ISSN: 2320-5407 Int. J. Adv. Res. 12(05), 1118-1137

Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/18827
DOI URL: http://dx.doi.org/10.21474/IJAR01/18827

ISSN NO. 2320-5407

RESEARCH ARTICLE

EVALUATING SPEECH ANALYSIS TECHNIQUES FOR PARKINSON'S DISEASE DETECTION: A COMPARISON OF MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

Anand Ratnakar
M.Tech Robotics and Artificial Intelligence, Dept. of Manufacturing Engineering and Industrial Management COEP Technological University Pune, India.

Manuscript Info

Manuscript History
Received: 31 March 2024
Final Accepted: 30 April 2024
Published: May 2024

Key words:-
Parkinson's Disease (PD) Diagnosis, Speech Analysis, Artificial Neural Networks (ANN), Support Vector Machine (SVM), Naive Bayes (NB), K Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Machine Learning (ML), Deep Learning (DL)

Abstract

Parkinson's disease (PD) presents a diagnostic challenge due to its often subtle and gradual onset. Speech analysis offers a promising avenue for early detection, enabling intervention before the disease significantly progresses. This study investigates the efficacy of supervised machine learning algorithms in identifying PD using speech features. We compared the performance of Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN) for PD classification. Our findings demonstrate the superiority of ANNs, achieving a test accuracy of 97.44%, which surpasses existing benchmarks and highlights their potential for PD diagnosis. This approach leverages readily available speech data, potentially reducing reliance on expensive and time-consuming clinical procedures. This research contributes to the development of non-invasive, speech-based diagnostic tools for PD, paving the way for earlier intervention and improved patient management.

Introduction:-

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by a decline in dopamine levels in the brain. This deficiency manifests in tremors, rigidity, and difficulties with balance and coordination. As the second-most common neurodegenerative disease after Alzheimer's disease, PD affects millions globally, with its prevalence expected to rise due to an aging population.

Early diagnosis of PD is crucial for optimizing patient outcomes and management. However, traditional diagnostic methods rely heavily on a patient's medical history and neurological examinations, which can lack sensitivity, especially in the early stages. Additionally, definitive diagnostic tests for PD are not currently available.

This necessitates the exploration of new and potentially more objective methods for PD detection. Machine learning (ML) and deep learning (DL) offer promising avenues to address this challenge. ML algorithms can learn complex relationships between features extracted from data (e.g., voice, gait) and disease status. Deep learning, a subfield of ML, utilizes artificial neural networks with multiple layers to automatically discover these relationships from raw data without the need for extensive feature engineering.

This research investigates the potential of both ML and DL techniques to analyze voice and potentially other relevant data modalities for improved PD diagnosis. We compare the performance of various ML algorithms with a

Corresponding Author:- Anand Ratnakar
Address:- M.Tech Robotics and Artificial Intelligence, Dept. of Manufacturing Engineering and Industrial Management COEP Technological University Pune, India.

deep learning model called Artificial Neural Networks (ANN) to identify the most effective approach for accurate PD classification. By leveraging the strengths of both ML and DL, we aim to develop a more objective and potentially earlier detection method for PD compared to traditional approaches.

Leveraging Machine Learning for Improved Diagnosis

Parkinson's disease (PD) diagnosis is most effective in its early stages, offering patients better treatment outcomes and improved quality of life. Traditionally, diagnosis relies on neurological history and motor assessments, which can lack sensitivity. Machine learning (ML), a subfield of Artificial Intelligence (AI), offers promising avenues for improved detection. By combining traditional methods with ML-based analysis, clinicians may achieve a more comprehensive understanding of the disease in patients.

One readily observable aspect of PD is gait (walking pattern). As walking is a fundamental part of daily life, gait analysis has emerged as a potential non-intrusive tool for PD detection, with the advantage of being deployable in home settings. Researchers have explored various gait analysis approaches, with some focusing on combining ML techniques for autonomous and offline operation.

Speech patterns can also be indicative of PD, particularly in early stages. Speech problems associated with PD include dysphonia (weak vocal quality), diplophonia (repetitive echoes), and hypophonia (impaired vocal muscle coordination). These subtle changes in speech can be detected and analyzed using computational methods to aid in PD diagnosis.

Research Motivation and Proposed Approach

This research investigates the potential of multivariate data analysis (MVDA) combined with ML techniques for early and accurate PD detection. Current research in this area often focuses on single-source data (text, speech, video, or images). This study highlights the limitations of such an approach and proposes MVDA for more comprehensive multimodal data processing. By analyzing a wider range of data points, including gait, speech, and potentially other relevant information, MVDA has the potential to improve disease detection accuracy.

This work specifically investigates the effectiveness of MVDA powered by ML in processing multimodal data for PD diagnosis. Existing research utilizes various ML algorithms like Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN) for PD detection based on vocal features. This study builds upon these advancements by leveraging large datasets and diverse ML approaches for improved disease identification. The proposed MVDA framework encourages the inclusion of a wide range of data points, such as multivariate acoustic characteristics, from a large patient population. This objective approach, aided by ML techniques, aims to achieve a more accurate and reliable diagnosis of PD compared to traditional subjective methods.

Research Contribution

This research explores various machine learning algorithms employed in speech analysis for PD diagnosis. The strengths and limitations of these algorithms for PD detection are evaluated, while also considering potential shortcomings in existing comparative studies. Artificial Neural Networks (ANN) have demonstrated promising accuracy in speech analysis for PD diagnosis compared to other classifiers.

The key contributions of this paper are as follows:

1. **Comparative Analysis of Machine Learning Algorithms:** This research aims to identify which ML algorithms, including SVM, KNN, Random Forest, Naive Bayes, and ANN, offer the most accurate classifications for PD diagnosis.
2. **Statistical Evaluation for Improved Diagnosis:** This study proposes the development of statistical evaluations for PD diagnosis. These evaluations aim to identify the optimal training and testing parameters, ultimately contributing to future research efforts.
3. **Comprehensive Machine Learning Model Exploration:** The proposed system utilizes seven different machine learning and deep learning models, including Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN). This comprehensive approach allows for identifying the model that performs best for PD diagnosis based on training and testing results.

4. **Feature Selection Methodology:** This study employs a comprehensive methodology to explore the effectiveness and efficiency of various feature selection approaches to improve PD prediction accuracy.
5. **Benchmarking Model Performance:** By training all seven machine learning and deep learning models on the same dataset, this research facilitates a direct comparison of their performance in PD diagnosis.

Related Works:-

In order to distinguish PD cases from healthy controls, a variety of modern machine learning algorithms, including support vector machines, artificial neural networks, logistic regression, naïve Bayes, etc., have been successfully used. In this study, numerous databases, including Web of Science, Elsevier, MDPI, Scopus, Science Direct, IEEE Xplore, Springer, and Google Scholar, were utilized to survey relevant papers on Parkinson's disease. The Table Below shows the details about the previous work.

Reference	Feature	Machine Learning Algorithms Used	Objective	Tools Used	Source of Data	No. of Subjects	Outcomes
Sakar et al., 2019	Speech	Naïve Bayes, SVM (RBF and Linear), KNN, Random Forest,MLP	Classification of PD from HC	JupyterLab With Python Programming Language	Collected from participants	252 (188 PD + 64 HC)	Highest accuracy obtained from SVM (RBF)-86%
Yasar A. et al., 2019	Speech	Artificial Neural Network	Classification of PD from HC	MATLAB	Collected from participants	120 (40 HC + 80 PD+)	Accuracy of ANN-94.93%
Avuçlu, E., Elen, A, 2020	Speech	KNN, Random Forest, Naïve Bayes, SVM	Classification of PD from HC	JupyterLab With Python Programming Language	UCI machine learning repository	31 (23 PD+8 HC)	Accuracy from Naïve Bayes-70.26%
Marar et al., 2018	Speech	Naïve Bayes,ANN, KNN, Random Forest, SVM, Logistic Regression	Classification of PD from HC	R programming	Collected from participants	31 (23 PD+8 HC)	Highest accuracy obtained from ANN-94.87%
Sheibani R et al., 2019	Speech	Ensemble Based Method	Classification of PD from HC	JupyterLab With Python Programming Language	UCI machine learning repository	31 (23 PD+8 HC)	Accuracy obtained from ensemble learning-90.6%
John M. Tracy et al., 2020	Speech	Logistic Regression (L2-regularized), Random Forest, Gradient Boosted Trees	Classification of PD from HC	Python	mPower database	2289 (246 PD + 2023 HC)	Highest accuracy obtained from gradient boosted trees: Recall-79.7%, Precision-90.1%, F1-score-83.6%
Cibulka et al., 2019	Handwriting Patterns	Random Forest	Classification of PD from HC	Not Mentioned	Collected from participants	270 (150 PD + 120 HC)	Classification error for rs11240569, rs708727, rs823156 is 49.6%, 44.8%, 49.3% respectively

