Gait Data-Driven Analysis of Parkinson's Disease Using Machine Learning

Archana Panda^{1*}, Prachet Bhuyan²

^{1,2}School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar India

Abstract

INTRODUCTION: Parkinson's disease is a progressive and complex neurological condition that mostly affects coordination and motor control. Parkinson's disease is most commonly associated with its motor symptoms, which include tremors, bradykinesia (slowness of movement), rigidity, and postural instability.

OBJECTIVES: Determine any minor alterations in walking patterns that could be early signs of Parkinson's disease. Track the course of Parkinson's disease over time by using gait data.

METHODS: In this study, we applied three types of VGRF datasets ("Dual Tasking, RAS, and Treadmill Walking") and developed an ML-based model using six different classifier methods. The datasets were analysed using 16 sensors, of which 8 were applied to each foot and the total pressure of the left and right foot. The aforementioned three distinct gait patterns movement disorders were the sources of the dataset. The gait signals dataset benefited by the participant demographic data.

RESULTS: Then, we passed the outcome of applying the model and measuring performance through a cross-validation operator to check the accuracy and decision-making of the five algorithms i) Deep Learning, ii) Neural Networks, iii) Support Vector Machine (SVM), iv) Gradient Boost Tree (GBT), v) Random Forest". The following findings compare the effectiveness of the various algorithms utilized and the observed PD very well.

CONCLUSION: The different ML classifier algorithms demonstrated good detection capability with different accuracy. Our proposed ensemble model is superior to compare with the existing models. Because we can observe the proposed ensemble model result and accuracy better than the other classifier model. The other classifier model's highest accuracy is 92.08% whereas our ensemble model got 92.31%. So, it has proved that our proposed ensemble model is excellent and robust.

Keywords: Parkinson's disease, VGRF, Machine Learning, Dual Tasking, RAS, Treadmill Walking.

Received on 18 December 2023, accepted on 12 March 2024, published on 19 March 2024

Copyright © 2024 A. Panda *et al.*, licensed to EAI. This is an open access article distributed under the terms of the <u>CC BY-NC-SA</u> 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetpht.10.5467

*Corresponding author. Email: <u>2081011@kiit.ac.in</u>

1. Introduction

Parkinson's disease is a progressive and complex neurological condition that mostly affects coordination and motor control. The medical profession has given the disease a great deal of attention since English physician James Parkinson first described it in 1817 [1]. Parkinson's disease, one of the most common neurological illnesses, is estimated to afflict millions of individuals worldwide. A feature of Parkinson's disease is the loss of dopamineproducing cells in the substantia nigra. Dopamine is a

neurotransmitter that helps brain cells communicate with one another, which is necessary for coordinating smooth and deliberate movement [2]. When dopamine-producing cells are compromised, Parkinson's disease patients' quality of life is significantly impacted by both motor and non-motor symptoms. Parkinson's disease is most commonly associated with its motor symptoms, which include tremors, bradykinesia (slowness of movement), rigidity, and postural instability. However, it is now understood that Parkinson's disease is a complex condition that can cause a wide range of additional symptoms, including issues with cognition, emotion, sleep, and the autonomic nervous system [3] [4The

interaction of these symptoms adds to the complexity of managing the condition and its treatment. Parkinson's disease currently has no known cure, although there are treatments that can help control symptoms and offer comfort to those who have the condition. Medication, physical therapy, lifestyle modifications, and occasionally even surgery are used as forms of treatment to manage symptoms and halt the disease's course. This study aims to identify patients with Parkinson's disease (PD) by analysing gait data. This analysis was conducted using data from Physionet's Gait Analysis Database. After walking for around two minutes, participants' VGRF records various forces in Newton as a function of time. Each subject has a total of sixteen sensors; eight of them are placed on the bottom of each foot to measure the pressure of the left and right foot over time (in seconds).

