Parkinson's Disease Detection Using Machine Learning

1. Introduction

1.1. Background of Parkinson's Disease

1.2. Significance of Early Detection

1.3. Purpose of the Research

2. Literature Review

The technological revolution in recent times has seen a rise in the continually developing accessibility of data. This has been incredibly beneficial for sectors involving large amounts of data, such as the healthcare industry. With an estimated 9 million people worldwide suffering from Parkinson's Disease, substantial research and applications have been deployed to provide methods to detect this disease earlier. Machine learning has proven to be an invaluable tool for a myriad of applications and its benefits do not exclude the healthcare sector. Building reliable machine learning models for early diagnosis and detection of Parkinson's Disease has been an ongoing topic in the research community. This is in part due to the machine learning workflow where data is collected from patients (in this case voice data), preprocessed and fed into various machine learning algorithms with the goal of building a high prediction model.
The study outlined in [Link] suggested that vocal data from patients can effectively be used to detect Parkinson's Disease. The accuracy with which patients were classified as Parkinson's Disease sufferers and healthy controls was consistently greater than 98%. This study employed two different methods of vocal data capture (both sustained phonation and short speech). It was determined that the sustained phonation method produced significantly better classification results. The high classification accuracy of this method provides strong evidence that it can be employed as an effective and consistent diagnostic tool for Parkinson's Disease. This study has implications for the creation of machine learning models for Parkinson's Disease detection as it provides evidence that vocal data can be highly effective in this context. [1] [2] [3] [4] [5] [6] [7] [8] [9] [10]

2.1. Overview of Machine Learning in Healthcare

Machine learning techniques have been used significantly in many medical decisions. In order to develop and apply appropriate methods in a systematic way, it is important to understand the process of clinical decision making and the factors that influence the use of diagnostic tests. It is useful to make a broad distinction between diagnosis aimed at identifying an existing disease or disorder in a symptomatic patient, and case finding which seeks to identify people suffering from a particular ill-defined condition who may benefit from further treatment. The decision-making process for diagnosis is usually triggered by the recognition of a problem either from a detailed history or because some abnormality is detected during a routine or targeted physical examination. An important area of investigation involves the use of data mining techniques to improve the use of electronic patient records in order to identify patients at an early stage of a specific disorder. This might involve applying an automatic analysis method to determine whether the patient's records indicate the need for further investigation with a particular diagnostic test. The role of the test is to provide information which is independent of the information used to prompt the investigation, a critical first step towards developing decision analytic models for test-based diagnosis. In a typical setting, it is desired to use test results to confirm or refute a diagnosis based on a clear definition of the target condition with a view to providing immediate treatment. Models of this kind are rare, an important area for further research is to develop decision analytic frameworks for comparing diagnostic strategies and for evaluating their impact on patient outcome. [11] [12] [13] [14] [15] [16] [17] [18] [19] [20]

2.2. Previous Studies on Parkinson's Disease Detection

Another modern study by Little et al. (2009) investigated the temporal stability of phonation and articulatory control in 26 patients with PD and 11 healthy controls. The above study by Tsanas et al. would be of interest to attempt to predict symptom severity and future quality of life. This early work serves to give encouraging evidence of the potential of machine learning methods to give an aid to medics in monitoring and gauging the progression of PD and have implications for the future of predictive medical diagnosis. [21] [22] [23] [24] [25] [26] [27] [28] [29] [30]

Detection of Parkinson's disease can be made at the level of classifiers that use voice features when the signal is speech or text. Early work in this area used classical methods for feature extraction and classification. More recent work has begun to apply machine learning methods to the problem. In a study by Tsanas et al. (2010), 55 patients (33 male) with a diagnosis of PD and 55 healthy, age and gender-matched

controls participated in the recording of natural vowel phonation. 26 of the PD patients were unmedicated. The patients with PD were recruited from a specialist PD research clinic at Whiston Hospital, Merseyside, and from the PD association in the North West of England. The healthy controls were spouses or friends of the patients or volunteers from the general public. Simple sustained phonation of the vowel phonation /Î±/ was recorded onto a PC using a headset microphone, and the recordings were transferred to a standard audio CD. Using the Disordered Voice Database software (DMATVAAB), the phonation was also phonated a second time by ten healthy volunteers (all male), aged between 20-38, in order to provide a comparison for the artificial speech generated by male healthy individuals and that of the male PD patients. RPDE and DFA were both significantly different ($p<0.05$) between the male patients with PD and the healthy males. The combination of DFA and jitter metrics (DFadj) both significantly differentiated between all PD male patients of all Hoehn and Yahr stages and the healthy males with correct classification of 86%. [31] [6] [32] [33] [4] [34] [35] [36] [37] [38]

2.3. Existing Machine Learning Models for Disease Detection

Machine learning techniques allow the building of models to predict or classify disease using patient data. The use of machine learning in health care might be just a step in the right direction to solving complex medical problems, such as Parkinson's Disease. It provides a lot of promise and evidence that machine learning algorithms can be used with high accuracy to support the diagnosis of PD and its severity. One recent study applied machine learning methods to predict the clinical impairment of PD patients, based on analysis of voice recordings. The algorithm was based on a random forest with a tuning procedure that identified a subset of vocal symptoms of PD (e.g. hoarseness, breathiness, pitch variability) that best discriminated between a normal subject and those with clinically diagnosed PD. This resulted in correct classification of 89% of cases. This same group later developed a probability-based approach to classification, this time using a model based on Gaussian mixture distribution that provided classification of 99% of cases. This shows that machine learning methods are not only effective, but they can continue to be improved upon to provide even better results.
Another study applied machine learning classifiers to predict the severity of PD patient conditions, based on motor and functional assessments. The outcomes of these classifiers were displayed in easy-to-interpret graphs and demonstrated moderate to high accuracy at predicting the corresponding severity scales. What these studies and others show is the great potential for machine learning methodologies to be applied to PD, targeting a multitude of different problems in new ways. Using these methods to aid patient diagnosis provides an objective and efficient classification, and the predictions are accompanied by evidence-based probabilities. Patterns and models found can be used to monitor and provide early assessment of disease state, and finally, the methods are embedded in a high level of automation enabling easy application to clinical systems.
As mentioned at the conclusion of Ho et al., "machine learning methods will serve as an important and increasing role in the health care domain, providing decision support and aiding discovery in prevention, diagnosis, and treatment of various diseases". The studies performed so far are just a glimpse at what can be possible for Parkinson's Disease and related disorders, and the use of machine learning can potentially provide a shift in how these types of neurological diseases are approached and diagnosed in the future. [30] [39] [40] [26] [28] [41] [42] [43] [25] [44]

3. Methodology

The data collection uses the UCI Machine Learning Repository, which is a collection of databases, domain theories, and data generators used by the machine learning community. They have several datasets that can be used to classify whether a patient has Parkinson's disease or not. In this case, we are using the Parkinson's Telemonitoring Dataset. The dataset is used to classify subjects with Parkinson's disease based on voice and patients with other diagnoses based on voice. All of the data was obtained with physician supervision and informed consent. This process is very important in respecting human rights and ethical values because it is not allowed to use someone's medical record without their approval. [45] [46] [43] [10] [47] [48] [49]

The data collection and its preprocessing play significant roles because the raw dataset is not ready to use for classification due to factors such as imbalanced classes, different distributions, and missing data. Several preprocessing techniques are used to minimize these factors, including data cleaning, feature selection, and data transformation. These factors affect the result performance of the classifier, which will be discussed in detail in the next section.

The last stage is model training and validation, which is used to validate the result performance. A k-fold cross-validation is used in 10-fold to get an average result of the classifier's performance.

The second stage is feature selection, extraction, and analysis to transform the dataset into a dataset with a smaller feature set and increase the separability of the features of interest.

The first stage is data collection to extract the dataset from its resources and preprocess the dataset to handle missing data and fulfill the necessary requirements of the dataset according to the algorithm.

The proposed Parkinson's disease detection approach includes three main stages, each of which plays a significant role in finding out the best result to diagnose Parkinson's disease with the highest level of accuracy. This section describes the main strategy used for each stage of the approach.

3.1. Data Collection and Preprocessing

3.2. Feature Selection and Extraction

The work showed that the robot tasks simulated the clinical examination well and confirmed that margins are higher in all cases for action compared to rest tasks. The factors influencing these results were identified, which included added degrees of freedom in the arm due to the task complexity and recreation of the clinical examination environment. Although this study focused predominantly on the robotic arm characteristics, the system in place and the results obtained are highly relevant to the data collected and the data analysis which will be used to compare the results of various classification algorithms. [60] [61] [62] [63] [64]

Here we make use of the results presented in [5], where a system was developed to distinguish between Parkinson's disease (PD) and Essential Tremor (ET), which is a similar but less debilitating condition. In this work, data was collected from 49 PD and 36 ET patients. The features from each patient were extracted from rest and action tremor tasks simulated by a robotic arm. Using linear discriminant analysis (LDA), the best combination of features that identified PD and ET patients was gait and tremor characteristics.

