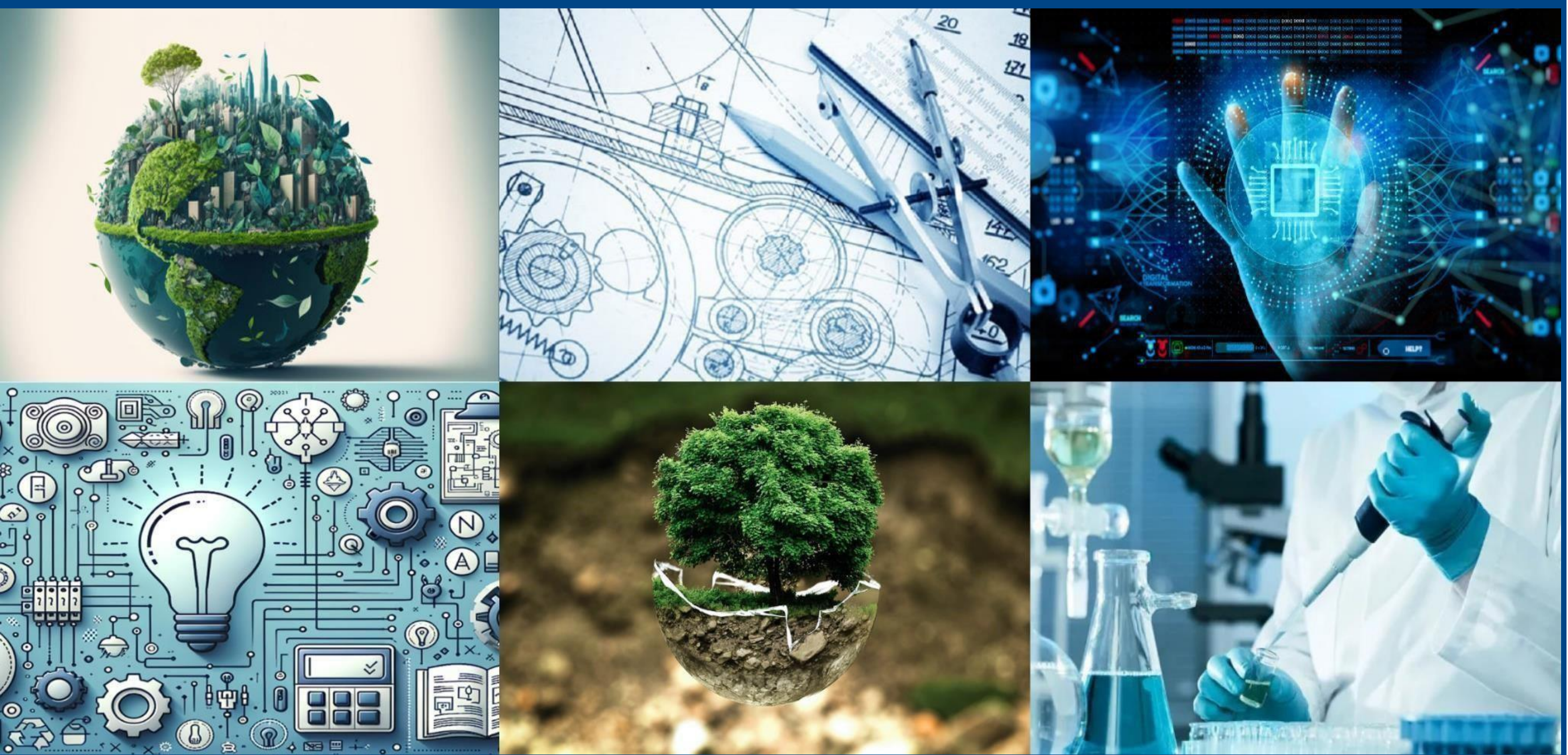
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 8, August 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

PERFORMANCE EVALUATION LEANING ALGORITHM IN PARKINSON'S DISEASE CLASSIFICATION

M. Prabhu, P.V. Sankar Ganesh

Assistant Professor, PG Department of Computer Applications, Mohamed Sathak College of Arts and Science, Chennai, India

Assistant Professor, PG Department of Computer Applications, Mohamed Sathak College of Arts and Science, Chennai, India

ABSTRACT: Health Parkinsonism disease is a neurodegenerative movement disease where the symptoms progressively develop start with a slight tremor in one hand and a feeling of stiffness in the body and it became inferior over time. Dopamine hypotheses acts as a communication between the parts of the brain and nervous system that assistance control and co-ordinate body activities. As dopamine normally neurons in the measures begin to experience trouble in speaking, writing, walking or completing other simple task. Around, 90% affected people with Parkinsonism have speech disorders. It affects over 5 million people worldwide. At present-day there is no conclusive result for this disease by non-specialist clinicians, particularly in the early stage of the disease where identification of the indications are very difficult in its earlier stages. The proposed predictive analytics background is a grouping of K-means clustering and Support vector machine which is used to gain insights from patients. By using supervised and unsupervised learning techniques, the problem can be solved with nominal error level.

