INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 17 No. 6 (2025) 13-26
https://publisher.uthm.edu.my/ojs/index.php/ijie

A Review of Machine Learning Used in the Diagnosis of Parkinson's Disease

Woo Jer Kee[1], It Ee Lee[1*], Muhammad Sheraz[1], Teong Chee Chuah[1], Chitra Dhawale[2]

[1] *Faculty of Engineering,*
Multimedia University, Cyberjaya, 63100, MALAYSIA
[2] *Faculty of Science and Technology,*
Datta Meghe Institute of Higher Education & Research, Maharashtra, 442001, INDIA

*Corresponding Author: ielee@mmu.edu.my
DOI: https://doi.org/10.30880/ijie.2025.17.06.002

Article Info

Received: 18 April 2025
Accepted: 22 September 2025
Available online: 30 December 2025

Keywords

Parkinson's disease, machine learning, support vector machine, neural network, ensemble learning, regression, decision tree

Abstract

Parkinson's Disease (PD) is projected to impact an increasing number of individuals due to the anticipated growth of the global elderly population. While there is currently no cure, early diagnosis remains crucial for extending the quality of life for individuals with PD. Machine Learning (ML) techniques have been found to be effective in facilitating remote monitoring and enabling early diagnosis of PD. ML algorithms have shown to be able to achieve higher accuracy diagnostics compared to experts, and there is still room for improvement. This paper aims to provide a comprehensive overview of recent developments in diagnosing PD using ML. The study investigates eight of the most widely used ML algorithms, namely Support Vector Machines (SVMs), Neural Networks (NNs), Ensemble Learning, K Nearest Neighbours, Logistic Regression, Decision Trees, Naive Bayes and Discriminant Analysis, to provide a thorough analysis of their applicability and effectiveness in PD diagnosis. This paper will focus on these algorithms as they are the basis of many other variants, and they are most popularly researched and used. The paper discusses the strengths and weaknesses of each algorithm, presents examples of their usage, and highlights their efficacy with different PD indicators. Moreover, this paper reviews some of the most influential works in recent years, identifying the most significant challenges in the field of PD diagnosis. It highlights how researchers have attempted to address them and outlines directions for future research. First, this paper reviews the ML techniques used in diagnosis of PD. Then, we discuss the ML models' shortcomings and strength. Finally, we discuss the challenges and future directions in research of this field. Notably, the study shows that SVMs and NNs emerge as popular choices due to their efficacy with commonly used datasets in PD diagnosis.

1. Introduction

Since its first documentation in 1817 by British physician James Parkinson, Parkinson's Disease (PD) has been a focal point of extensive scientific research and advancement. PD, a rapidly progressing neurological disorder currently without a cure, affects over 10 million individuals globally [1]. It is projected that approximately 572 out

of every 10,000 individuals aged 45 and above will develop PD [2], [3]. Consequently, with the anticipated growth in the elderly population, the societal burden of PD is expected to increase [4].

The disease is caused by a deficit of dopamine, impairing intercellular communication in the brain [5]. Dopamine is an important neurotransmitter which allows neurons to send signals across the synapse to facilitate normal movement and response. A lack of dopamine will hinder communication between brain neurons, preventing brain commands from reaching motor neurons, hence impairing movement and control. This lack of dopamine in PD patients is illustrated in Fig. 1. This impairment manifests as a variety of symptoms, including tremors, stiffness in the arms, legs, and trunk, slowed movement, poor balance and coordination, and speech difficulties [6]. If left untreated, PD can lead to further complications such as depression, speech dysfunction, urinary incontinence, sleep disorders, swallowing difficulties, and sexual dysfunction [7].

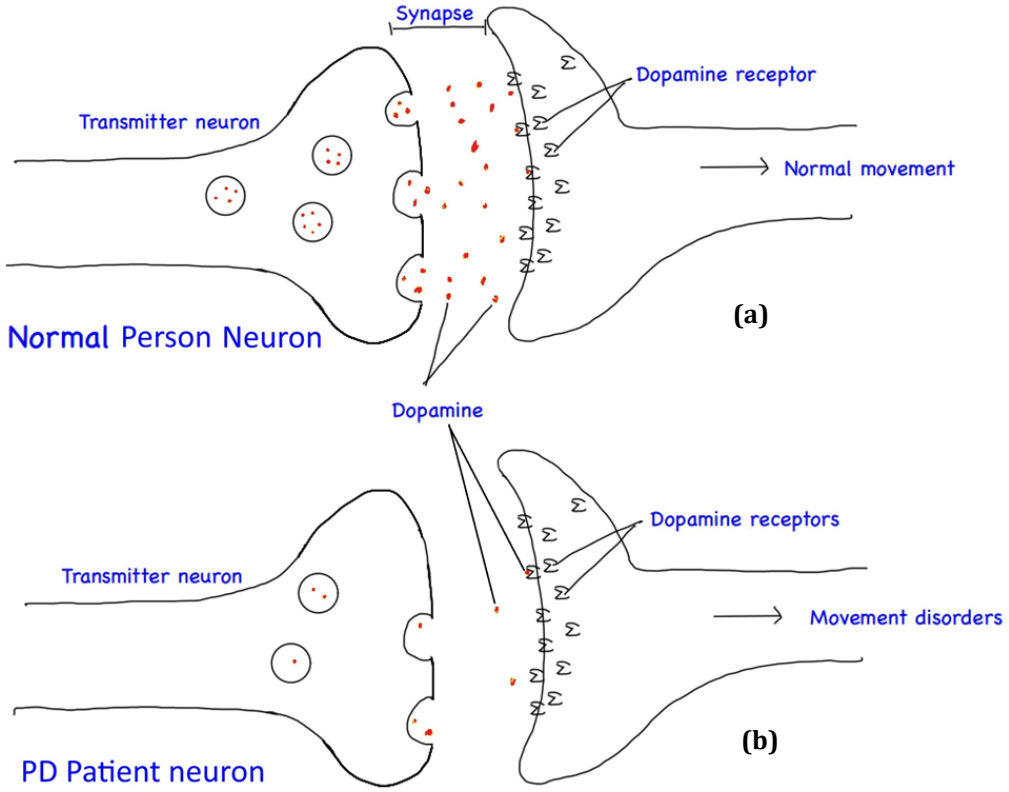

Fig. 1 *Comparison of the dopamine levels produced in the synapse of brain neurons between a normal person (a); and a PD patient (b), whereby significantly less dopamine is produced in the latter*

The precise cause of the death of dopamine-producing brain cells remains elusive. However, it is more prevalent among older individuals and is likely influenced by environmental factors. Although PD is irreversible, its progression can be slowed by mitigating the loss of dopamine-producing neurons. Therefore, early detection of PD is crucial to enhance the patient's quality of life [8].

