International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 10 | Oct 2023 www.irjet.net p-ISSN: 2395-0072

Enhancing Parkinson's Disease Detection Through Machine Learning

SK. Abdul Sattar[1],T. Guru Preetham[2], V.Kalyan[3],P.Venu[4],B.Avinash[5]

[1,2,3,4]B.Tech. Students, Department of IT, Vasireddy Venkatadri Institute of Technology, Guntur
[5]Assistant Professor, Department of IT, Vasireddy Venkatadri Institute of Technology, Guntur

---

ABSTRACT:

Parkinson's Disease(PD) is a progressive neurodegenerative illness that primarily affects motor functions, reducing the affected individuals' quality of life. Its defining characteristics include bradykinesia, rigidity, tremor, and postural instability. The disease results from the death of dopamine-producing brain neurons. While Parkinson's disease has no known cure, early identification and treatment can significantly enhance a patient's quality of life. We used voice signals in our proposed model to identify Parkinson's disease. To further clean and improve the data quality, we applied data preprocessing techniques such as data imputation and data standardization. We employed several machine learning classifiers, including SVM and Random Forest, for Parkinson's disease identification. Our proposed model demonstrated an efficiency of over 90%. After experimenting with various algorithms, we achieved an accuracy of 90% for the SVM algorithm and 95% for the Random Forest algorithm. Comparing the two, we found that the Random Forest algorithm is the most effective for identifying Parkinson's disease.

Key Words: **Parkinson's disease**, **Neurodegenerative**, **Data preprocessing, Machine learning classifiers, Random Forest algorithm**

1.INTRODUCTION:

This paper delves into Parkinson's disease, a progressive neurodegenerative condition profoundly impacting motor functions and quality of life. Characterized by bradykinesia, rigidity, tremor, and postural instability, it results from the degeneration of dopamine-producing neurons. While a cure remains elusive, early diagnosis and intervention are pivotal. This study introduces a novel approach utilizing voice signals for Parkinson's disease identification. It employs rigorous data preprocessing techniques, including imputation and standardization, alongside machine learning classifiers like SVM and Random Forest. Notably, the Random Forest algorithm emerges as a preferred choice for effective disease detection.

2. LITERATURE REVIEW:

The prediction of Parkinson's disease has previously been done through MRI scans, gait analysis, genetic data, but there has been less research on audio impairment for early detection.

Using an SVM model, they examined genetic data to forecast when Parkinson's disease (PD) might manifest in elderly people. In they trained an SVM model to obtain an accuracy of 0.889. These findings support the advantages of using auditory data rather than genetic data for the categorization of Parkinson's disease[1].

They used a linear classification model to describe the shuffling movement of Parkinson's disease patients. Their study concentrated on the patient's gait, and subsequent research promoted the inclusion of audio and sleep data to enhance the outcomes[2].

Using ensemble learning approaches, another study used several classifiers to identify PD with an accuracy of 86% [3].

With an efficiency of 73.8%, this study introduces a voice-based model for Parkinson's disease identification. The research leverages a substantial dataset, encompassing both Parkinson's-affected individuals and those without the condition, and it employs machine learning algorithms, including XGBoost, Decision tree classifier and Naive Bayesian methods[4].

our study shows that our model is more accurate, efficient, and robust . As a result, it predicts the target outcome more accurately and runs more efficiently. Furthermore ,our model proves reliable in various scenarios, making it a practical choice for real-life situations.

3. OBJECTIVE:

The goal of this implementation is to effectively classify people as having Parkinson's disease or being healthy by using machine learning techniques, primarily a Random Forest Classifier. It aims to achieve high accuracy and strong predictive performance for PD diagnosis at an early stage.

4. METHODOLOGY:

4.1 Data Collection and Preparation:

I meticulously gathered voice recordings from individuals with and without Parkinson's Disease, addressing data quality through meticulous handling of missing values and outliers.

4.2 Machine Learning Model Implementation:

I implemented the Random Forest algorithm for Parkinson's Disease detection, rigorously training it on the preprocessed dataset to unveil crucial patterns and relationships.

4.3 Model Evaluation:

The model's performance was thoroughly assessed using precision, recall, F1 score, and accuracy.

4.4 Results Analysis and Implications:

Extensive analysis was conducted to measure the model's accuracy in Parkinson's Disease detection, revealing significant implications for early diagnosis and treatment.