Our proposed system provides accurate results by integrating spiral drawing inputs of normal and Parkinsonism affected patients. From these drawings Support Vector Machine algorithm. The Parkinsonism disease genetic factor prediction can be measured as an organization assignment with two tags. We proposed Support Vector Machine (SVM) algorithm to solve this bi- classification problem. For classification, SVM constructs a set of hyperplanes in a high-dimensional space to classify genes with different labels. SVM is useful in solving classification as well as regression problematic statements. In the organization problem, two periods are separated or classified by a hyper plane. SVM creates two marginal lines along with a hyperplane having some distance so that they will be simply linearly independent for together the classification points.

KEYWORDS: Health Parkinsonism Disease, K-Means Clustering, Support Vector Machine, Techniques.

I. INTRODUCTION

Health Parkinsonism disease (HPD) is one of the most common neurodegenerative diseases with a occurrence rate of 2% in the population above 62 years old, affecting 1–2 people per 1,000 (Tysnesand Storstein, 2017). The estimated global population affected by HPD has more than doubled from 1990 to 2016 (from 2.5 million to 6.1 million), which is a result of increased number of aged publics and age-standardized prevalence rates (Dorsey et al., 2019). HPD is a progressive neurological disorder associated with motor and non-motor features (Jankovic, 2008) which comprises multiple aspects of movements, including planning, initiation and execution (Contreras- Vidal and Stelmach, 1995).

Supervised and unsupervised learning techniques are being increasingly applied in the healthcare zone. As its name indicates, Supervised and unsupervised learning allows for a computer platform to study and extract meaningful illustration from data in a semi-automatic method. For the diagnosis of HPD, machine learning models have been applied to a multitude of data modalities, including handwritten patterns. In recent years, the number of publications on the application of machine learning to the diagnosis of HPD has increased. Although previous studies have reviewed the use of machine learning in the diagnosis and assessment of HPD, they were limited to the analysis of motor symptoms, kinematics, and wearable sensor data.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

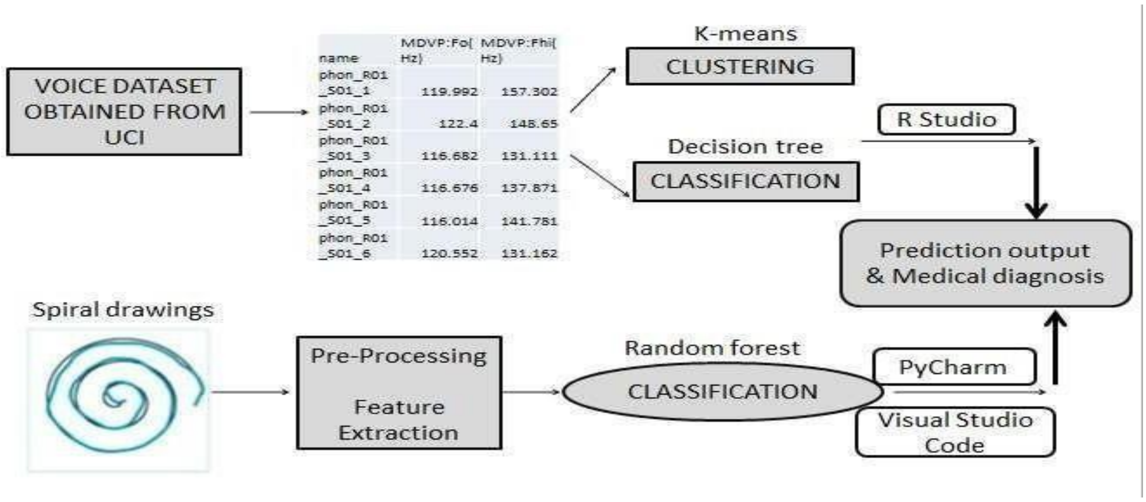
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Diagnosis is clearly a difficulty in HPD management, and an effective screening process, particularly one that doesn‘t require a clinic visit, would be beneficial. Since HPD patients exhibit characteristic vocal features, voice recordings are a useful and noninvasive tool for diagnosis. If machine learning algorithms could be applied to a voice recording dataset to accurately diagnosis HPD, this would be an effective screening step prior to an appointment with a clinician. Actually, this sort of binary classification problem is common in many areas of medical diagnosis, and techniques that work well within one domain are likely to be applicable to others.

Parkinson’s disease voice dataset analysis

HPD voice dataset is collected from UCI machine learning repository and these are stored into the R Tool environment as Testing and Training datasets. These are stored into the tool environment as Testing and Training datasets. R tool is a programming language and software environment for statistical analysis, graphics representation, data analysis and as well as machine learning. It involves the following steps and procedures.