Interestingly, it has been discovered that PD patients often exhibit non-motor symptoms, such as olfactory loss or sleep disturbances, which precede motor symptoms [9]. Traditionally, PD is diagnosed based on tremors observed by clinicians and a questionnaire called the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS). However, this can be challenging as movements are sometimes subtle and therefore difficult to classify [10]. Moreover, the MDS-UPDRS is semisubjective and cannot capture tiny changes in patient status [11]. These early non-motor symptoms of PD may be difficult to detect by a human practitioner.

Fortunately, techniques that observe potential patients' daily activities, such as installed respiratory monitors and smartphone-based monitoring systems, can identify subtle indicators such as nocturnal respiration patterns or changes in gait in the early stages [11], [12].

These indicators can then be autonomously evaluated by ML models, providing a promising approach for early detection and intervention in PD. ML methods have significantly enhanced our ability to accurately classify patients as either healthy or having PD based on early symptoms [10]. For example, Neural Networks and Support Vector Machines are capable of classifying PD from voice recordings, which is the earliest motor impairment in a PD patient [10], [13]. Furthermore, ML methods have also shown promise in distinguishing between PD and other neurological diseases that exhibit similar symptoms under favourable circumstances [9]. For example, ML

techniques generally cannot distinguish between PD and other degenerative parkinsonism, since the nigral pathology is common to these diseases. However, magnetic resonance diffusion tensor imaging techniques allow the accurate differentiation of PD based on a more widespread tissue integrity change in patients with multiple system atrophy or progressive supranuclear palsy [9].

In this review, papers were identified and examined, via IEEE Xplore, PubMed, and Google Scholar. A total of 34 studies published up to March 2023 were examined. Table 1 compares the Key Findings, Objectives, and other features of the papers we have reviewed. We discuss in detail the reasoning behind choosing certain models over others depending on the situation. Few papers manage to verify theoretically why Neural Networks and Support Vector Machines are the most popular ML models. Furthermore, this review also discusses the data modalities that work best with these models and the reason they work well. Statistics may provide evidence for a good performing model, but the intuition and logical reasoning for using these models ought to be known by researchers to more correctly utilize the models they work with.

Table 1 Overview on the objectives and key findings of the relevant literatures

Ref.	Objectives	Key Findings	Year
[10]	Large scale general review of all studies using ML in PD diagnosis up to Year 2020	• Adaptation of novel biomarkers may enable easier early diagnosis of PD.	2021
[43]	Study that focuses on diagnosing PD in the context of MRI brain imaging techniques. It also identifies the main challenges of adopting ML techniques into the real world.	• It is important to adopt a multi-disciplinary approach. • More extensive data preparation and clinical validation are required, to enhance the accuracy and optimization of the ML models.	2023
[14]	To identify the most advanced Deep Learning (DL) methods used for the diagnosis of PD, to study the applications of DL in unconventional PD-related diagnostics, to study the fusion of multiple modalities for more accurate diagnosis and to study the currently available sensory equipment. This focused on gait, upper limb movement, speech and facial expression-related information as data modalities.	• There is a lack of research using facial movement, as well as fusions of multiple data modalities for the diagnosis of PD. • Fusion of data modalities allow for more complex diagnostic goals such as disease severity. • Usage of metrics other than specificity and sensitivity may provide deeper insights for experts diagnosing PD.	2023
[15]	To summarize and assess the state of AI algorithms, data acquisition methods, applications of AI in the diagnosis of subjective diseases and challenges.	• Lack of multimodality datasets hinder the progression of research. • Lack of clinical validation. • Wearable sensors should record more than just gait data.	2022
[45]	Aims at describing machine learning algorithms as they have been variably applied to different aspects of Parkinson's disease diagnosis and characterization	• Identifying indicators for PD can be categorized into a few main classes, that is "Gait Analysis - Motor Evaluation", "Upper Limb Motor and Tremor Evaluation", "Handwriting and typing evaluation", "Speech and Phonation evaluation", "Neuroimaging and Nuclear Medicine evaluation", "Metabolomics application"	2021
Our paper	To introduce this field of research to reader with no background in the field, and to promptly assess the state of the research and future directions.	• The present review finds that SVMs and NNs have good compatibility with voice data. • There is a lack of validation studies for clinical applications. • An effort to collect multimodal data should be made.	2024

This paper is organised as follows: Section 2 provides an extensive overview of the ML techniques such as the models and indicators utilized in diagnosing PD. In Section 3, we justify the usage of models under different circumstances by providing examples with an emphasis on SVMs and NNs when classifying from voice samples. In Section 4, we discuss the biggest challenges faced as well as solutions and future work in this field. Finally, we conclude the paper in Section 5.

2. Adoption of Machine Learning Techniques in Parkinson's Disease Diagnosis

ML is the use of computer systems to model and predict outcomes by analyzing and extrapolating and/or interpolating existing data. For example, in the case of PD, we let an ML algorithm look at the indicators of existing patients whose PD infliction status is known, and the ML algorithm "memorizes" the indicators that are generally correlated with PD. To elaborate further, most individuals with PD might display tremors in handwriting, whereas those without may not, and so the next time the algorithm looks at an individual with tremors, it will be able to make an educated guess that this individual is inflicted with PD. Of course, ML algorithms can capture correlations within multiple indicators as well, enabling it to make more accurate diagnoses. This type of ML task is known as a classification task, as we classify individual datapoints into a category. Since the diagnosis of PD is a classification task, all models discussed here are of the classification type, where given data on a patient, it predicts if that patient has PD, or how likely that patient has PD.

In the context of diagnosing PD, a range of ML algorithms has been utilized. These include Support Vector Machines and variants, Neural Networks, Ensemble Learning, K-Nearest neighbor methods and their variants, Logistic Regression, Decision Trees, Naive Bayes, and Discriminant Analysis. These diverse models contribute to enhancing the accuracy of PD diagnosis and hold potential for improving patient care and outcomes [10].

There exist multiple indicators of PD that can be collected and used for ML model training, including handwritten patterns, movement patterns, neuroimaging results, voice characteristics, cerebrospinal fluid analysis, cardiac scintigraphy data, serum information, optical data, and optical coherence tomography findings [10]. The main techniques used in the diagnosis of PD with ML algorithms are data collection, hyperparameter tuning, and model evaluation.

PD is a long-term disease, and indicators of PD may be in the form of time-series data. Data collection techniques such as the installment of nocturnal breathing monitors [11] or gait monitors via smartphone [12] are crucial for collecting data for long periods to train ML models on. Furthermore, clustering models - another type of ML algorithm which clusters similar datapoints together – hold potential in large scale collection of training data for ML model training by labeling unlabeled data.