Hsu S-Y et al., 2019	Handwriting Patterns	SVM with RBF Kernel, Logistic Regression	Classification of PD from HC	Weka	PACS	202,94 Severe PD +102 mild PD + 6 HC	Highest accuracy obtained from SVM-RBF 83.2% having sensitivity 82.8%, specificity 100%
Drotár, P et al., 2016	Handwriting Patterns	K-NN, Ensemble Classifier (AdaBoost), Support Vector Machine	Classification of PD from HC	Python [scikit-learn library]	PaHaw database	37 PD and 38 HC	Accuracy-81.3%
Fabian Maass et al., 2020	Handwriting Patterns	SVM	Classification of PD from HC	Weka	UCI machine learning repository	157,82 PD + 68 HC + 7 Normal Pressure Hydrocephalus (NPH))	Sensitivity-80%, and specificity-83%
J. Mucha et al., 2018	Handwriting Patterns	Random Forest	Classification of PD from HC	Python	PaHaw database	69, 33 PD + 36 HC	accuracy-90% with sensitivity 89%, and specificity 91%
Wenzel et al., 2019	Handwriting Patterns	CNN	Classification of PD from HC	MATLAB	PPMI database	645, 438 PD + 207 HC	accuracy-97.20%
Segovia, F. et al., 2019	Handwriting Patterns	SVM with 10 Cross Validation	Classification of PD from HC	Python	Virgen De La Victoria Hospital, Malaga, Spain	189, 95 PD + 94 HC	accuracy-94.25%
Ye, Q. et al., 2018	Gait	Least Square (LS)-SVM, Particle Swarm Optimization (PSO)	Classification of PD, ALS, HD from HC	Not mentioned	Neurology Outpatient Clinic at Massachusetts General Hospital, Boston, MA, USA	64, 15 PD + 16 HC +13 [Amyotrophic lateral sclerosis Disease (ALS)] + 20 [Huntington's Disease (HD)]	Accuracy to diagnose PD from HC- 90.32%, Accuracy to diagnose HD from HC-94.44%, Accuracy to diagnose ALS from HC- 93.10%

Klomsae, A et al., 2018	Gait	Fuzzy KNN	Classification of PD, ALS, HD from HC	Not mentioned	Neurology Outpatient Clinic at Massachusetts General Hospital, Boston, MA, USA	64, 15 PD + 20 HD +13 ALS + 16 HC	Accuracy to diagnose PD from HC- 96.43%, Accuracy to diagnose HD from HC-97.22%, Accuracy to diagnose ALS from HC-96.88%
J. P. Félix et al., 2019]	Gait	SVM, KNN, Decision Tree, Naïve Bayes, LDA	Classification of PD from HC	MATLAB R2017a	Neurology Outpatient Clinic at Massachusetts General Hospital, Boston, MA, USA	31, 15 PD+ 16 HC	Highest Accuracy obtained from SVM,KNN and Decision Tree - 96.80%
Andrei et al., 2019	Gait	SVM	Classification of PD from HC	Not mentioned	Laboratory for Gait and Neurodynamics	166, 93 PD+ 73 HC	Accuracy- 100%
Priya SJ et al., 2021	Gait	ANN	Classification of PD from HC	MATLAB R2018b	Laboratory for Gait and Neurodynamics	166 ,93 PD+ 73 HC	Accuracy- 96.28%
Oğul, et al., 2020	Gait	ANN	Classification of PD from HC	MATLAB	Laboratory for Gait and Neurodynamics	166 ,93 PD+ 73 HC	Classification accuracy - 98.3%
Li B et al., 2020	Gait	Deep CNN	Classification of PD from HC	Not mentioned	Collected from participants	20, 10 PD + 10 HC	Accuracy- 91.9%

Table 1:- Comparative Studies of Machine Learning Approaches to diagnose Parkinson’s Disease.

Proposed System:-

This research investigates the potential of machine learning (ML) and deep learning (DL) algorithms for classifying Parkinson's disease (PD) and healthy controls (HC) using voice analysis.

System Architecture:

The proposed work involves methods with several key components that work together to achieve PD classification:

Data Acquisition:

1. We retrieved voice recordings from the publicly available Max Little dataset.
2. This dataset contains recordings from individuals diagnosed with PD and healthy controls, along with 22 pre-extracted features related to various aspects of the speaker's voice.

Data Preprocessing:

Depending on initial data exploration, preprocessing steps might have been applied to the data, including:

1. Handling missing values using techniques like mean/median imputation or more sophisticated methods.
2. Scaling features (normalization or standardization) to ensure all features contribute equally during model training (applicable for specific algorithms).
3. Encoding categorical variables (if present) through appropriate techniques.

Model Training:
We implemented and trained various ML and DL models on the preprocessed data. This involved:
1. Selecting a diverse range of algorithms, such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Naive Bayes, K-Nearest Neighbors, and potentially an Artificial Neural Network architecture.
2. Splitting the data into training and testing sets. The training set is used to train the models, allowing them to learn the underlying patterns that differentiate PD from HC recordings in the voice data.
3. Training each chosen model on the training set.
4. Employing hyperparameter tuning (optional) to optimize the models' performance by adjusting their internal settings.

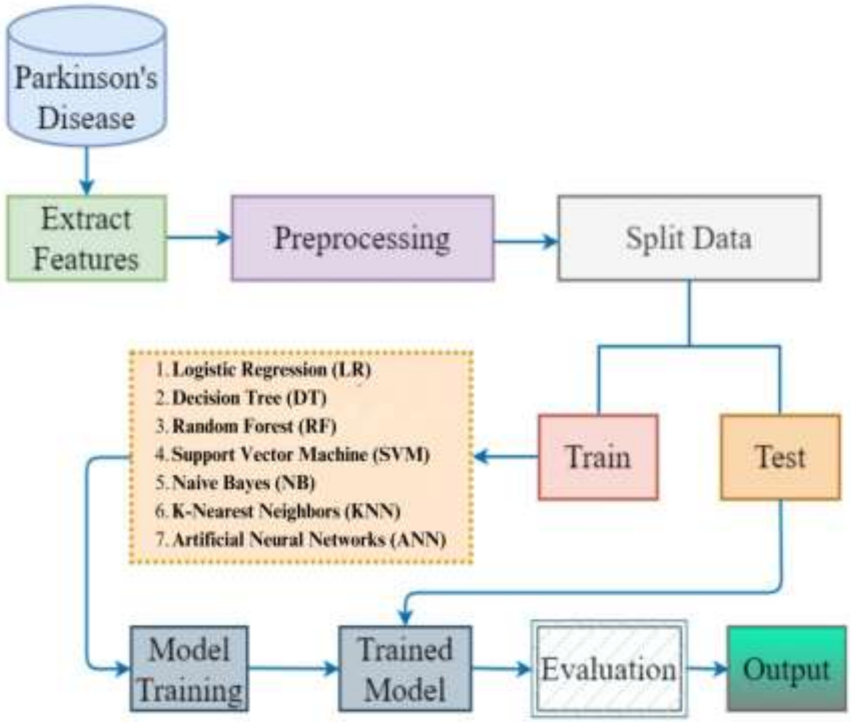

Figure 1:- Flowchart of the proposed work.

Model Evaluation:
1. The performance of each trained model was rigorously evaluated on unseen data using the testing set.
2. Established classification metrics like accuracy, precision, recall, and F1-score were used to assess the models' ability to correctly classify PD recordings.

Model Selection:
By comparing the evaluation metrics across all models, we identified the model that demonstrated the most robust and accurate performance in classifying PD recordings from HC recordings within the Max Little dataset.

Methodology:-

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. The original study published the feature extraction methods for general voice disorders.

Dataset 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.