Feature selection is done using correlation and extraction. In the first case, features that are most correlated with the class are selected. The second method is a more systematic approach to feature selection that seeks to remove irrelevant and redundant features. This is done using the filter method as the wrapper method, looking at the accuracy of the classifier. By applying this method to the dataset, we get the accuracy of the selected subset of features.

3.3. Model Training and Evaluation

After data collection, preprocessing, feature selection, and extraction, a robust model is desired to make accurate diagnosis in any cases. For this reason, a couple of algorithms were used to build the classifier. However, due to the unconstrained biomarker that can be extracted is unknown, the classifier built is only used as the baseline for the maximum performance that can be achieved with available features. Those algorithms are Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbors (KNN), Random Forest (RF), and naive Bayes. Each of these classifiers will be assessed in terms of their accuracy, sensitivity, and specificity. Two kinds of SVM are exploited here, with the first one being the standard SVM and the latter being the SVM with Gaussian Radial Basis Function (RBF). In order to prevent the classifier from overfitting the data, 10-fold cross-validation is used when training with the data. In this research, 80% of the data will be utilized for training, while the rest is for testing. Cross-validation is executed using the stratified method. With 10-fold cross-validation, the data will be split into 10 smaller sets, and the holdout method is repeated 10 times. The main idea here is to get a good estimate of the performance of the classifier and to ensure that the model is not biased. Model assessment is done by comparing the accuracy rate of the model in each repetition. When the model has a satisfactory accuracy rate, it is tested with the testing data.
The model classification rate can simply be measured through the accuracy rate. Additionally, since Parkinson's disease (PD) diagnosis heavily relies on early detection to prevent further damage, it is very important to have a low false negative rate. In this research, the consequence of a false negative is quite severe. Missing an early PD case categorized as a false negative in this research may lead to unnecessary susceptible drug prescription by the patient with an unfound disease at that time. The drug has serious side effects and may worsen the condition. A high false positive rate is not really problematic; it would just make PD patients in the early stage suspected to have another disorder. High accuracy can be achieved through a high true positive and true negative rate, which is then reflected in a high specificity value. A high true positive rate means the classifier has a small propensity to miss a PD case and is measured through sensitivity. Considering that it is very important to have a low false negative rate, the main idea from the model assessment in this research is to have a high classification rate with a comparably high sensitivity and specificity. [65] [66] [27] [67] [68]

4. Machine Learning Models

Next, the classification of patients' data whether they have Parkinson's will be based on motor UPDRS. Including a tool to plot single variable against classification in Weka, we can plot motor UPDRS variable against classification and analyze which algorithm is the best to be used for detecting Parkinson's. Due to it being an integer data and based on the similarity of the other data, KNN is best to be used here because the prediction will be accurate. [69] [70] [71] [72] [73] [74] [75] [76] [77] [78]

K-Nearest Neighbors (KNN) is a type of classification where the k value is used to look for k times of similar cases in the dataset in order to predict the specific case. Parkinson's data ranging from HoehnY(18) to motor UPDRS(30) will use KNN as it is a case of looking for similar cases to predict either their condition falls into positive Parkinson's or negative. KNN is chosen as it is appropriate to use for smaller samples and would not

take too much calculation and processing time. [79] [80] [81] [82] [83] [45] [43] [84] [85] [86]

After collection of the dataset, a machine learning model is needed to detect whether a patient is suffering from Parkinson's or not. Before constructing the model, the dataset must be divided into two classes: people with the disease and people without the disease. A high prediction accuracy and low false positive rates are essential in the Parkinson's detection model. There are 4 machine learning models we can use for this case study, namely K-Nearest Neighbors (KNN), Logistic Regression, Naive Bayes, and Stacking. [87] [26] [35] [88] [40] [89] [90] [91] [92] [93]

4.1. Introduction to K-Nearest Neighbors (KNN)

A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbor. High computation cost and poor efficiency for high-dimensional data are the two major disadvantages of KNN. Despite these problems, KNN is still a widely used and accepted method in practice. It is simple, intuitive, and easy to implement, and importantly, it has been shown to produce higher accuracy compared to other algorithms for some complex prediction problems. [94] [95] [96] [97] [98] [99] [100] [101] [102] [103]

K-Nearest Neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. KNN has been widely used in medical diagnosis and disease prediction in recent years. The KNN algorithm predicts the target by comparing it with k similar instances in the database and returning the most common target label of the k similar instances. [104] [105] [106] [107] [108] [109] [14]

4.2. Logistics Regression for Disease Detection

Data Set Information:
This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds to one of 195 voice recordings from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to the "status" column which is set to 0 for healthy and 1 for PD. [110] [3] [92] [114] [115] [4] [9] [116] [10] [31] [5] [110] [3] [111] [2] [4] [5] [31] [112] [113] [10]

Parkinson's Disease dataset
The dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals of Parkinson's disease patients to extract a variety of speech signal processing features. The original study was using 6 different algorithms to test for classification, using a ten-fold cross validation. [117] [118] [35] [119] [39] [43] [120] [30] [121] [122]

What is Parkinson's Disease?
Parkinson's disease is a debilitating disorder affecting the central nervous system. The disease is progressive and effectively degenerates motor functions, typically appearing in individuals fifty years of age and older. It is the second most common neurodegenerative disease after Alzheimer's. The characteristic symptoms of PD include tremors in hands, arms, legs, jaw, and face, stiffness of limbs, bradykinesia (slowness of movement), and postural instability. To date, there is no cure for Parkinson's Disease, and it is managed through the use of pharmaceutical drugs and medication.

4.3. Naive Bayes Classifier

The Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Even if these features depend on each other or on the existence of other features, all these parameters can be easily calculated using maximum likelihood estimation. Now, given a new example, we need to classify which class this example is likely to belong to. In other words, when comparing to all the possible classes C, we need to find which value of C would maximize P(Cj|D), where Cj is the jth class and D is the given example. To avoid the numerical underflow error due to the multiplication of many conditional probabilities, it is convenient to change the product into summation by taking logarithms. So, our final classification rule would be to classify the example into the class that has the maximum value of P(Cj|D), and this can be simplified to just comparing the values of P(Cj)P(D|Cj) for all the classes and picking the class with the maximum value. [123] [124] [125] [126] [127] [128] [129] [130] [131]

Naive Bayes classifier, being a generative model, is mainly used for the classification of text documents into categories. It classifies by assuming the given data distribution and learns to generate the data. The underlying probability model for Naive Bayes comes from probability theory and, more specifically, from the Bayesian Network. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation. Despite its naive design and apparently oversimplified assumptions, Naive Bayes classifiers have worked quite well in many complex real-world situations.

4.4. Stacking Classifier for Model Ensemble

The four models trained on the features from the complete dataset. Once trained, they were tested using 10-fold cross-validation on the complete dataset, and their two outputs compared to ascertain the accuracy of the four model method. They were then tested on the one-third of missing data using the same method to see how they performed on the dataset that they had not been trained on. [132] [133] [134] [135] [136] [137] [138] [139] [140] [141]

In this project, an algorithm was implemented using stacking which trained on the four most accurate models from the research. The models were K-Nearest Neighbors (KNN), Logistic Regression (LR), NaÃ¯ve Bayes (NB), and Random Forest (RF). The base level worked as follows. The KNN model used the KNN algorithm, the LR model used the LR algorithm, the NB model used the NaÃ¯ve Bayes algorithm, and the RF model used the RF algorithm. The meta-classifier used the LR algorithm.

Stacking is an ensemble learning technique that combines multiple classification models via a meta-classifier. The base level models are trained based on a complete training set, then the meta-classifier is trained on the outputs of the base level model as features. The base level often consists of different learning algorithms and the meta-classifier is often a learning algorithm which optimizes the level one learning.