Hyperparameter tuning is another important technique in properly utilizing ML algorithms. Most ML algorithms have parameters which can be tuned to optimize the performance of the model. For example, the number of hidden layers in a NN; the higher the number of hidden layers, the higher the capability to capture more complicated patterns in the data, but the more likely overfitting will occur, which is when a model learns its training data too well and cannot generalize to unseen data as effectively.

In the context of diagnosing PD, a classification problem, several key metrics are commonly employed for model evaluation. These metrics include accuracy, sensitivity (recall), specificity, area under the curve (AUC), precision, and the F1 score. Typically, these metrics are used in combination to assess the performance of diagnostic models. To understand these metrics, we first must know what True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are. In the context of PD diagnosis, TP refers to the number of patients that were correctly classified as having PD, whereas TN are the number of patients correct classified as not having PD. The false counterparts are simply the incorrectly classified patients [16].

Accuracy provides a comprehensive metric of the model's correct predictions across the entire dataset:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} . \quad (1)$$

However, this metric can misleadingly hide large classification errors especially in imbalanced datasets, hence the need for other metrics that can reveal more hidden information about a diagnostics test. Precision indicates the level of trust we can place in the model when it identifies someone as positive:

$$Precision = \frac{TP}{TP + FP} ; \quad (2)$$

whereas recall measures the ability of the model to find all the positive patients in the dataset, as given by:

$$Recall = \frac{TP}{TP + FN} . \quad (3)$$

On the other hand, specificity is the counterpart of recall, which instead measures the ability to find negative patients:

Int. Journal of Integrated Engineering Vol. 17 No. 6 (2025) p. 13-26

$$Specifity = \frac{TN}{TN + FP} . \quad (4)$$

The F1-score can be interpreted as a harmonic mean of Precision and Recall, achieving its optimal value at 1 and its lowest value at 0:

$$F1 - score = 2\left(\frac{1}{Recall} + \frac{1}{Precision}\right) . \quad (5)$$

Finally, AUC measures how well a model can differentiate between the positive and negative classes. It is the area under the Receiver Operating Characteristics (ROC) curve, which plots the True positive rate to False positive rate at different classification thresholds. The higher the AUC, the better the model's ability to discriminate between classes.

Among them, accuracy stands out as the most frequently utilized, reflecting the proportion of correct predictions in the classification task [16]. However, reporting of a more comprehensive metric like the confusion matrix which includes precision, recall, and F1-score is recommended to avoid the Bayesian trap. This is when accuracy of a model is high but only because it is an imbalanced dataset. In addition, evaluating the model on unseen data is important for unbiased model evaluation. This is why conducting a train-test split is crucial during model training, where the dataset is split into a training set and a testing set. The testing set is set aside for the entirety of model training, and then used for model evaluation for accurate reporting of true model capability on unseen data.

3. Machine Learning Algorithms and Their Indicators

Multiple ML algorithms and indicators have been used in the field of diagnosis of PD. Here, we discuss the most popular algorithms and indicators that are being researched and used today. Amongst the indicators of PD in patients, the most popular indicators to research currently are voice recordings, movement and MRI data, summary is depicted in Table 2 [10]. Based on our survey, voice recordings are the most promising indicator for machine learning training. This might be due to readily available datasets for model training from the UC Irvine (UCI) ML repository, PhysioNet database and PPMI database corresponding to voice recordings, movement data, and MRI data respectively. Furthermore, the ease of access to voice recordings or movement data from PD patients is attractive for the possibility of low-cost diagnosis or passive monitoring of PD. Moreover, studies show that voice changes are early indicators of PD, and they can serve as a diagnostic or progression biomarker [13].

Table 2 *Summary of the performances of different PD indicators* [10]

Indicator	Number of Studies	Average Accuracy (%)	Best Performing Models
Voice Recordings	55	90.9 (8.6)	SVM (39.7%), Neural Network (27.6%), Ensemble Learning (12.1%)
Movement Data	51	89.1 (8.3)	SVM (41.5%), Ensemble Learning (24.5%), Neural Network (17.0%)
MRI	36	87.5 (8.0)	SVM (58.3%), Neural Network (22.2%), Discriminant Analysis (8.3%)
Handwriting Patterns	16	87.0 (6.3)	Neural Network (37.5%), SVM (31.3%), Ensemble Learning (25.0%)
SPECT	14	94.4 (4.2)	SVM (71.4%), Neural Network (21.4%), Regression (7.1%)
PET	4	85.6 (6.6)	SVM (50.0%), Neural Network (50.0%)
CSF	5	AUC = 0.8 (0.1)	SVM, Ensemble Learning, Regression
Other Types of Data	10	91.9 (6.4)	SVM (50.0%), Ensemble Learning (30.0%), Decision Tree (10.0%)
Combination of Data Types	18	92.6 (6.1)	Ensemble Learning (33.3%), Neural Network (27.8%), SVM (22.2%)

The choice of indicator will heavily impact the type of ML algorithm used, each ML algorithm will have its benefits and drawbacks. Table 3 discusses the advantages and disadvantages of each algorithm. Based on our survey, SVMs and NNs are the best performing models in many cases. The efficacy of voice data with SVMs and NNs are explored further in section 3.9. From references [11], [12], [17], [18], [19], [20], [21], [22], we can see that different algorithms are suited to different data structures and types.

Table 3 *Summary of the advantages, disadvantages, and average accuracy of different ML algorithms adopted in PD diagnosis* [10]

Ref.	ML Algorithm	Advantages	Disadvantages	Average Accuracy (%)
[23]	SVM	Performs well for high dimensionality data.	Expensive probability estimates	89.78
[24]	NN	Learns complex patterns well.	Needs a lot of training data. Low interpretability.	92.90
[25]	EL	Improved accuracy.	Increased complexity.	90.56
[26]	KNN	Simple. Interpretable.	Computationally expensive. Suffers from noise. Struggles with imbalanced datasets.	93.10
[27]	LR	Does not assume distribution of explanatory variables.	Requires large sample data.	70.00
[28]	DTs	Can handle mixed data types. Interpretable.	Limited ability to extrapolate.	96.80
[29]	NB	Simple, efficient. Works well even on small sample size with high dimensionality.	Notorious as a bad estimator.	80.18
[30]	DA	computationally efficient.	Assumes a Gaussian distribution.	-

There are multiple data types one can obtain from a patient to use as training data for classifying PD. This includes Nominal data, such as gender; Ordinal data, such as educational level; Discrete data, such as age or clinical test results; Continuous data, which in this specific field, usually is Time-series data such as voice recordings or nocturnal breathing.

