

Parkison's Disease Detection using Machine Learning Algorithms: a brief review

Manjunath Patil¹, S.B.Patil²

^{1,2} Department of Electronics and Telecommunication Engg., D.Y. Patil College of Engineering & Technology, Kolhapur

Abstract— Parkinson's disease (PD) is the second most common neurological disease after Alzheimer's disease, affecting 2–3% of the population older than 65 years of age. Parkinson's disease is a neurological disorder. This disease breaks down the cells in the nervous system. This disease reduces the dopamine level in the brain. Mainly it affects the fore-brain region. It is one of the chronic degenerative diseases with progressive illness, which means it develops new symptoms over time. Parkinson's disease (PD) is one of the major public health diseases in the world which is progressively increasing day by day and had its effect on many countries. So, it is very important to predict it in the early stage which has been a challenging task among researchers because the symptoms of disease come into existence in either middle or late middle stage. There are different Parkinson's disease detection techniques are proposed and these depend upon the voice of a person, handwriting, and some techniques are proposed using images. Machine learning algorithms like KNN, Logistic Regression, Decision Tree, and Random Forest shows better prediction about Parkinson's disease. In this paper, we have surveyed different Parkinson disease detection techniques which depend on with and without machine learning algorithms.

Keywords— Decision Tree, Electroencephalography, KNN, Logistic Regression, Machine learning, Parkinson's disease, Random Forest

I. INTRODUCTION

Neurodegenerative disorders are one of the results of the progressive tearing and neurons or some part of neuron loss in different areas of the nervous system. Neurons are the basic functional unit of brain activity. They are contiguous rather than continuous. A good healthy looking neuron as shown in fig. 1 has extensions called dendrites or axons, a cell body and a nucleus that contains our DNA.

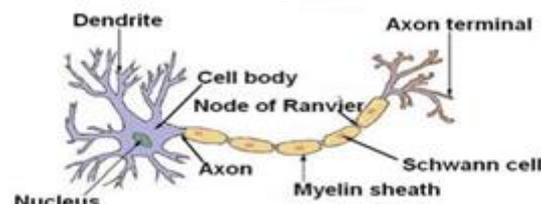


Fig.1 : Structure of neuron present in human brain

Neurodegenerative disorders are one of the results of progressive tearing and neurons or some part of neuron loss in different areas of the nervous system. Neurons are the basic functional unit of brain activity. They are contiguous rather than continuous. A good healthy looking neuron is as shown in fig 1 which has extensions called dendrites or axons, a cell body, and a nucleus that contains our DNA. DNA system is our genome and a hundred billion neurons contain our entire genome which is

packaged into it. When a neuron gets sick, then it loses its extension to neighbor neurons and hence its ability to communicate which is not good for it and its metabolism becomes low so it starts to accumulate junk and it tries to contain the junk into the little packages in little pockets presents in the brain. When things become worse and if the neuron is a cell culture it completely loses its communication, becomes round in nature, and full of vacuoles. Our work deals with the prediction of Parkinson's disorder in the early-stage which is nowadays is a tremendously increasing incurable disease. Parkinson's disease is likely to be the most spreading disease. Parkinson's disease gets its name from James Parkinson who earlier described it as a paralysis agitans and later gave his surname was known as Parkinson Disease (PD). It generally affects the nervous system and basically on neurons which is responsible for overall body movements or metabolism. The main chemicals involved in human brain activity are dopamine and acetylcholine which affect the human brain. Parkinson's Disease is symbolize by the loss of dopaminergic neurons in the substantia nigra.

There are various environmental factors that have been implicated in Parkinson's disease. Below is some of the listed factor which caused Parkinson's disease in an individual.