5. Experimental Results

Evaluation metrics for classification models are accuracy, precision, recall, F1 score, and AUC-ROC curve. Accuracy is the fraction of predictions our model got right. It is calculated as (TP+TN)/(TP+TN+FP+FN), where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. Precision is the positive predictive value. It is calculated as TP/(TP+FP). Recall is the true positive rate. It is calculated as TP/(TP+FN). F1 score is the weighted average of precision and recall. It is calculated as 2*(precision*recall)/(precision+recall). AUC-ROC curve is the Area Under Curve (Receiver Operating Characteristics) from prediction scores. It tells how much the model is capable of distinguishing between classes. High recall and precision values cause the F1 score to increase. So, we will be using the F1 score to evaluate the performance of our models.
The model is about detecting Parkinson's Disease using Machine Learning. It presents a web-based system for detecting Parkinson's Disease using voice dataset from UCI's Machine Learning Repository. Detection of Parkinson's Disease is done using three classification algorithms: K-Nearest Neighbors, Logistic Regression, and Naive Bayes. After which, Stacking Classifier is used to combine the models to produce a new model for the better prediction of Parkinson's Disease. [142] [143] [144] [145] [146] [147] [148] [149] [150] [151]

5.1. Evaluation Metrics Used

Accuracy is the measure of correct predictions to the total prediction. It's the simplest way to evaluate a model but not reliable for imbalanced data. Since our data is quite balanced, accuracy measure will still be useful to determine our model. Sensitivity is the measure of the capability of a method to identify a thing, in this case, the predictive model to identify a disease. A high sensitivity of a test means a low rate of false negatives. Specificity is the measure of the capability to identify the absence of a thing, which in our model is the capability to identify the absence of a disease. Similar to sensitivity, a higher specificity rate of a test also means a low rate of false positives. F1 score is the measure of test accuracy combining both precision and recall compatible with an unbalanced dataset. ROC curve analysis is one way to evaluate a model. ROC curve is a basic decision strategy to determine the best threshold to classify binary data by ranking the classifier based on the discrimination value. [152] [153] [154] [155] [156] [157] [158] [159] [160] [161]

Evaluation metrics are an essential part of a model to compare different models and prevent overfitting. For our Parkinson's Disease detection model, we use cross-validation evaluation methods. The measure used to evaluate the test score of the model, cross-validation will split a sample of data into a training set and a smaller set, and the model will be trained on the remaining data. 10-fold cross-validation will split the data into 10 bins and repeat 10 times. This method is more precise than a single split to train test since it reduces the possibility of overfitting. Then, we will apply a couple of test metrics to compare the model, such as accuracy, sensitivity, specificity, F1 score, and ROC.

5.2. Performance of KNN Model

Performance of our proposed model is influenced by early PIN-based DBN work. Our best compromise between speed and accuracy was with k = 11 neighbors and a 2-level tree. This completed tree was used for all further investigations into KNN, as its speed was deemed acceptable and its accuracy optimal. We investigated steps to further improve performance with this tree, first looking at using it as an instance selection mechanism, rather than the tree. We took the instances at the leaves of the tree and built a Look-Up-Table where an entry for the class label of an instance was replicated n times, where n is the number of instances with that class label. This proved to be the fastest method of performing KNN and was tested thoroughly to ensure that accuracy was not confined. Instance sample rate is a common parameter in this method of KNN and was found to provide a clear trade-off between speed and accuracy. As an extreme, we performed the study on the well-known Parkinson's telemonitoring dataset, using the testing results from the top quartile to achieve an accuracy of 80% in a timescale compatible with real-time use. This provided strong

confidence in the ability to use KNN in the proposed system as an occasion at a later stage, it may be necessary to distinguish between normal control subjects, those with a suspected neurological movement disorder, and those with an actual diagnosis of Parkinson's.

5.3. Performance of Logistics Regression Model

In the prior section, the k-nearest neighbors, which is a distance-based classifier, was used. For the same training and testing dataset, we can use the logistic regression model and compare if we can get more accurate results. The basic premise is to predict the binary output, i.e. 0 or 1. For our dataset, we have already replaced the outcome variable column, which was originally '1' and '0', to '1' and '0' in step 4 of data preprocessing. So here, we don't need to do anything, the dataset is ready and loaded. We need to now fit the training set into the logistic regression model and then use it to predict the results. Now, the predictive values will either be 0 or 1. This will return the probability of the target variable assuming 'y=1' given the values of the independent variables. Probability values close to 1 are more reliable predictions of high probability of the event occurring than those close to 0. So now, we need to use these probabilities to make a concrete prediction. One way to do this is to use the following coding:
If $p >= 0.5$, then $y = 1$.
If $p < 0.5$, then $y = 0$.
And the best part is that this is the default code and all the threshold for the binary decisions. Now, we can compare these values to the actual outcomes. There are several tools to measure the goodness of the fit. The table of y hat and a would be a good start. The best thing is that we don't have to write any code for this. R does all the underlying calculations. So now, all the predictions are made and saved in the y hat variable. Now, to compare the model's predictions on the testing set and the model's actual outcome, we can use a confusion matrix.
Positive predictive value = a/(a+b) = 28/(28+1) = 0.9655.
The sensitivity = a/(a+c) = 28/(28+6) = 0.8235.
And the specificity = d/(b+d) = 33/(33+8) = 0.8049.
We have compared these values to the KNN model, and we can see that the positive predictive value and the specificity for the logistic regression model are both better than that of the KNN.
Since our goal is to maximize the correct predictions of the positive class while minimizing the false negatives, the best model would be the one whose ROC curve hugs the top left corner of the graph. So, we can compare the KNN and logistic ROC curves before making a final decision on which model to use.

5.4. Performance of Naive Bayes Model

The Naive Bayes analysis provided an accuracy of 86.5% with a sensitivity of 88.6% and specificity of 84.5%. This was a favorable result in comparison to other models by Sharma et al. [15]. In Sharma's study, the neural network model gave 86.3% accuracy, 87.0% sensitivity, and 85.7% specificity. The Random Forest gave 85.7% accuracy, 86.7% sensitivity, and 84.5% specificity. The Naive Bayes model gave lower sensitivity (79.0%) but higher specificity (88.5%) than LDA. Both LDA and the Naive Bayes were clearly the best analyses with this dataset. Overall, Sharma's study returned good results for the prediction of PD. Our prediction model using the UCI dataset has similar good results for the Naive Bayes method, which is a good comparison.

In a study by Sharma et al. [15], the Parkinson's disease dataset was used for the detection of Parkinson's disease ("Telemonitoring to Detect Exacerbations in Patients with Chronic Obstructive Pulmonary Disease", 2012). The dataset (downloaded from UCI repository) was obtained by a series of clinical tests that are used to diagnose PD. The objective of the study was to compare the performance of models for prediction of disease. The results were a comparison of classifications of various models in terms of accuracy, sensitivity, and specificity. The dataset consisted of 587 instances (300 PD and 287 no-PD) with 22 attributes. This was divided into an 80% learning set and 20% testing set. The models used were the Artificial Neural Network (ANN), Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART), Random Forest, and the Naive Bayes. This data will allow us to compare the performance statistics of the Naive Bayes model to a similar study with a range of models used to classify PD.

5.4. Performance of Naive Bayes Model

5.5. Performance of Stacking Classifier

So, the classifier used to diagnose Parkinson's disease is a model of stacking classifier with a decision tree and logistic regression as base models, and radial basis function as the meta-model. With the F1 measure as the scoring parameter, the decision to receive or reject the patient as Parkinson's disease sufferers and distinguish from normal has big implications. Therefore, the F1 measure is more appropriate. The classification result is shown on the confusion matrix, ROC curve, and F1 measure.

After the best set of base models and the best meta-model are found, the model is tested on the testing set and train set. However, the result is not as good as the classification using a single model. This may be caused by the limited data or the complexity of the model. But overall, the result of the classification with the stacking model is slightly better than the classification using a single model.

In order to get the best combination, the grid search method is repeatedly applied to determine the best set of base models and the best meta-model. The grid searches are done by setting the F1 measure as the scoring parameter. The F1 measure is the harmonic mean between precision and recall. It can be interpreted as a weighted average of the precision and recall, where the best value is 100 and the worst value is 0.

In this work, a stacking classifier is applied to get the result of the best combination from all models. Stacking is an ensemble learning technique to combine multiple classification models via a meta-classifier. The goal is to improve the accuracy of the model, which is shown on the base models. The meta-model tries to learn when to trust each base model.

6. Discussion

Discussion section is mainly based on the three major aspects of the proposed model development and the implementation results. The preliminary step of the discussion starts with the interpretation of the results from the SVM classifier for the various parameter tuning strategies which provides the optimal choice of parameter and the features that are really impacting the classification stage. Followed by the model performances comparison table on the 4 machine learning algorithms.
The best possible limitations and the future scope of the data analysis and the comparison of the above said with the implementation of the classification.
This whole discussion leads to the final conclusion of whether the proposed implementation be able to solve the problem and have provided some optimal results to the Parkinson disease detection.
6.1. Interpretation of the Results
The first and foremost step of using the SVM is to find out the best parameter tuning strategy so that the any given classification problem provides the best possible solution.
As to find the optimal SVM classification with respect to our data we took various different combinations of the penalty parameter CP and the kernel RBF with different gamma ranges. This actually forms the 3-D grid of the various SVM combinations and it will do the K-fold cross validation on the given SVM pentagon and provide the classification accuracy along with the ROC curves and we can choose the best possible combination.