Dataset Characteristic	Multivariate
No. of Instances	197
Attributes Characteristic	Real
No. of Attributes	23
Missing Values	N/A
Made by	Max Little of the University of Oxford
Associated Tasks	Classification
Types of Classification	Binary {0 for healthy and 1 for PD patient}

Table 2:- Detail of Parkinson's Dataset.

Dataset Attributes:

Attribute Name	**Description**
name	Unique identifier for each subject recording (e.g., "subject1_recording02")
MDVP:Fo(Hz)	Average vocal fundamental frequency (perceived pitch) in Hertz (Hz)
MDVP:Fhi(Hz)	Maximum vocal fundamental frequency (Hz)
MDVP:Flo(Hz)	Minimum vocal fundamental frequency (Hz)
MDVP:Jitter(%)	Variation in fundamental frequency over time (%)
MDVP:Jitter(Abs)	Absolute variation in fundamental frequency
MDVP:RAP	Ratio of Average Period to Average Amplitude variation
MDVP:PPQ	Normalized logarithmic measure of variation in fundamental frequency
Jitter:DDP	Local detrended fluctuation in fundamental frequency
MDVP:Shimmer	Variation in amplitude of the voice signal over time
MDVP:Shimmer(dB)	Amplitude variation in decibels (dB)
Shimmer:APQ3	Amplitude variation based on the 3rd quartile
Shimmer:APQ5	Amplitude variation based on the 5th quartile
MDVP:APQ	Amplitude variation measure
Shimmer:DDA	Local detrended fluctuation in amplitude variation
NHR	Ratio of noise to tonal components in the voice
HNR	Harmonic-to-Noise Ratio
status	Health status (1: Parkinson's, 0: Healthy)
RPDE	Nonlinear complexity measure 1
D2	Nonlinear complexity measure 2
DFA	Signal complexity measure (fractal scaling exponent)
spread1	Nonlinear measure of fundamental frequency variation 1
spread2	Nonlinear measure of fundamental frequency variation 2
PPE	Normalized log-area variation measure of fundamental frequency variation

Table 3:- Detail of Parkinson's Dataset Attributes.

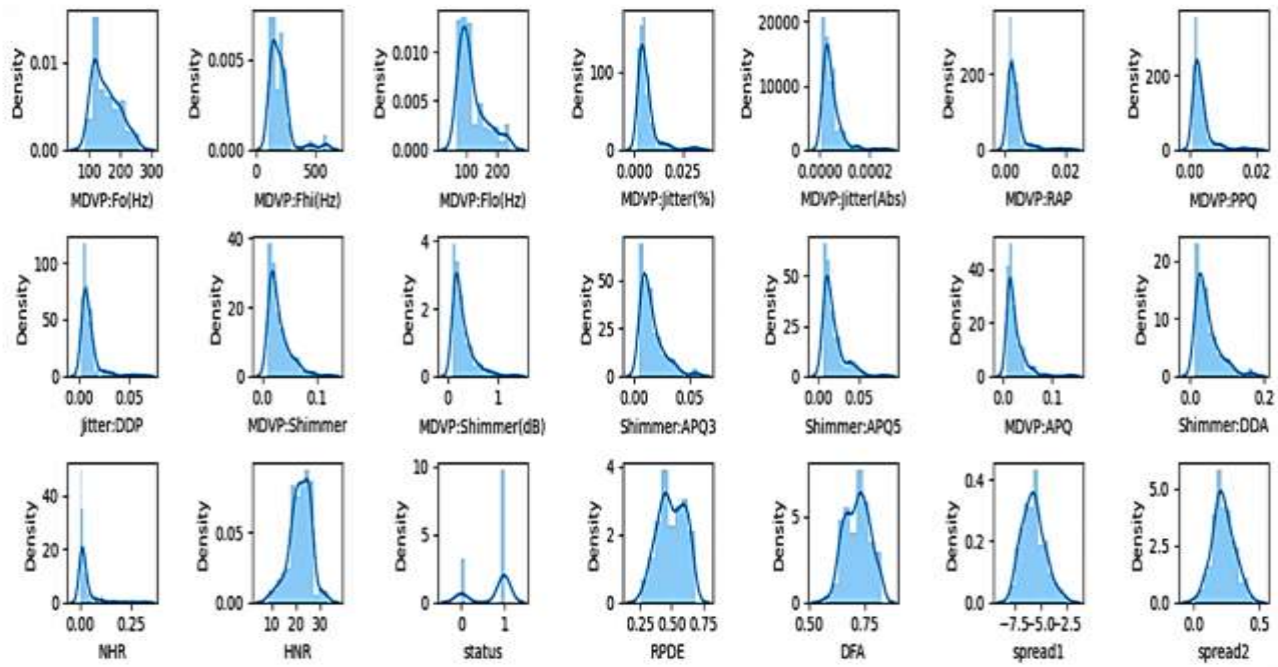

Figure 2:- Distribution plot displays a distribution and range of a set of numeric values plotted against a dimension.

Machine Learning And Deep Learning Classification Algorithms

This section explores the application of various machine learning algorithms for the diagnosis of Parkinson's Disease (PD) based on voice analysis data. We investigate the following algorithms and their potential suitability for this task:

Logistic Regression (LR):

1. **Strengths:** LR is a well-established linear classification algorithm. It excels at interpreting the coefficients of the model, providing insights into the features that most significantly contribute to PD classification. Additionally, LR offers efficient training and is relatively less prone to overfitting compared to more complex models.
2. **Limitations:** LR assumes a linear relationship between features and the target variable (presence/absence of PD). If the underlying relationships are non-linear, LR might not achieve optimal performance.

Decision Tree (DT):

1. **Strengths:** DT is a flexible and interpretable algorithm that can handle both continuous and categorical features without extensive data preprocessing. It builds a tree-like structure where each node represents a decision based on a specific feature. This structure allows for easy visualization and understanding of the decision-making process.
2. **Limitations:** DTs can be susceptible to overfitting, particularly with deep trees and high dimensionality. Additionally, they may be sensitive to small variations in the training data.

Random Forest (RF):

1. **Strengths:** RF addresses the overfitting limitations of DTs by constructing an ensemble of multiple decision trees trained on random subsets of features and data points. This approach reduces variance and enhances the model's generalizability to unseen data. RF also offers robustness to outliers and missing values.
2. **Limitations:** While interpretability is lower compared to individual DTs, techniques like feature importance analysis can still provide insights into the most influential features. RF can be computationally expensive to train compared to simpler models like LR.

Naive Bayes (NB):

1. **Strengths:** NB is a probabilistic classifier based on Bayes' theorem. It assumes conditional independence between features, which can be a simplifying assumption but may be suitable for certain types of voice data. NB is efficient for training and can handle high dimensionality.
2. **Limitations:** The conditional independence assumption may not always hold true for real-world data, potentially impacting classification accuracy. Additionally, NB may struggle with imbalanced datasets where one class (e.g., PD) has significantly fewer samples compared to the other (healthy controls).

K-Nearest Neighbors (KNN):

1. **Strengths:** KNN is a simple and intuitive non-parametric classification algorithm. It classifies a new data point based on the majority vote of its K nearest neighbors in the training data. KNN requires minimal training and can handle various feature types.
2. **Limitations:** KNN's performance is highly dependent on the choice of the K parameter (number of neighbors) and the distance metric used. Additionally, KNN can be computationally expensive for large datasets due to the need to compare new data points with all data points in the training set.

Support Vector Machine (SVM):

1. **Strengths:** SVMs are powerful algorithms that excel in high-dimensional feature spaces and can handle non-linear relationships through the use of kernel functions. They are also robust to outliers and efficient in terms of memory usage during training.
2. **Limitations:** SVMs can be challenging to tune due to the presence of hyperparameters (kernel type, regularization parameter). Additionally, they may not provide clear interpretability of the model's decision-making process.

Artificial Neural Networks (ANN):

1. **Strengths:** ANNs are powerful learning models inspired by the structure and function of the human brain. They consist of interconnected nodes (artificial neurons) arranged in layers. ANNs can learn complex non-linear relationships between features and the target variable, potentially achieving high accuracy on classification tasks.
2. **Limitations:** ANNs are often considered "black boxes" due to their complex internal structure. This can make interpretability challenging. Additionally, training ANNs can be computationally expensive and requires careful hyperparameter tuning to avoid overfitting.