Importing data

Organize the data in an Excel worksheet to include column names in the first row (i.e. people voice collected at various time zones) and each subsequent row contains all the information (i.e. set of 22 parameter is taken into consideration and the peoples voice range for those parameters is tested and then noted), finally the status column shows two values 0 (healthy) and 1(affected). Import data into tool, using the "Import data..." feature.

K-Means clustering algorithm

Unsupervised learning algorithm that tries to cluster data based on their similarity, and just tries to find patterns in the data. Here, we have to specify the number of clusters we want the data to be grouped into and then the algorithm randomly assigns each observation to a cluster, and finds the centroid of each cluster and then, it iterates by reassigning data points to the cluster whose centroid is closest and calculates new centroid of each cluster.

Decision Tree method

It is also called a prediction tree. It uses a structure to specify sequences of decisions and consequences, the goal is to predict a response or output. The forecast can be accomplished by creating a decision tree with test points and branches. At each check point, a decision is made to pick a particular branch and cross the trees and can be used in a variety of disciplines, on the basis of individual characteristics.

Predicted Output

The predicted output for voice data analysis based on clustering and classification is with an accuracy of 86%

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808086

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

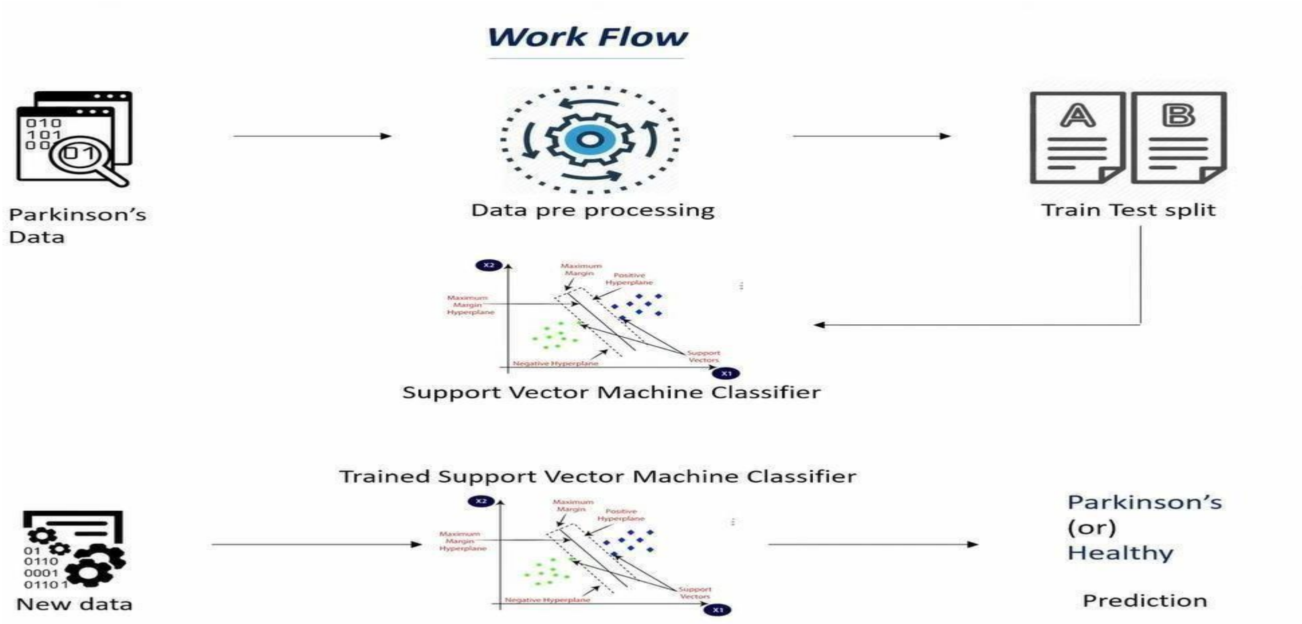
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. DATA PREPROCESSING

Dataset
The primary dataset used in this paper is from the supervised and unsupervised Learning and contains data from voice recordings of 18 subjects with health Parkinsonism disease and 6 control subjects. There are a total of 165 recordings, from which 18 different voice measure features have been extracted. Each example also includes a subject identifier and a binary classification attribute which indicates whether or not the subject has HPD.

Data collection
Collecting and analyzing data from many different sources is known as data collection. This means that the data we collect must be acquired and kept in a way that makes sense for the professional challenge at pointer. The reference process of data mapping divides the collected dataset into 70 percent training data and 30 percent testing data. The data is divided for the displaying dataset into training and testing sets is to assign data points to the previous and the remaining to the later. A typical is therefore trained using a training dataset, at that moment applied to a test set. My paper could then be evaluated based on its performance.