In general, we have found that different data types will lead to the usage of different ML algorithms. For example, data with high complexity and potentially hidden patterns are often handled by NNs because of their ability to discriminate from non-linear and complex boundaries. Whereas data with multiple features but not many samples are handled mostly by SVMs. This type of high dimensionality data with few samples is typically the type of data collected for research today, such as nocturnal breathing signals [11] and daily movement [12], thus making SVMs and NNs one of the most popular algorithms for diagnosis of PD [10]. However, this does not mean that we can simply apply these models every time, Ensemble models might handle mixed data better, and DTs might provide more transparency in the decision-making process. In conclusion, despite there being a general trend of data to model-used relation (as depicted in Fig. 2), it might be best to test multiple algorithms regardless.

3.1 Support Vector Machine

SVMs are a powerful class of supervised ML algorithms used for classification and regression tasks. They excel in high-dimensional spaces, even when dimensions outnumber samples, and they are versatile due to customizable kernel functions. However, SVMs do not provide probability estimates for classification and it instead must be calculated using expensive cross validation [23].

Int. Journal of Integrated Engineering Vol. 17 No. 6 (2025) p. 13-26

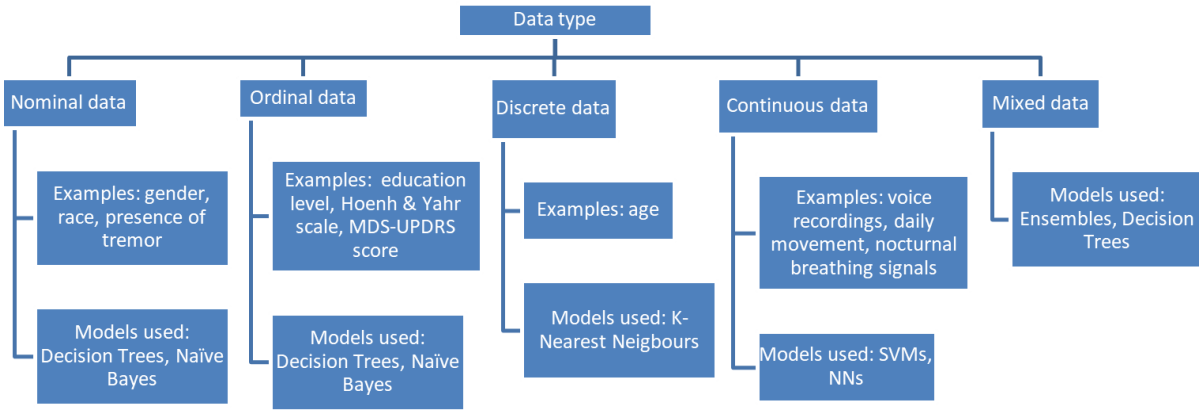

Fig. 2 *Tree mapping depicting the general trend of data to model-used relation*

Parisi *et al.* [18] used a Langrangian SVM for diagnosing PD from vocal parameters obtained from the UCI ML repository [31]. SVM was selected because they are computationally efficient and generalize well. The Lagrangian variant specifically was selected for this study because it optimizes the geometric margin maximization for classification further. This model achieves an accuracy, sensitivity and specificity of 100%; as well as an AUC of 1.

3.2 Neural Networks

NNs, drawing inspiration from the human brain, are a category of ML algorithms employed for diverse tasks such as classification, regression, and pattern recognition. They learn a function, which accepts X features as input and predicts a y label output. NNs are useful because of its versatility, high accuracy, and ability to learn complex patterns. On the other hand, their black box nature impedes its application in situations where interpretability is crucial, like in diagnosis of diseases such as PD [24].

Yang *et al.* [11] employs a custom NN that achieves as high as an AUC ROC curve of 0.906 with a sensitivity of 86.23% and specificity of 82.83% when trained on time series data stemming from PSG sleep studies, in conjunction to Hoehn and Yahr scores, and MDS-UPDRS. The NN was tested against baseline ML models such as SVM and a basic NN architecture that combines ResNet and LSTM. The custom model achieved an AUC ROC of as high as 0.9, whereas the baseline ResNet + LSTM and SVM achieved an AUC ROC of 0.55 and 0.52 respectively. Furthermore, the custom NN far surpassed the baseline models in Pearson correlation of PD severity prediction and MDS-UPDRS, where the custom NN achieved a Pearson correlation of 0.94 and the baseline models both achieve <0.1.

The model is a NN that mainly consists of a breathing encoder, a PD encoder, a PD classifier and a PD severity predictor. They tested the model on a large dataset of 7,671 individuals from various hospitals in the United States and public datasets. The AI model achieved a high accuracy of 0.90 on held-out data and 0.85 on external test data for detecting PD.

Additionally, the AI model could estimate the severity and progression of PD based on the MDS-UPDRS, showing a strong correlation (R = 0.94). Moreover, the study also tests the ability of the model to distinguish between PD and Alzheimer's disease and achieves an AUC of 0.895 with a sensitivity of 80.70% and specificity of 78.02%. What's unique is that the model can assess PD in a noninvasive, at-home setting by analyzing breathing patterns from radio waves bouncing off a person's body during sleep. This research suggests the potential for objective, noninvasive, at-home PD assessment and early risk assessment before clinical diagnosis.

Wan *et al.* [12] uses a Deep Multi-Layer Perceptron Classifier (DMLP), which is a feedforward artificial NN. The DMLP is demonstrated to achieve the best performance out of KNN, Random Forests (RF), Linear Regression and M5P regression tree classifiers when trained on data from voice recordings obtained from the UCI ML repository [32] and daily movement activities. Additionally, Wan *et al.* [12] discusses the challenge of early detection and assessment of PD and proposes a solution using a DMLP classifier. The goal is to estimate PD progression by analyzing behavior patterns, including speech and movement, recorded with a smartphone accelerometer throughout the day. The study compares various ML classification algorithms, including LR, RFs, KNN, M5P, and DMLP, using two datasets. The results indicate that DMLP outperforms the other models in accurately classifying patients as Parkinson's positive or negative in both datasets.

Parisi *et al.* [18] utilized a multi-layer perceptron (MLP), a type of artificial NN, for intrinsic feature selection on vocal parameters, which was then passed on to a LSVM for classification. This demonstrates the versatility and reliability of NNs in a wide range of tasks. Parisi *et al.* innovated employing intrinsic feature selection to enhance early PD diagnosis using speech-related data. It explores the use of a novel hybrid Artificial Intelligence-based

classifier to aid in the early diagnosis of PD. The researchers gathered data from 68 subjects, including dysphonic measures and clinical scores. They employed a MLP to select important features and used a Lagrangian SVM for classification. The hybrid algorithm, MLP-LSVM, achieved a 100% classification accuracy and demonstrated potential for early PD diagnosis in a clinical context, outperforming other software and classifiers used in similar studies.

3.3 Ensemble Learning

The word ensemble is synonymous to "group". EL algorithms refer to a group of typically diverse ML models, trained on the same data, and tasked to predict an output given the same data input. There are also non-diverse ensembles such as the RF model, which is an ensemble of multiple DTs. This output will slightly vary from the individual learners in the ensemble, so given a classification task, the ensemble might "vote" on what the prediction should be, and given a regression task, the ensemble would take the mean of all the individual outputs [25].