Machine Learning Classification for Parkinson's Disease

This section explores the use of machine learning (ML) classifiers for PD classification. We begin by identifying the target variable, which in this case is the health status of the patient (presence or absence of PD). We then analyze the distribution of health statuses within the dataset and visualize this data graphically. A common approach involves splitting the data into two sets: a training set (typically 80%) used to train the ML model, and a testing set (20%) used to evaluate the model's performance on unseen data.

Figure 3 depicts the distribution of health statuses in our dataset. A value of "0" represents healthy individuals, with a count of 48. A value of "1" represents patients diagnosed with PD, with a count of 147. This translates to a prevalence of PD in the dataset of 75.38% (147 out of 195) and a healthy control group of 24.62% (48 out of 195).

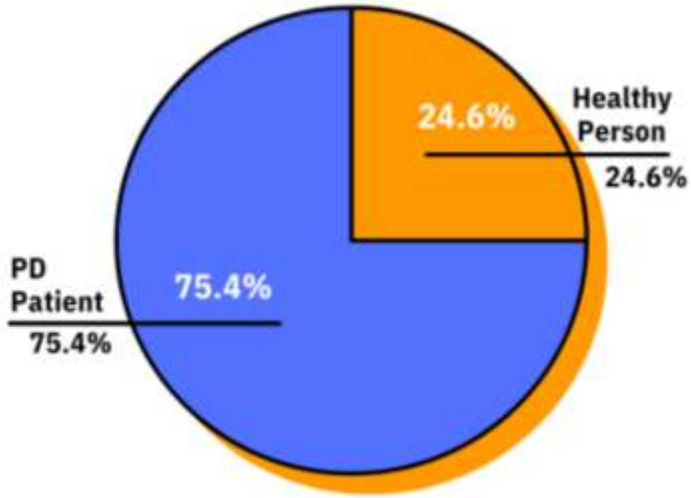

Figure 3:- Health Status of PD Patient.

Evaluation of Machine Learning Models for Parkinson's Disease Diagnosis
This section explores the evaluation of machine learning models employed for Parkinson's disease (PD) diagnosis.

Confusion Matrix
A confusion matrix is a visualization tool that summarizes a classification model's performance on a set of test data. It allows us to understand how well the model distinguishes between different classes (healthy vs. PD in this case). The matrix displays the number of correctly and incorrectly classified instances based on the model's predictions.

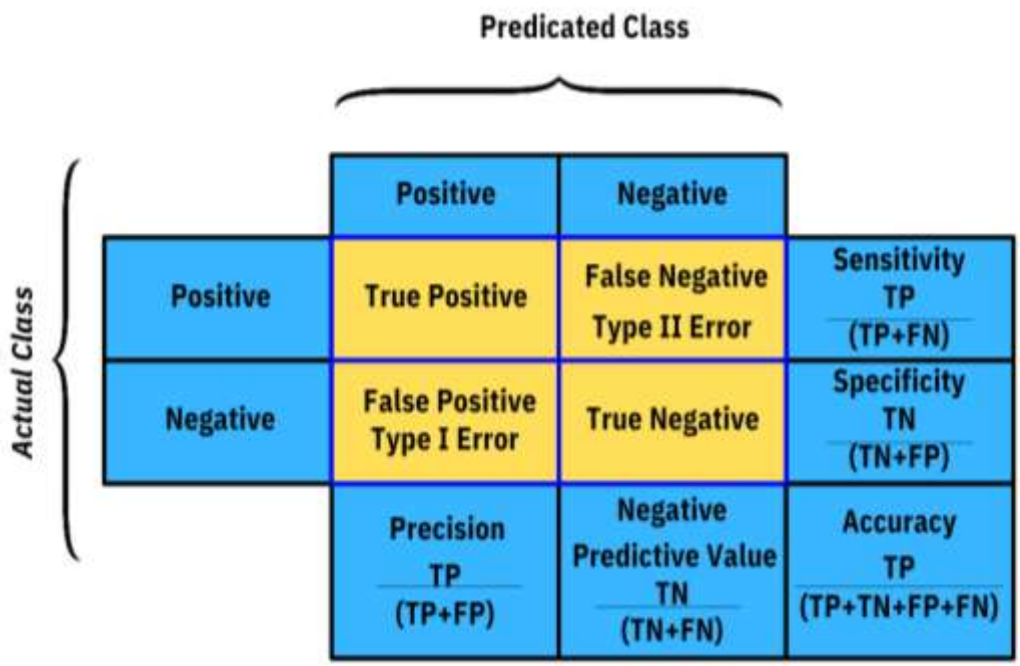

Figure 4:- Confusion Matrix for Model Evaluation.

Key Metrics Derived from the Confusion Matrix:
1. **True Positives (TP):** These represent instances where the model correctly identifies a patient with PD (positive class).
2. **True Negatives (TN):** These represent instances where the model correctly identifies a healthy individual (negative class).

3. **False Positives (FP):** These represent instances where the model incorrectly classifies a healthy individual as having PD (incorrect positive prediction).
4. **False Negatives (FN):** These represent instances where the model incorrectly classifies a patient with PD as healthy (incorrect negative prediction).

By analyzing these values within the confusion matrix, we have calculated various performance metrics to assess the effectiveness of the machine learning models for PD diagnosis.

Performance metrics for PD diagnosis:

1. **Accuracy:** Overall proportion of correctly classified cases
   Accuracy = (TP + TN) / (Total samples)
2. **Precision:** Proportion of true positives among all predicted positive cases
   Precision = (TP / (TP + FP))
3. **Recall:** Proportion of true positives among all actual positive cases
   Recall = (TP / (TP + FN))
4. **F1-Score:** Harmonic mean of precision and recall, providing a balanced view of model performance
   F1-score = 2 * (Precision * Recall) / (Precision + Recall)

By evaluating these metrics for different machine learning models, we have identified the model that achieves the most accurate and reliable classification for PD diagnosis based on the chosen features.

Kappa Statistic for Evaluating Inter-Rater Reliability in PD Diagnosis

While the confusion matrix provides valuable insights into a machine learning model's performance for PD diagnosis, it doesn't necessarily address the question of agreement between the model's predictions and a potential "gold standard" diagnosis by a human expert. Here, the Kappa statistic (κ) emerges as a valuable tool for assessing inter-rater reliability.

Understanding Kappa:

Kappa is a statistical measure that goes beyond simple agreement between two raters (model and human expert in this case). It considers the agreement that occurs by chance and focuses on the agreement beyond this random chance. Unlike the percentage agreement, which can be misleading, Kappa provides a more robust measure of agreement, ranging from -1 to 1.

Interpreting Kappa Values:

1. **$\kappa < 0$:** Indicates disagreement worse than chance.
2. **$0 \leq \kappa \leq 0.20$:** Represents slight agreement.
3. **$0.21 \leq \kappa \leq 0.40$:** Indicates fair agreement.
4. **$0.41 \leq \kappa \leq 0.60$:** Represents moderate agreement.
5. **$0.61 \leq \kappa \leq 0.80$:** Suggests substantial agreement.
6. **$0.81 \leq \kappa \leq 1.00$:** Indicates almost perfect agreement.

Formula for Kappa Score:

The Kappa statistic is calculated using the following formula:

$\kappa = (P(A) - P(E)) / (1 - P(E))$

Where:

1. **P(A):** Represents the observed agreement between the model and the human expert. This is calculated as the sum of the diagonal elements of the confusion matrix divided by the total number of samples.
2. **P(E):** Represents the expected agreement by chance. This is calculated by summing the product of row and column totals in the confusion matrix (excluding diagonal elements) and then dividing by the total number of samples squared.

Experiments and Results:-

The proposed work, The Machine Learning algorithms including Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbors (KNN) are implemented in Python 3.11.4: Jupyter Notebook And Deep Learning Algorithm Artificial Neural Networks (ANN) is implemented in Python 3.10.12: Google Colab. Here we detail the experimental setup and the results of the Total Seven machine learning and Deep Learning classification methods.

Logistic Regression (LR):

Title	Results
Training Accuracy	87.18%
Testing Accuracy	84.62%
Precision	80.65%
Recall	78.13%
F1-Score	79.37%
Kappa Score	0.3546

Table 4:- Performance Analysis for Logistic Regression (LR) Classifier.