Data Sampling
A quantity of strategies were discovered for dealing with an unbalanced dataset. In the oversampling approach, samples from the minority (healthy) class were replicated to create a dataset with equal proportions of equally periods. In the under sampling approach, fewer samples were taken from the majority (HPD) class, leading to a smaller but balanced dataset. The assessment segment contains results from the most operational of these approaches for each model.



Training and Testing
Training and testing use of data mapping, the collected dataset is divided into two portions: 70 percent training data, and 30 percent testing data. In direction to allocate data points to the previous and the later in the demonstrating dataset, in the data has been divided into training and testing sets. A typical is therefore trained using a training set, it have applied to a test set our application may be evaluated in this manner.

Prediction and Comparison
The techniques ready to detect Parkinsonism disease and predict based on the given dataset. The data features acquired from test is associated .Supervised learning algorithms can only be properly compared if they are unsupervised Learning is basically that field of new technology with the help of which systems can provide sense to data in much the

same way as human beings ensure. The fundamental focus of learning is to allow systems to learn from experience without being explicitly programmed or human intervention.

III. CLASSIFICATION

Learning algorithm can be classified into three types:
1. Supervised learning
2. Unsupervised learning
Supervised Learning

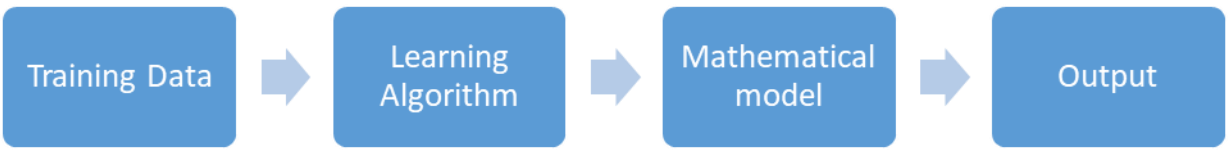
Supervised learning method in which we provide sample labeled data to the supervised learning system in order to train it, and on that beginning, it predicts the result. The techniques generates a model using labeled data to understand the datasets and learn about each data, when the training and processing are done then we test the model by providing a sample data to check whether it is predicting the particular output or not. The goal of supervised learning is to map input at a with the output data. The example of supervised learning is unwanted data filtering.

Unsupervised Learning
Unsupervised learning method in which learns without any organizing. The training is provided to the mechanism with the set of data that has not been labeled, categorized and the algorithm needs to act on that data without any observation. The aim of unsupervised learning is to reorganize the input data into new features or a group of objects with similar patterns. In unsupervised learning, we don't have a predetermined result. The mechanism tries to find useful understandings from the large amount of data.

Reinforcement learning agent gets a reward for each right action and gets a penalty for each incorrect achievement. The agent learns repeatedly with these feedbacks and expands its performance. In reinforcement learning, the agent interacts with the environment and explores it. The aim of representative is to get the most compensation opinions, and in future, it expands its performance.

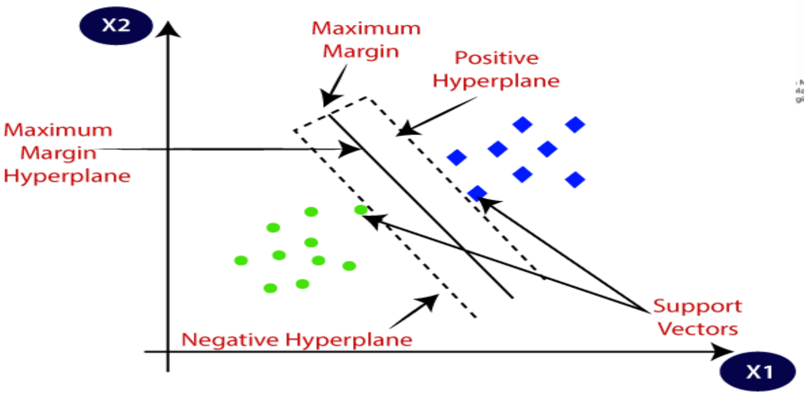
IV. DESIGNING LEARNING SYSTEM

The Training Data to Learning Algorithm, this algorithm will produce an exact model and with the help of the measured model, the mechanism will make a prediction and take an assessment without being explicitly programmed. At the time of during the training data, the more system will work with it the additional it will get experience and it will result the more efficient result is produced.

