Design and Implementation of ML Model for Early Diagnosis of Parkinson's Disease using Gait Data Analysis in IoT Environment

Navita Mehra[1]
Ph.D Scholar
Dept. of Comp. Sci. & App
Maharshi Dayanand University, Rohtak, India

Pooja Mittal[2]
Assistant Professor
Dept. of Comp. Sci. & App
Maharshi Dayanand University, Rohtak, India

***Abstract*—Parkinson's disease (PD) is the world's second most neurodegenerative disorder that results in a steady loss of movement. The symptoms in patients occur slowly with the passage of time are and very hard to identify in its initial stage. So, early diagnosis of PD is the foremost need for timely treatment to people. The introduction of smart technologies like the Internet of Things (IoT) and wearable sensors in the healthcare domain offers a smart way of identifying the symptoms of PD patients. In which smart sensors are worn on the patient's body which continuously monitor the symptoms in patients and track their possible health status. The major objective of this work is to propose a machine learning-based healthcare model that best classifies the subjects into healthy and Parkinson's patients by extracting the most important features. A step regression-based feature selection method is followed to improve the classification of PD. A Shapiro Wilk test is adopted to check the normality of the gait dataset. This model is implemented on three publicly available Parkinson's datasets collected from three different studies available on Psyionet. All these data sets contain VGRF recordings obtained from eight different sensors placed under each foot. Experimentation is done on the Jupyter notebook by utilizing Python as a programming language. Experimental results revealed that our proposed model with effective pre-processing, feature extraction, and feature selection method achieved the highest accuracy result of 95.54%, 98.80%, and 94.52% respectively when applied to three datasets. Our research inducts knowledge about significant characteristics of a patient suffering from PD and may help to diagnose and cure at an early stage.**

Keywords*—Internet of things (IoT); sensors; parkinson's disease (PD); machine learning (ML); vertical ground reaction force (VGRF)

I. INTRODUCTION

The old age people of today's society suffer from a large number of neurodegenerative disorder diseases like Alzheimer's disease, dementia and Parkinson's Disease (PD), etc. PD is the second most affecting neurodegenerative disease after Alzheimer's disease affecting people worldwide. A report generated by the Parkinson's Foundation states that 10 million people live with Parkinson's disease (PD) worldwide and among these, approximately one million people are from the United States (US). This report also states that men are 1.5 times more affected by PD as compared to women [1]. 1.04 million people were diagnosed with PD in the US in 2017 and it is estimated to be 1.6 million by 2037 [2]. The progression of PD differs from one person to another person and there does not exist any standard test for the diagnosis of PD and detection is done only based on observations of its symptoms. The clinical demonstrations of PD mainly include tremors in hands, slowing movement, limb rigidity, altered taste in smell, posture instability, etc. [3]. The symptoms in patients occur slowly with time so it's very hard to detect or diagnose this. PD mainly arises due to the degradation of dopamine cells in the brain which control the movement of limbs in the body. By the time symptom appears 50 to 80% of dopamine neurons have already died [4].

Recent discoveries reveal that analysis of gait patterns can be considered a better approach for the detection of neurodegenerative diseases like PD [5] [6]. In the past few years, the invention of smart motion analysis systems and sensors-based motion-capturing devices offers an opportunity for researchers to work on more advanced gait analysis techniques. Smart wireless sensor devices like smart watches, smart bracelets, stand-alone video cameras, smartphones, insole sensors, and other portable devices provide the easiest way of capturing the motion and movement disorders among patients with PD [7]. Major approaches utilized for analysis of gait dataset include smart motion capturing cameras, inertial measurement unit (IMU) based sensors to discover sharp motion and body force in a particular direction [8], plantar pressure determined by utilizing planter sensors, some force subtle platforms for measuring VGRF, High-resolution Electromyography (EMG) devices to capture muscles actions [9]. The neuroimaging approach involves expensive optical cameras and force platforms, while foot-worn sensors offer a reliable, fast, simple, and reasonable gait analysis approach. However, gait analysis faces lots of challenges like high data dimensionality, nonlinear data dependency, and complex correlation.

In recent years, ML techniques have revealed the impressive capability to support clinicians not in identifying the existence of PD but also supports in categorizing the states of PD on the basis of the motor indicators of the subjects [10] [11].

For the detection of PD, ML techniques have been applied to different kinds of measured data like handwritten patterns

[12-14], speech data, neuroimaging data [15], smell identification data [16], spontaneous cardiovascular oscillations [17], and gait data [18]. Many researchers are now working on the early, precise, and timely detection of PD, particularly when ML techniques are applied to learn major strategies. UPDRS and H&Y are two important rating scales utilized for monitoring the progression of PD [19] [20]. H&Y scale is the most commonly utilized rating scale for effective validation of severity level according to functional disability. Therefore, the major objective of the current study was to develop an ML model that could help physicians to diagnose PD by utilizing gait data generated from IoT-based wearable sensors. After that performance of that model is analyzed by using different performance evaluation measures. The major contribution of this work is as follows:

1) This work provides a significant way of detecting/ dosing the PD in patients using an IoT environment.

2) This work proposed a PD diagnosis model by utilizing effective feature extraction and selection method.

3) This work also adopts various performance evaluation metrics to predict healthy and PD patients with the help of collected data patterns.

4) The work also focuses on the comparison of the proposed model on the basis of selected and unselected features set in classifying healthy and PD patients in an IoT environment.

This paper is organized as follows: Section II describes the major causes, symptoms, and measurable indicators of PD. Section III introduces materials and methods utilized for our experimental purpose. Section IV presents different performance metrics utilized for evaluating the performance of the proposed model. Section V describes the result obtained through the implementation of our model. Conclusions are drawn in Section VI. In last, limitations and future directions are described in Section VII.

II. Outline About Major Causes, Symptoms and Measurable Indicators of PD

Parkinson’s disease occurs due to brain disorder and is a progressive neurodegenerative disease that results in inadvertent or non-controllable movements, such as shaking, toughness, and difficulty with steadiness and coordination as shown in Fig. 1.

Categorization of major symptoms of PD is described Major symptoms of PD are described below:

1) Motor symptoms (movement-based symptoms)

- slowed movement (bradykinesia) means muscles weakness,
- tremor means muscles are at rest
- rigidity or stiffness
- Unbalanced postures etc.