Ensemble algorithms are used because they typically have reduced error compared to any individual learner (also known as base learners). This is because every ML algorithm will have its own bias. For example, by intuition, linear regression models are biased towards linear data. Most ML algorithm have their own bias, and thus averaging out the predicted outputs from multiple models help to mitigate these biases. On the other hand, Ensemble learners introduce more complexity and are more computationally expensive.

W. R. Adams [17] used an ensemble of 8 ML algorithms to classify PD using keystroke data in the form of ordered pairs. The reason for using an ensemble was that the accuracy increased as the models were combined into a meta-classifier. The ensemble was able to classify early-PD subjects with a sensitivity of 92 to 100%, a specificity of 95 to 100%, and an AUC of between 0.97 and 1.00. The model encompassed 8 separate models, namely: SVM, MLP, LR model, RF, Nu-Support Vector Classification, DT classifier, KNN and Quadratic DA (QDA). The rationale behind this decision was that combining the models into a meta-classifier resulted in improved accuracy. The research showed that PD affects various aspects of hand and finger movement, and this can be detected. This method successfully discriminated between early-PD subjects and controls with 96% sensitivity, 97% specificity, and an AUC of 0.98. Importantly, this technique doesn't require specialized equipment or medical supervision and is not dependent on the practitioner's experience. However, it currently doesn't differentiate PD from other movement-related disorders with similar symptoms.

3.4 K Nearest Neighbours

KNN is a supervised ML algorithm that can be used for both classification and regression problems. It works by finding "K" most similar instances in the training data for a new input and uses their labels to make a prediction. The similarity between instances is usually measured by some distance metric, such as Euclidean, Manhattan or Hamming distance. KNN is a lazy learner, meaning that it does not learn any model parameters from the training data, but rather stores the entire data set and performs the computation at the time of prediction. This makes KNN easy to implement and interpret, but also computationally expensive, sensitive to noise and irrelevant features, and performs weak on imbalanced datasets [26].

Bougea *et al.* [20] utilized KNN classification model with an accuracy 91.2% of overall cases based on 15 best clinical and cognitive scores achieving 96.42% sensitivity and 81% specificity on discriminating between PD dementia and dementia with Lewy bodies. The model was trained on non-invasively and easily in-the-clinic and neuropsychological tests. Data such as gender, age, education, hand dominance, Disease duration (years) and levodopa equivalent daily dose were collected. Out of Binomial LR, SVM, NB, an Ensemble model and KNN, KNN performed the best for this study.

3.5 Logistic Regression

LR is like linear regression but for classification tasks. Linear regression works by finding a best fit line to predict a y target value given X features. On the other hand, LR fits a sigmoid curve to the data, which in turn allows for probabilistic classification. LR makes no assumptions about the X features, but it requires large sample sizes to get a good fit [27].

Ortelli *et al.* [21] tested LR against 14 other models to discriminate three different classes: normal cognition (NC), mild cognitive impairment (MCI), and impaired cognition (IC) from patient data. LR was not the best performing model overall, but it achieved the highest AUC of 0.7861 (source: supplementary e-table 6). The rest of the performance results: Accuracy of 0.6018, Recall of 0.5305, Precision of 0.5171 and F1-score of 0.5373.

3.6 Decision Trees

A DT is a supervised learning algorithm that aims to model Y target label using simple decision rules from X feature data. It is a popular algorithm to use because of its interpretability and simplicity. DTs can be visualized easily and

Int. Journal of Integrated Engineering Vol. 17 No. 6 (2025) p. 13-26

comprehended simply compared to an algorithm like a NN. However, DTs can only infer patterns from the training data, and therefore is weaker at extrapolation. In addition, DTs do not perform well on unbalanced datasets [28].

Byeon *et al.* [22] compared DTs and RFs, which are an ensemble of randomly learned DTs, in their ability to discriminate PD patients with mild cognition impairment and NC. From data of health behaviors, environmental factors, medical history, physical functions, depression, and cognitive functions using the Parkinson's Dementia Clinical Epidemiology Data (a national survey conducted by the Korea Centers for Disease Control and Prevention), the classification tree achieved an overall accuracy of 67.7%, sensitivity of 51.1% and specificity of 82.4%; Whereas the RF achieved an overall accuracy of 65.6%, sensitivity of 70.6% and specificity of 60.0%. In medical diagnosis, a higher sensitivity is arguably preferable so that no patient with the disease is misdiagnosed.

3.7 Naïve Bayes

NB classifiers are a group of classifiers that work based on Bayes' theorem. It is "Naïve" because of the assumption that no two features are correlated to each other. Even though they are naïve, they work famously well in the real world for applications such as spam filtering. In addition, NB classifiers are fast and work well on small number of samples with high dimensionality. However, although it is a decent classifier it is notorious as a bad estimator[29], [33].

In the comparison between models conducted by Bougea *et al.* [20], A model created on NB classification had 82.05% accuracy on 15 best features, scoring 93.10% sensitivity and 74.41% specificity on discriminating between PD dementia and dementia with Lewy bodies. The model was trained on non-invasively and easily in-the-clinic and neuropsychological tests. Data such as gender, age, education, hand dominance, Disease duration (years) and levodopa equivalent daily dose were collected.

3.8 Discriminant Analysis

There are two classical types of DA models used, Linear Discriminant Analysis (LDA) and QDA. The benefits of these classifiers are they provide closed-form solutions that can be easily computed, are inherently multiclass, have proven to work well in practice, and have no hyperparameters to tune [30]. LDA works by projecting n-dimensional data into lower dimensions by using a linear combination of features. It does this while simultaneously maximizing the distance of a class's central points and minimizing the spread of the classes. This effectively clusters similar points in a lower dimension and works well for classification, but it is most often used for reducing dimensionality for later classification.

'The curse of dimensionality' was likely for the dataset used by W. R. Adams [17], thus LDA was utilized to prevent overfitting the training data. LDA can also be used for classification tasks. Yang S *et al.* [19] utilized a LDA variant, Fisher's Linear DA, which doesn't assume normal distributions or equal class covariances, for the classification of PD patients from healthy controls. The FLDA provides a linear classification with an accuracy of 79%, sensitivity of 0.857, specificity of 0.583, and an area of 0.83 under ROC curve from voice measurements.

QDA is also used for classification tasks, and because it can learn quadratic boundaries, it is therefore more flexible. Ortelli *et al.* [21] tested QDA against 14 other models to discriminate three different classes: NC, MCI, and IC from patient data. QDA performed the best overall and achieved micro-average ROC curve, AUC = 0.81; and AUC = 0.85 for NC, 0.67 for MCI, and 0.83 for IC.