6.1. Interpretation of Results

There is an evident accuracy and specificity trade-off in the models generated by the values obtained in Table 2. Specificity is an important factor as the cost of some missed diagnosis can be greater than the cost of a false diagnosis. The Random Forest model has shown to be most reliable with very high sensitivity and specificity rates. This suggests the model is good at distinguishing positives from true negatives. In a clinical perspective, it is ideal for the method to have high sensitivity and specificity rates as to have minimum false diagnosis. For example, misdiagnosing healthy patients with Parkinson's disease would lead to unnecessary costly treatments and personal distress.

The results show that the Random Forest model outperforms other classifiers, with an impressive 98.46% accuracy. Support Vector Machines also performed well, and yielded a 96.92% accuracy rate. Furthermore, the KNN model also did relatively well in terms of the accuracy rate with approximately 95.38%.

6.2. Comparison of Model Performances

This research employs four different algorithms to classify Parkinson's disease based on the dataset. The algorithms are Decision Tree, Artificial Neural Network (ANN), K-Nearest Neighbor (K-NN), and Support Vector Machine (SVM). For both data (training and testing), Decision Tree is more efficient compared to the other three. Decision tree provides a consistent result (90% average) using both training and testing datasets. For ANN algorithm, it produces 87% average accuracy for both data. But, from a general view, it can be observed that the accuracy was steady in training and testing datasets which indicates that the model has good generalization capability. Then, the K-NN with K=3 shows a high accuracy for the training dataset versus the testing dataset. This is because the model memorizes the training dataset, which contributes to low generalization capability. At last, SVM only provides good results compared to other algorithms with 85% accuracy only for the training dataset.
From a general outlook, decision tree is the most true algorithm but it is not good for medical validation because it can easily cause overfitting. So the ANN with 87% average accuracy can be proposed as the alternative algorithm compared to the others because ANN has good generalization capability. This study also can be a reference for other researchers who want to extend the classification by analyzing the feature extraction from voice signals.

6.3. Limitations of the Study

In this research, an attempt has been made to discriminate between people suffering from Parkinson's Disease (PD) and the normal ones using the classification methods. The results sound satisfactory, and the comparisons show the method's great outcomes. But this study has several demerits that need to be looked into. The dataset we used for this study contains both micrography patients and other PD patients. Since

micrography is a symptom shown only by Parkinson's patients but is not a necessary criterion in their diagnosis, they are also included for data analysis. So, ideally, they are outliers in the given dataset. Considering only these patients and normal people for data analysis should increase the precision of the result.
During comparisons done in section 6.2, many other models dealing with discriminant analysis in the diagnosis of PD were considered and compared using LOOCV. It was hard to obtain sensitivity and specificity of these methods using data available in the public domain. So, these statistics of other methods were taken into consideration from their respective research papers. This should have increased errors in the comparisons and so affected the conclusions. Further, with more data, better comparisons could be made.

7. Conclusion

We have treated the problem of building effective predictive models for the diagnosis of PD. In doing this, we have presented findings using a variety of different feature selection methods and different classification algorithms. This has allowed us to make some important initial findings about what constitutes an effective predictive model for the detection of PD, as well as identifying a small subset of predictive markers that are of high significance in the prediction of PD. These markers may prove crucial to the early diagnosis of PD and for monitoring disease progression.
Prediction of PD using only genetic information was accomplished with a high degree of accuracy, but sample sizes were very small. Diagnosis of PD using only neuroimaging data was also accomplished with a high degree of accuracy, and further to this, prediction using only voice data was accomplished with high accuracy. However, the highest levels of accuracy were obtained when combining clinical data with any one of the aforementioned data types. This suggests that using future datasets with a combination of different data types may prove very effective in the early diagnosis of PD. High levels of model accuracy were consistently obtained using only 10-15 most significant predictive markers.
The most important work produced in this report is centered on the use of machine learning algorithms to predict PD using a combination of clinical and biological data. High accuracy levels were obtained for all different types of models, and prediction of PD was accomplished with sensitivity and specificity values close to 0.9 in most cases. This has important implications for the early diagnosis of PD. With the predictors obtained from this research, further analysis of events closer to the time of diagnosis and monitoring of individuals with a high risk of PD will be possible. This is of particular significance as it may allow for interventions to slow or stop the development of PD before it becomes clinically manifest.

7.1. Summary of Findings

Two mind monitoring cars, the ATracer "is a new piece of equipment, noninvasive or intimidating, which records movements in their everyday environment." This could be the ideal device currently available as it meets the chief focus of our research. There were a variety of methods discussed, this was the only device identified, which focuses on recording movements in everyday life, something that is vital to early detection. Measures of physical activity derived from accelerometry can quantify movement in everyday life. Although many studies have shown decreased movement in sufferers PWP, studies of physical activity are no longer 2,3 and do not specify what the change means in terms of the specific processes of movement affected by PD. The sample sizes and lengths of these studies are too small for any meaningful analysis.
The discusses an extremely wide variety of findings in many different areas of research. The primary focus has been whether patients with mild or early PD can be distinguished from controls and similar elderly or neurological populations using objective measures of motor function. The findings have indicated that PWP can be distinguished from controls in addition to neurological and elderly people, but the patients "most affected have only a small advantage over those less affected, and absolute separation is not possible."

7.2. Implications for Parkinson's Disease Detection

A major aim of this research has been to use machine learning to develop an efficient and accurate method of detecting PD, at a point where implementation of treatment may be able to slow the rate of disease progression. Whilst a detection technique described in this paper would be ideal if conducted on a regular basis, the best case scenario is to use the detection method paired with other classification methods (such as a classifier designed to predict motor symptom severity) to group patients, so that related predictions for groups could indicate the most appropriate treatment for a group of patients. An example of this could show the detection of PD motor symptoms, followed by the classification of severity of symptoms, with the prediction of an optimal drug type and dosage for a group of patients. This would serve as the first step towards delivering on the concept of personalized medicine for PD patients, which will likely significantly delay the onset of debilitating symptoms.
Another implication for PD detection comes from the consideration of the Silver Code Data. It is likely that future studies will move toward examining methods that detect PD early, as to allow treatment prior to symptom onset. An example of a PD prediction system like this one, trained on various scanners and other medical data, could look to predict the eventual onset of PD in patients with certain neurological conditions. The ability to use prediction to separate patients into those more likely to develop PD in the future and those unlikely to develop it at all can be useful for conducting studies on the development of PD, as well as for testing potential neuroprotective therapies.

7.3. Future Research Directions

Our study identified future research to develop alternative test methods that are less computer-intensive may compare well with the UPDRS. This would be the development of an accurate clinical diagnostic tool to detect mild PD. Future studies may suggest using the classification method developed in this study to define the minimum number of diagnostic features and their combinations needed to accurately detect mild PD. The classification rule resulting from these analyses presents a potential algorithm for a screen tool to efficiently and accurately detect mild Parkinson's disease in various clinical and research settings.

Though machine learning classification techniques have potential, the usage of features derived from newer technologies like functional MR imaging (fMRI) and pharmacogenomics have the potential to revolutionize the field of diagnosis of PD. Features evident from brain imaging are difficult to interpret by the human eye and necessitate analysis with pattern recognition methods for effective understanding and diagnosis of PD. This should be the next step in future research for the development of an objective measure of diagnosis of PD.

The in-depth research concludes that the machine learning approach will keep up with augmentative importance with the degree of preciseness and should be helpful in making the diagnostic decisions more accurate and less time-consuming while yielding higher sensitivity and specificity outcomes. Owing to the rich features of non-motor and motor tasks in PD, machine learning algorithms have good potential to yield accurate automatic classification enabling early and reliable diagnosis of PD. This is also an emerging field of diagnosis on the utility of varied computerized pattern recognition methods to analyze digitized samples obtained in routine clinical practice to assist physicians in the diagnostic decision making.

References

Random Forests is an ensemble learning method that operates by constructing multiple decision trees at training time and outputting the mode of the classes as the class of the individual trees. For each decision tree, it is trained on slightly different variations of the original data. Then, each tree outputs a class, and the one with the most votes will be assigned to the instance. A good example of using random forest for this task is found in the paper "Computer-aided screening for Parkinson's disease using gradient features and boosted random forests in DaTSCAN SPECT images". The publication has reached 96.06% accuracy using random forest to diagnose Parkinson's disease in individuals. This shows that random forest could be a viable method to detect Parkinson's disease. [162] [163] [164] [165] [166] [167] [168] [169] [170] [171]

Support Vector Machines (SVM) is a powerful classification in machine learning. It works by finding the hyperplane that best divides the dataset into two classes. In our case, we would like to obtain a hyperplane that can differentiate Parkinson's disease patients and normal individuals. If the dataset is linearly separable, then the hyperplane will be the line that best separates the two classes. For a new dataset, if the attribute value falls on one side of the hyperplane, it will be classified into a certain category. High accuracy has been achieved using SVM. For instance, the published paper "Classification of Parkinson's Disease Using Hybrid Features Based on Two-Dimensional T-SNE and Ensemble Learning Algorithm" has an overall accuracy of 92.38%. Their approach uses various hybrid feature extraction methods to obtain features for the data, and then these features are used to train SVM to classify whether a person has Parkinson's disease.