2) Non-Motor symptoms (non-movement-based symptoms)

- depression
- loss of smell sense
- sleep disorder
- Inconvenience in thinking and focusing etc.

Even though, PD is incurable because the symptoms of PD usually start steadily and become worse over time. So, early detection of PD can assist to follow proper medication/surgical treatment so that the symptoms can be alleviated. H & Y and UPDRS are two generally utilized clinical ranking scales for monitoring the progression of PD [22] [23]. The former rating scale considers only motor symptoms while the latter assesses both motor and non-motor symptoms.

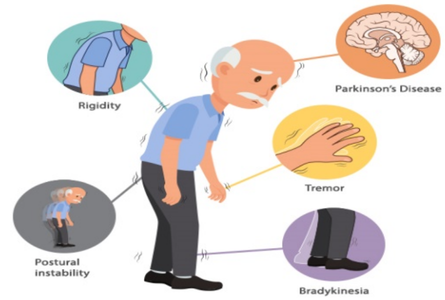

Fig. 1. Parkinson ’s Disease Symptoms Appearance [21].

III. IoT Approach Toward PD Diagnosis

In the traditional medical scenario, diagnosis of PD is a very hard task because it involves proper tracking of the patient's tasks throughout the day. Because there can be a gradual variation in symptoms throughout the day. But patients have no proper sources to track their activities and may lose the most important observations. Another way that the patient moves to the clinician for proper assessment of their postural steadiness and rigidity level which requires transportation cost and time.

The advancement of various smart sensor-based technologies and their integration into the healthcare system reduces the pressure of treatment of various neurological diseases like Parkinson’s disease. “Kevin Ashton” 1999 introduced the concept of the Internet of Things (IoT) which offers connectivity among various wearable sensors and internet platforms that result in amazing remote monitoring services for patients [24]. So, with the introduction of IoT tracking the major symptoms of PD becomes easy [25]. Symptoms collection is possible by just simply placing a few small wireless sensors on the patient body and remotely monitoring the symptoms and details of the patient. So, quality of life can be improved by importing new smart technologies.

Fig. 2 illustrates how a PD patient is treated in an IoT environment while performing any activity. Smart sensors were worn on patients' bodies while they performed their activities. The recording of smart sensors was continuously transferred to the system-based storage or cloud using various transmitting technologies like Bluetooth, Wi-Fi, etc., and after that data is analyzed by utilizing various machine intelligence techniques.

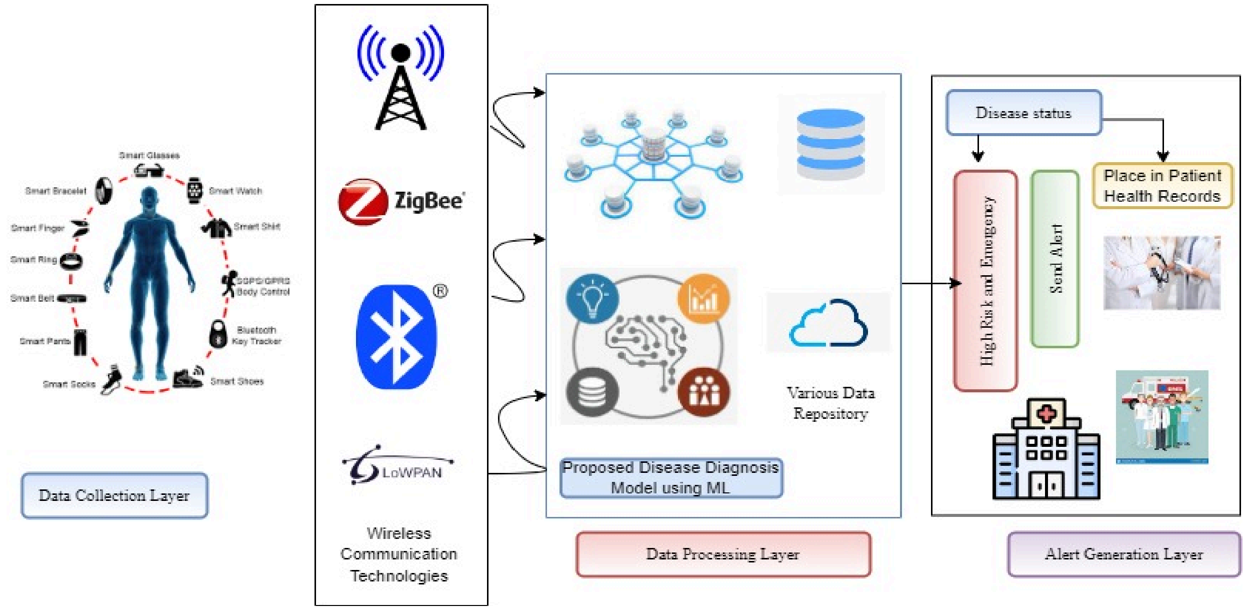

Fig. 2. PD Patients Monitoring and Classification using Wearable Sensors in IoT Environment.

IV. MATERIALS AND METHODS

This section will describe the data set and techniques utilized for the implementation of our work.

A. PD Dataset Description

Parkinson's dataset utilized in this study is publicly available at Psyionet [23]. This dataset was collected via a team of three scientists at the "Movement disorder unit of the Tel-Aviv Sourasky Medical Center, Israel" and named after that Yogev et al. [24] consist of the dataset recordings when subjects walking on level ground, Frenkel-Toledo et al. [25] contains dataset recordings when the patient walking on a treadmill and Hausdorff et al. [26] contains recordings when subjects moving at a comfortable place with RAS. This dataset recording was taken from 73 healthy subjects and 93 subjects affected by Parkinson's disease. In this work, three datasets are separately considered for the identification of healthy and Parkinson's patients.

Table I details the total number of PD and healthy patients from three datasets with their associated physical and clinical characteristics. For small and simple depiction these datasets are denoted as Ga [24], Si [25], and Ju [26]. Each shoe had pressure sensors shown in Fig. 3. Table II depicts the absolute position of sensors in the X-Y coordinate framework. The VGRF signal representation of PD and the non-PD patients is shown in Fig. 4.

The mean and Standard Deviation of each physical feature's age, height, and, weight are taken out to know the average of each characteristic and dispersion of the dataset relative to its mean value.