3.9 Efficacy of NNs and SVMs in Classifying PD from Voice Data

We have established that voice data is the most popular indicator for early diagnosis of PD, and that NNs and SVMs are of the most popular models used for classification [10]. In this section, we would like to dive deeper into the interaction between these models and voice data, to further justify their popularity.

Many studies use the voice data obtained from the UCI ML repository, so we will be looking into the data from here too. There are two popularly used datasets from the repository. First, Oxford PD Telemonitoring Dataset [34]. The data consists of subject age, subject gender, time interval from baseline recruitment date, motor MDS-UPDRS, total MDS-UPDRS, and 16 biomedical voice measures aimed at extracting distinct characteristics of the speech signals from 5875 voice recordings. Thus, resulting in a 5875 × 16 feature matrix. The voice measures result in a single number; therefore, this dataset only has integer and real data types. As voice measures were taken over time, this data is of the time-series data type. However, the biomedical voice measures are closely correlated, and many correlated features serve only to complicate the model, while providing not much new information [35]. This is why feature selection methods should be applied to this dataset before model training. The authors of this dataset suggest in their paper [35] that least absolute shrinkage and selection operator (LASSO) regression, which reduces prediction error whilst reducing the number of input features, is a good choice for this.

Second, Parkinson Speech Dataset with Multiple Types of Sound Recordings [36]. The data collected belongs to 20 People with Parkinsonism (6 female, 14 male) and 20 healthy individuals (10 female, 10 male). From all subjects, 26 voice samples including sustained vowels, numbers, words, and short sentences were recorded.

Furthermore, MDS-UPDRS score and presence of PD were also recorded. In total, there were 1040 instances of data recorded, resulting in a 1040 x 29 data matrix for training.

SVMs work by projecting data into higher dimensional space and finding an optimal separating hyperplane. The optimal separating hyperplane is found by maximizing the distance between the boundaries of the closest points of the groups of data, these points are also called support vectors. The distance between the separating hyperplane and the closest support vector is called the margin. However, this alone is ineffective in creating the optimal separating hyperplane as it is greatly affected by outliers or misclassified points in the training data. Therefore, SVMs allow some misclassification of data by allowing a specific number of datapoints inside the margin [37]. This is a famous problem that plagues all ML algorithms called the bias-variance tradeoff [38]. By allowing some amount of misclassification, SVMs allow better generalization to new data. SVMs can also achieve different separating hyperplanes through mathematical functions known as kernel. Linear, polynomial, radial basis function, sigmoid, etc., are the different kernel functions used in SVM classifier [39].

NNs are composed of layers: an input layer, one or more hidden layers, and an output layer. Each layer is comprised of neurons, or nodes, that perform crucial computations. Neurons receive input data from the previous layer and apply activation functions to introduce non-linearity. This step is crucial for learning complex relationships within the data. Neurons also conduct weighted sums of the inputs, adjusting their influence through weights and adding bias terms to account for offsets. The weights and biases are adaptable and evolve during training through a technique called backpropagation. A loss or cost function quantifies the disparity between the predictions and the actual target values, and backpropagation guides the adjustment of weights and biases using optimization algorithms like stochastic gradient descent to minimize the loss function. NNs have better performance compared to traditional ML methods especially when learning from large datasets [39].

Both SVMs and NNs benefit from feature engineering techniques such as principal component analysis (PCA), LASSO or feature selection [35]. This is to help the models learn patterns more efficiently. In addition, normalization is also beneficial for the training of these models, as it improves the performance of models that utilize distance metrics like the separating hyperplane of SVMs, and it also allows for more efficient gradient descent for NNs.

Since multiple recordings of the same subject are recorded, instead of using each recording independently, it has been found that taking the mean and standard deviation of the vocal features as a summarizing representation is an efficient method for model training and provides better results [40]. This is because summarizing the voice samples of subjects decreases the effect of variations between voice samples of a subject. As for validation of the models, using conventional methods such as boot strapping or leave-one-out results in biased predictive models as some samples of an individual will be split between training and testing sets, resulting in a leak between training and testing sets [40]. In response, most studies use leave-one-subject-out validation scheme in which all the voice samples of one individual are left out for testing. This effectively circumvents the leaking problem mentioned. SVMs performed better than KNN classifiers when using summarized (taking of mean and standard deviation as representative) leave-one-out methods, producing the highest accuracy of 77.50% [40]. Furthermore, using summarized leave-one-out methods extend feature dimension whilst shortening sample space, this plays to the strengths of SVMs, hence the massive popularity of SVMs in this field [40].

In a different comparative study focusing on dimension reduction techniques such as PCA and kernel principal component analysis (KPCA) which still used the dataset from the UCI ML Repository, Deep NNs were compared to other prominent ML algorithms [41]. Deep NNs produced a second-highest accuracy of 92.91% from PCA dimension reduction and fourth-highest accuracy of 89.624% from KPCA dimension reduction, whilst reducing the features used in training through both PCA and KPCA [41].

4. Challenges in Early Diagnosis of Parkinson's Disease

4.1 Technological Challenges

Although analyzing biomarkers such as cerebrospinal fluid, blood biochemical, and neuroimaging are effective and accurate in diagnosis of PD, they are neither cheap nor scalable for large scale usage [11], [42]. There is no conclusive test currently available for diagnosis, so the identification of the disease relies solely on clinical and observational criteria [17]. Furthermore, differences between laboratory and daily-life conditions present challenges for the implementation of reliable detection systems [46]. Despite having an abundance of research and new methods proposed, there is a lack of research that validates and verifies these new studies [10]. Researchers should expand their scope of validation for new methods and models in diagnosis, to enable quicker and safer adoption in hospitals. Much of the data used for training currently collects only one indicator from each sample. However, there are multiple biomarkers for PD. This makes collection of high-quality multimodal training data difficult [43]. More effort should be directed to the collection and preparation of data used in model training.

Int. Journal of Integrated Engineering Vol. 17 No. 6 (2025) p. 13-26

4.2 Disease Challenges

PD is hard to differentiate from other diseases that display the same symptoms (atypical parkinsonian disorders), especially in the early stages [44][10]. In response to this difficulty, ML methods have been applied for differential diagnosis with promisingly high accuracy [9]. It is difficult to differentiate symptoms caused by aging and other comorbidities from those of PD, resulting in a complicated identification of a set of control and diseased subjects, as well as trouble controlling experiment variables due to unknown effects of comorbidities on PD development [43]. To mitigate the effects of comorbidities, researchers often employ a large dataset to account for the variability in the population and disease [43]. PD progresses very slowly, with a period of 20 years before the clinical phase and patients surviving up to 20 years in the clinical phase. This makes studies on patients difficult as they are prone to drop out of the studies. Moreover, assumed control subjects can contract PD later into the studies, leading to ascertainment bias [43]. Researchers can utilize at-home monitoring technologies, reducing the needs for in-person visits [43].

