WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Parkinson's Disease Detection Using Machine Learning

[1]Dr. V. Lokesh Raju, [2]S. Venkata Sai Kiran, [3]Korada Harika, [4]A. Dharani, [5]K. Satish, [6]M. Lavanya, [7]G. Tharun

[1]Associate Professor, Ph.D., [2,3,4,5,6,7] Students, Dept. of Electronics and Communication, Aditya Institute of Technology and Management, Tekkali.
DOI: https://doi.org/10.55248/gengpi.5.0424.0946

ABSTRACT:

Millions of people worldwide suffer with Parkinson's disease (PD), a progressive neurological illness. One of the biggest challenges in diagnosing Parkinson's disease (PD) is still early detection, which is why machine learning (ML) techniques are being investigated. Using machine learning (ML) techniques such as logistic regression, support vector machines (SVM), random forests, gradient boosting, K-Nearest Neighbors (KNN), and XGBoost, the goal of this project is to construct a system for detecting Parkinson's illness. These algorithms are trained and assessed to determine which model best separates persons with and without Parkinson's disease. They do this by using a dataset that includes voice-based biomarkers from both affected and healthy subjects. The application of Exploratory Data Analysis (EDA) approaches facilitates feature selection and preprocessing stages that are essential for model creation by providing insights into the distribution, correlations, and underlying patterns of the data.

Keywords: Parkinson's Disease, Early detection, Logistic Regression, Support Vector Machines, Random Forest, Gradient Boosting, K-Nearest Neighbors, XGBoost

1. INTRODUCTION

Parkinson's disease is a degenerative neurological condition that primarily affects older age groups and impacts millions of people globally. named for James Parkinson, a British physician who originally reported the condition's symptoms in 1817. It has an impact on the neurological system and the body's nerve-controlled regions. With time, the condition's symptoms get more severe as a portion of your brain deteriorates. This illness can affect your senses, cognitive function, mental health, and many other areas, even though its effects on balance, movement, and muscular control are the most well-known. Parkinson's disease usually manifests at age 60, and the disease's risk increases with advancing age. It is slightly more common in men or those categorized as male at birth (DMAB) than in women or those classed as female at birth (DFAB).

One specific area of the brain that degenerates in Parkinson's disease patients is the basal ganglia. As this area deteriorates, you lose the authority that those places once held. Usually age-related, Parkinson's disease can affect people as young as 20 (though this is extremely rare; most patients have a parent, sibling, or child who also has the condition). Generally speaking, Parkinson's disease is the second most common degenerative brain condition linked to aging. It is also the most common motor (movement-related) brain disease. Experts estimate that it affects 1% or more of persons over 60 worldwide.

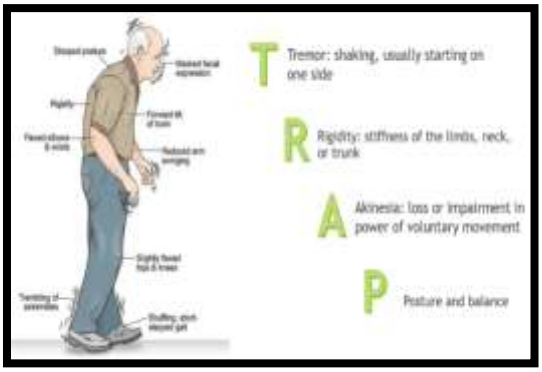

Figure 1: Symptoms of Parkinson's Disease

2. LITERATURE SURVEY

Aditi Govindu et al. (2023) [1] studied the use of machine learning techniques in telemedicine for early Parkinson's disease identification. They used four models: Support Vector Machine, Random Forest, K-Nearest Neighbors, and Logistic Regression. The Random Forest classifier was found to be the most accurate, with a 91.83% detection accuracy.

Dr. C K Gomathy et al. (2021) [3] developed a prediction model for early Parkinson's disease identification using advanced algorithms like XGBoost, Naive Bayes, and Decision Trees. The model achieved an accuracy of 94.87%, outperforming other models, including Decision Trees and Naive Bayes, demonstrating the potential for early detection and treatment.

Shrihari K. Kulkarni and K. R. Sumana (2021) [4] used machine learning and deep learning algorithms to predict early Parkinson's disease performance parameters. They used Decision Trees, Logistic Regression, Naive Bayes, and Recurrent Neural Networks. Their multimodal approach, combining traditional and deep learning techniques, achieved accuracies ranging from 88.75% to 94.5%, demonstrating the potential of predictive modeling in Parkinson's disease diagnosis.