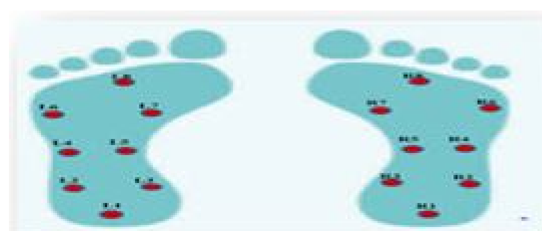

Fig. 3. Pressure Sensors Positioned under each Foot to get the Best Collection.

TABLE I. NUMBER OF SUBJECTS IN THREE SENSORS DATASETS WITH THEIR ASSOCIATED PHYSICAL AND CLINICAL CHARACTERISTICS

Dataset	**Group**		**Subjects**		**Subject**		**Avg. (Yrs.) Mean±SD**		**Height (Mtr.)**		**Weight (Kg.)**	
			Healthy		***PD***		***PD***	***Healthy***	***PD***	***Healthy***	***PD***	***Healthy***
	Healthy	***PD***	***F***	***M***	***F***	***M***						
Ga [24]	18	29	8	10	9	20	61.6±8.8	57.9±6.7	1.67±.07	1.68±.08	73.1 ±11.2	74.2±12.7
Si [25]	29	35	11	18	13	22	67.2±9.1	64.5±6.8	1.66±.07	1.69±.07	70.3±8.4	71.5±11.0
Ju [26]	26	29	14	12	13	16	66.80±10.8	39.31±18.5	1.87±.15	1.83±.08	75.1±11.0	66.8±11.07

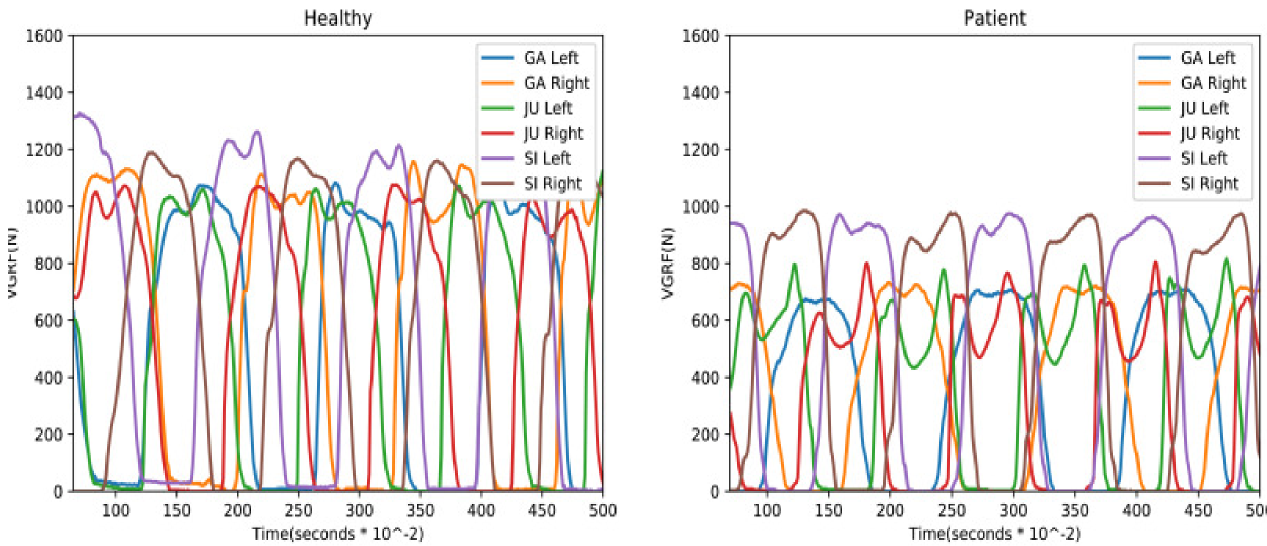

Fig. 4. VGRF Signal Representation of PD Patient and Healthy Person.

TABLE II. PLACEMENT OF LEFT AND RIGHT SENSORS RELATIVE TO X AND Y DIRECTION UNDER EACH FOOT (HERE R DENOTES RIGHT SENSOR, L DENOTES LEFT SENSOR)

Sensor name	**Distance in X- direction (cm)**	**Distance in Y-directionn (cm)**
L1	50	80
L2	70	40
L3	30	40
L4	70	0
L5	30	0
L6	70	40
L7	30	40
L8	50	80
R1	50	80
R2	70	40
R3	30	40
R4	70	0
R5	30	0
R6	70	40
R7	30	40
R8	50	80

B. Proposed ML-based Parkinson's Disease Diagnosis Model

Fig. 5 shows the proposed graphical representation for the diagnosis of PD. First of all, data is collected from wearable eight-foot sensors worn on both left and right feet. After that data may pass through the data pre-processing phase, feature extraction/ selection phase, and final classification phase in which different ML models are applied [31] [32] [33] [34].

- Data Preprocessing

The dataset was collected from different walking tests to avoid the gait starting and ending effect the initial twenty and last twenty data from each gait cycle were removed [35]. The progression of PD among a person can be retrieved through the variations of gait because the walking patterns of individuals changed over time. Therefore, a better study regarding the disease's significant features can provide the best way to understand gait disorders and can be considered significant biomarkers.