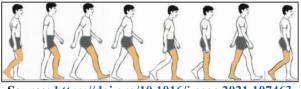
2. Related Work

Following a review of pertinent literature, the emphasis shifted to the use of gait analysis to differentiate between individuals with Parkinson's disease and healthy individuals. As the primary regulator in humans, the brain is crucial. Even one component being damaged might have far-reaching effects. One such instance is Parkinson's disease (PD) [5]. The majority of those over 50 will get Parkinson's disease, a neurological disorder [6]. Its symptoms progress gradually, and its onset is gradual. For both motor and nonmotor PD symptoms, there are two primary categories [7]. Nonmotor symptoms include mood problems, anxiety, depression, and dysfunction; motor symptoms include cognitive movement disorders, shaking, walking difficulty, stiffness, and postural instability. In order to distinguish PD patients from healthy controls, multiple technologies are applied. In one study, the feet of PD patients and healthy individuals were fitted with wireless inertial sensor systems (AHRS) in order to identify aberrant gait characteristics and distinguish PD patients from controls. In order to classify Parkinson's disease (PD) patients with or without gait impairment and healthy individuals, the authors examined for physical kinematic aspects of pitch, roll, and yaw rotations of the foot during walking [8]. A single mobile inertial sensor placed over the shoes was shown by Barth et al. to have great sensitivity (88%) and specificity (86%) in differentiating between healthy individuals and early Parkinson's disease sufferers [9]. Another work also suggested using a video infrared camera system to collect data for a Bayesian gait detection method. The collected data is used to create a picture frame matrix, skeleton numbering, depth frame matrix, and depth frame contour. Subsequently, MATLAB was employed to do extra analyses on the gathered data [10]. Eight sensors were placed under each foot and extrapolated features from vertical ground reaction force (VGRF) data acquired while individuals walked were used in the unique mathematical method proposed by Alam et al. to evaluate gait in people with

PD and healthy controls. The Gait pattern of healthy people and PD patients was evaluated using different machine learning classifiers and had an average accuracy of 85.21% "k-NN", 95.7% "SVM" sensitive 94.4% and specificity of 96.6% [11]. Automatic diagnosis of bradykinesia, the cardinal symptom of PD was the primary focus of Samà et al. study. They suggested a mathematical approach based on an SVM classifier and got a high accuracy of more than 90%, a high sensitivity of 92.52%, and a specificity of 89.07% [12]. Using the triaxial accelerometer found in the LG Optimus S. Arora et al. researched the viability and accuracy of creating a mobile app to objectively assess PD patients and differentiate them from healthy participants. After observing PD patient's controls for a month and performing gait tests using an accelerometer, they extracted 23 features in the frequency and time domain and used a random forest method to differentiate between PD and controls. The system achieved a combined accuracy of 98%, with a sensitivity of 98.5% and a specificity of 97.6% [13]. Dopaminergic-depleted PD patients walked more slowly and with shorter stride lengths, same cadence, and longer double support durations compared to healthy control subjects [14]. Dopamine in the substantia nigra can be identified in MRI images with the application of ML algorithms. However, due to the high price of MRI-based PD diagnosis, gaitbased PD analysis has become increasingly popular as an alternative method. Voice analysis [15], foot pressure systems [16], RGB-D cameras [17], optoelectronic motion analysis systems [18] and wearable sensors like accelerometers or inertial measurement units have all been studied using ML techniques in the medical field.

3. Methodology

In this study, we applied three types of VGRF datasets ("Dual Tasking, RAS, and Treadmill Walking") and developed an ML-based model using six different classifier methods. The datasets were analysed using 16 sensors, of which 8 were applied to each foot and the total pressure of the left and right foot. The aforementioned three distinct gait patterns movement disorders were the sources of the dataset. The gait signals dataset benefited by the participant demographic data.



Source: https://doi.org/10.1016/j.asoc.2021.107463

Figure 1: Different walking patterns for VGRF

Vertical Ground Reaction Force (VGRF) at a 100 Hz sampling rate is seen in the above figure. The subjects

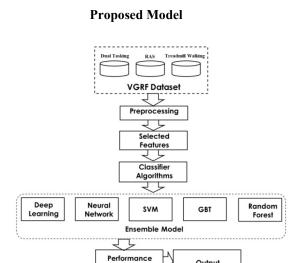


were directed to walk at their normal walking speed on level ground for two to five minutes over distances of twenty-five to seventy-seven metres while wearing a motorised treadmill equipped with dual tasking capabilities, a similarly wheeled walker, and Rhythmic Auditory Stimulation (RAS). It is crucial to recognise that those with Parkinson's disease or those in excellent health were the only ones included in the three studies; people with other walking difficulties were not included [19].