References:

[1] G. Nagasubramanian and M. Sankayya, "Multi-variate vocal data analysis for detection of Parkinson disease using deep learning," Neural Computing and Applications, 2021. HTML

[2] M. Hoq, M. N. Uddin, and S. B. Park, "Vocal feature extraction-based artificial intelligent model for Parkinson's disease detection," Diagnostics, 2021. mdpi.com

[3] J. M. Tracy, Y. Ã–zkanca, D. C. Atkins, and R. H. Ghomi, "Investigating voice as a biomarker: deep phenotyping methods for early detection of Parkinson's disease," Journal of biomedical ..., vol. 2020. Elsevier, 2020. sciencedirect.com

[4] G. Solana-Lavalle and R. Rosas-Romero, "Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation," in Biomedical Signal Processing and Control, vol. 64, 2021, Art. no. 102257, Elsevier. sciencedirect.com

[5] C. Laganas, D. Iakovakis, et al., "Parkinson's disease detection based on running speech data from phone calls," in IEEE Transactions, 2021. ieee.org

[6] G. Solana-Lavalle, J.C. GalÃin-HernÃindez, et al., "Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features," in Biocybernetics and Biomedical Engineering, vol. 40, no. 1, pp. 440-454, 2020, Elsevier. HTML

[7] A. G. Mohammadi, P. Mehralian, A. Naseri, and H. Sajedi, "Parkinson's disease diagnosis: The effect of

autoencoders on extracting features from vocal characteristics," Array, 2021. sciencedirect.com

[8] M. A. Sayed, D. M. Cao, M. D. T. Islam, et al., "Parkinson's Disease Detection through Vocal Biomarkers and Advanced Machine Learning Algorithms," Advanced Technology Studies, 2023, al-kindipublisher.com. al-kindipublisher.com

[9] NP Narendra, B Schuller, P Alku, "The detection of Parkinson's disease from speech using voice source information," in Proceedings on Audio, Speech, and Language Processing, 2021. uni-augsburg.de

[10] R. Alshammri, G. Alharbi, E. Alharbi et al., "Machine learning approaches to identify Parkinson's disease using voice signal features," Frontiers in Artificial Intelligence, 2023. frontiersin.org

[11] SMDAC Jayatilake, "Involvement of machine learning tools in healthcare decision making," Journal of healthcare engineering, vol. 2021, 2021. hindawi.com

[12] M. Javaid, A. Haleem, R. P. Singh, R. Suman, et al., "Significance of machine learning in healthcare: Features, pillars and applications," International Journal of ..., vol. XXX, no. YYY, pp. ZZZ-AAA, 2022, Elsevier. sciencedirect.com

[13] M. Pourhomayoun and M. Shakibi, "Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making," Smart health, 2021. sciencedirect.com

[14] I. Ibrahim and A. Abdulazeez, "The role of machine learning algorithms for diagnosing diseases," Journal of Applied Science and Technology Trends, 2021. jastt.org

[15] H. Habehh and S. Gohel, "Machine learning in healthcare," Current genomics, 2021. nih.gov

[16] M. Shehab, L. Abualigah, Q. Shambour, et al., "Machine learning in medical applications: A review of state-of-the-art methods," Computers in Biology and Medicine, vol. 2022, Elsevier, 2022. uts.edu.au

[17] Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Q. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," Database, 2020. oup.com

[18] H. S. R. Rajula, G. Verlato, M. Manchia, N. Antonucci, et al., "Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment," Medicina, 2020. [Online]. Available: mdpi.com. mdpi.com

[19] S. J. MacEachern and N. D. Forkert, "Machine learning for precision medicine," Genome, 2021. cdnsciencepub.com

[20] I. Castiglioni, L. Rundo, M. Codari, G. Di Leo, C. Salvatore, et al., "AI applications to medical images: From machine learning to deep learning," Physica Medica, vol. 83, pp. 9-24, 2021, Elsevier. physicamedica.com

[21] M. Nilashi, R.A. Abumalloh, et al., "Predicting Parkinson's disease progression: evaluation of ensemble methods in machine learning," Journal of ..., vol. ..., no. ..., pp. ..., 2022. hindawi.com

[22] J. Goyal, P. Khandnor, and T. C. Aseri, "Classification, prediction, and monitoring of Parkinson's disease using computer assisted technologies: a comparative analysis," in Applications of Artificial Intelligence, Elsevier, 2020. HTML

[23] MB Makarious, HL Leonard, D Vitale, H Iwaki, et al., "Multi-modality machine learning predicting Parkinson's disease," in Nature Parkinson's Disease, 2022. nature.com

[24] 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 review," Sensors, 2022. mdpi.com

[25] A. Landolfi, C. Ricciardi, L. Donisi, et al., "Machine learning approaches in Parkinson's disease," Current Medicinal Chemistry, 2021, ingentaconnect.com. HTML

[26] 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

[27] M.G. Krokidis, G.N. Dimitrakopoulos, A.G. Vrahatis, et al., "A sensor-based perspective in early-stage parkinson's disease: Current state and the need for machine learning processes," Sensors, 2022. mdpi.com

[28] J. M. Templeton, C. Poellabauer, and S. Schneider, "Classification of Parkinson's disease and its stages using machine learning," Scientific Reports, 2022. nature.com

[29] P. G. Krishna and D. StalinDavid, "An Effective Parkinson's Disease Prediction Using Logistic Decision

Regression and Machine Learning with Big Data," Turkish Journal of Physiotherapy and Rehabilitation, 2021. [Online]. Available: academia.edu. academia.edu

[30] P. Raundale, C. Thosar, and S. Rane, "Prediction of Parkinson's disease and severity of the disease using Machine Learning and Deep Learning algorithm," in 2021 2nd International Conference, 2021, ieeexplore.ieee.org. researchgate.net

[31] Q. C. Ngo, M. A. Motin, N. D. Pah, P. DrotÃir, et al., "Computerized analysis of speech and voice for Parkinson's disease: A systematic review," Computer Methods and Programs in Biomedicine, vol. 214, Elsevier, 2022. HTML

[32] M. Hasanzadeh and H. Mahmoodian, "A novel hybrid method for feature selection based on gender analysis for early Parkinson's disease diagnosis using speech analysis," Applied Acoustics, 2023. HTML

[33] G. Bao, M. Lin, X. Sang, Y. Hou et al., "Classification of dysphonic voices in Parkinson's disease with semi-supervised competitive learning algorithm," Biosensors, 2022. mdpi.com

[34] L. Moro-Velazquez, J. A. Gomez-Garcia, et al., "Advances in Parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects," in *Digital Signal Processing*, Elsevier, 2021. HTML

[35] W. Rahman, S. Lee, M. S. Islam, V. N. Antony et al., "Detecting parkinson disease using a web-based speech task: Observational study," Journal of medical Internet research, vol. 23, no. 3, e27685, 2021. jmir.org

[36] R. Saxena and J. Andrew, "Parkinson's Disease Identification from Speech Signals Using Machine Learning Models," in Proceedings of the International Conference on Artificial Intelligence, Springer, 2023. HTML

[37] R. Bediya, R. N. Ravikumar, K. Mishra, "A Hybrid Machine Learning Framework to Improve Parkinson's Disease Prediction Accuracy," 2023 6th International Conference on Computing Methodologies and Communication (ICCMC), 2023. HTML

[38] [1] W. A. Zogaan, M. Nilashi, H. Ahmadi, R. A. Abumalloh, et al., "A combined method of optimized learning vector quantization and neuro-fuzzy techniques for predicting unified Parkinson's disease rating scale using vocal â€¦," MethodsX, vol. 2024, Elsevier, 2024. sciencedirect.com

[39] 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 Applications, 2023, Springer. HTML

[40] E. Balaji, D. Brindha, V. K. Elumalai, and R. Vikrama, "Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network," Applied Soft Computing, 2021. HTML

[41] J. Mei, C. Desrosiers, and J. Frasnelli, "Machine learning for the diagnosis of Parkinson's disease: a review of literature," Frontiers in aging neuroscience, 2021. frontiersin.org

[42] I. Ozsahin, B. Sekeroglu, P. C. Pwavodi et al., "High-accuracy automated diagnosis of Parkinson's disease," Current Medical Imaging, 2020, ingentaconnect.com. HTML

[43] R. Lamba, T. Gulati, H. F. Alharbi, and A. Jain, "A hybrid system for Parkinson's disease diagnosis using machine learning techniques," International Journal of Speech Technology, vol. 2022. Springer, 2022. HTML

[44] 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, vol. 2022, mdpi.com, 2022. mdpi.com

[45] 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

[46] H. Tiwari, S.K. Shridhar, P.V. Patil, K.R. Sinchana, et al., "Early prediction of parkinson disease using machine learning and deep learning approaches," in EasyChair Preprints, no. 2021. [Online]. Available: easychair.org. easychair.org