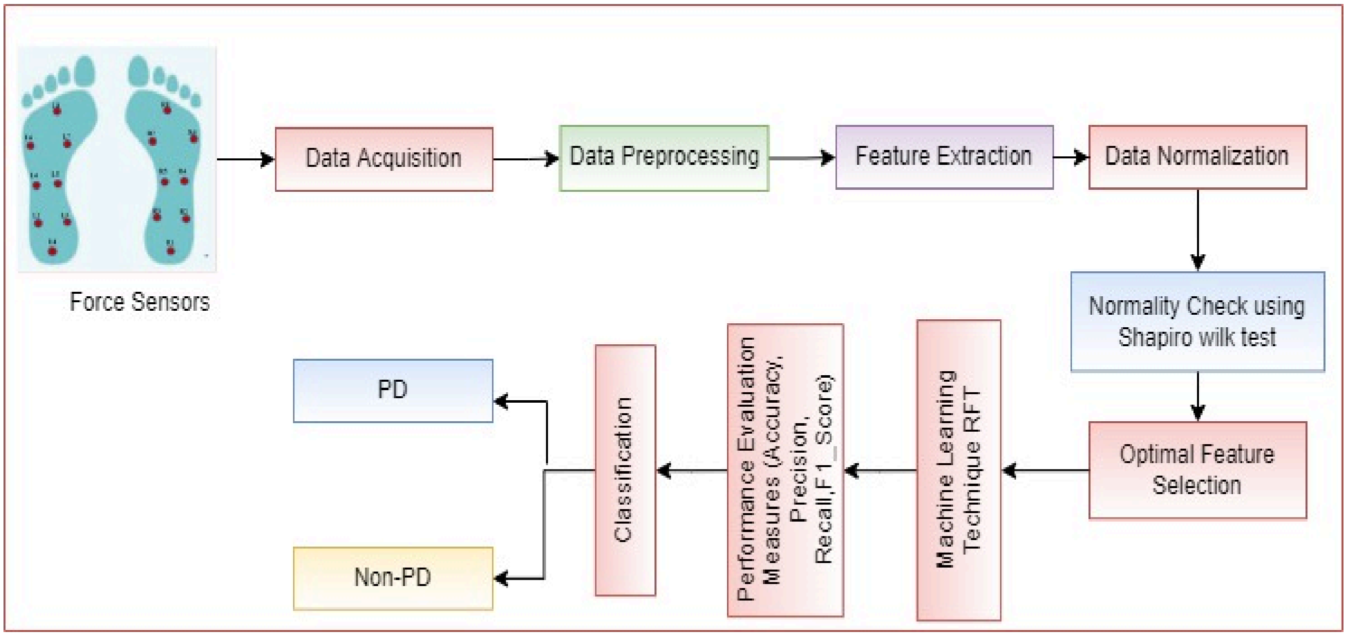

Fig. 5. Proposed ML-based Parkinson's Disease Diagnosis Model.

- Feature Extraction

The most important features extracted from raw sensors are depicted in Table III. A Shapiro-Wilk test is utilized to test the normal distribution of features with a confidence bound of 6% for the hypothesis test.

- Data Normalization

The data set was collected from different sensors. Multiple regression techniques are followed to reduce the distribution in the gait data set. That is represented by the equation:

$$Xi = \beta_0 + \sum_{i=1}^{n} \beta_j \, Y_{i,j} + \in i \quad (1)$$

Where X_i denotes the dependent spatiotemporal features of the i^{th} observation, $X_{i,j}$ denotes the j^{th} physical features like age, weight, height, and walking speed, β specifies the unknown regression coefficient, and $\in i$ denotes the residual error observed for i^{th} iteration. Further feature extraction is performed after normalizing the data.

TABLE III. EXTRACTED SET OF FEATURES WITH THEIR CATEGORY AND NAME

Category	Feature Name	Description
Time	Stance-duration	The period for which one foot is in direct contact with the floor
	Swing- duration	The period for which the body is completely on the support of one leg
	Stride- duration	The time b/w two continuous events of a similar foot
	Step-time	the time gap b/w starting interaction of one foot to starting contact of contralateral foot
Length	Stride-length	Distance b/w two consecutive ground contacts of the same foot
	Step-length	The gap b/w starting contact of one foot to starting contact of the other foot
Frequency	Cadence	No. of steps occupied per unit of time
Temporal	Swing stance ratio	The proportion of swing to stance interval
	Standardized stance duration (std stn-dur)	The ratio of stance duration to the stride time i.e., (stn-dur/str-dur)
	Standardized swing duration (std-sw-dur)	The ratio b/w swing duration to stride duration, i.e., (sw-dur/str-dur)
	Standardized double limb support (std-DLS-dur)	The ratio among DLS to the stride duration stride time, i.e., (DLS-dur/str-dur)
Force	Heel-strike force	Sensor values mean underneath the heel for initial 5% sample points instance interval of the total gait cycle
	Toe-of force (To-force)	Mean of sensor values underneath the toe for the last 5% sample points instance interval of the total gait cycle
	Centre of pressure (x,y) (COP_x, COP_y)	The total amount of pressure field acting on a body causing a force to act on the ground

- Optimal Feature Selection

Feature selection is a significant step that must be followed before the classification process because it improves overall classification performance and results in less computational time and complexity [30]. A stepwise regression method is applied to select the optimal feature set for classifying the patients into healthy and PD classes. First of all, correlation among various autonomous variables (like gender(G), weight (w), height (h), and walking speed (s)) are calculated by utilizing the Spearman correlation coefficient. Reduction in data dispersion is calculated by utilizing the coefficient of variation with a 95% confidence level (CDL) and a standard error (SE) [31]. The statistical measurable significance of the outcome is evaluated by the value of p as ($p<0.001$). Table IV: describes all selected sets of features.

- Random Forest Tree (RFT)

RFT is introduced by Bierman [31]. RFT is used as a classification technique for this model because analysis of different data mining reveals the RFT as the best one when computed on different datasets in paper [33]. This model works well for both classification and regression-based problems. This method also comes under the ensemble approach as it combines multiple Decision trees. This method was mainly introduced to resolve the pruning problem that occurred in the decision tree approach. Besides searching for the most significant feature while distributing a node, this algorithm looks for the best feature amongst a random set of attributes. RFT method follows the bootstrap aggregating or bagging approach for training the learners. The working procedure of RFT is described in Fig. 6.