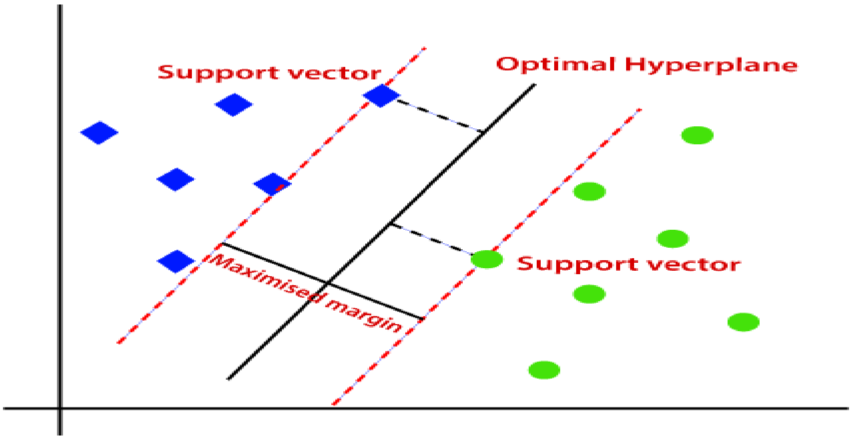


Support Vector Machine method
Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. It is mainly used for Classification problems in learning method. The aim of the SVM method is to create the best line or decision boundary that can separate n-dimensional space into periods so that we can easily put the new data point in the right category in the upcoming. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. In these great cases are called as support vectors and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

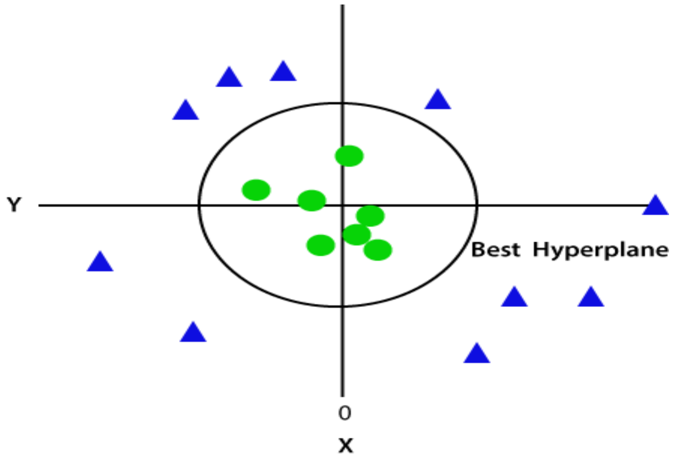
© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808086
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Linear SVM
Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.



Non-linear SVM
Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier. To separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension z. It can be calculated as: $Z=x^2+y^2$



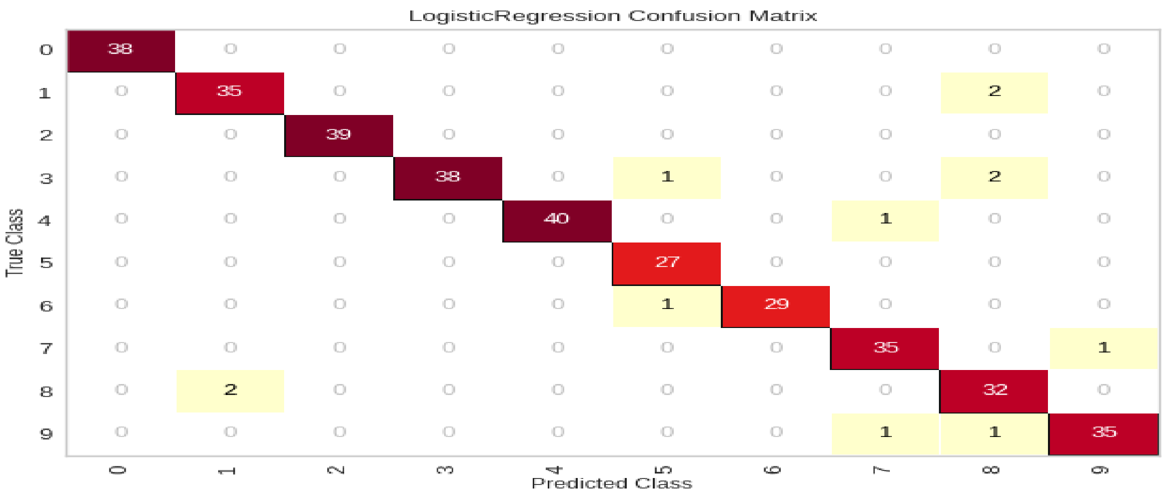
Hyperplane and Support Vectors

It can be multiple lines/decision boundaries to separate the modules in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane SVM. We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

V. PERFORMANCE METRICES CONFUSION

MATRIX

The Confusion Matrix visualizer is a Score Visualizer that takes a stable learn classifier and a set of test x and y values and returns a report showing how each of the test values predicted classes associate to their actual classes. Data researchers use confusion matrices to understand which classes are most easily confused. It is provide comparable statistics as what is accessible in a Classification report, but rather than top-level scores, they provide deeper insight into the classification.