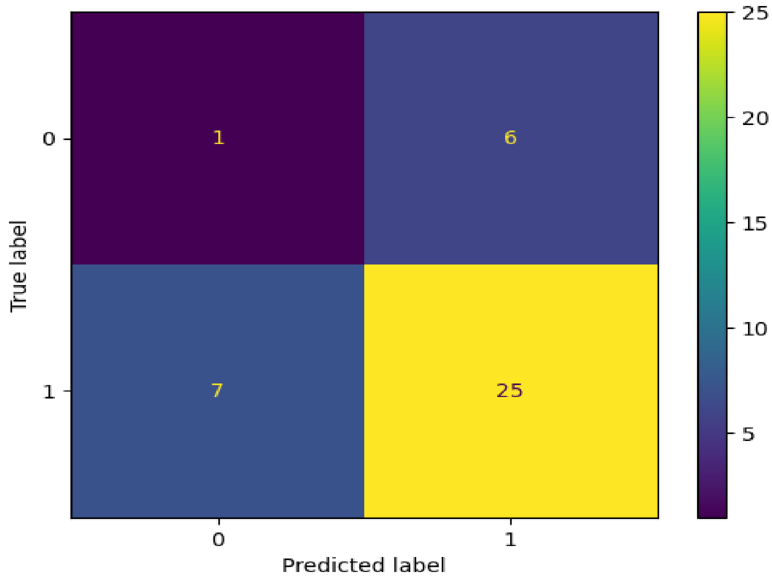

Figure 5:- Confusion Matrix and Heatmap for Logistic Regression (LR) Classifier.

Decision Tree (DT):

Title	Results
Training Accuracy	100%
Testing Accuracy	100%
Precision	77.42%
Recall	75.00%
F1-Score	76.16%
Kappa Score	1.00

Table 5:- Performance Analysis for Decision Tree (DT) Classifier.

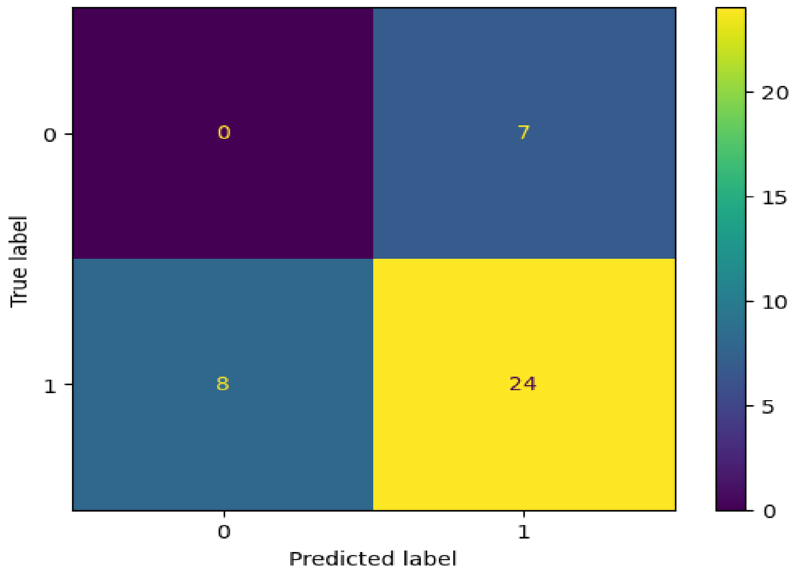

Figure 6:- Confusion Matrix and Heatmap for Decision Tree (DT) Classifier.

Random Forest (RF):

Title	**Results**
Training Accuracy	100%
Testing Accuracy	89.74%
Precision	81.25%
Recall	81.25%
F1-Score	81.25%
Kappa Score	0.6855

Table 6:- Performance Analysis for Random Forest (RF) Classifier.

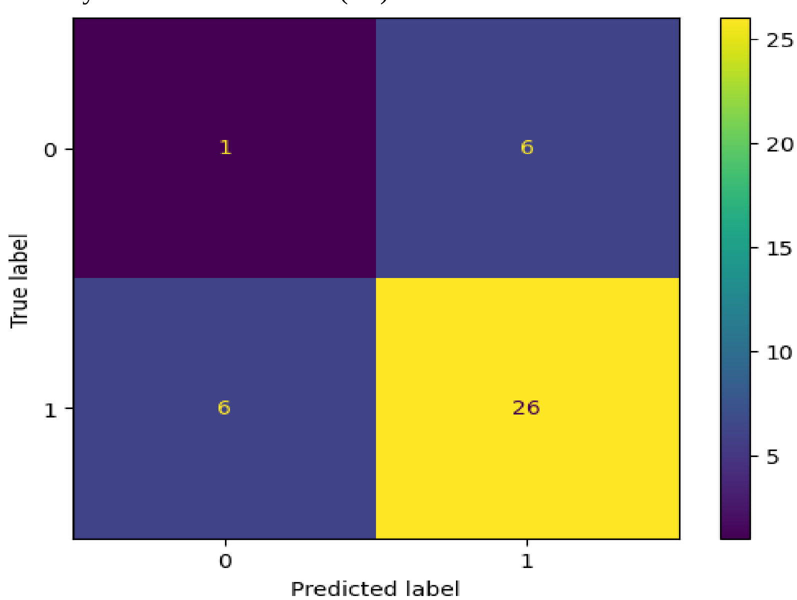

Figure 7:- Confusion Matrix and Heatmap for Random Forest (RF) Classifier.

Naive Bayes (NB):

Title	Results
Training Accuracy	71.79%
Testing Accuracy	69.23%
Precision	84.21%
Recall	50.00%
F1-Score	62.75%
Kappa Score	0.0414

Table 7:- Performance Analysis for Naive Bayes (NB) Classifier.

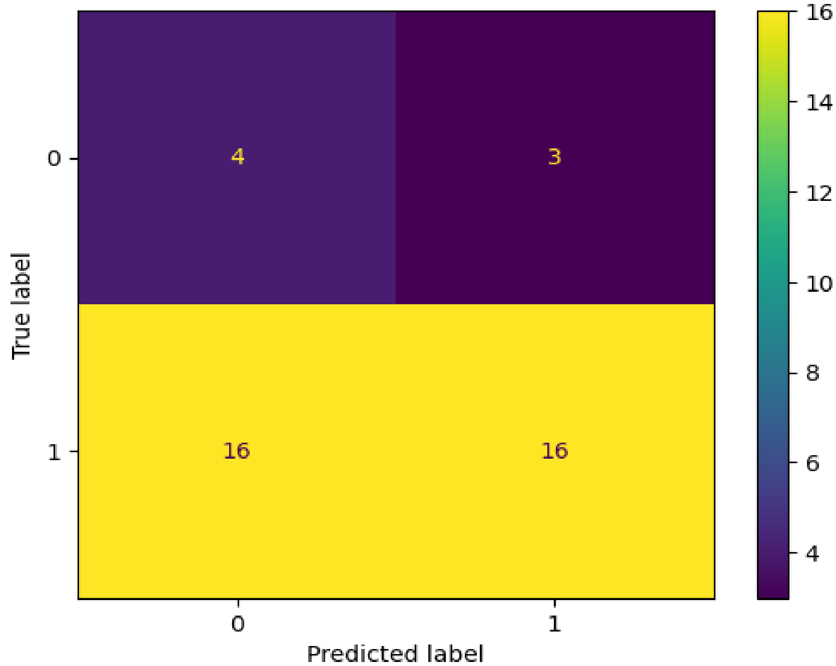

Figure 8:- Confusion Matrix and Heatmap for Naive Bayes (NB) Classifier.

K-Nearest Neighbors (KNN):

Title	Results
Training Accuracy	89.74%
Testing Accuracy	82.05%
Precision	87.88%
Recall	90.63%
F1-Score	89.23%
Kappa Score	0.3546

Table 8:- Performance Analysis for K-Nearest Neighbors (KNN) Classifier.