TABLE IV. SELECTED SET OF FEATURES WITH NORMALIZED VALUES

Coefficient of Variation (%)	**Raw /Un-normalized Data**			**Standardized Data**		
	ME	***90% CL***	***SE***	***ME***	***90% CL***	***SE***
Cadence	9.28	[8.45:15.47]	1.32	4.98	[4.41:6.55]	0.68
Stride interval	10.45	[8.62:12.26]	0.91	5.41	[5.09:7.71]	0.67
Stride length	17.33	[14.00:20.68]	1.68	5.75	[4.96:6.51]	0.62
Stance interval	13.89	[12.32:15.46]	0.78	6.75	[5.21:7.31]	0.59
Swing interval	11.00	[9.35:12.66]	0.82	9.77	[9.19:11.32]	0.38
Step time	24.21	[21.15:27.33]	1.54	12.29	[11.72:13.81]	0.79
Step length	17.02	[15.67:19.01]	0.83	6.76	[5.98:7.59]	0.58
Double Limb Support	28.70	[27.06: 31.45]	1.41	13.6	[11.07:14.44]	0.35

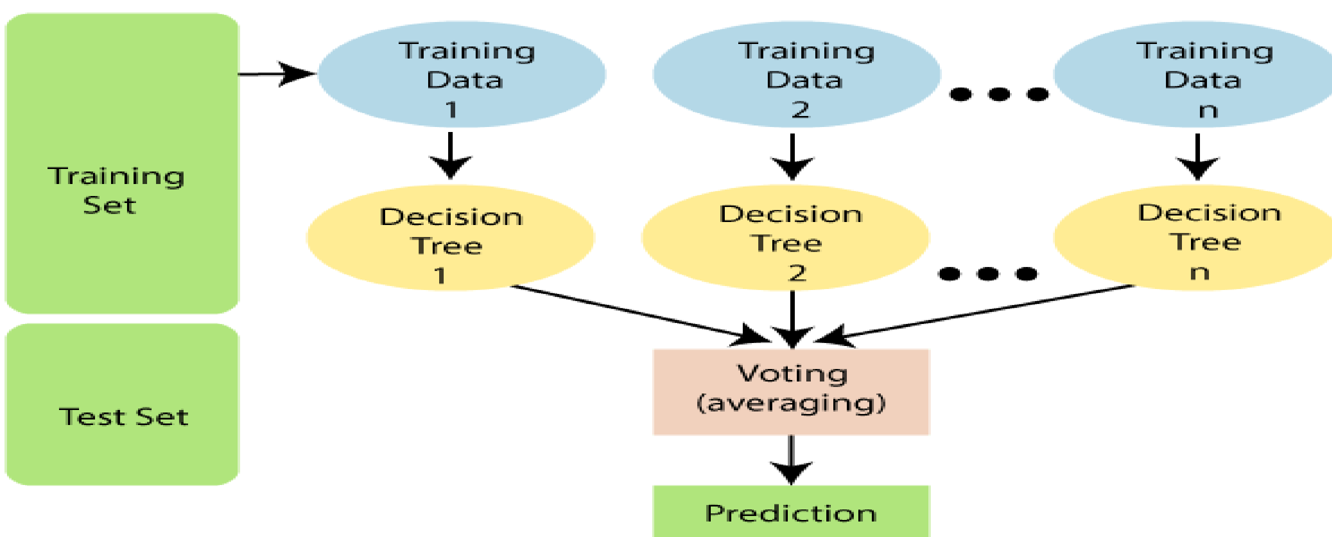

Fig. 6. Working Procedure followed by RFT.

RFT method follow the following steps:

1) For a given training set $S=s_1, s_2,\ldots, s_n$ with classes $Z=z_1,z_2,\ldots,z_n$, bagging method repetitively (B times) selects a random sample with replacement and applied these samples to fit the tree. For b=1,... B: Selects n training samples from S and Z; and call them S_b, Z_b.

2) Train the classification tree R_b on S_b, Z_b.

3) After training phase, predictions for unknown sample S' can be done by taking the majority votes from each classification tree.

4) For a given training set $S=s_1, s_2,\ldots, s_n$ with classes $Z=z_1,z_2,\ldots,z_n$, bagging method repetitively (B times) selects a random sample with replacement and applied these samples to fit the tree. For b=1, ..,B: Selects n training samples from S and Z; and call them S_b, Z_b.

5) Train the classification tree R_b on S_b, Z_b

6) After training phase, predictions for unknown sample S' can be done by taking the majority votes from each classification tree.

C. Implementation Details

Implementation of the proposed model is done on Jupyter IDE an open-source software developed to support highly interactive data science and scientific computing using Python as a programming language. Most important data computing, visualization, and performance measures and machine learning-based libraries including such as (pandas, NumPy, Matplotlib, sns, metrics, and sklearn) are utilized to support various built-in functionality for computation purposes. Random forest tree (RFT) is utilized as classification techniques to classify subjects into healthy and PD classes and tuned with specific hyperparameters (Max_depth=20, n-estimators=550, criteria=entropy) using Grid search cross validation method [34].

V. Performance Evaluation Metrics

The performances of different ML models can be evaluated by using Accuracy, Recall, Precision, and F1_Score.

- Accuracy

Accuracy mainly refers to the fraction of truly classified samples to the total no. of samples.

$$\text{Accuracy} = \frac{\text{Tn+Tp}}{\text{Tn+Tp+Fn+Fp}} \tag{2}$$

- Recall/Sensitivity

Recall mainly refers to the correctly classified samples by the ML model.

$$\text{Recall} = \frac{Tp}{Tp+Fn} \tag{3}$$

- Precision or Positive Predicted value (PPR)

Precision mainly refers to the proportion of truly classified samples among all positive samples.

$$\text{Precsion} = \frac{Tp}{Tp+Fp} \tag{4}$$

- F1_Score

F1_score ranges between the value 0(refers to worst) and 1(refers to best) and it specifies the balance between recall and precision. It is also called a "weighted harmonic means of precision and recall" and results in an accurate mean of performance of the test.

$$\text{F1\_}Score = 2 \times \frac{\text{Precision}\times\text{Sensitivity}}{\text{Precision}+\text{Sensitivity}} \times 100\% \tag{5}$$

VI. Result and Discussion

To avoid the statistically unbiased and overfitting problem, a five-fold CV (Cross-Validation) method is applied. After that, it was arbitrarily divided into five equivalent parts and among five, four subsets are utilized to train the model and the remaining subsets are utilized to test the model. This study addressed a classification problem on three different datasets by utilizing two different feature sets. Table V, VI will show the performance results from three different datasets. The graphical visualization of the performance results is shown in Fig. 7-10. It also shows that there is a great improvement in model performance when is trained with an extracted set of features. For the Ga sub dataset, there is an improvement of approximately 2% can be identified when utilizing feature selection. In the same manner, for the Ju sub dataset, there is an improvement of 4% can be identified when utilizing feature selection technique. Almost the same improvement can be found, for the Si sub dataset, there is an improvement of approximately 2% can be identified when utilizing feature selection. The result revealed that the best combination of related features obtained through feature selection improve the overall performance of proposed ML model. Along with that, it will also reduce execution time both in the training and testing phases.