Time	11	12	13	14	15	16	17	18	R1	R2	R3	84	R5	Rő	R7	RB	TLP	TRP
0	176.22	149.82	88.55	50.16	10.78	62.15	63.36	36.08	194.81	191.18	89.87	67.98	19.03	96.25	54.45	64.35	637.12	777.92
0.01	176.22	149.82	88.55	50.16	10.78	62.15	63.36	36.08	194.81	191.18	89.87	67.98	19.03	96.25	51.45	64.35	637.12	777.92
0.02	174,13	149.82	88.55	50.16	10.78	62.15	65.67	36.08	192.39	191.18	89.87	70.4	19.03	98.67	54.45	64.35	637.34	780.34
0.03	172.15	147.62	86.24	50.16	10.78	62.15	63.09	38.28	190.08	189.09	89.87	70.4	19.03	98.67	56.54	67.32	635.47	781
0.04	170.06	145.42	86.24	50.16	12.98	62.15	70.4	38.28	187.77	187	87.67	70.4	19.03	101.09	58.85	70.18	635.69	781.99
0.05	168.08	143.22	86.24	50.16	12.98	64.46	72.82	38.28	185.46	187	87.67	72.93	19.03	103.4	61.27	76.01	636.24	792.77
0.06	166.1	141.02	84.04	50.16	12.98	64.46	77.55	38.28	185.46	184.91	87.67	72.93	21.12	103.4	61.27	78.98	634.59	795.74
0.07	162.03	138.82	81.84	50.16	15.18	64.46	79.86	38.28	183.04	184.91	87.67	75.46	21.12	105.82	63.69	81.84	630.63	803.55
0.08	158.07	134.42	81.84	50.16	15.18	66.77	84.59	40.59	180.73	182.82	87.67	75.46	21.12	108.24	66.11	90.53	631.62	812.68

Figure 2: Sample dataset of VGRF

This data contains 19 attributes, The first attribute is time in seconds, L1 to L8 measured left-foot pressure in Newton per second, R1 to R8 measured right-foot pressure in Newton per second, the TLP is the total pressure of the left foot and TRP is the total pressure of the right foot.



Output

Figure 3: The proposed model for gait data segmentation

Evaluation

As per our proposed model in Figure 3, we have collected the total number of data 32467 from the physioNet website. After that, we prepared the data with selected attributes, then we used the Ensemble model with "Deep Learning, Neural Networks, SVM, Decision Tree, and Random Forest" to classify the VGRF data of PD patients. Finally, we measured and evaluated the Accuracy, Kappa and specificity through our proposed model, we can quantify the efficacy of PD detection efforts with gait data analysis.

4. Ensemble Model in Machine Learning

In machine learning, an ensemble model is a method that takes the output of numerous distinct models and uses it to generate a single prediction that is both more accurate and more robust. The fundamental premise underlying ensemble approaches is that by integrating the benefits of many models, the total performance can be enhanced, and the shortcomings of individual models can be compensated for. From classification and regression to anomaly detection and beyond, ensemble approaches have proven to be highly effective in the machine learning field. Ensemble methods can be broadly categorized into two main types: bagging and boosting.

- Bagging, also known as Bootstrap Aggregating, is a method that involves training multiple models using different portions of the entire dataset. To create these subgroups, the training data is usually sampled at random using model replacement. Every is trained independently during prediction, and the outcome is determined by average (in regression) or voting (in classification) the predictions made by the several models. Random Forest is a wellliked ensemble approach in the bagging style. It combines multiple decision trees to create a model that is accurate and resistant to overfitting.
- Boosting: By training a large number of weak models (models that perform marginally better than random guessing) sequentially and giving them more weight, boosting gives priority to cases that were misclassified by several models. This adaptive approach is designed to improve model accuracy by addressing shortcomings in previous iterations. Well-known boosting techniques include XGBoost (Extreme Gradient Boosting), Gradient Boosting (GB), and AdaBoost (Adaptive Boosting).