5.DESIGN:

• ARCHITECTURE DIAGRAM

• DATA-FLOW-DIAGRAM

ARCHITECTURE DIAGRAM:

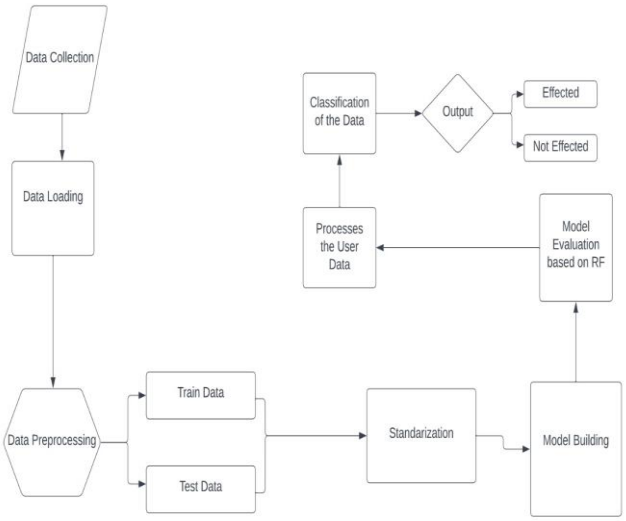

Figure 1: Architecture Diagram

DATA-FLOW-DIAGRAM:

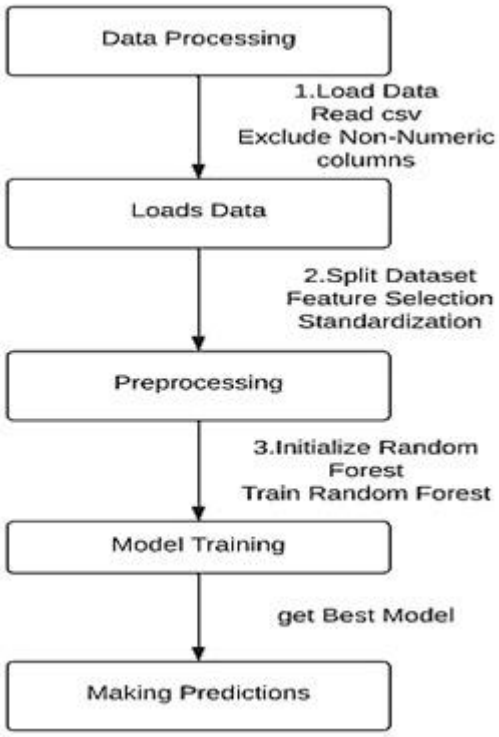

Figure 2: Data Flow Diagram

6.INPUT:

Screenshots:

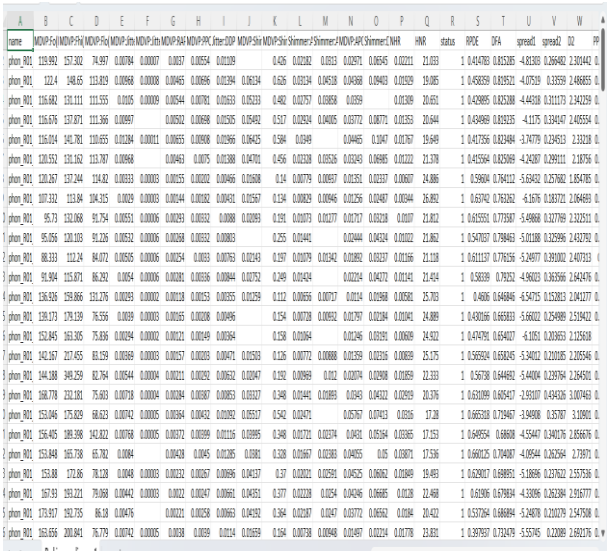

Figure 3: DATASET

```
# checking for missing values in each column
parkinsons_data.isnull().sum()
```

```
name                0
MDVP:Fo(Hz)         0
MDVP:Fhi(Hz)        0
MDVP:Flo(Hz)        0
MDVP:Jitter(%)      0
MDVP:Jitter(Abs)    4
MDVP:RAP            0
MDVP:PPQ            0
Jitter:DDP          0
MDVP:Shimmer        4
MDVP:Shimmer(dB)    0
Shimmer:APQ3        0
Shimmer:APQ5        5
MDVP:APQ            0
Shimmer:DDA         4
NHR                 0
HNR                 0
status              0
RPDE                0
DFA                 0
spread1             0
spread2             0
D2                  0
PPE                 0
dtype: int64
```

Figure 4: Checking missing values

7.IMPLEMENTATION:

7.1 Data Collection:

A collection of 31 biomedical voice measurements, 23 of whom have been diagnosed with Parkinson's disease (PD), make up the dataset we have assembled. In this dataset, the "name" column identifies each of the 195 voice recordings collected from these people, and each row corresponds to a particular voice measurement. This dataset's main goal is to make it possible to distinguish between people who are healthy and those who have Parkinson's disease (PD) using the "status" column, where a value of 0 indicates that a person is healthy and a value of 1 indicates that a person has PD. The ASCII CSV format of the data is accessible.