The ground-breaking study by [5] V. Ulagamuthalvi et al. (2020), which used machine learning classifiers to identify this crippling condition, in an effort for Parkinson's disease diagnosis. In order to categorize individuals with Parkinson's disease, their study carefully used eXtreme Gradient Boosting (XGBoost) and Logistic Regression classifiers, obtaining an astounding 96% accuracy.

Our study analyzing four literature reviews on Parkinson's disease detection using machine learning revealed that advanced algorithms can achieve high accuracy in early diagnosis, with accuracies ranging from 91.83% to 96%. The XGBoost algorithm was the frontrunner in multiple studies, highlighting the potential of machine learning in transforming Parkinson's disease diagnosis, providing early detection and intervention pathways, enhancing patient outcomes and quality of life. The dataset used was provided by MAX Little [6] in 2008.

3. OBJECTIVES

1. Develop a machine learning-based screening tool to distinguish between healthy individuals and those suffering from Parkinson's Disease using voice-based biomarkers.

2. Enhance the diagnostic process by providing a reliable and efficient screening method, reducing the need for frequent medical visits and complementing traditional diagnostic methods.

3. Employ various machine learning algorithms and detection techniques to enhance diagnostic accuracy, reliability, and efficiency.

4. Create a robust and dependable tool for Parkinson's Disease detection that offers a non-invasive screening option.

5. Benefits:

a) Improved patient outcomes through early detection and intervention.

b) Reduction in the burden of frequent medical visits for screening purposes.

c) Complementary tool to traditional diagnostic methods, potentially streamlining the diagnostic process.

6. Develop a screening tool that enhances diagnostic accuracy, reliability, and efficiency, ultimately improving patient outcomes in Parkinson's Disease management.

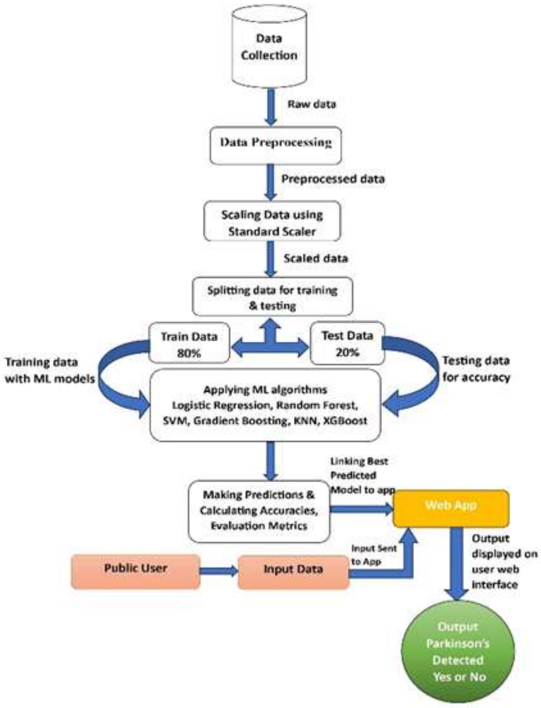

Figure 2: Architecture Diagram

4. METHODOLOGY

In this work, we offer a thorough process for creating a machine learning-based system that uses biological speech measurements to identify Parkinson's disease.

1. Data Collection: Compile a large dataset of biomedical voice measurements from both healthy and Parkinson's disease-diagnosed people. We have utilized the dataset by Max Little from UCI Machine Learning Repository [6].

2. Data Preprocessing: To get the dataset ready for modelling, carry out preprocessing operations including feature selection, data cleaning and normalization.

Exploratory Data Analysis (EDA) is a statistical method used to analyze and visualize data sets, aiming to understand their structure, patterns,