ACCURANCY SCORE

In this learning methods accuracy plays an important role. Accuracy is a mirror of the effectiveness of our method. Not even this accuracy communicates the percentage of correct predictions. It is just a scientific term, in this learn provides some function for it to use in addition to get the accuracy of the method. Accuracy score, report, confusion metrix are some of them. In this part, we will recognize the accuracy, the scientific background of accuracy and how to predict it with hands-on code.

Accuracy= Correct Predictions/All Predictions=TP+TN/TP+TN+FP+FN

If y_i is the predicted value of the i-th sample and y_i is the corresponding true value then the fraction of correct predictions over $n_{data\ sample}$ is defined as

$$\text{Accuracy }(y.y1)=1/n_{data} \sum_{i=0}^{n\ data\ -1} 1(yi = y1)$$

Where 1(x) is the indicator function.

Precision

To overcome the limits of Accuracy, researchers typically use Precision, Recall and Specificity. Precision communicates what proportion of positive predictions was actually correct. It achieves this by counting the samples correctly predicted as positive (TP) and dividing it by the total positive predictions, correct or incorrect (TP, FP).

Precision = TP/ (TP+FP)

Recall = Sensitivity = True Positive Rate = Hit Rate

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808086

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

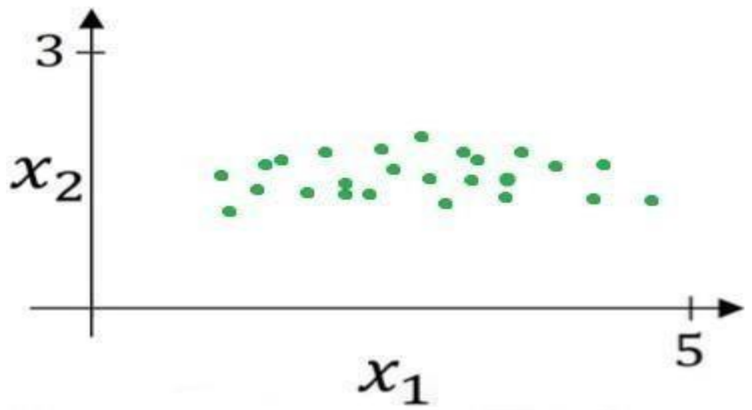
In the same way to Precision, Recall targets at measuring what proportion of actual positives was identified correctly. It does so by dividing the correctly predicted positive samples (TP) by the total number of positives, either correctly predicted as positive or incorrectly predicted as negative (TP, FN).

$$\text{Recall} = TP/ (TP + FN)$$

VI. NORMALIZATION

Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to use a common scale, without changing differences in the ranges of values or losing information. Normalization is also required for some algorithms to model the data correctly.
For example, assume your input dataset contains one column with values ranging from 0 to 1, and another column with values ranging from 1000 to 10,000. The great difference in the *scale* of the numbers could cause problems when you attempt to combine the values as featuresduring modeling.

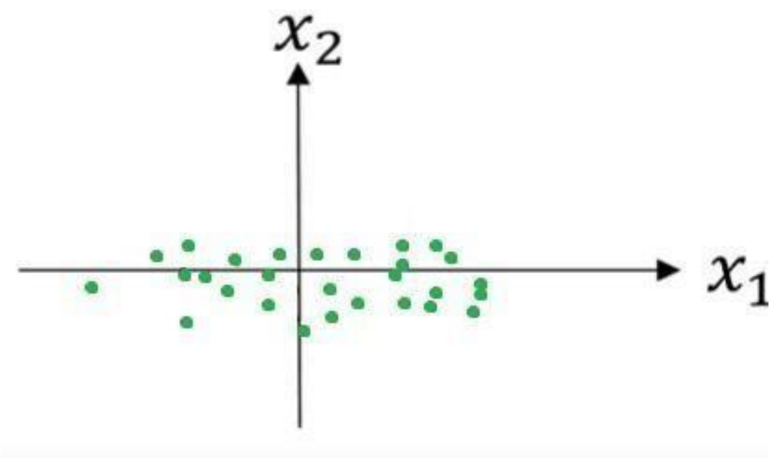
Normalization avoids these problems by creating new values that maintain the general distribution and percentages in the source data, while keeping values within a scale applied diagonally numeric columns used in the method.



The mean of the dataset is calculated using the formula and then is subtracted from each individual training example effectively shifting the dataset so that it has zero mean.

$$\text{Mean}(X) = \overline{x} = 1/n \sum_{i=1}^{n} X_i$$

The ranges of both features on their respective axes will be same and the result.



© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808086

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

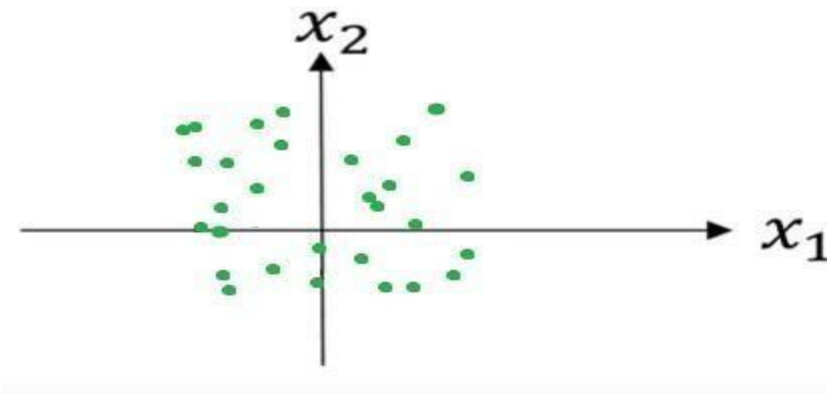
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Notification in the plot above that the feature X_1 has a larger variance (look at the range on the x-axis), than the other feature X_2.We normalize the variances by applying the formula for variance shown below and then divide each training example with the value of variance thus obtained.

$\sigma^2 = 1/n \sum X^{(i)2}$

A quick-thinker might have noticed that by dividing each training example with this value, the variance for both x_1 and x_2 will be equal to 1. The scatterplot will now look like the following:



Normalization supports expressively affect the performance of our method for the enhanced. When the data is normalized, this outline mapping is balanced and gradient succession finds it simpler to spread the mean. It converges quicker than when the data is not normalized it takes more steps to find its way to the minimum.

VII. TOOLA AND TECHNIQUES

Several methods are applied for normalization and widely used techniques as follows:

Rescaling: It is known as –min –max normalization, it is the simplest of all methods and calculated as:

$$x' = x\text{-}min(x)/max(x)\text{-}min(x)$$

Mean normalization: In this method uses the mean of the observations in the transformation process:

$$x' = x\text{- } average(x)/max(x)\text{-}min(x)$$

Z-score normalization: It is also known as standardization, in this method uses Z-score or –standard score. It is widely used in learning algorithms such as SVM and logistic regression

$$Z = x - \mu/ \sigma$$

Here, z is the standard score, μ is the population mean and σ is population standard deviation
The data points in the left and the right figures you might be able to see which data points shifted from pre-normalized to post- normalized method. These changes are often at the boundaries rather than at either end of the spectrum in the distribution

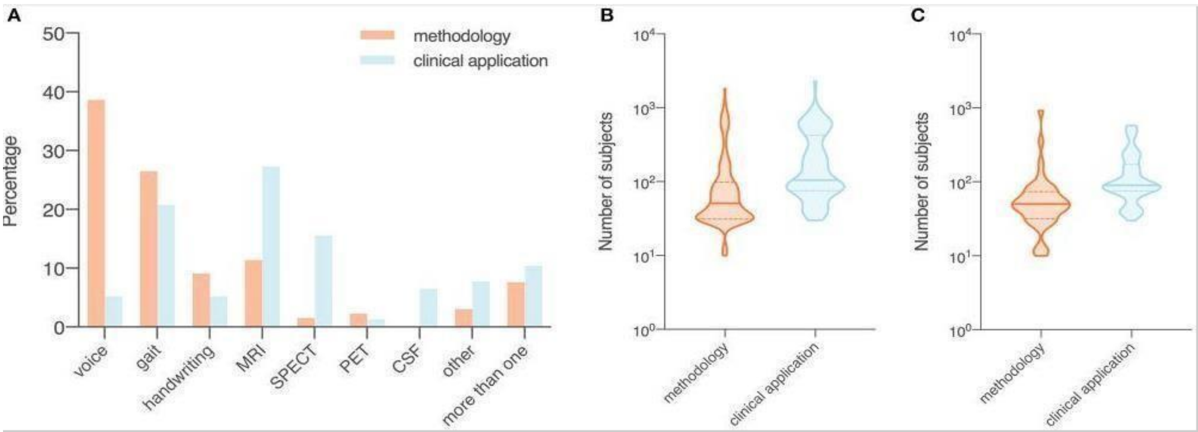
Data modality (A) and number of subjects (B, C) of included studies, summarized by objectives (procedure or clinical application). Orange, studies with a focus on the development of an innovative technical methodology to be used in the diagnosis of Parkinsonism disease; blue, studies that investigate the use of distributed learning methods or novel data modalities (A) Proportion of data modalities in included studies displayed as percentages. (B) Sample size in all included studies. (C) Sample size in studies that collected data from recruited human participants. Data shown are means (SD).