7.2 Data Preprocessing:

Data preprocessing is a vital machine learning technique for enhancing data quality. This process includes crucial steps to clean, format, and eliminate erroneous values from the dataset. There are a few steps in data pre-processing that we are going to talk about in order for it to be cleaned up, formatted, and free of any garbage values.

7.2.1 Handling Missing Values:

In this implementation, we adopted median imputation as our approach to manage missing values, ensuring data completeness and quality.

```
median_values = parkinsons_data[['MDVP:Shimmer', 'MDVP:Jitter(Abs)', 'Shimmer:APQ5', 'Shimmer:DDA']].median()
parkinsons_data['MDVP:Shimmer'].fillna(median_values['MDVP:Shimmer'], inplace=True)
parkinsons_data['MDVP:Jitter(Abs)'].fillna(median_values['MDVP:Jitter(Abs)'], inplace=True)
parkinsons_data['Shimmer:APQ5'].fillna(median_values['Shimmer:APQ5'], inplace=True)
parkinsons_data['Shimmer:DDA'].fillna(median_values['Shimmer:DDA'], inplace=True)
```

Figure 5: Handling Missing Values

7.2.2 Splitting Data:

The dataset was partitioned into training and testing subsets using scikit-learn's "train_test_split" function, facilitating robust model evaluation and performance assessment.

```
X_train, X_test, y_train, y_test = train_test_split(X_numeric, y, test_size=0.1, random_state=42, stratify=y)
```

Figure 6: Splitting data

7.2.3 Feature Scaling:

In this implementation, we implemented the "StandardScaler" from scikit-learn to standardize numerical features, ensuring uniform feature scaling for robust model performance.

```
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

Figure 7: Feature Scaling

7.3 Random Forest Classifier:

The Random Forest Classifier, an ensemble learning technique, is versatile for classification and regression. By constructing numerous decision trees and aggregating their predictions, it enhances accuracy while mitigating overfitting risks.

7.3.1 Model Training:

In this implementation, we trained the Random Forest Classifier on preprocessed training data, enabling the model to make informed predictions using the dataset's features.

```
best_rf_classifier = RandomForestClassifier(random_state=42)
best_rf_classifier.fit(X_train, y_train)
rf_predictions = best_rf_classifier.predict(X_test)
```

Figure 8: Model Training

7.3.2 Model Evaluation:

The model's effectiveness is gauged through pivotal metrics: precision, F1 score, recall score, and accuracy, serving as crucial benchmarks for performance evaluation.

```
rf_accuracy = accuracy_score(y_test, rf_predictions)
precision = precision_score(y_test, rf_predictions)
recall = recall_score(y_test, rf_predictions)
f1 = f1_score(y_test, rf_predictions)
```

```
print("Random Forest Accuracy",rf_accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1-Score:", f1)
```

```
Random Forest Accuracy 0.95
Precision: 1.0
Recall: 0.9333333333333333
F1-Score: 0.9655172413793104
```

Figure 9: Model Evaluation

7.4 Support Vector Machine:

Support Vector Machines (SVM) are a class of versatile machine learning algorithms applied to both classification and regression tasks. SVMs excel at finding the optimal decision boundary that maximizes class separation.

7.4.1 Model Training:

In this implementation, we trained the Support Vector Machine (SVM) with an RBF kernel, known for its effectiveness in handling non-linear data patterns and classifying with precision.

```
# Initialize and fit an SVM classifier
svm_classifier = SVC(probability=True, random_state=42)
svm_classifier.fit(X_train, y_train)
```

Figure 10: Model Training

7.4.2 Model Evaluation:

In this implementation, the model's effectiveness is measured using vital metrics: precision, F1 score, recall score, and accuracy, collectively providing a comprehensive evaluation of its performance.

```
svm_accuracy = accuracy_score(y_test, svm_predictions)
svm_precision = precision_score(y_test, svm_predictions)
svm_recall = recall_score(y_test, svm_predictions)
svm_f1 = f1_score(y_test, svm_predictions)

print("SVM Accuracy:", svm_accuracy)
print("SVM Precision:", svm_precision)
print("SVM Recall:", svm_recall)
print("SVM F1-Score:", svm_f1)
```