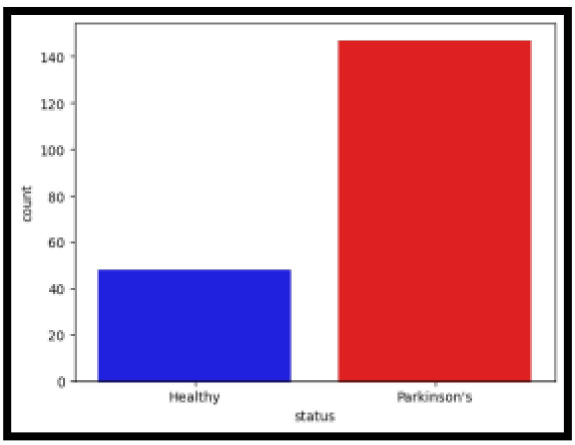

Figure 3: Plot between Healthy & Parkinson's from the dataset

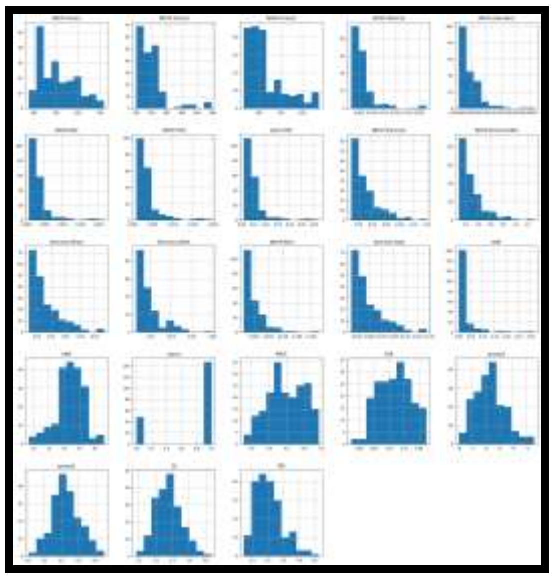

Figure 4: Analyzing the Data Attributes

The dataset displays the status of Parkinson's disease affected individuals shown in Figure 3, with red and blue bars representing affected and healthy individuals, with the "status" column representing the target variable.

Histogram graphs shown in Figure 4 visualize data attributes, providing insights into frequency and range. They help identify patterns or abnormalities, and can be compared across groups like Parkinson's disease patients. This exploratory data analysis technique aids in modeling and interpretation.

A correlation matrix shown in Figure 5 displays pairwise correlations between variables in a dataset, with each cell representing a correlation coefficient. Analyzing this helps identify patterns and relationships, guiding data analysis and decision-making.

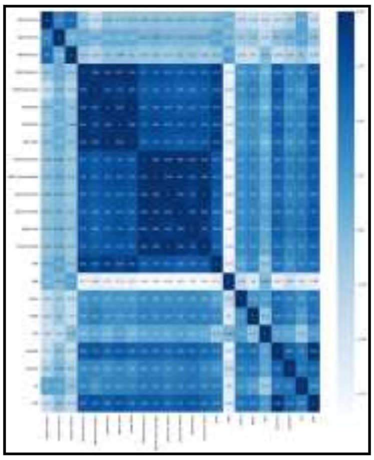

Figure 6: Correlation Matrix of the data attributes

3. Model Development: Put into practice different ML algorithms, techniques including Random Forest, Gradient Boosting, K-Nearest Neighbor, Support Vector Machine (SVM), Logistic Regression, and XGBoost are employed.

4. Model Assessment: Use cross-validation strategies to evaluate the performance of each model according to pertinent metrics, such as F1-score, accuracy, precision, and recall.

5. Local Web Application Development: Develop an intuitive online platform for inputting voice measurements and receiving Parkinson's Disease probability predictions. Deploy thoroughly tested web application with trained models for reliability, verified through real data testing.

5.1 MACHINE LEARNING MODELS

- **Logistic Regression:**

Logistic regression is a statistical technique used for binary classification problems, specifically in Parkinson's disease detection, to model the likelihood of a given input being a member of a specific class.

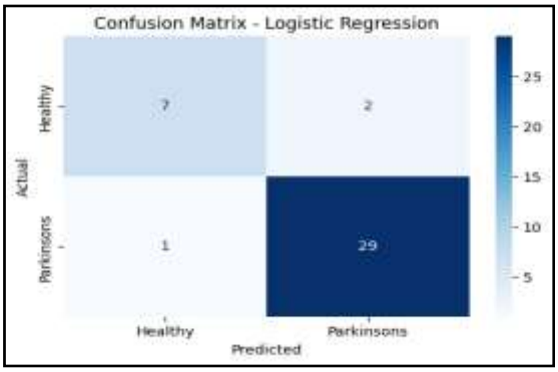

Figure 7: Confusion Matrix of Logistic Regression model

```
Accuracy score of test data:  0.9230769230769231
F1 score of test data:  0.9508196721311476
precision:  0.9354838709677419
recall:  0.9666666666666667
```

Figure 8: Evaluation metrics of Logistic Regression model

➢ **Support Vector Machine (SVM):**

Support Vector Machine (SVM) is a supervised learning technique used to identify the hyperplane in feature space, effectively dividing classes and categorizing individuals into Parkinson's disease or healthy controls.