4.3 Patient Challenges

Patients who suffer from PD are typically elderly, have difficulty travelling, and spread geographically. Therefore, they cannot easily receive care that is concentrated in urban areas [11]. Individuals afflicted with PD typically don't realize they have PD until the later stages when more serious symptoms show. PD detection should be more widespread and common for elderly even if they do not show symptoms. A PD detection model could be implemented in a governmental health app to monitor gait or speech to detect the disease earlier. Table 4 summarizes the key challenges associated with the early diagnosis of PD, which are categorized in terms of technological, disease-related, and patient-specific barriers, along with their possible mitigating alternatives.

Table 4 *Summary of the challenges and possible solutions in early diagnosis of PD*

Category	Challenges	Possible Solutions
Technological Challenges	• Biomarker analysis is accurate but expensive and not scalable. • No conclusive test exists; diagnosis is based on clinical observation. • Differences between lab and real-world conditions hinder detection system implementation. • Lack of validation for new diagnostic methods. • Difficulty in collecting high-quality multimodal training data.	• Expand validation studies for new methods. • Improve data collection for model training.
Disease Challenges	• Hard to differentiate from atypical parkinsonian disorders. • Aging and comorbidities complicate diagnosis. • Slow progression leads to patient dropout and ascertainment bias.	• Use ML for differential diagnosis. • Employ large datasets to account for variability. • Utilize at-home monitoring to reduce in-person visits.
Patient Challenges	• Elderly patients struggle with travel and access to urban healthcare. • Many patients realize they have PD only in later stages.	• Implement widespread early screening, even for asymptomatic elderly. • Use government health apps to monitor gait or speech for early detection.

5. Conclusions

In conclusion, PD is hard to diagnose early and accurately traditionally. However, ML techniques can be used to overcome this issue. From our literature review, we have found that SVMs and NNs are especially suited to the task of PD diagnosis. Moving forward, future research in this field should focus on conducting large-scale validation studies to confirm the effectiveness of ML algorithms in real-world clinical settings. Ultimately, the application of ML in PD diagnosis holds great promise, and continued research in this area has the potential to significantly impact patient care and outcomes.

Acknowledgement

This research work is supported by Telekom Research and Development Sdn. Bhd. (RDTC/231096) and Multimedia University, Malaysia.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *It Ee Lee, Woo Jer Kee;* ***data collection:*** *Woo Jer Kee, It Ee Lee;* ***analysis and interpretation of results:*** *Muhammad Sheraz, Teong Chee Chuah, Chitra Dhawale;* ***draft manuscript preparation:*** *Woo Jer Kee, It Ee Lee, Muhammad Sheraz, Teong Chee Chuah, Chitra Dhawale. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Richens, J.G., Lee, C.M. & Johri, S. (2020). Improving the accuracy of medical diagnosis with causal machine learning. Nat Commun, 11(3923), 1-9, https://doi.org/10.1038/s41467-020-17419-7.

[2] Marras, C., Beck, J.C., Bower, J.H. et al. (2018). Prevalence of Parkinson's disease across North America. NPJ Parkinsons Dis., 4(21), 1-7, https://doi.org/10.1038/s41531-018-0058-0.

[3] Willis, A.W., Roberts, E., Beck J.C. et al. (2022). Incidence of Parkinson disease in North America. NPJ Parkinsons Dis., 8(170) 10.1038/s41531-022-00410-y.

[4] Nerius, M., Fink, A., & Doblhammer, G. (2017). Parkinson's disease in Germany: Prevalence and incidence based on health claims data. Acta Neurol Scand, 136(5), 386-392, 10.1111/ANE.12694.

[5] Senturk, Z.K. (2020). Early diagnosis of Parkinson's disease using machine learning algorithms. Med Hypotheses, 138(109603), https://doi.org/10.1016/j.mehy.2020.109603.

[6] Hoehn, M.M. & Yahr, M.D. (1967). Parkinsonism: Onset, progression and mortality. Neurology, 17(5), 10.1212/WNL.17.5.427.

[7] Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry, 79(4), 368-376, 10.1136/JNNP.2007.131045.

[8] Pagan, F.L. (2012). Improving outcomes through early diagnosis of Parkinson's disease. Am J Manag Care, 18(7 Suppl), S176-S182, https://www.ajmc.com/view/a408_12sep_pagan_s176to182.

[9] Tolosa, E., Garrido, A., Scholz, S.W., and Poewe, W. (2021). Challenges in the diagnosis of Parkinson's disease. Lancet Neurol, 20(5), 385-397, 10.1016/S1474-4422(21)00030-2.

[10] Mei, J., Desrosiers, C., & Frasnelli, J. (2021). Machine learning for the diagnosis of Parkinson's disease: A review of literature. Front Aging Neurosci, 13(633752), 10.3389/fnagi.2021.633752.

[11] Yang, Y., Yuan, Y., Zhang, G. et al. (2022). Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. Nat Med, 28, 2207–2215, https://doi.org/10.1038/s41591-022-01932-x.

[12] Wan, S., Liang, Y., Zhang, Y. et al. (2018). Deep multi-layer perceptron classifier for behavior analysis to estimate Parkinson's disease severity using smartphones. IEEE Access, 6, 36825-36833, 10.1109/ACCESS.2018.2851382.

[13] Ma, A., Lau, K.K., & Thyagarajan, D. (2020). Voice changes in Parkinson's disease: What are they telling us?. J Clin Neurosci, 72, 1-7, 10.1016/J.JOCN.2019.12.029.

[14] Skaramagkas, V., Pentari, A., Kefalopoulou, Z. et al. (2023). Multi-modal deep learning diagnosis of Parkinson's disease – A systematic review. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 2399-2423, 10.1109/TNSRE.2023.3277749.

[15] Rana, A., Dumka, A., Singh, R. et al. (2022). A computerized analysis with machine learning techniques for the diagnosis of Parkinson's disease: Past studies and future perspectives. Diagnostics, 12(11), 10.3390/DIAGNOSTICS12112708.

[16] Grandini, M., Bagli, E., & Visani, G. (2020, August 13). Metrics for multi-class classification: An overview. https://doi.org/10.48550/arXiv.2008.05756.

[17] Adams, W.R. (2017). High-accuracy detection of early Parkinson's Disease using multiple characteristics of finger movement while typing. PLoS One, 12(11), 10.1371/JOURNAL.PONE.0188226.

[18] Parisi, L., RaviChandran, N., & Manaog, M.L. (2018). Feature-driven machine learning to improve early diagnosis of Parkinson's disease. Expert Syst Appl, 110, 182-190, 10.1016/J.ESWA.2018.06.003.