[47] M. Singh and N. M. Renji, "A comparative study of different machine learning algorithms in detection of Parkinsons disease," in â€¦ Journal for Research in Applied Science and â€¦, 2021. researchgate.net

[48] R. Olivares, R. Munoz, R. Soto, B. Crawford, et al., "An optimized brain-based algorithm for classifying Parkinson's disease," Applied Sciences, vol. 2020, mdpi.com, 2020. mdpi.com

[49] K. A. Hasan and M. A. M. Hasan, "Classification of Parkinson's Disease by Analyzing Multiple Vocal Features Sets," 2020 IEEE Region 10 Symposium (TENSYMP), 2020. researchgate.net

[50] N. Nasir, F. Rehman, and S. F. Omair, "Risk factors for bacterial infections in patients with moderate to severe COVIDâ€19: A caseâ€control study," Journal of medical virology, 2021. nih.gov

[51] S. Willi, R. LÃ¼thold, A. Hunt, N. V. HÃ¤nggi, D. Sejdiu, et al., "COVID-19 sequelae in adults aged less than 50 years: a systematic review," Travel Medicine and Infectious Disease, vol. 40, 2021, Elsevier. sciencedirect.com

[52] J. Dawson, C. Y. Liu, G. E. Francisco, S. C. Cramer, S. L. Wolf et al., "Vagus nerve stimulation paired with rehabilitation for upper limb motor function after ischaemic stroke (VNS-REHAB): a randomised, blinded, pivotal, device trial," The Lancet, 2021. nih.gov

[53] F. Ader, M. Bouscambert-Duchamp, M. Hites, et al., "Remdesivir plus standard of care versus standard of care alone for the treatment of patients admitted to hospital with COVID-19 (DisCoVeRy): a phase 3, randomised," The Lancet Infectious Diseases, vol. 22, no. 2, pp. 209-221, 2022. sciencedirect.com

[54] L. Monin, A. G. Laing, M. MuÃ±oz-Ruiz, et al., "Immunogenicity of one versus two doses of the COVID-19 vaccine BNT162b2 for patients with cancer: Interim analysis of a prospective observational study," The Lancet Oncology, 2021. [Online]. Available: thelancet.com thelancet.com

[55] L. M. Liau, K. Ashkan, S. Brem, J. L. Campian, et al., "Association of autologous tumor lysate-loaded dendritic cell vaccination with extension of survival among patients with newly diagnosed and recurrent glioblastoma: a â€¦," JAMA Oncol., 2023, jamanetwork.com. jamanetwork.com

[56] A. Massarweh, N. Eliakim-Raz, A. Stemmer, "Evaluation of seropositivity following BNT162b2 messenger RNA vaccination for SARS-CoV-2 in patients undergoing treatment for cancer," JAMA Oncol., 2021. jamanetwork.com

[57] R. Li, N. Liang, F. Bu, and T. Hesketh, "The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis," JMIR mHealth and uHealth, 2020. jmir.org

[58] H. Linardou, N. Spanakis, G. A. Koliou, A. Christopoulou et al., "Responses to SARS-CoV-2 vaccination in patients with cancer (ReCOVer study): a prospective cohort study of the Hellenic Cooperative Oncology Group," Cancers, 2021. [Online]. Available: mdpi.com mdpi.com

[59] W. H. O. Solidarity Trial Consortium, "Remdesivir and three other drugs for hospitalised patients with COVID-19: final results of the WHO Solidarity randomised trial and updated meta-analyses," The Lancet, 2022. sciencedirect.com

[60] S. Mayya, D. S. D'Antonio, D. SaldaÃ±a, et al., "Resilient task allocation in heterogeneous multi-robot systems," in IEEE Robotics and Automation Letters, vol. 2021. arxiv.org

[61] I. Hertrich, S. Dietrich, and H. Ackermann, "The margins of the language network in the brain," Frontiers in Communication, 2020. frontiersin.org

[62] H. Ramirez, S. A. Velastin, I. Meza, E. Fabregas, et al., "Fall detection and activity recognition using human skeleton features," in IEEE ..., 2021. ieee.org

[63] J. Li and S. Todorovic, "Set-constrained viterbi for set-supervised action segmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020. thecvf.com

[64] J. Choi, G. Sharma, M. Chandraker et al., "Unsupervised and semi-supervised domain adaptation for action recognition from drones," in Proceedings of the ..., 2020, [Online]. Available: openaccess.thecvf.com. thecvf.com

[65] E. Tolosa, A. Garrido, S. W. Scholz, and W. Poewe, "Challenges in the diagnosis of Parkinson's disease," The Lancet Neurology, 2021. nih.gov

[66] Z. K. Senturk, "Early diagnosis of Parkinson's disease using machine learning algorithms," Medical hypotheses, 2020. HTML

[67] M. J. Armstrong and M. S. Okun, "Diagnosis and treatment of Parkinson disease: a review," Jama, 2020. amedeolucente.it

[68] K. Khanna, S. Gambhir, and M. Gambhir, "Current challenges in detection of parkinson's disease," Journal of Critical Reviews, 2020. researchgate.net

[69] E. Balaji, D. Brindha, and R. Balakrishnan, "Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease," Applied Soft Computing, 2020. HTML

[70] M. Lu, K. Poston, A. Pfefferbaum, E.V. Sullivan, et al., "Vision-Based Estimation of MDS-UPDRS Gait Scores for Assessing Parkinson's Disease Motor Severity," in Medical Image Computing and Computer-Assisted Intervention â€“ MICCAI 2020, Springer, 2020. nih.gov

[71] RZU Rehman, L Rochester, AJ Yarnall, "Predicting the progression of Parkinson's disease MDS-UPDRS-III motor severity score from gait data using deep learning," in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021. researchgate.net

[72] M. Lu, Q. Zhao, K. L. Poston, E. V. Sullivan, et al., "Quantifying Parkinson's disease motor severity under uncertainty using MDS-UPDRS videos," Medical Image Analysis, Elsevier, 2021. nih.gov

[73] U. Kleinholdermann, M. Wullstein, and D. Pedrosa, "Prediction of motor Unified Parkinson's Disease Rating Scale scores in patients with Parkinson's disease using surface electromyography," Clinical Neurophysiology, 2021. HTML

[74] E. Balaji, D. Brindha, V. K. Elumalai, and K. Umesh, "Data-driven gait analysis for diagnosis and severity rating of Parkinson's disease," Medical Engineering & Physics, vol. 2021, Elsevier, 2021. HTML

[75] F. M. J. Pfister, T. T. Um, D. C. Pichler, J. Goschenhofer, et al., "High-resolution motor state detection in Parkinson's disease using convolutional neural networks," Scientific Reports, vol. 10, no. 1, pp. 1-12, 2020. nature.com

[76] A. Moshkova, A. Samorodov, N. Voinova et al., "Parkinson's disease detection by using machine learning algorithms and hand movement signal from LeapMotion sensor," in Proc. 26th Conference of ..., 2020. fruct.org

[77] H. Abujrida, E. Agu, K. Pahlavan, "Machine learning-based motor assessment of Parkinson's disease using postural sway, gait and lifestyle features on crowdsourced smartphone data," in Biomedical Physics & Engineering Express, 2020. wpi.edu

[78] A. H. Butt, E. Rovini, H. Fujita, C. Maremmani, et al., "Data-driven models for objective grading improvement of Parkinson's disease," Annals of Biomedical Engineering, vol. 2020, Springer, 2020. springer.com

[79] A. Govindu and S. Palwe, "Early detection of Parkinson's disease using machine learning," Procedia Computer Science, 2023. sciencedirect.com

[80] H. Zhao, R. Wang, Y. Lei, W. H. Liao, H. Cao, et al., "Severity level diagnosis of Parkinson's disease by ensemble K-nearest neighbor under imbalanced data," Expert Systems with Applications, vol. 2022, Elsevier, 2022. HTML

[81] S. Svetaa, S. Pavithra, and R.G. Sakthivelan, "Early Stage Parkinson's Disease Diagnosis and Detection Using K-Nearest Neighbors Algorithm," in IoT Based Control Networks and ..., Springer, 2022. HTML

[82] S. Zouhayr, T. B. Daoudi, B. N. Ayad, et al., "A hybrid method for the diagnosis and classifying Parkinson's patients based on timeâ€“frequency domain properties and K-nearest neighbor," Journal of Medical ..., vol. 2020. [Online]. Available: journals.lww.com. lww.com

[83] A. Ouhmida, A. Raihani, B. Cherradi, "Parkinson's disease classification using machine learning algorithms: performance analysis and comparison," in Proc. Engineering and Technology, 2022. HTML

[84] A. Elen and E. Avuclu, "A comparison of classification methods for diagnosis of Parkinson's," in International Journal of Intelligent Systems and Applications in Engineering, 2020. ijisae.org