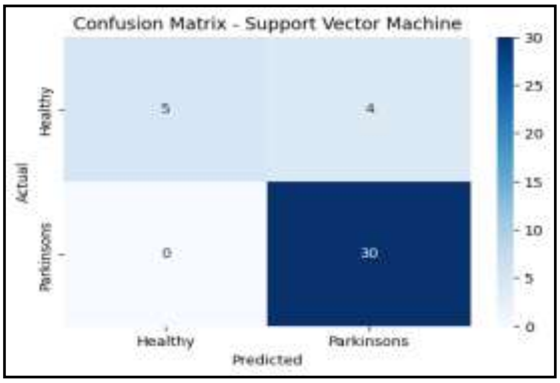

Figure 9: Confusion Matrix of SVM

```
Accuracy score of test data:  0.8974358974358975
F1 score of test data:  0.9375
precision:  0.8823529411764706
recall:  1.0
```

Figure 10: Evaluation metrics of SVM model

➢ **Random Forest Classifier:**

Random Forest ensemble learning technique builds decision trees for Parkinson's disease diagnosis, aggregating predictions through voting or averaging. Each tree identifies different patterns.

Figure 11: Confusion Matrix of Random Forest Classifier model

```
Accuracy score of test data:  0.8717948717948718
F1 score of test data:  0.9180327868852459
precision:  0.9032258064516129
recall:  0.9333333333333333
```

Figure 12: Evaluation metrics of Random Forest Classifier model

➢ **Gradient Boosting Classifier:**

Gradient boosting is an ensemble learning method that creates decision trees to improve prediction performance, particularly for PD identification using voice measures as input features.

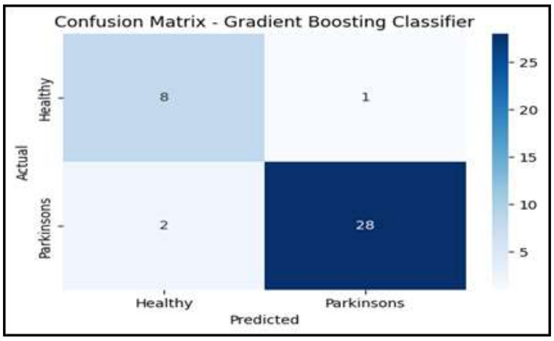

Figure 13: Confusion Matrix of Gradient Boosting Classifier model

```
Accuracy score of test data:  0.9230769230769231
F1 score of test data:  0.9491525423728815
precision:  0.9655172413793104
recall:  0.9333333333333333
```

Figure 14: Evaluation metrics of Gradient Boosting Classifier

- **K-Nearest Neighbor Classifier:**

K-Nearest Neighbors is a powerful supervised learning technique used for Parkinson's disease detection, using voice measures as input features and learning from training data points and classes.

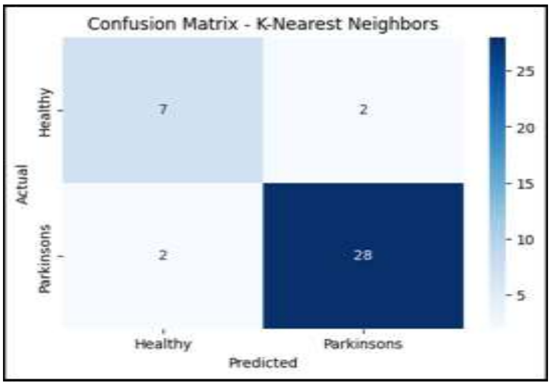

Figure 15: Confusion Matrix of KNN Classifier

```
Accuracy score of test data:  0.8974358974358975
F1 score of test data:  0.9333333333333333
precision:  0.9333333333333333
recall:  0.9333333333333333
```

Figure 16: Evaluation metrics of KNN Classifier

- **XGBoost Classifier:**

XGBoost is a machine learning library optimized for efficiency and scalability, used for Parkinson's disease detection. It optimizes decision tree parameters to minimize prediction errors.

