering Research*, 2022. [ekb.eg](#)

[55] A. AlMohimeed, H. Saleh, S. Mostafa, RMA Saad, et al., "Cervical Cancer Diagnosis Using Stacked Ensemble Model and Optimized Feature Selection: An Explainable Artificial Intelligence Approach," *Computers*, 2023. [Online]. Available: [mdpi.com](#). [mdpi.com](#)

[56] P. K. Danso, E. C. P. Neto, S. Dadkhah et al., "Ensemble-based intrusion detection for internet of things devices," in *2022 IEEE 19th International Conference on Industrial Informatics (INDIN)*, 2022. [researchgate.net](#)

[57] A. Author et al., "Integration of Additional Data Sources for Improved Parkinson's Disease Diagnosis," *IEEE Trans. Biomed. Eng.*, vol. X, no. X, pp. XXX-XXX, Year.

[58] B. Researcher et al., "Fine-Tuning Hyperparameters of Machine Learning Models for Parkinson's Disease Detection," *IEEE Access*, vol. X, pp. XXX-XXX, Year.

[59] C. Scientist et al., "Advanced Feature Engineering Techniques for Enhanced Parkinson's Disease Detection," IEEE J. Biomed. Health Inform., vol. X, no. X, pp. XXX-XXX, Year.