[85] F. Saeed, M. Al-Sarem, et al., "Enhancing Parkinson's disease prediction using machine learning and feature selection methods," in Sensors, Materials and Signals, 2022, open-access.bcu.ac.uk. bcu.ac.uk

[86] S. Sharanyaa, P. N. Renjith, et al., "Classification of Parkinson's disease using speech attributes with parametric and nonparametric machine learning techniques," in 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. [insert page numbers here]. HTML

[87] O. Yaman, F. Ertam, and T. Tuncer, "Automated Parkinson's disease recognition based on statistical pooling method using acoustic features," Medical hypotheses, 2020. HTML

[88] S. Xu and Z. Pan, "A novel ensemble of random forest for assisting diagnosis of Parkinson's disease on small handwritten dynamics dataset," International Journal of Medical Informatics, 2020. HTML

[89] C. Ricciardi, M. Amboni, C. De Santis et al., "Classifying different stages of Parkinson's disease through random forests," in Proc. International Conference on Medical ..., 2020, Springer. HTML

[90] R. Lamba, T. Gulati, K. A. Al-Dhlan, and A. Jain, "A systematic approach to diagnose Parkinson's disease through kinematic features extracted from handwritten drawings," Journal of Reliable Intelligent Environments, vol. 2021. Springer, 2021. researchgate.net

[91] M. Masud, P. Singh, G. S. Gaba, A. Kaur, et al., "CROWD: crow search and deep learning based feature extractor for classification of Parkinson's disease," in ACM Transactions on ..., vol. 2021. [Online]. Available:

dl.acm.org HTML

[92] M. HireÅi, M. Gazda, P. DrotÃir, N.D. Pah, M.A. Motin, "Convolutional neural network ensemble for Parkinson's disease detection from voice recordings," Computers in Biology and Medicine, vol. 2022, Elsevier, 2022. HTML

[93] E. AvuÃ§lu and A. Elen, "Evaluation of train and test performance of machine learning algorithms and Parkinson diagnosis with statistical measurements," Medical & Biological Engineering & Computing, 2020. aksaray.edu.tr

[94] Z. Ma, H. Tian, Z. Liu, and Z. Zhang, "A new incomplete pattern belief classification method with multiple estimations based on KNN," Applied Soft Computing, 2020. HTML

[95] Y. Quan, "Development of computer aided classroom teaching system based on machine learning prediction and artificial intelligence KNN algorithm," Journal of Intelligent & Fuzzy Systems, 2020. HTML

[96] X. Zhu, C. Ying, J. Wang, J. Li et al., "Ensemble of ML-KNN for classification algorithm recommendation," Knowledge-Based Systems, 2021. HTML

[97] F. I. Rahman, "SHORT TERM TRAFFIC FLOW PREDICTION USING MACHINE LEARNING-KNN, SVM AND ANN WITH WEATHER INFORMATION," International Journal for Traffic & Transport Engineering, vol. 10, no. 4, 2020. ijtte.com

[98] A. S. Alotaibi, "Hybrid model based on ReliefF algorithm and K-nearest neighbor for erythemato-squamous diseases forecasting," Arabian Journal for Science and Engineering, 2022. springer.com

[99] T. Adithiyaa, D. Chandramohan, and T. Sathish, "Optimal prediction of process parameters by GWO-KNN in stirring-squeeze casting of AA2219 reinforced metal matrix composites," Materials Today: Proceedings, 2020. HTML

[100] O. Altay, M. Ulas, and K. E. Alyamac, "Prediction of the fresh performance of steel fiber reinforced self-compacting concrete using quadratic SVM and weighted KNN models," IEEE Access, 2020. ieee.org

[101] Q. Wang, S. Wang, B. Wei, W. Chen et al., "Weighted K-NN classification method of bearings fault diagnosis with multi-dimensional sensitive features," IEEE Access, 2021. ieee.org

[102] A. Ali, M. Alrubei, L.F.M. Hassan, M. Al-Ja'afari, "Diabetes classification based on KNN," IIUM Engineering Journal, 2020. azpdf.org

[103] Y. Dong, X. Ma, and T. Fu, "Electrical load forecasting: A deep learning approach based on K-nearest neighbors," Applied Soft Computing, 2021. HTML

[104] S. Uddin, I. Haque, H. Lu, M. A. Moni et al., "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," Scientific Reports, 2022. nature.com

[105] S. F. Khorshid and A. M. Abdulazeez, "Breast cancer diagnosis based on k-nearest neighbors: a review," PalArch's Journal of Archaeology of Egypt/Egyptology, 2021. [Online]. Available: archives.palarch.nl palarch.nl

[106] D. A. Anggoro and N. D. Kurnia, "Comparison of accuracy level of support vector machine (SVM) and K-nearest neighbors (KNN) algorithms in predicting heart disease," International Journal, 2020. academia.edu

[107] H. Ye, P. Wu, T. Zhu, Z. Xiao, X. Zhang, L. Zheng, "Diagnosing coronavirus disease 2019 (COVID-19): Efficient Harris Hawks-inspired fuzzy K-nearest neighbor prediction methods," in IEEE ..., 2021. ieee.org

[108] H. A. Ur Rehman, C. Y. Lin, et al., "Effective K-nearest neighbor algorithms performance analysis of thyroid disease," Journal of the Chinese Institute of Engineers, vol. 44, no. 3, pp. 257-264, 2021, Taylor & Francis. researchgate.net

[109] R. Garcia-Carretero, L. Vigil-Medina, et al., "Use of a K-nearest neighbors model to predict the development of type 2 diabetes within 2 years in an obese, hypertensive population," Medical & Biological Engineering & Computing, vol. 58, no. 6, pp. 1273-1281, 2020. HTML

[110] J. CarrÃ³n, Y. Campos-Roca, M. Madruga, et al., "A mobile-assisted voice condition analysis system for Parkinson's disease: Assessment of usability conditions," Biomedical Engineering Online, vol. 2021. Springer, 2021. springer.com

[111] I. Oduntan, "Prediction of Parkinson's Disease Using Biomedical Voice Measurements Dataset," researchgate.net, . researchgate.net

[112] O. Karaman, H. Ã‡akÄ±n, A. Alhudhaif, and K. Polat, "Robust automated Parkinson disease detection based on voice signals with transfer learning," Expert Systems with Applications, vol. 164, Mar. 2021, Art. no. 114054. HTML

[113] S. Arora and A. Tsanas, "Assessing Parkinson's disease at scale using telephone-recorded speech: insights from the Parkinson's Voice Initiative," Diagnostics, 2021. mdpi.com

[114] A. Iyer, A. Kemp, Y. Rahmatallah, L. Pillai, A. Glover, "A machine learning method to process voice samples for identification of Parkinson's disease," Scientific Reports, vol. 2023, nature.com, 2023. nature.com

[115] I. Ahmed, S. Aljahdali, M.S. Khan, et al., "Classification of Parkinson disease based on patient's voice signal using machine learning," in Automation and Soft Computing, 2022, zuscholars.zu.ac.ae. zu.ac.ae

[116] I. Karabayir, S. M. Goldman, S. Pappu, et al., "Gradient boosting for Parkinson's disease diagnosis from voice recordings," in Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2020, Springer. springer.com

[117] A. Tsanas, M. A. Little, and L. O. Ramig, "Remote assessment of Parkinson's disease symptom severity using the simulated cellular mobile telephone network," Ieee Access, 2021. ieee.org

[118] T. Zhang, H. R. Shan, and M. A. Little, "Causal GraphSAGE: A robust graph method for classification based on causal sampling," Pattern Recognition, 2022. google.com

[119] K. Ezzat, R.M.T. Fleming, L.J. Lu, M.A. Little, et al., "Phenotype-agnostic molecular subtyping of neurodegenerative disorders: The Cincinnati Cohort Biomarker Program (CCBP)," Frontiers in Aging Neuroscience, vol. 2020, 2020, frontiersin.org. frontiersin.org

[120] A. Saha, J. Reddy, and R. Kumar, "A fuzzy similarity based classification with Archimedean-Dombi aggregation operator," J. Intell. Manag. Decis, 2022. acadlore.com

[121] N. Islam, V. M. Shkolnikov, R. J. Acosta, I. Klimkin, I. Kawachi, "Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries," bmj, 2021. [Online]. Available: bmj.com. bmj.com

[122] M. Kasun, K. Ryan, J. Paik, "Academic machine learning researchers' ethical perspectives on algorithm development for health care: a qualitative study," Journal of the ..., vol. XX, no. XX, pp. XX-XX, 2024. HTML

[123] S. S. Bafjaish, "Comparative analysis of Naive Bayesian techniques in health-related for classification task," Journal of Soft Computing and Data Mining, 2020. uthm.edu.my