An ensemble method integrates the forecasts from several separate models to provide a final prediction that is more reliable and accurate. Let T be the total number of base models, and let M1, M2,...., MT be the individual base models in the ensemble. A hypothesis function ht (X), which receives input characteristics X and generates predictions yt, is linked to each base model Mt. For a given input X, each base model Mt predicts a class label yt. The final prediction (\ensemble) of the ensemble is obtained by adding together the predictions of each of the base models. The class label that receives the most votes will be chosen as the final prediction in a majority voting process.

Because of their numerous advantages, ensemble models have become widely used in machine learning. The aim of ensemble models, a potent machine-learning technique, is to use many models to provide predictions that are more



robust and accurate. Due to their capacity to tackle difficult assignments and enhance performance in numerous applications, these methods have become commonplace in modern machine learning.

5. Data Description

Gait data comprises an extensive array of measurements and information about an individual's ambulatory pattern. This resource provides significant insights into diverse facets of human movement and health. Gait data encompasses various factors, including step length, step width, walking speed, cadence (measured in steps per minute), and stride duration. These metrics collectively offer a comprehensive depiction of an individual's walking pattern. The utilisation of sophisticated technologies such as wearable sensors, inertial measurement units, and motion capture systems facilitates the precise acquisition and examination of gait data. We have collected the VGRF signal pairs of three different walking patterns data i) Dual-Tasking, ii) Rhythmic Auditory Stimulation (RAS), and iii) Treadmill Walking refers to the ability to perform two different tasks simultaneously, which involves cognitive and motor coordination. In the context of Parkinson's disease.

- **Dual-Tasking**: It presents a significant challenge due to the complex interplay of motor and cognitive impairments that characterize the condition. Parkinson's disease often leads to difficulties in executing smooth and controlled movements, known as motor deficits, as well as cognitive impairments that can affect attention, memory, and executive functions. When individuals with Parkinson's disease engage in dual-task activities, such as walking while talking or carrying an object, their limited cognitive resources and compromised motor control can result in impaired performance. This phenomenon is known as dual-task interference.
- Rhythmic Auditory Stimulation (RAS): It is a therapy that uses rhythm and music to help people with neurological diseases, especially Parkinson's disease, improve their motor coordination and movement. RAS involves coordinating physical actions, like walking or exercises for the upper limbs, with an outside sound stimulus, usually a rhythmic beat or musical rhythm. The rhythmic cues from the auditory input work as an outside clock that helps control and improve the consistency of motor actions. In Parkinson's disease, which can make it hard to control your muscles, RAS has shown that it can help improve walking, reduce times when walking stops, and improve the quality of movements overall. By giving people, a structured beat to move to, RAS may be able to tap into neural pathways that help people move

more smoothly and in sync with each other. RAS is not only a healing way to deal with motor symptoms, but it also shows the amazing link between music and the brain's ability to control movement. It is a fun and non-invasive way to help people with neurological conditions.

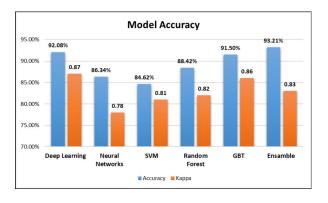
Treadmill Walking: People with Parkinson's disease can benefit from using motorised treadmills because they offer a structured and controlled setting in which to work on motor [] related problems. People with Parkinson's disease often have trouble walking and moving around. Motorised treadmills are used in rehabilitation situations because the speed and incline can be changed to help with these problems. Motorised treadmills help improve gait pattern, stride length, and cadence by giving you visual and audible cues, like goals and metronome beats. They also help reduce the times when stop moving. Also, the constant movement of the treadmill makes it easier to start taking steps, which can be especially helpful for people who have trouble thinking. People with Parkinson's disease who use motorised treadmills regularly may be able to improve their walking, balance, and general mobility. This could help improve their quality of life.