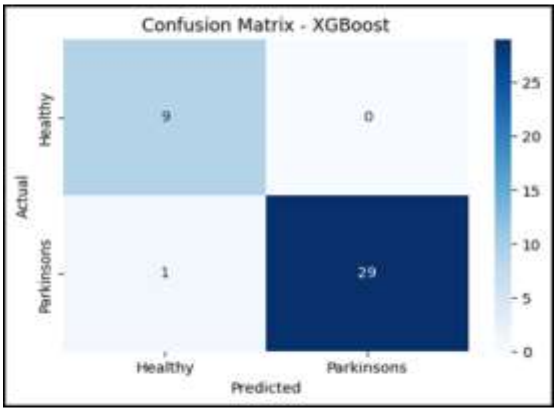

Figure 17: Confusion Matrix of XGBoost Classifier

```
Accuracy score of test data:  0.9743589743589743
F1 score of test data:  0.983050847457627
precision:  1.0
recall:  0.9666666666666667
```

Figure 18: Evaluation metrics of XGBoost Classifier

6. RESULTS

When comparing machine learning models for Parkinson's disease identification, XGBoost was the clear winner with the best accuracy rate of 97.44%. This remarkable result highlights how well it can distinguish between people who have Parkinson's disease and those who do not, using the feature set that is given.

Summary of all Machine Learning models:

	Accuracy	Recall	Precision	F1 score
Logistic Regression	0.923077	0.966667	0.935484	0.950820
SVM	0.897436	1.000000	0.882353	0.937500
Random Forest	0.871795	0.933333	0.903226	0.918033
Gradient Boosting	0.923077	0.933333	0.965517	0.949153
KNN	0.897436	0.933333	0.933333	0.933333
XGBoost	0.974359	0.966667	1.000000	0.983051

With 100% precision, XGBoost is particularly good at reducing false positive results, which is important in medical diagnostics because inaccurate results can have serious consequences.

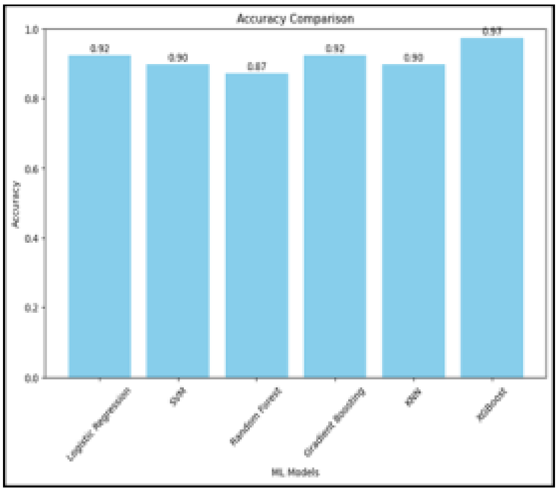

Figure 19: Comparison of all ML models

ROC: A graph is given that shows all the classifiers performance across all potential thresholds as a ROC curve. The true positive rate is plotted on the Y-axis, and the false positive rate is plotted on the x-axis. We obtained an excellent accuracy for the XGBoost classifier based on that.

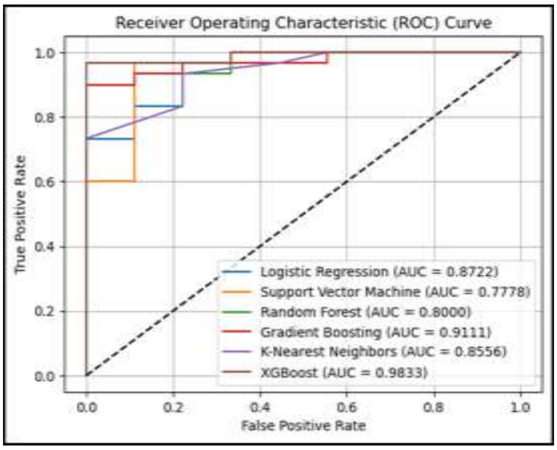

Figure 20: ROC Curves of all ML Models

Local Web Application: The online application for PD prediction that uses Flask intends to provide users with an interface where they may enter attribute values that are pertinent to the diagnosis of PD. The study paper's extensive investigations yielded attribute ranges that were previously computed and utilized by the program. The Flask backend will handle the attribute values that users enter into the web interface.

Figure 21: Our Web Interface Home Page

We are testing the web app using data from a Parkinson's affected person data, with the predicted results shown in figure 23.