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808086

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

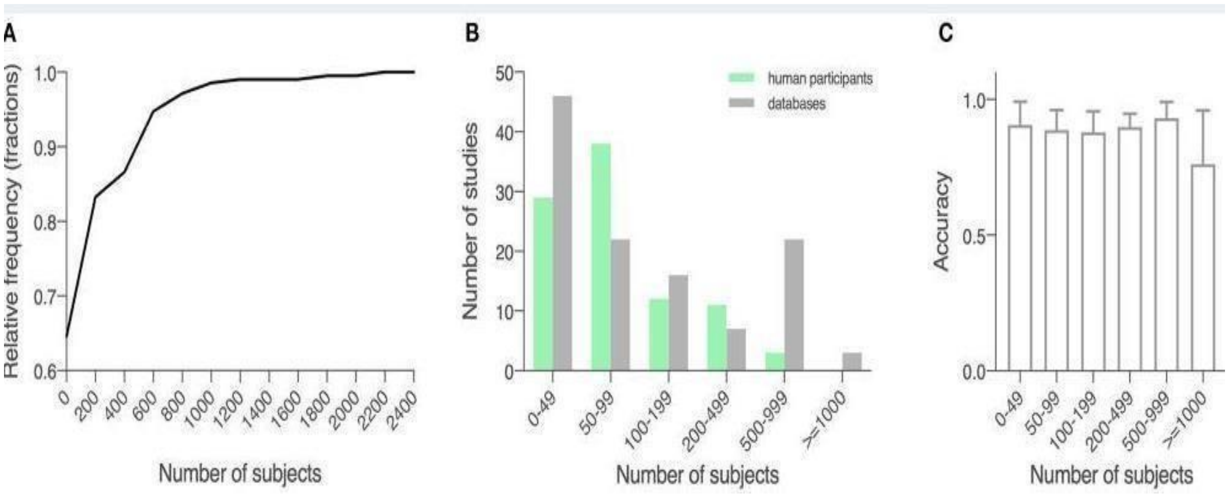
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Data sample size of the involved trainings. (A) Cumulative relative frequency graphdepicting the frequency of the sample sizes considered.

(B)Histogram depicting the frequency of a sample size of 0, 50, 5,100, 200, 500 and over 1,000 for trainings using nearby recruited human participants and trainings using earlier distributed open databases. Green, trainings using nearby recruited humanparticipants; gray, trainings using data sourced from open databases.

(C)Method performance as measured by accuracy in relation to data sample size, shown in means (SD).



VIII. CONCLUSION

The prediction of Health Parkinsonism disease is most important and challenging problem for biomedical engineering researchers and doctors. In this techniques, minimum redundancy maximum relevance feature selection algorithms was used to select the most essential feature between all the features to predict the Parkinsonism diseases. This paper predicts health disease through machine learning technique. In this techniques the outcome provides hopeful results by analyzing both voice and spiral illustration analysis. Earlier review papers have focused only on a particular imaging based analysis such as MRI or PET, or on one specific type of dementia only such as AD. This proposed aim to provide learning technique to predict and provide early diagnostics for Parkinsonism disease. Recently, the field of Speech processing and its recognition have been broadly recognized for their diverse applications. Most importantly speech processing has great potential in the detection, classification and diagnosis of health Parkinsonism disease.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The use of several instance learning for detecting Parkinson disease symptoms is studied. Proposed work speak to the invention of HPD symptom detection from weakly labeled data as a semi-supervised various learning problem. The features were chosen to address the subject and symptom specific nature of the problem.

IX. FUTURE WORK

In future work, we plan to increase our subject pool and utilize optimal feature selection strategies under learning frameworks for developing artificial intelligence specific methods. These techniques can potentially be adapted to various other physiological sensing and monitoring applications as well. The current learning approach that works to detect mitochondrial impairments, a potential mechanism underlying the pathogenesis of Parkinsonism disease (HPD), accurately distinguished between patients with HPD and healthy controls.

REFERENCES

1.Muhlenbach F. and Rakotomalala, R.(2015) Discretization of Continuous Attributes, Wang, J., Ed., Encyclopedia of Data Warehousing and Mining,

2.M. Liu, D. Zhang, and D. Shen, Relationship induced multi- template learning for diagnosis of Alzheimer’s disease and mild cognitive impairment, IEEE transactions on medical imaging, (2016).

3.Abdullah Caliskan, Hasan Badem, A. Baştürk, M.E. Yüksel, Diagnosis Of The Parkinson Disease By Using Deep Neural Network Classifier, Journal Of Electrical And Electronics Engineering, (2017).

4.M. Abdar and M. Zomorodi-Moghadam, Impact of Patients‘Gender on Health disease using Classification Algorithms‖ Journal of AI and Data Mining (2018).

5.Sleigh, J.W.; Olofsen, E.; Dahan, effects of general anaesthesia. In Proceedings of the 5th International Conference on Memory, Awareness and Consciousness (2019).

6.Tsuda, M., Asano, S., Kato, Y., Murai, K., and Miyazaki, M. (2020). Differential diagnosis of multiple system atrophy with predominant Parkinson and Parkinson's disease using neural networks. J. Neurol. Sci.

INNO SPACE
SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com