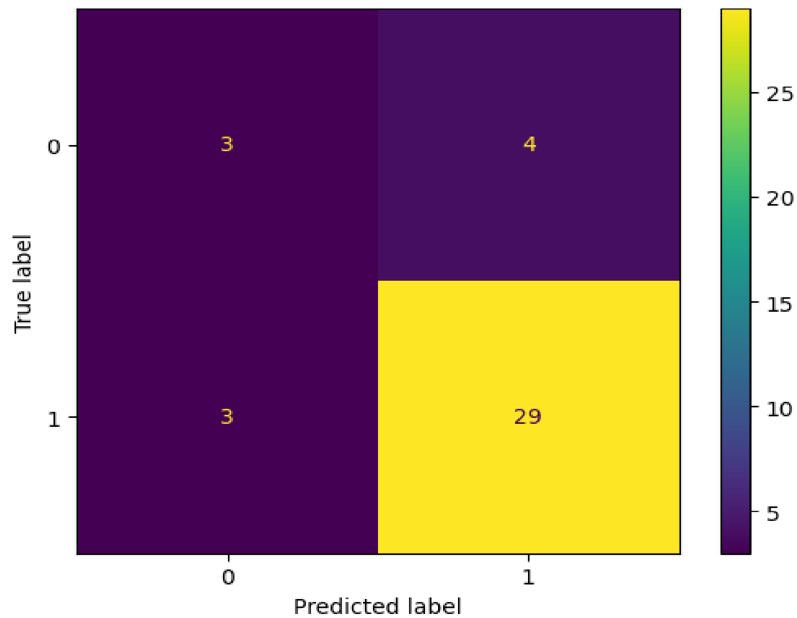

Figure 9:- Confusion Matrix and Heatmap for K-Nearest Neighbors (KNN) Classifier.

Support Vector Machine (SVM):

Title	**Results**
Training Accuracy	86.54%
Testing Accuracy	87.18%
Precision	88.57%
Recall	96.88%
F1-Score	92.54%
Kappa Score	0.4772

Table 9:- Performance Analysis for Support Vector Machine (SVM) Classifier.

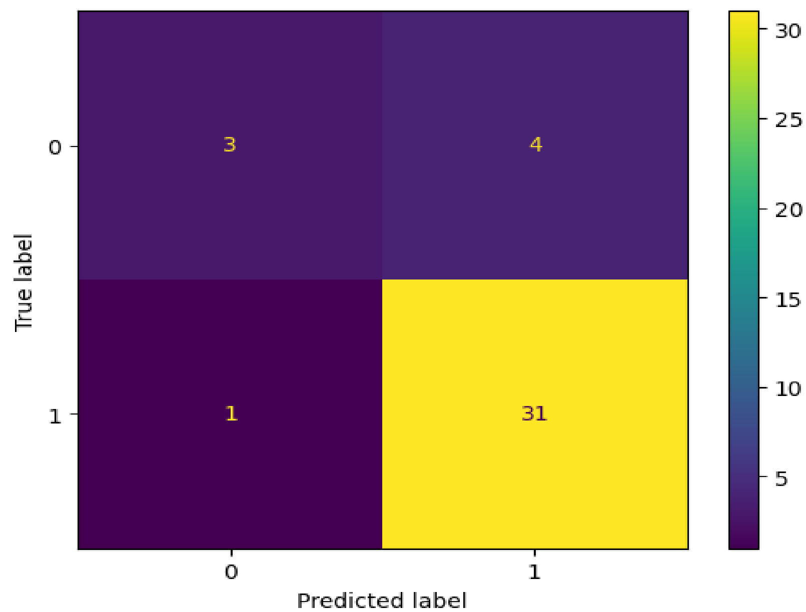

Figure 10:- Confusion Matrix and Heatmap for Support Vector Machine (SVM) Classifier.

Artificial Neural Networks (ANN):

Title	Results
Training Accuracy	100%
Testing Accuracy	97.44%
Precision	96.67%
Recall	100%
F1-Score	98.31%
Kappa Score	0.9305

Table 10:- Performance Analysis for Artificial Neural Networks (ANN) Classifier.

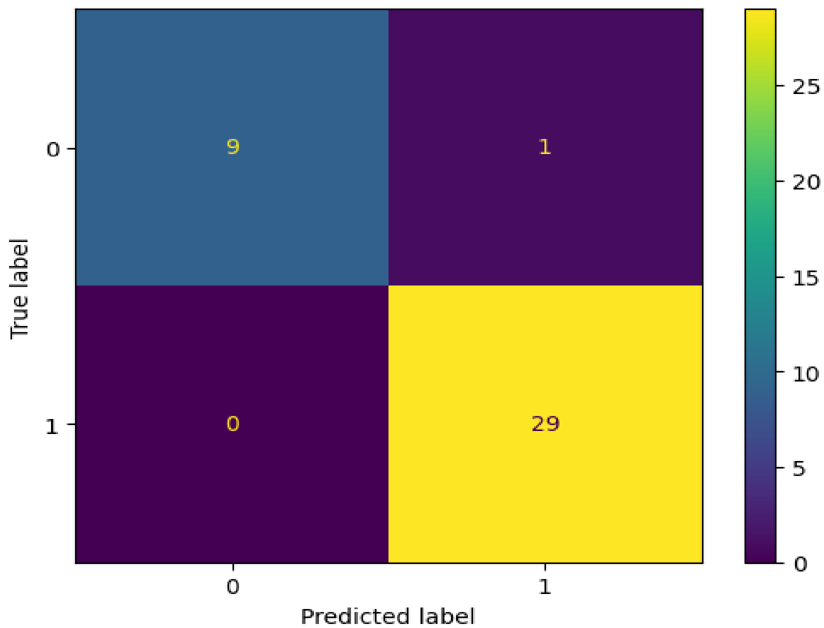

Figure 11:- Confusion Matrix and Heatmap for Artificial Neural Networks (ANN) Classifier.

Comparative Study of Machine Learning Algorithms Used in Proposed Work

Classifier	Training Accuracy	Testing Accuracy	Precision	Recall	F1-Score	Kappa Score
Logistic Regression (LR)	87.18%	84.62%	80.65%	78.13%	79.37%	0.3546
Decision Tree (DT)	100%	100%	77.42%	75.00%	76.19%	1.00
Random Forest (RF)	100%	89.74%	81.25%	81.25%	81.25%	0.6855
Naive Bayes (NB)	71.79%	69.23%	84.21%	50.00%	62.75%	0.0414
K-Nearest Neighbors (KNN)	89.74%	82.05%	87.88%	90.63%	89.23%	0.3546
Support Vector Machine (SVM)	86.54%	87.18%	88.57%	96.88%	92.54%	0.4772

Artificial Neural Networks (ANN)	100%	97.44%	96.67	100%	98.31%	0.9305

Table 11:- An overview of evaluation results and Performance Analysis for all Classifiers used in Proposed Work.

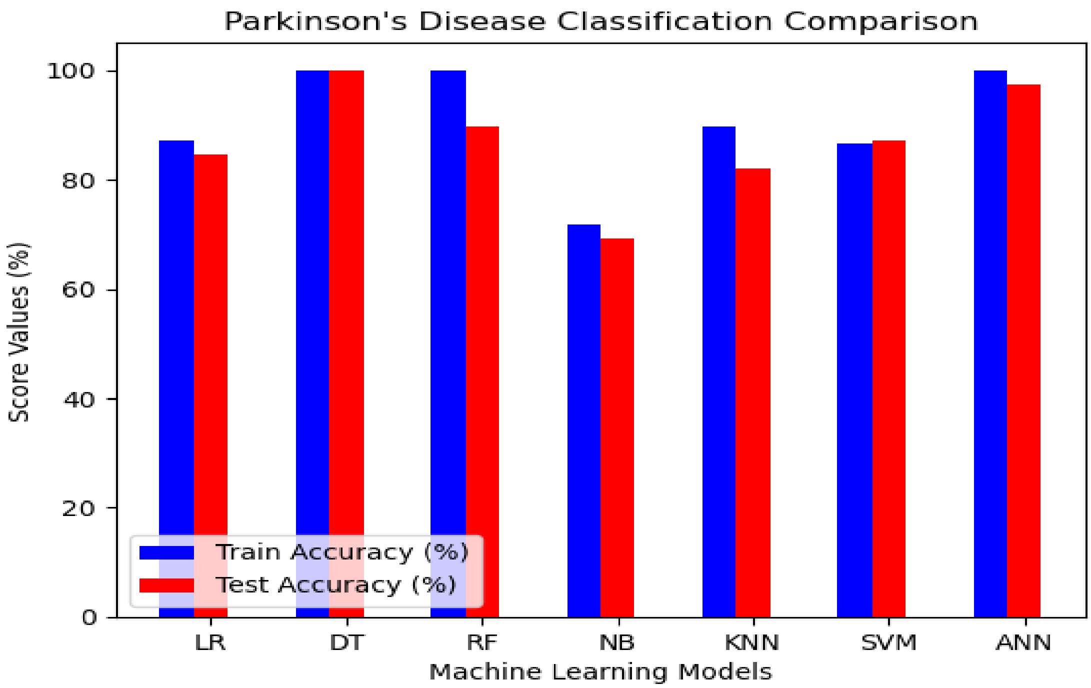

Figure 12:- Graphical Representation of Comparison of Training And Testing Accuracy for all Classifiers.