1.1. Environmental factors:

Environment is nothing but the surrounding or the place in which an individual lives .So environment is one of the major factors that will not only affects the human's brain but also affects all the living organisms that lives in the vicinity of it. Many researches and evidences have already proved that the environment is one of the big hands in the development of neurodegenerative disorders mainly like Alzheimer's and Parkinson's. These two diseases are most common and affected neurological diseases. There are certain environmental factors that are influencing neurodegenerative disorder with high pace are-

- a. Exposure to heavy metals (like lead, mercury, cadmium and aluminium) and pesticides like fungicides, herbicides.
- b. Air Quality: Pollution in the air due to different unwanted gases, materials results in respiratory diseases i.e. polluted air also effects on the neurological system through respiratory system.
- c. Water quality: Water is one of the biotic parameter. Biotic and Abiotic contaminants present in water leads to water pollution.
- d. Unhealthy lifestyle: It leads to obesity and sedentary lifestyle.
- e. Psychological stress: It increases level of stress hormone that depletes the functions of Neurons.

1.2 Brain injuries or Biochemical Factors:

Brain is a control central of our complete body system. Due to certain trauma affected peoples have a brain injury which leads some biochemical enzymes to come into picture which provides neurons stability and provide support to some chromosomes and genes in maintenance.

1.3 Aging Factor:

Aging is associated with different changes in dynamic behavioral, biological, environmental, physiological, psychological, and social processes. Aging is one of the main reasons for the development of the Parkinson's disease.

1.4 Genetic factors:

Genetic factor is considered as the main and mostly affected molecular physiological cause which leads to neurodegenerative disorders. The size, depth and effect of actions of different genes define the status or level of neurodegenerative disease which increases itself gradually overtime. Mainly the genetic factors which lead to Neurodegenerative disorders are categorized into pharmacokinetics and pharmacodynamics.

1.5 Speech Articulation factors:

Due to the condition associated with Parkinson's disease Like rigidity and bradykinesia, some speech language pathology such voice , articulation and Swallowing alterations are found . There are various ways in which Parkinson's disease (PD) might affect the individual like-

- (i) The voice gets breathy and softer.
- (ii) Speech may be smeared.
- (iii) The person finds difficulty in finding the right words due to which speech becomes slower.

In the world, Ten percent of people aged 65 or more do have a neurodegenerative disease, and there are no cures for them. Almost 30% of the people are facing this incurable disease. Current treatment, if available at all, only reduces symptoms and that too for a limited period of time. The main cause for Parkinson's disease is the accumulation of protein molecules in the neuron which gets mis-folded and hence causing Parkinson's disease. So till now, researchers got the symptoms and the root causes i.e. from where this disease had evolved. But very few have come to its cure. So in this era where Parkinson's disease is progressing at a double pace, it is very important to find the solution which can detect it in its early phases.

Motor symptoms

This is a symptom where any voluntary action is involved. It indicates movement-related disorders like tremor, rigidity, freezing, Bradykinesia, or any voluntary muscle movement.

Non-motor symptoms

Non-motor symptoms include disorders of mood and affect with apathy, cognitive dysfunction as well as complex behavioral disorders. There are two other categories of Parkison's Disease which are divided by doctors: Primary symptoms and Secondary symptoms.

Primary symptoms

It is the most important symptom. Primary symptoms are rigidity, tremor, and slowness of movement.

Secondary symptoms

It is a symptom that directly impacts the life of an individual. These can be either motor or non-motor. Its effect depends on person to person.

Machine learning is one of the new era where machines are learn for provided data called as test data and predict the output for other data called as test data. Deep learning is a subset of machine learning where data is learns by using different layers called as neurons. In this paper, we did a survey on the prediction of Parkison's disease using different Machine learning and Deep learning algorithms. Also, we mentioned the accuracy of each algorithm for a particular dataset.

II. LITERATURE REVIEW

Prediction of Parkinson disorder is one of the most important problem that has to be detected in the early phases of the commencement of the disease so as to reduce the disease progression rate among the individuals .Various researches have been made to find the basic cause and some have reached to the heights by proposing a system which differentiates the healthy people from those with any ND'S using various machine learning techniques. Lots of pre-processing feature selection and classification techniques have been implemented and developed in the past decades. Following is the given work done in the prediction of Parkinson's disorders. We have categories the review into three parts i.e.

- (i) Review on Pre-processing techniques.
- (ii) Review on classification methods.
- (iii) Review on different computational methodology.