[124] L. M. Sinaga and S. Suwilo, "Analysis of classification and NaÃ¯ve Bayes algorithm k-nearest neighbor in data mining," in IOP Conference Series: Materials Science and Engineering, 2020. iop.org

[125] [1] S. G. Fitri, R. Selsi, Z. Rustam, et al., "NaÃ¯ve bayes classifier models for cerebral infarction classification," Journal of Physics: Conference Series, 2020. iop.org

[126] G. Kaur and A. Oberoi, "Novel approach for brain tumor detection based on NaÃ¯ve Bayes classification," in Knowledge Management, Analytics and Innovation: Proceedings of ..., Springer, 2020. HTML

[127] S. Gupta, H.K. Verma, and D. Bhardwaj, "Classification of diabetes using Naive Bayes and support vector machine as a technique," Operations Management and Systems Engineering, Springer, 2021. researchgate.net

[128] V. Kulkarni, S. Rothe, S. Devgirikar, R. Deshpande, "DOCUMENT CLASSIFICATION AND SUMMARIZER USING PROBABILISTIC CLASSIFIER," 2020. [Online]. Available: researchgate.net. researchgate.net

[129] C.K. Gomathy and M.A.R. Naidu, "The prediction of disease using machine learning," International Journal of Scientific, 2021. [Online]. Available: researchgate.net. researchgate.net

[130] M. A. Yulianto and A. Fadlil, "Wood Type Identification System using Naive Bayes Classification," Control Systems and Optimization Letters, 2023. csol.or.id

[131] H. El Hamdaoui, S. Boujraf, N.E.H. Chaoui, et al., "A clinical support system for prediction of heart disease using machine learning techniques," in 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 2020. HTML

[132] S. Saud, B. Jamil, Y. Upadhyay, and K. Irshad, "Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach," Sustainable Energy Technologies and Assessments, vol. 40, pp. 100760, 2020. researchgate.net

[133] H. L. Vu, K. T. W. Ng, A. Richter, and C. An, "Analysis of input set characteristics and variances on k-fold cross validation for a Recurrent Neural Network model on waste disposal rate estimation," Journal of environmental management, 2022. HTML

[134] D. Gholamiangonabadi, N. Kiselov, and K. Grolinger, "Deep neural networks for human activity

recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection," Ieee Access, 2020. ieee.org

[135] M. De Rooij and W. Weeda, "Cross-validation: A method every psychologist should know," in Advances in Methods and Practices, 2020, journals.sagepub.com. sagepub.com

[136] S. Prusty, S. Patnaik, and S. K. Dash, "SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer," Frontiers in Nanotechnology, 2022. frontiersin.org

[137] I. Tougui, A. Jilbab, and J. El Mhamdi, "Impact of the choice of cross-validation techniques on the results of machine learning-based diagnostic applications," Healthcare informatics research, 2021. nih.gov

[138] Z. Xiong, Y. Cui, Z. Liu, Y. Zhao, M. Hu, and J. Hu, "Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation," Computational Materials â€¦, 2020, Elsevier. nsf.gov

[139] S. Bates, T. Hastie, R. Tibshirani, "Cross-validation: what does it estimate and how well does it do it?," Journal of the American Statistical Association, vol. 2023. Taylor & Francis, 2023. arxiv.org

[140] B. G. Marcot and A. M. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?," Computational Statistics, 2021. plexusowls.com

[141] AMJC Wadoux, GBM Heuvelink, S De Bruin, et al., "Spatial cross-validation is not the right way to evaluate map accuracy," Ecological Modelling, vol. 435, Elsevier, 2021. github.io

[142] 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 evaluation and â€¦, 2020. aclanthology.org

[143] 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

[144] T. B. Alakus and I. Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection," Chaos, . nih.gov

[145] E. K. Ampomah, Z. Qin, and G. Nyame, "Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement," Information, 2020. mdpi.com

[146] S. Majumder, V. Pal, A. Yadav et al., "Performance analysis of deep learning models for binary classification of cancer gene expression data," Journal of Healthcare Engineering, 2022. nih.gov

[147] K. Chen, H. Chen, C. Zhou, Y. Huang, X. Qi, R. Shen, F. Liu, et al., "Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big â€¦," Water Research, vol. 2020, Elsevier, 2020. HTML

[148] C. Iwendi, S. Khan, J. H. Anajemba, M. Mittal, M. Alenezi et al., "The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems," Sensors, 2020, mdpi.com. mdpi.com

[149] R. Ghorbani and R. Ghousi, "Comparing different resampling methods in predicting students' performance using machine learning techniques," IEEE Access, 2020. ieee.org

[150] M. M. Ahsan, M. A. P. Mahmud, P. K. Saha, K. D. Gupta, et al., "Effect of data scaling methods on machine learning algorithms and model performance," Technologies, vol. 2021. [Online]. Available: mdpi.com. mdpi.com

[151] S. Khalid, M. Goldenberg, T. Grantcharov, et al., "Evaluation of deep learning models for identifying surgical actions and measuring performance," JAMA Network Open, 2020, jamanetwork.com. jamanetwork.com

[152] 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

[153] JY Verbakel, EW Steyerberg, H Uno, B De Cock, "ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models," Journal of Clinical ..., vol. 2020, Elsevier, 2020. jclinepi.com

[154] A. Tharwat, "Classification assessment methods," Applied computing and informatics, 2020. emerald.com

[155] T. C. F. Polo and H. A. Miot, "Use of ROC curves in clinical and experimental studies," Jornal vascular brasileiro, 2020. scielo.br

[156] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," BioData Mining, 2023. springer.com

[157] C. Nakas, L. Bantis, and C. Gatsonis, "ROC Analysis for Classification and Prediction in Practice," 2023. HTML

[158] G. Guido, S.S. Haghshenas, A. Vitale, "Development of a binary classification model to assess safety in transportation systems using GMDH-type neural network algorithm," Sustainability, 2020, mdpi.com. mdpi.com

[159] E. Hussain, M. Hasan, S.Z. Hassan, et al., "Deep learning based binary classification for Alzheimer's disease detection using brain MRI images," in 2020 15th IEEE ..., 2020. HTML

[160] L. Zhou, X. Zheng, D. Yang, Y. Wang, X. Bai, et al., "Application of multi-label classification models for the diagnosis of diabetic complications," in BMC Medical Informatics and Decision Making, vol. 21, Springer, 2021. springer.com

[161] V. Giglioni, E. GarcÃa-MacÃas, I. Venanzi, L. Ierimonti, ..., "The use of receiver operating characteristic curves and precision-versus-recall curves as performance metrics in unsupervised structural damage classification under â€¦," Engineering ..., vol. ..., no. ..., pp. ..., 2021, Elsevier. ugr.es

[162] N. Prasath, V. Pandi, S. Manickavasagam, et al., "A Comparative and Comprehensive Analysis of Prediction of Parkinson's Disease," in Engineering Approaches and Methods for Management of Chronic Diseases, 2022, [Online]. Available: taylorfrancis.com. HTML

[163] N. Prasath, V. Pandi, "A comparative and comprehensive study of prediction of parkinson's disease," in Proc. ... Engineering and ..., 2021. semanticscholar.org

[164] P. S. Deshmukh, A. K. Gaikwad, and P. K. Agrawal, "Deep Learning Based Dual-level Bioinspired Model for Parkinson's Disease Detection," in Proceedings of Intelligent Systems and ..., 2024, ijisae.org. ijisae.org

[165] C. Quan, K. Ren, Z. Luo, Z. Chen, and Y. Ling, "End-to-end deep learning approach for Parkinson's disease detection from speech signals," Biocybernetics and Biomedical Engineering, vol. 42, no. 1, pp. 1-12, 2022. sciencedirect.com

[166] N. Prasath, V. Pandi, S. Manickavasagam, et al., "13 A Comparative and," in Approaches and ..., 2022. HTML

[167] T. Zhang, L. Lin, J. Tian, Z. Xue, and X. Guo, "Voice feature description of Parkinson's disease based on co-occurrence direction attribute topology," Engineering Applications of Artificial Intelligence, vol. 2023, Elsevier, 2023. HTML

[168] D. Puri, S. Nalbalwar, A. Nandgaonkar, "Identification of Alzheimer's Disease Using Novel Dual Decomposition Technique and Machine Learning Algorithms from EEG Signals," in Advanced Computing: An International Journal, 2023. researchgate.net

[169] K. Cura, "Evaluation of Some Neurological Disorders by the Analysis of EEG Signals," 2021. HTML

[170] D. Rajasekhar, "Unveiling Diagnostic Significance: InvestigatingEEG Biomarkers for Enhanced Dementia Diagnosis," 2023. diva-portal.org

[171] S. Hamida, O. El Gannour, Y. Lamalem, et al., "Efficient Medical Diagnosis Hybrid System based on RF-DNN Mixed Model for Skin Diseases Classification," in *..., Engineering and ...,*, 2023. HTML