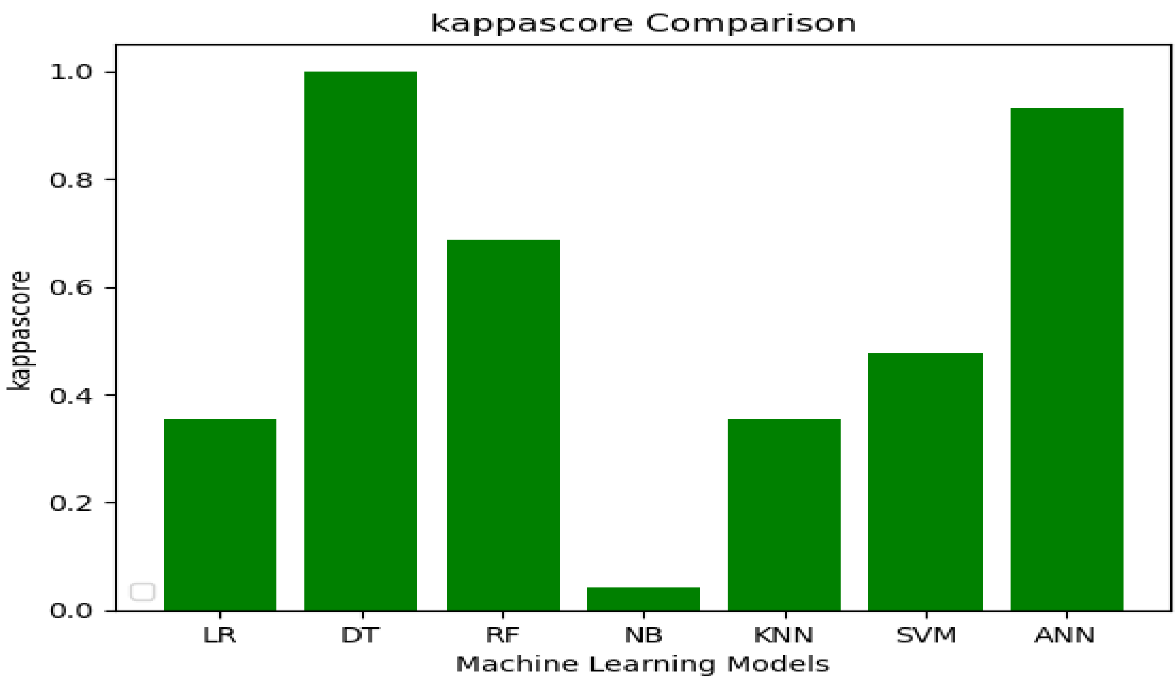

Figure 13:- Graphical Representation of Comparison of Kappa Score for all Classifiers.

Conclusion:-

This research investigated the potential of automated machine learning (ML) and deep learning (DL) techniques to classify Parkinson's disease (PD) from healthy controls (HC) based on non-invasive speech biomarkers. Our study focused on comparing the performance of various classifiers in handling the challenges of noisy and high-dimensional speech data, common in real-world applications. The findings demonstrate that achieving clinical-level accuracy for PD detection is feasible with careful feature selection and appropriate model selection.

Among the evaluated algorithms, Logistic Regression (LR) achieved an accuracy of 84.62%, Decision Tree (DT) achieved 100% accuracy, Random Forest (RF) achieved 89.74% accuracy, Naive Bayes (NB) achieved 69.23% accuracy, K-Nearest Neighbors (KNN) achieved 82.05% accuracy, Support Vector Machine (SVM) achieved 87.18% accuracy, and Artificial Neural Networks (ANNs) achieved the highest accuracy of 97.44%. It's important to note that while the Decision Tree classifier achieved the highest reported accuracy, it is susceptible to overfitting, which can lead to poor performance on unseen data.

This research highlights the significant advantage of Artificial Neural Networks (ANNs) for PD classification using speech analysis. The deep learning model achieved an impressive accuracy of 97.44%, significantly outperforming other methods. ANNs' inherent ability to learn complex, non-linear relationships within the data offers a clear advantage for this specific task.

References:-

[1] DeMaagd, G., & Philip, A. (2015). Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis. Pharmacotherapy, 40, 504–532.

[2] Rizek, P., Kumar, N., & Jog, M. S. (2016). An update on the diagnosis and treatment of Parkinson disease. Canadian Medical Association Journal, 188(16), 1157–1165.

[3] World Health Organization. (n.d.). Parkinson's disease. Retrieved October 30, 2022, from https://www.who.int/news-room/fact-sheets/detail/parkinson-disease

[4] de Rijk, M. C., Launer, L. J., Berger, K., Breteler, M. M., Dartigues, J. F., Baldereschi, M., ... & Trenkwalder, C. (2000). Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurology, 54(Suppl. 5), S21–S23.

[5] Cantürk, İ., & Karabiber, F. (2016). A machine learning system for the diagnosis of Parkinson's disease from speech signals and its application to multiple speech signal types. Arabian Journal for Science and Engineering, 41, 5049–5059.

[6] Singh, N., Pillay, V., & Choonara, Y. E. (2007). Advances in the treatment of Parkinson's disease. Progress in Neurobiology, 81, 29–44.

[7] Rana, A., Rawat, A. S., Bijalwan, A., & Bahuguna, H. (2018). Application of multi-layer (perceptron) artificial neural network in the diagnosis system: A systematic review. In Proceedings of the 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE) (pp. 1–6). San Salvador, El Salvador.

[8] Lakany, H. (2008). Extracting a diagnostic gait signature. Pattern Recognition, 41, 1627–1637.

[9] Figueiredo, J., Santos, C. P., & Moreno, J. C. (2018). Automatic recognition of gait patterns in human motor disorders using machine learning: A review. Medical Engineering & Physics, 53, 1–12. https://doi.org/10.1016/j.medengphy.2017.12.006.

[10] Hazan, H., Hilu, D., Manevitz, L., Ramig, L. O., & Sapir, S. (2012). Early diagnosis of Parkinson's disease via machine learning on speech data. In Proceedings of the 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel (pp. 1–4). Eilat, Israel. https://doi.org/10.1109/eeei.2012.6377065.

[11] Karan, B., Sahu, S. S., & Mahto, K. (2019). Parkinson disease prediction using intrinsic mode function based features from speech signal. Biocybernetics and Biomedical Engineering, 40, 249–264. https://doi.org/10.1016/j.bbe.2019.05.005.

[12] Frid, A.; Safra, E.J.; Hazan, H.; Lokey, L.L.; Hilu, D.; Manevitz, L.; Ramig, L.O.; Sapir, S. Computational diagnosis of Parkinson's Disease directly from natural speech using machine learning techniques. In Proceedings of the 2014 IEEE International Conference on Software Science, Technology and Engineering, Washington, DC, USA, 11–12 June 2014; pp. 50–53.

[13] Rawat, A. S., Rana, A., Kumar, A., & Bagwari, A. (2018). Application of multi layer artificial neural network in the diagnosis system: a systematic review. IAES International Journal of Artificial Intelligence, 7(3), 138.