6. Data Analysis and Findings

We used data from a diagnosed PD patient to assess the significance of the symptoms. Parkinson's disease is characterized by the severity of gait symptoms. The current results are inferior to those published for several ML-based classification models for diagnosis. The decision tree for PD early detection is given below, and the accuracy of five alternative classifiers and ensemble algorithms is defined in the model. Several phases and operators have been built into the proposed model that we have constructed. In this model, we first used the selected attribute operator to extract the important attributes from the aforementioned dataset and then used the set role operator to pick a categorical variable from among the important attributes. Then, we passed the outcome of applying the model and measuring performance through a cross-validation operator to check the accuracy and decision-making of the five algorithms i) Deep Learning, ii) Neural Networks, iii) Support Vector Machine (SVM), iv) Gradient Boost Tree (GBT), v) Random Forest". The following findings compare the effectiveness of the various algorithms utilized and the observed PD very well.



6.1 Findings



The above figure shows the individual six ML algorithms of the Gait Data analysis and can conclude the Deep Learning accuracy is 92.08% the kappa is 0.87, the Neural Networks accuracy is 86.34% and the kappa is 0.78, the Support Vector Machine (SVM) accuracy is 84.62%, the kappa is 0.81, the Random Forest accuracy is 88.42%, the kappa is 0.82, the Gradient Boosted Trees (GBT) accuracy is 91.50% the kappa is 0.86, and the Ensemble model shows the highest accuracy is 93.21% the kappa is 0.83. With the said result, we can say our Ensemble model is robust in the Gait Data analysis for Parkinson's Disease (PD). The following figures are shown as per the conditions detected classification of healthy and PD patients from Gait Data. Figure 5 shows the Dual-Tasking left and Right foot to identify healthy and PD patients. Figure 6 shows the RAS left and Right foot to identify healthy and PD patients. Figure 7 shows the Treadmill Walking left and Right foot to identify healthy and PD patients. The blue colour line indicates healthy, and the orange colour line indicates PD patients.

Dual Tasking

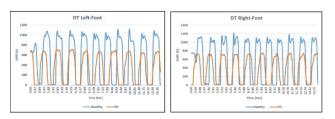


Figure 5: VGRF Dual Tasking Left and Right Foot

Rhythmic Auditory Stimulation (RAS)

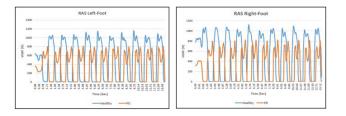
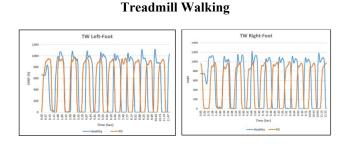
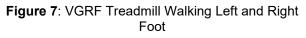


Figure 6: VGRF RAS Left and Right Foot





Three distinct walking patterns are transmitted by VGRF signals in pairs

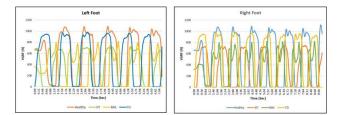


Figure 8: Three distinct walking patterns Left and Right Foot

Figure 8 shows the three distinct walking patterns Dual-Tasking, RAS, and Treadmill Walking left and Right foot to identify healthy and PD patients.

7. Conclusion

In this study, we used machine learning methods to investigate the relationship between Parkinson's disease and to develop a model for identifying Parkinson's disease. The training data is denoted by {(X1, y1),, (Xn, yn), and the classifiers "Deep Learning, Neural Networks, Support Vector Machine (SVM), Random Forest, Gradient Boost Tree (GBT), and Ensemble Model" are used in our study. The classification of Parkinson's disease is critical for gaining a better understanding of the disease's causes, initiating treatment methods, and developing appropriate treatments. Based on the Gait data, this study proposed an empirical model for automatically distinguishing between healthy and PD patients.

The different ML classifier algorithms demonstrated good detection capability with different accuracy. Our proposed ensemble model is superior to compare with the existing models. Because we can observe the proposed ensemble model result and accuracy better than the other classifier model. The other classifier model's highest accuracy is 92.08% whereas our ensemble model got 92.31%. So, it has proved that our proposed ensemble model is excellent and robust.