```
SVM Accuracy: 0.9
SVM Precision: 0.8823529411764706
SVM Recall: 1.0
SVM F1-Score: 0.9375
```

Figure 11: Model Evaluation

8.RESULT ANALYSIS:

The analysis showcases a successful application of machine learning in Parkinson's disease prediction. The Random Forest algorithm demonstrated superior accuracy, achieving 95%, while the SVM also performed well with a 90% accuracy rate. This shows that machine learning in conjunction with data preprocessing can be an effective tool for Parkinson's disease early detection and therapy, thereby improving the quality of life for patients. The implementation's two classifiers' respective performance discrepancies are depicted in a visual comparison chart.

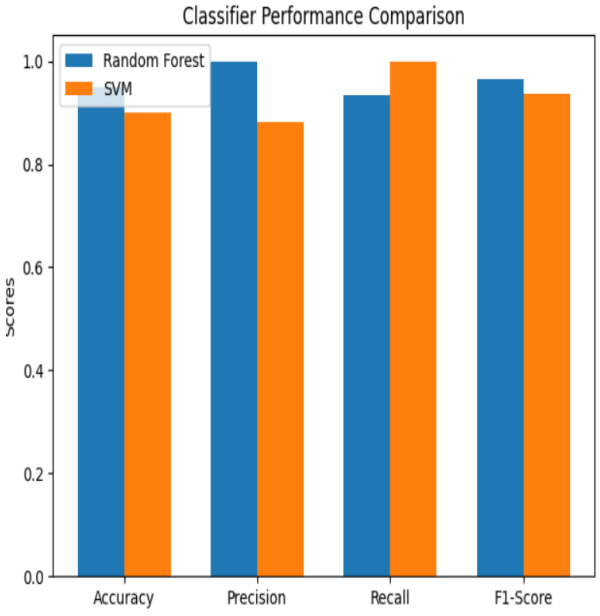

Figure 12: Comparison of Classifiers

9.CONCLUSION:

Detecting Parkinson's disease in its early stages is crucial for improving patient outcomes, but existing methods often lack accuracy and efficiency. In response to this challenge, we have developed a novel approach utilizing machine learning predictive models for the precise

identification of Parkinson's disease. Our method exhibits a high level of accuracy in distinguishing individuals with the condition from those without it, as demonstrated by our experimental results, thereby offering a promising solution for more reliable and efficient early diagnosis.

10.FUTURE WORK:

In our future work, we plan to expand the dataset's diversity and size, striving to enhance the model's robustness. We aim to incorporate additional data modalities, such as neuroimaging and wearable device data, to gain comprehensive diagnostic insights. Collaborative efforts with medical professionals will be essential as we validate the model's clinical applicability in real-world settings. Moreover, we will continue refining and validating the model on larger, more diverse patient populations to pave the way for practical clinical implementation. As a promising extension, we intend to investigate the model's potential in calculating levels of depression and voice tremor for a more comprehensive health assessment.

11.REFERENCES:

[1] Alatas Bilal, Moradi Shadi, Tapak Leili, Afshar Saeid (2022), "Identification of Novel Noninvasive Diagnostics Biomarkers in the Parkinson's Diseases and Improving the Disease Classification Using Support Vector Machine", BioMed Research International, Hindawi.

[2] R. Alkhatib, M. O. Diab, C. Corbier and M. E. Badaoui, "Machine Learning Algorithm for Gait Analysis and Classification on Early Detection of Parkinson," in IEEE Sensors Letters, vol. 4, no. 6, pp. 1-4, June 2020, Art no. 6000604, doi: 10.1109/LSENS.2020.2994938.

[3] 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 q-factor wavelet transform. Applied Soft Computing 74, 255–263 (2019).

[4] Gomathy, C K. (2021). THE PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING TECHNIQUES.

[5] Moro-Velazquez, Laureano, et al. "Advances in Parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects." Biomedical Signal Processing and Control 66 (2021): 102418.

[6] M. Wodzinski, A. Skalski, D. Hemmerling, J. R. Orozco-Arroyave and E. Nöth, (2019) "Deep Learning Approach to Parkinson's DiseaseDetection Using Voice Recordings and Convolutional Neural Network Dedicated to Image Classification," 2019 41st Annual InternationalConference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 717-720, doi: 10.1109/EMBC.2019.8856972.

[7] Naranjo, Lizbeth, Carlos J. Perez, Yolanda Campos‐Roca, and Jacinto Martin. "Addressing voice recording replications for Parkinson's disease detection." Expert Systems with Applications 46 (2016): 286-292.

[8] Lahmiri, S., & Shmuel, A. (2019). Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. Biomedical Signal Processing and Control, 49, 427-433.

[9] Prashanth R, Dutta Roy S, Early Detection of Parkinson's Disease through Patient Questionnaire and Predictive Modelling, International Journal of Medical Informatics(2018),https://doi.org/10.1016/j.ijmedinf.2018.09.008

[10] Pahuja, G., & Nagabhushan, T. N. (2021). A comparative study of existing machine learning approaches for Parkinson's disease detection. IETE Journal of Research, 67(1), 4-14.