[14] Little, M.A., McSharry, P.E., Roberts, S.J. et al. Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection. BioMed Eng OnLine 6, 23 (2007). https://doi.org/10.1186/1475-925X-6-

23
-[15] Little, M.; McSharry, P.; Hunter, E.; Spielman, J.; Ramig, L. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Nat. Preced. 2008, 1-1. https://doi.org/10.1038/npre.2008.2298.1.
[16] Lichman, M. UCI Machine Learning Repository; University of California, School of Information and Computer Science: Irvine, CA, USA. Available online: http://archive.ics.uci.edu/ml (accessed on 25 September 2022).
[17] Rana, Arti & Dumka, Ankur & Singh, Rajesh & Rashid, Mamoon & Ahmad, Nazir & Panda, Manoj. (2022). An Efficient Machine Learning Approach for Diagnosing Parkinson's Disease by Utilizing Voice Features. Electronics. 11. 3782. 10.3390/electronics11223782.
[18] Sakar, B.E.; Isenkul, M.E.; Sakar, C.O.; Sertbas, A.; Gurgen, F.; Delil, S.; Apaydin, H.; Kursun, O. Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings. IEEE J. Biomed. Health Inform. 2013, 17, 828–834. https://doi.org/10.1109/jbhi.2013.2245674.
[19] Vadovský, M.; Paralič, J. Parkinson's disease patients classification based on the speech signals. In Proceedings of the 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), Herlany, Slovakia, 26–28 January 2017; pp. 000321–000326.
[20] Ouhmida, A., Raihani, A., Cherradi, B., & Terrada, O. (2021). A Novel Approach for Parkinson's Disease Detection Based on Voice Classification and Features Selection Techniques. International Journal of Online & Biomedical Engineering, 17(10).
[21] Mabrouk, R.; Chikhaoui, B.; Bentabet, L. Machine Learning Based Classification Using Clinical and DaTSCAN SPECT Imaging Features: A Study on Parkinson's Disease and SWEDD. IEEE Trans. Radiat. Plasma Med. Sci. 2018, 3, 170–177. https://doi.org/10.1109/trpms.2018.2877754.
[22] Benba, A.; Jilbab, A.; Hammouch, A. Using Human Factor Cepstral Coefficient on Multiple Types of Voice Recordings for Detecting Patients with Parkinson's Disease. Irbm 2017, 38, 346–351. https://doi.org/10.1016/j.irbm.2017.10.002.
[23] Sakar, C.O.; Serbes, G.; Gunduz, A.; Tunc, H.C.; Nizam, H.; Sakar, B.E.; Tutuncu, M.; Aydin, T.; Isenkul, M.E.; Apaydin, H. A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable factor wavelet transform. Appl. Soft Comput. 2019, 74, 255–263.
[24] Yasar, A.; Saritas, I.; Sahman, M.A.; Cinar, A.C. Classification of Parkinson disease data with artificial neural networks. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Wuhan, China, 10–12 October 2019; Volume 675, p.012031.
[25] Avuçlu, E.; Elen, A. Evaluation of train and test performance of machine learning algorithms and Parkinson diagnosis with statistical measurements. Med. Biol. Eng. Comput. 2020, 58, 2775–2788. https://doi.org/10.1007/s11517-020-02260-3.
[26] Marar, S.; Swain, D.; Hiwarkar, V.; Motwani, N.; Awari, A. Predicting the occurrence of Parkinson's Disease using various Classification Models. In Proceedings of the 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), Bhopal, India, 28–29 December 2018; pp. 1–5.
[27] Nikookar, E.; Sheibani, R.; Alavi, S.E. An ensemble method for diagnosis of Parkinson's disease based on voice measurements. J. Med. Signals Sens. 2019, 9, 221–226. https://doi.org/10.4103/jmss.jmss_57_18.
[28] Tracy, J.M.; Özkanca, Y.; Atkins, D.C.; Ghomi, R.H. Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease. J. Biomed. Inform. 2019, 104, 103362. https://doi.org/10.1016/j.jbi.2019.103362.
[29] Cibulka, M.; Brodnanova, M.; Grendar, M.; Grofik, M.; Kurca, E.; Pilchova, I.; Osina, O.; Tatarkova, Z.; Dobrota, D.; Kolisek, M. SNPs rs11240569, rs708727, and rs823156 in SLC41A1 Do Not Discriminate Between Slovak Patients with Idiopathic Parkinson's Disease and Healthy Controls: Statistics and Machine-Learning Evidence. Int. J. Mol. Sci. 2019, 20, 4688. https://doi.org/10.3390/ijms20194688. for Parkinson
[30]. Hsu, S.-Y.; Lin, H.-C.; Chen, T.-B.; Du, W.-C.; Hsu, Y.-H.; Wu, Y.-C.; Tu, P.-W.; Huang, Y.-H.; Chen, H.-Y. Feasible Classified Models Disease https://doi.org/10.3390/s19071740. from 99mTc-TRODAT-1 SPECT Imaging. Sensors 2019, 19, 1740.
[31] Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., & Faundez-Zanuy, M. (2016). Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. Artificial intelligence in Medicine, 67, 39-46.
[32] Maass, F.; Michalke, B.; Willkommen, D.; Leha, A.; Schulte, C.; Tönges, L.; Mollenhauer, B.; Trenkwalder, C.; Rückamp, D.; Börger, M.; et al. Elemental fingerprint: Reassessment of a cerebrospinal fluid biomarker for Parkinson's disease. Neurobiol. Dis. 2019, 134, 104677. https://doi.org/10.1016/j.nbd.2019.104677.
[33] Mucha, J.; Mekyska, J.; Faundez-Zanuy, M.; Lopez-De-Ipina, K.; Zvoncak, V.; Galaz, Z.; Kiska, T.; Smekal, Z.; Brabenec, L.; Rektorova, I. Advanced Parkinson's Disease Dysgraphia Analysis Based on Fractional Derivatives

of Online Handwriting. In Proceedings of the 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Moscow, Russia, 5–9 November 2018; pp. 1–6. https://doi.org/10.1109/icumt.2018.8631265.
[34] Wenzel, M.; Milletari, F.; Krüger, J.; Lange, C.; Schenk, M.; Apostolova, I.; Klutmann, S.; Ehrenburg, M.; Buchert, R. Automatic classification of dopamine transporter SPECT: Deep convolutional neural networks can be trained to be robust with respect to variable image characteristics. Eur. J. Pediatr. 2019, 46, 2800–2811. https://doi.org/10.1007/s00259-019-04502-5.
[35] Segovia, F.; Gorriz, J.M.; Ramirez, J.; Martinez-Murcia, F.J.; Castillo-Barnes, D. Assisted Diagnosis of Parkinsonism Based on the Striatal Morphology. Int. J. Neural Syst. 2019, 29, 1950011. https://doi.org/10.1142/s0129065719500114.
[36] Ye, Q.; Xia, Y.; Yao, Z. Classification of Gait Patterns in Patients with Neurodegenerative Disease Using Adaptive Neuro-Fuzzy Inference System. Comput. Math. Methods Med. 2018, 2018, 9831252. https://doi.org/10.1155/2018/9831252.
[37] Klomsae, A.; Auephanwiriyakul, S.; Theera-Umpon, N. (2018). String grammar unsupervised possibilistic fuzzy c-medians for gait pattern classification in patients with neurodegenerative diseases. Comput. Intell. Neurosci. 2018, 2018, 1869565.
[38] Felix, J.P.; Vieira, F.H.T.; Cardoso, A.A.; Ferreira, M.V.G.; Franco, R.A.P.; Ribeiro, M.A.; Araujo, S.G.; Correa, H.P.; Carneiro, M.L. A Parkinson's Disease Classification Method: An Approach Using Gait Dynamics and Detrended Fluctuation Analysis. In Proceedings of the 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, 5–8 May 2019. https://doi.org/10.1109/ccece.2019.8861759. Bioengineering
[39] Andrei, A.-G.; Tautan, A.-M.; Ionescu, B. Parkinson's Disease Detection from Gait Patterns. In Proceedings of the 2019 E-Health and Conference (EHB), https://doi.org/10.1109/ehb47216.2019.8969942. Iasi, Romania, 21–23 November 2019; pp. 1–4.
[40] Priya, S.J.; Rani, A.J.; Subathra, M.S.P.; Mohammed, M.A.; Damaševičius, R.; Ubendran, N. Local Pattern Transformation Based Feature Extraction for Recognition of Parkinson's Disease Based on Gait Signals. Diagnostics 2021, 11, 1395. https://doi.org/10.3390/diagnostics11081395.
[41] Yurdakul, O.C.; Subathra, M.; George, S.T. Detection of Parkinson's Disease from gait using Neighborhood Representation Local Binary Patterns. Biomed. Signal Process. Control 2020, 62, 102070. https://doi.org/10.1016/j.bspc.2020.102070.
[42] Li, B.; Yao, Z.; Wang, J.; Wang, S.; Yang, X.; Sun, Y. Improved Deep Learning Technique to Detect Freezing of Gait in Parkinson's Disease Based on Wearable Sensors. Electronics 2020, 9, 1919. https://doi.org/10.3390/electronics9111919.