TABLE V. Performance Results before the Application of the FS Method for Individual Sub-Dataset Utilizing 5-Fold CV Method

Data Sets	**Accuracy**	**Precision**	**Recall**	**F1_Score**
Ga [27]	93.5	88.12	87.4	89.42
Ju [28]	94.42	92.50	90.21	95.60
Si [29]	92.52	90.20	91.21	89.1

TABLE VI. Performance after the Application of FS Method for Individual Sub-Dataset Utilizing 5-Fold CV Method

Data Sets	**Accuracy**	**Precision**	**Recall**	**F1_Score**
Ga [27]	95.54	91.12	89.4	91.42
Ju [28]	98.80	96.50	95.12	95.24
Si [29]	94.52	91.20	92.12	89.97

It is observed that when proposed model applied on three datasets provide the best accuracy result in the classification of patient into healthy and PD class. Along with that, by evaluating the differences in results between three datasets, it can be notified that proposed ML models show their best results in the case of the Ju dataset. It is also notified that results obtained on Ga are much better as compared to Si. Result can also present the fact that the smaller number of patients with little severity and high severity is more important. It is highly notifying that the patients with little severity may be considered healthy by ML models in some cases and may be considered patients in some cases. High severity patients can be easily separated from healthy subjects.

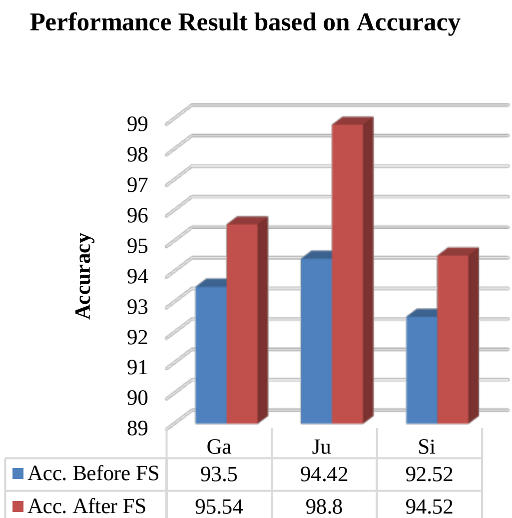

Fig. 7. Accuracy Score with and without Feature Selection.

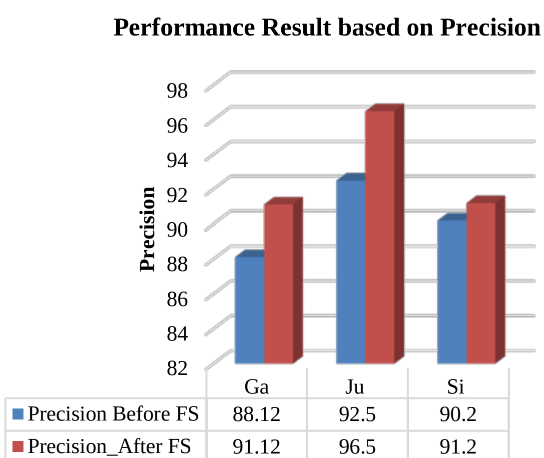

Fig. 8. Precision Score with and without Feature Selection.

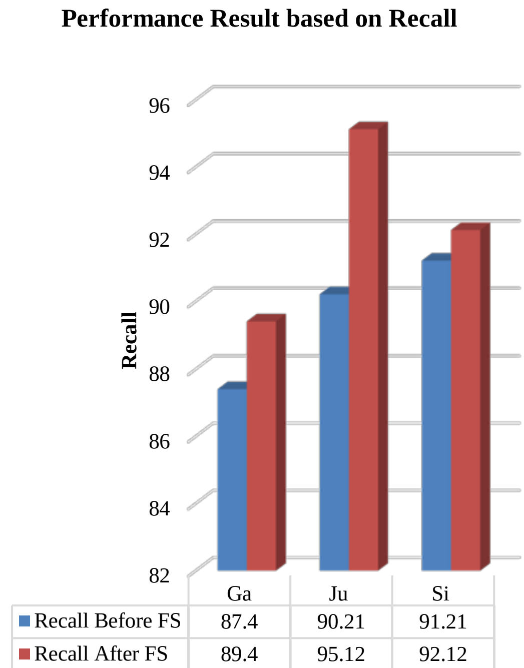

Fig. 9. Recall Score with and without Feature Selection.

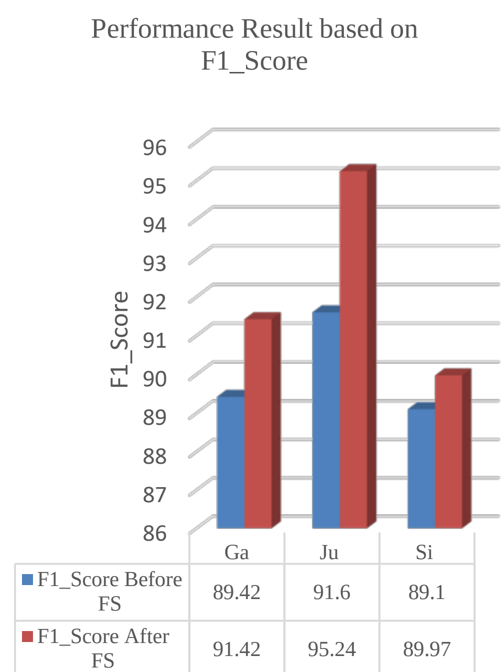

Fig. 10. F1_Score with and without Feature Selection.

Table 7 shows the comparison between the proposed work and the existing work. Results revealed that our proposed work is more precious than the existing work and will support maximum accuracy of 98.80%.