8. References

- Parkinson J. (1817). An Essay on the Shaking Palsy. London: Sherwood, Neely, and Jones.
- Kalia, L.V., Lang, A.E. (2015). Parkinson's disease. Lancet., 386(9996), 896-912. doi:10.1016/S0140-6736(14)61393-3
- [3] Chaudhuri, K. R, Odin, P., Antonini, A., Martinez-Martin, P. (2011). Parkinson's disease: the non-motor issues. Parkinsonism Relat Disord. 17(10), 717-723. doi:10.1016/j.parkreldis.2011.08.018
- Jankovic., J. (2008). Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry, 79(4), 368-376. doi:10.1136/jnnp.2007.131045
- [5] Alissa, M. (2021). Parkinson's Disease Diagnosis Using Deep Learning. arXiv:2101.05631
- [6] Burke, R.E., O'Malley, K. (2013). Axon degeneration in Parkinson's disease. Exp. Neurol. 246, 72–83.
- [7] Gunduz, H. (2019). Deep Learning-Based Parkinson's Disease Classification Using Vocal Feature Sets. IEEE Access, 7, 115540–115551.
- [8] Tien I., Glaser S. D., Aminoff M. J. (2010). Characterization of gait abnormalities in Parkinson's disease using a wireless inertial sensor system. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology, 3353–3356.
- [9] Barth J., Klucken J., Kugler P., Kammerer T., Steidl R., Winkler J., Hornegger J., Eskofier B. (2011). Biometric and mobile gait analysis for early diagnosis and therapy monitoring in Parkinson's disease. Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 868–871.
- [10] Procházka A., Vyšata O., Vališ M., Upa O., Schätz M., Mařík V. (2015). Bayesian classification and analysis of gait disorders using image and depth sensors of Microsoft Kinect. Digit. Signal Process. A Rev. J., 47, 169–177. doi: 10.1016/j.dsp.2015.05.011.
- [11] Alam M.N., Garg A., Munia T.T.K., Fazel-Rezai R., Tavakolian K. (2017). Vertical ground reaction force marker for Parkinson's disease. PLoS ONE., 12, e0175951. doi: 10.1371/journal.pone.0175951.
- [12] Samà A., Pérez-López C., Rodríguez-Martín D., Català A., Moreno-Aróstegui J.M., Cabestany J., de Mingo E., Rodríguez-Molinero A. (2017). Estimating bradykinesia severity in Parkinson's disease by analysing gait through a waist-worn sensor. Comput. Biol. Med., 84, 114–123. doi: 10.1016/j.compbiomed.2017.03.020
- [13] Arora S., Venkataraman V., Donohue S., Biglan K.M., Dorsey E.R., Little M.A. (2014). High accuracy discrimination of Parkinson's disease participants from healthy controls using smartphones. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3641–3644.
- [14] M. Svehlík, E.B. Zwick, G. Steinwender, et al. (2009). Gait analysis in patients with Parkinson's disease off dopaminergic therapy. Arch Phys Med Rehabil, 90,1880-1886. doi: 10.1016/j.apmr.2009.06.017
- [15] Ozkan, H. (2016). A comparison of classification methods for telediagnosis of Parkinson's disease. Entropy 18, 115. doi: 10.3390/e18040115
- [16] Abdulhay, E., Arunkumar, N., Narasimhan, K., Vellaiappan, E., and Venkatraman, V. (2018). Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson's disease. Future Gener.

Comput. Syst. 83, 366–373. doi: 10.1016/j.future.2018.02.009

- [17] Buongiorno, D., Bortone, I., Cascarano, G. D., Trotta, G. F., Brunetti, A., and Bevilacqua, V. (2019). A low-cost vision system based on the analysis of motor features for recognition and severity rating of Parkinson's disease. BMC Med. Inform. Decis. Mak., 19, 243. doi: 10.1186/s12911-019-0987-5
- [18] Varrecchia, T., Castiglia, S. F., Ranavolo, A., Conte, C., Tatarelli, A., Coppola, G., et al. (2021). An artificial neural network approach to detect the presence and severity of Parkinson's disease via gait parameters. PLoS One 16, e0244396. doi: 10.1371/journal.pone.0244396
- [19] Frenkel-Toledo S., Giladi N., Peretz C., Herman T., Gruendlinger L., Hausdorff J.M. (2005). Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease. Mov. Disord., 20, 1109–1114. doi: 10.1002/mds.20507