[19] Yang, S., Zheng, F., Luo, X. et al. (2014). Effective dysphonia detection using feature dimension reduction and kernel density estimation for patients with Parkinson's disease. PLoS One, 9(2), 10.1371/journal.pone.0088825.

[20] Bougea, A., Efthymiopoulou E., Spanou, I. et al. (2022). A novel machine learning algorithm predicts dementia with Lewy bodies versus Parkinson's disease dementia based on clinical and neuropsychological scores. J Geriatr Psychiatry Neurol, 35(3), 317-320, 10.1177/0891988721993556.

[21] Ortelli, P., Ferrazzoli, D., Versace, V. et al. (2022). Optimization of cognitive assessment in Parkinsonisms by applying artificial intelligence to a comprehensive screening test. NPJ Parkinsons Dis., 8(42), https://doi.org/ 10.1038/s41531-022-00304-z.

[22] H. Byeon, 'Is the Random Forest Algorithm Suitable for Predicting Parkinson's Disease with Mild Cognitive Impairment out of Parkinson's Disease with Normal Cognition?', Int J Environ Res Public Health, vol. 17, no. 7, Apr. 2020, doi: 10.3390/IJERPH17072594.

[23] Scikit-Learn Developers (BSD License). 1.4. Support Vector Machines – scikit-learn 1.4.1 documentation. Accessed: 2024, March 17. [Online]. Available: https://scikit-learn.org/stable/modules/svm.html

[24] Scikit-Learn Developers (BSD License). 1.17. Neural network models (supervised) – scikit-learn 1.4.1 documentation. Accessed: 2024, March 17. [Online]. Available: https://scikit-learn.org/stable/modules/neural_networks_supervised.html

[25] Scikit-Learn Developers (BSD License). 1.11. Ensembles: Gradient boosting, random forests, bagging, voting, stacking – scikit-learn 1.4.1 documentation. Accessed: 2024, March 17. [Online]. Available: https://scikit-learn.org/stable/modules/ensemble.html

[26] Scikit-Learn Developers (BSD License). 1.6. Nearest Neighbors – scikit-learn 1.4.1 documentation. Accessed: 2024, March 17. [Online]. Available: https://scikit-learn.org/stable/modules/neighbors.html#nearest-neighbors-classification

[27] Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. Crit Care, 9(1), 112-118, 10.1186/CC3045/TABLES/7.

[28] Scikit-Learn Developers (BSD License). 1.10. Decision Trees – scikit-learn 1.4.1 documentation. Accessed: 2024, March 17. [Online]. Available: https://scikit-learn.org/stable/modules/tree.html

[29] Scikit-Learn Developers (BSD License). 1.9. Naive Bayes – scikit-learn 1.4.1 documentation. Accessed: 2024, March 17. [Online]. Available: https://scikit-learn.org/stable/modules/naive_bayes.html

[30] Scikit-Learn Developers (BSD License). 1.2. Linear and Quadratic Discriminant Analysis – scikit-learn 1.4.1 documentation. Accessed: 2024, March 17. [Online]. Available: https://scikit-learn.org/stable/modules/lda_qda.html

[31] UCI Machine Learning Repository. Parkinson speech dataset with multiple types of sound recordings. Accessed: 2023, December 27. [Online]. Available: https://archive.ics.uci.edu/dataset/301/parkinson+speech+dataset+with+multiple+types+of+sound+recordings

[32] UCI Machine Learning Repository. Parkinsons telemonitoring. Accessed: 2023, December 27. [Online]. Available: https://archive.ics.uci.edu/dataset/189/parkinsons+telemonitoring

[33] Zhang, H. (2004). The optimality of Naive Bayes. The Florida AI Research Society.

[34] UCI Machine Learning Repository. Parkinsons telemonitoring. Accessed: 2024, January 18. [Online]. Available: https://archive.ics.uci.edu/dataset/189/parkinsons+telemonitoring

[35] Tsanas, A., Little, M.A., McSharry, P.E. et al. (2010). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Trans Biomed Eng, 57(4), 884-893, 10.1109/TBME.2009.2036000.

[36] UCI Machine Learning Repository. Parkinson speech dataset with multiple types of sound recordings. Accessed: 2024, January 18. [Online]. Available: https://archive.ics.uci.edu/dataset/301/parkinson+speech+dataset+with+multiple+types+of+sound+recordings

[37] Cortes, C. & Vapnik, V. (1995). Support-vector networks. Mach Learn, 20(3), 273-297, 1995, https://doi.org/ 10.1007/BF00994018.

[38] Belkin, M., Hsu, D., Ma, S., et al. (2019). Reconciling modern machine-learning practice and the classical bias – variance trade-off. Proc Natl Acad Sci USA, 116(32), 15849-15854, https://doi.org/10.1073/pnas.1903070116.

[39] Sarker, I.H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN Comput Sci, 2(3), 1-21, https://doi.org/10.1007/s42979-021-00592-x.

[40] Sakar, B.E., Isenkul, M.E., Sakar, C.O. et al. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Inform., 17(4), 10.1109/JBHI.2013.2245674.

[41] Anand, A., Haque, M.A., Alex, J.S.R. et al. (2018). Evaluation of machine learning and deep learning algorithms combined with dimentionality reduction techniques for classification of Parkinson's disease. Proc. 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2018), 342-347, 10.1109/ ISSPIT.2018.8642776.

[42] Wagner, S.K., Romero-Bascones, D., Cortina-Borja, M. et al. (2023). Retinal optical coherence tomography features associated with incident and prevalent Parkinson disease. Neurology, 101(16), 1581-1593, 10.1212/WNL.0000000000207727.

[43] Garcia Santa Cruz, B., Husch, A., Hertel, F. (2023). Machine learning models for diagnosis and prognosis of Parkinson's disease using brain imaging: general overview, main challenges, and future directions. Front Aging Neurosci., 15, https://doi.org/10.3389/fnagi.2023.1216163.

[44] Rana, A., Dumka, A., Singh, R. et al. (2022). A computerized analysis with machine learning techniques for the diagnosis of Parkinson's disease: Past studies and future perspectives. Diagnostics, 12(11), 10.3390/diagnostics12112708.

[45] A. Landolfi et al., 'Machine Learning Approaches in Parkinson's Disease', Curr Med Chem, vol. 28, no. 32, pp. 6548–6568, Jan. 2021, doi: 10.2174/0929867328999210111211420/CITE/REFWORKS.

[46] L. Sigcha et al., 'Improvement of Performance in Freezing of Gait detection in Parkinson's Disease using Transformer networks and a single waist-worn triaxial accelerometer', Eng Appl Artif Intell, vol. 116, p. 105482, Nov. 2022, doi: 10.1016/J.ENGAPPAI.2022.105482.