Figure 22: Giving Parkinson's affected person inputs in the Local Web App

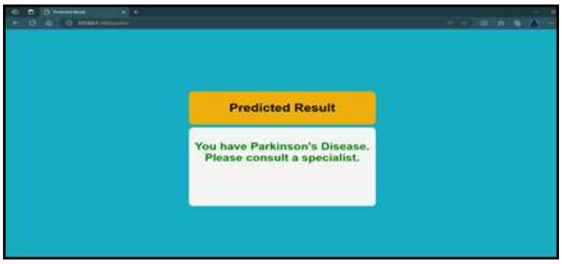

Figure 23: Predicted Result of a Parkinson's person

Now, we are testing the web app using data from a healthy person data, with the predicted results shown in figure 25.

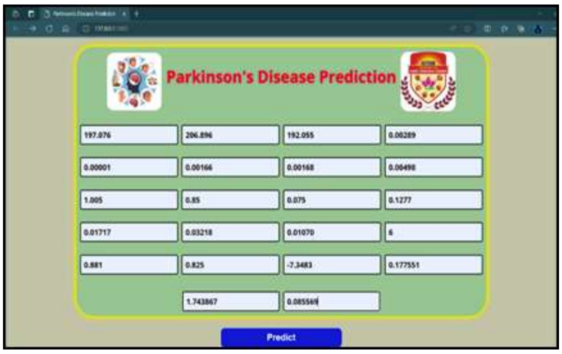

Figure 24: Giving Healthy Person's data in the Local web App

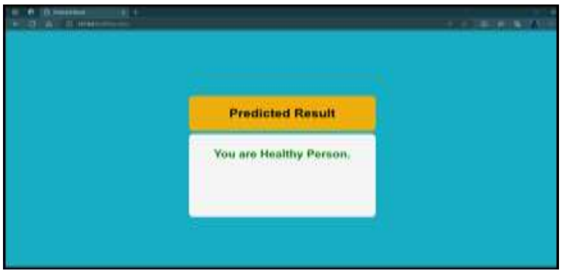

Figure 25: Predicted Result of Healthy Person

7. CONCLUSION AND FUTURE SCOPE

In conclusion, ML models for Parkinson's detection mark a significant advancement in diagnostics. Using various algorithms, including Logistic Regression, Support Vector Machines, Random Forests, Gradient Boosting, K-Nearest Neighbors, and XGBoost, the analysis yielded highly accurate predictive models. XGBoost was the top-performing classifier, achieving 97.44% accuracy and 100% precision. A Flask-based web application was implemented for real-time predictions, ensuring robustness and reproducibility.

Future Scope of this project is to enhance healthcare by integrating data sources, leveraging technologies like deep learning, and developing global user-friendly applications. Collaboration with healthcare providers and researchers is crucial for large-scale clinical trials and validation studies.

REFERENCES

[1] Aditi Govindu & Sushila Palwe. (2023). Early detection of Parkinson's disease using machine learning. Procedia Computer Science. 218. 249-261. DOI: https://doi.org/10.1016/j.procs.2023.01.007

[2] Uday Kumar G S, Dr. S. Baskaran, Dr. D. Sumathi. (2022). Implementing a Model to Detect Parkinson Disease using Machine Learning Classifiers. JOURNAL OF ALGEBRAIC STATISTICS. Volume 13, No. 1, 2022, p. 99-110, ISSN: 1309-3452

DOI: https://doi.org/10.52783/jas.v13i1.65

[3] Dr. C K Gomathy, Mr. B. Dheeraj Kumar Reddy, Ms. B. Varsha, and Ms. B. Varshini, (OCT-2021). THE PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING TECHNIQUES. International Research Journal of Engineering and Technology (IRJET).

https://www.researchgate.net/publication/357448942_THE_PARKINSON'S_DISEASE_DETECTION_USING_MACHINE_LEARNING_TECHNIQUES

[4] Shrihari K Kulkarni, K R Sumana. (AUG-2021). Detection of Parkinson's Disease Using Machine Learning and Deep Learning Algorithms. International Journal of Engineering Science Invention (IJESI). ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726 VOLUME 8 ISSUE:8, Page No :1189-1192.

https://www.researchgate.net/publication/355282729_Detection_of_Parkinson's_Disease_Using_Machine_Learning_and_Deep_Learning_Algorithms

[5] Ulagamuthalvi V, Kulanthaivel G, Reddy G. S. N, Venugopal G. (2020). Identification of Parkinson's Disease Using Machine Learning Algorithms. Biosc.Biotech.Res.Comm. Vol 13(2), Page No :576-579.

54DOI: http://dx.doi.org/10.21786/bbrc/13.2/32

[6] Max Little, (2008). Parkinsons. UCI Machine Learning Repository.https://doi.org/10.24432/C59C74