TABLE VII. ACCURACY COMPARISON BETWEEN PROPOSED WORK AND PREVIOUS WORK

Reference	Features Type	No. of Features	ML Technique	Accuracy Score (%)
Wu and Krishnan [35]	Time	3	SVM	90.32
Perumal and Sankar [36]	Frequency	10	SVM, NN, LDA	87.52-92.5
Abdul hay et al. [37]	Frequency and Time	3	MEDIUM Gaussian SVM	94.8
Khoury et al. [38]	Spatiotemporal	19	K-NN, DT, RF	83.3-92.86
Alam et al. [39]	Time	33	SVM, K-NN	93.6
Khera et al. [40]	Time Features	10	KNN, SVM, DT, RF	85-98.50
Proposed Work	Spatial, Time and Force	8	RFT	98.80 (Ju), 95.54 (Ga), 94.52 (Si)

VII. CONCLUSION

Wearable sensors offer a smart way of assessing the movement disorders among patients suffering from PD and provide a potentially vast quantity of informative data in order to quantify and monitor the progression of PD among patients. The current work, proposed a ML based PD diagnosis model and evaluated on VGRF dataset (Ga, Si, Ja) composed of different gait cycles that is publicly available on Psyionet. The performance of proposed model assessed by utilizing accuracy, precision, recall and F1_score as performance evaluation measures. The current work, presented that the proposed model outperformed other existing model to classify the patient into healthy and PD classes when subjected to preprocessed gait dataset with selected set of features. A five-fold cross validation method is implemented to achieve the accuracy result of 98.80 on Ju dataset, 95.54 on Ga dataset and 94.52 on Si dataset.

VIII. FUTURE DIRECTION

The conclusion reveals that the proposed model provides best results in the classification of PD but still faces some challenges. Future work concerns the introduction of more relevant features in order to enhance the diagnosis of PD and investigation of the cost-effective hybrid ML/Ensembles ML models and also handling the issue of imbalanced data.

ACKNOWLEDGMENT

I am highly thankful to Dr. Pooja Mittal for her comments and support in the preparation of this manuscript.

REFERENCES

[1] "Statistics|Parkinson's Foundation" , 2022.

[2] W. Yang, J. L. Hamilton, C. Kopil, C. B. James, M. T. Caroline, L A. Roger, E. R. Dorsey, N. Dahodwala, I. Cintina, P. Hogan, T. Thompson, "Current and projected future economic burden of Parkinson's disease in the U. S.," *npj (national publishing group) Parkinson's Dis.* , vol. 6, no. 15 , July 2020.doi: https://doi.org/10.1038/s41531-020-0117-1.

[3] Parkinson's disease: Causes, Symptoms, Stages, Treatment, Support (clevelandclinic.org), 2022.

[4] A. Z. Khan, F. Aamir, A. Kafeel, M. U. Khan, "Freezing of gait detection in parkinson's disease from accelerometer readings," Int. Conf. on Computing, July 2021.

[5] H. Zhonelue, Li. Gen, G. Chao, T. Yuyan, L.Jun, Z. Jin, L.Yun, Yu . Xiaoliu, R. Kang, C. Shengdi, " Prediction of freezing of gait in parkinson's disease using a random forest model based on an orthogonal experimental design: a pilot study, " Front. in Hum. Neurosci., vol. 15, 2021. doi: https://doi.org/10.3389/fnhum.2021.636414.

[6] G. Shalin, S. Pardoel, E. D. Lemaire, J. Nantel, J. Ofman, "Prediction and detection of freezing of gait in Parkinson's disease from plantar pressure data using long short-term memory neural-networks, " J. of Neuro Engi. Rehab. , vol. 18, no. 167, 2021. doi: 10.3390/s20154345.

[7] K. M Giannakopoulou, I. Roussaki, K. Demestichas, "Internet of things technologies and machine learning Methods for parkinson's disease diagnosis, monitoring, and management: a systematic review, " Sensors, vol. 22, no. 5, 2022. doi: 10.3390/s22051799.

[8] X. Jiang, C. Napier, B. Hannigan, J. J Eng., C. Menon, "Estimating vertical ground reaction force during walking using a single inertial sensor, " Sensors, vol. 20, no. 15, Aug. 2020. doi: 10.3390/s20154345.

[9] D. Castro, W. Coral, C. Rodriguez; J. Cabra, J. Colorado, "Wearable-Based Human Activity using an IoT approach, " J. Sens. Actuator Netw., vol. 6, no. 28, 2017. doi: https://doi.org/10.3390/jsan6040028.

[10] E. Abdulhay, N. Arunkumar, and N. Kumaravelu, E. Vellaiappan, "Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson's disease", Future Gen. Comp. Sys., vol. 83, June 2018. doi: https://doi.org/10.1016/j.future.2018.02.009.

[11] J. M. C. Desrosiers, and J. Frasnelli, "Machine Learning for the diagnosis of Parkinson's disease: A review of literature," Front. in aging neurosci., vol. 13, 6 May 2021. https://doi.org/10.3389/fnagi.2021.633752.

[12] P. Drotár, J. Mekyska, I. Rektorová, L., Masarová, Z. Smékal, and M. Faundez Zanuy, "Analysis of in-air movement in handwriting: a novel marker for Parkinson's disease," Comp. Methods and Prog. in Biomedicine, vol. 117, no. 3, pp. 405–411, 2014. Doi: https://doi.org/10.1016/j.cmpb.2014.08.007.

[13] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez Zanuy, "Decision support framework for Parkinson's disease based on novel handwriting markers, " IEEE Trans. on Neural Syst. and Rehailb. Eng., vol. 23, no. 15, pp. 508–516, May 2015.

[14] P. Drotár J. Mekyska, I, Rektorová, L. Masarová, Z. Smékal, and M. Faundez Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease, " Artif. Intell. in Med, vol. 67, 39–46, 2016.doi : 10.1109/TNSRE.2014.2359997.

[15] H. Choi, S. Ha, H. J. Im, S. H. Paek and D. S. Lee, "Refining diagnosis of Parkinson's disease with a deep learning-based interpretation of dopamine transporter imaging, " Neuroimage Clin., vol. 16, pp. 586–594, Sep. 2017. doi : 10.1016/j.nicl.2017.09.010.

[16] L. Silveira-Moriyama, A. Petrie, D. R. Williams, A. Evans, R. Katzenschlager, E. R. Barbosa, and A. J. Lees, "The use of a color-coded probability scale to interpret smell tests in suspected parkinsonism," Movement Disorders, vol. 24, no. 8, pp. 1144–1153, June 2009. doi: 10.1002/mds.22494.

[17] G. Valenza, S. Orsolini, S. Diciotti, L. Citi, E. P. Scilingo, M. Guerrisi, S. Danti, C. Luchetti, C. Tessa, R. Barbieri, et al., "Assessment of spontaneous cardiovascular oscillations in Parkinson's disease," Bio. Signal Process. and Control, vol. 26, pp. 80–89, April 2016. Doi: https://doi.org/10.1016/j.bspc.2015.12.001.

[18] F. Wahid, R. K. Begg, C. J. Hass, S. Halgamuge, and D. C. Ackland, ''Classification of Parkinson's disease gait using spatial-temporal gait features,'' IEEE J Biomed Health Inform. , vol. 19, no. 6, pp. 1794–1802. doi: 10.1109/JBHI.2015.2450232.

[19] C. G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. Counsell, & L. Seidl, " Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations the Movement Disorder Society Task Force on rating scales for Parkinson's disease" Mov Disord., vol. 19, no. 9, pp. 1020-1028, Sep. 2004. doi: 10.1002/mds.20213.

[20] L. J. W. Evers, J. H. Krijthe, M. J. Meinders, R. Bloem, T. M. Heskes "Measuring Parkinson's disease over time real-world within-subject reliability of the MDS-UPDRS, " Mov Disord., vol. 34, no. 10, pp. 1480-1487, Oct. 2019.

[21] B. Baker, W.Xiang and I. Atkinson, "Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities, in *IEEE Access*, vol. 5, pp. 26521-26544, 2017, doi: 10.1109/ACCESS.2017.2775180.

[22] M. Raza, M. Awais, S. Hussain, " Intelligent IoT framework for indoor healthcare monitoring of Parkinson's Disease Patient, " in *IEEE Journal on Selected Areas in Comm.* , vol. 39, no. 2, pp. 593-602, Feb. 2021, doi: 10.1109/JSAC.2020.3021571.

[23] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark & H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. vol. 101, no. 23, Jun 2000. doi: 10.1161/01.cir.101.23.e215. PMID: 10851218.

[24] G. Yogev, N. Giladi, C. Peretz, S. Springer, E.S. Simon, J. M Hausdorff, " Dual tasking, gait rhythmicity, and Parkinson's disease: Which aspects of gait are attention demanding?", Eur J Neuroscience, vol. 22 , no. 5, pp. 1248-56, Sep. 2005. doi: 10.1111/j.1460-9568.2005.04298.x.

[25] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, J. M. Hausdorff, " Treadmill walking as a pacemaker to improve gait rhythm and stability in Parkinson's disease", Mov Disord., vol. 20(9), pp.1109-1114. Sep. 2005. doi: 10.1002/mds.20507.

[26] J. M Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, N. Giladi," Rhythmic auditory stimulation modulates gait variability in Parkinson's disease", Eur J Neuroscience, vol. 26, pp. 2369-2375, 2007. doi: 10.1111/j.1460-9568.2007.05810.x.

[27] E. Balaji, D. Brindha, V. K. Elumalai and K. Umesh, "Data-Driven Gait Analysis for Diagnosis and Severity Rating of Parkinson's Disease, " Med.l Engi. and Physics, vol. 91, pp. 54-64, May 2021. doi: 10.1016/j.medengphy.2021.03.005.

[28] E. Balaji, D. Brindha and R. Balakrishnan, "Supervised Machine Learning based Gait Classification System or Early Detection and Stage Classification of Parkinson's Disease, " App Soft Compt J., vol. 94, 2020. doi: https://doi.org/10.1016/j.asoc.2020.106494.

[29] R. Prashanth, Sumantra Dutta Roy, Pravat K. Mandal, Shantanu Ghosh, "High-Accuracy detection of early Parkinson's disease through multimodal features and machine learning, " Int J Med Inform., vol. 90, pp. 13-21, 2016. doi: 10.1016/j.ijmedinf.2016.03.001.

[30] J. Sappakitkamjorn, S. A. Niwitpong, "Confidence intervals for the coefficients of variation with bounded parameters", Int J Math and Comput Sci., vol. 7, no. 9, pp. 1416–1421, 2013. doi.org/10.5281/zenodo.1087806.

[31] M. Krzywinski, N. Altman, "Classification and regression trees", Nat Methods, vol. 14, pp. 757–758, Aug. 2017. https://doi.org/10.1038/nmeth.4370.

[32] L. Breiman," Bagging predictors. Machine Learning, " vol. 24, pp.123–140, 1996.

[33] Navita, P. Mittal , "Healthcare Data Analysis using Data Mining Techniques for Disease Prediction", Indian Journal of Comp. Sci. & Eng., vol. 12, no. 5, Oct-Sep 2021.

[34] Scikit Learn. Available online: sklearn.model_selection.Grid Search CV — scikit-learn 1.1.1 documentation.

[35] Y. Wu, S. Krishnan," Statistical analysis of gait rhythm in patients with Parkinson's disease, " IEEE Trans on Neural Syst and Rehabil, vol. 18, no. 2, April 2010. doi: 10.1109/TNSRE.2009.2033062.

[36] S. V. Perumal, R. Sankar,"Gait and tremor assessment for patients with Parkinsons disease using wearable sensors, " ICT Express, vol. 2, no. 4, pp. 168-174, Dec. 2016. https://doi.org/10.1016/j.icte.2016.10.005.

[37] E. Abdulhay, N. Arunkumar, K. Narasimhan, E. Vellaiappan, V. Venaatraman, "Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease, " FGCS , vol. 83, pp. 366-373, 2018. doi: https://doi.org/10.1016/j.future.2018.02.009.

[38] N. Khoury, F. Attal, Y. Amirat, L. Oukhellou, S. Mohammed, "Data Driven based Approach to aid Parkinson's Disease Diagnosis, " Sensors, vol. 19, no. 2, Jan. 2019. doi: 10.3390/s19020242M.

[39] N. Alam, A. Garg, T.T. K Munia, R. Fazel-Rezai, K. Tavakolian, "Vertical ground reaction force marker for Parkinson's disease, " PloS One, vol. 12, no. 5, May 2017. https://doi.org/10.1371/journal.pone.0175951.

[40] P. Khera, N. Kumar, "Age-Gender Specific Prediction Model for Parkinson's Severity Assessment using Gait Biomarkers", Int.Journal of Engg. Sci. and Tech., vol. 27, March 2022. https://doi.org/10.1016/j.jestch.2021.05.009.