2.1 Review on pre-processing techniques used in Parkinson's disease prediction

Sahoo et. al (2012) [2] reported a study for the prediction of Parkinson's disease using Data mining techniques. The three methods used i.e. Decision stump, Logistic Regression, and Sequential

Minimization Optimization. The results inferred, the support vector machine model outstands among the others with an accuracy of 76%, sensitivity 0.97 while in terms of specificity statistical model has done well with 0.62 as compared with two other. Bonato et. al (2004) [3] have proposed evidence that data mining and artificial intelligence may help in recognizing the severity of motor fluctuations in PD patients. They collected the data using ACC (accelerometer) and EMG (electromyography) signals which were recorded while executing standardized sets of motor assessment tasks. In another study, Saritha .k et al (2017) [4] have implemented a javascript program to record the voice of the patient and later used Praat to convert that accepts input in .wav file and using a script yields a voice report. The decision tree gave the best results among the applied algorithm with the accuracy of 100% without feature selection and with feature selection, it is 94%.

2.2 Review on classification techniques used in prediction of neurodegenerative disorders

Nayan reddy Challa et al (2016) [1] have discussed the importance of non-motor systems which was neglected by many doctors over motor systems. In the study Rapid eye movement (REM), sleep behavior distortion, and olfactory loss were considered, and using four machine learning techniques i.e. Multilayer Perceptron, Bayes Net, RF, and Boosted Logistic Regression, prediction is performed. Among which Boosted logistic regression with an accuracy of 97.159% and area under ROC curve was 98.9%.is considered as a better method. Chandrayan et al. (2016) [5] proposed extreme learning machines to predict PD. Using ELM they have done a comparative analysis and inferred that unlike conventional Neural Network elm doesn't require an iterative variation of hidden neurons. So the simple architecture makes elm a more reliable method than others for prediction. Jennifer He et al. (2017) [6] observed that the best feature for the prediction of Parkinson's disease is fundamental frequency among all voice recording features. They have tested various machine learning methods which include Boosted decision tree, Decision jungle, Locally Deep SVM, Logistic regression, Neural Networks, and SVM on Microsoft azure machine learning studio amongst which the best is Two-class Boosted decision trees, which is an ensemble technique. Weitschek et al. (2014) [7] uses EEG Electroencephalography to diagnose brain abnormalities. They have given an automatic patient classification from the EEG biomedical signals involved in Alzheimer's disease and MCI in order to support medical doctors. The authors performed pre-processing using time-frequency transforms and then applied classification using machine learning. Rodrigues et al. (2012) [8] uses K-mean which obtain (EEG) temporal events in order to improve AD diagnosis. They achieved the sequence of EEG energy variations that are found more frequent in AD patients than in any healthy person. Fernandez-Ruiz et al. [9] found that Alzheimer's disease had shown a volume reduction in some regions of the brain. Some areas like precuneus start showing changes when measured through Magnetic Resonance Imaging. So in their study, they took precuneus as a biomarker to identify defects in the brain using machine learning techniques. Johnstone et al. [10] took the dataset which was collected from ADNI (Alzheimer's disease Neuroimaging Initiative) of plasma proteome. They applied combinatorial optimization i.e. feature selection. So they differentiated the MCI patient and the AD patient depending on whether APOE was included or some other factors were there. At last, they get an accuracy of 90% by generating the signature longitudinal rather than cross-sectional data which further improved the classification. Rathore et al. (2016) [16] used various machine learning techniques mainly regression techniques. After comparing the ML techniques, error rates have been calculated i.e. AAE and ARE. K fold validation is applied to validate the results. At last, Kruskal Wallis test and Dunn's multiple comparison test are used to doing a comparative analysis of techniques used In 2016, Kumar Tiwari [17] proposed a minimum redundancy maximum relevance feature selection algorithm to select the most important feature which alone can predict Parkinson's disease. He observed that the random forest provided an overall accuracy of 90.3% which is better in comparison to all other machine learning-based approaches such as bagging, rotation forest, random subspace, support vector machines, etc. Mamoshina et al (2016) [18] represented his work using deep learning as he states that it is different from traditional feature learning techniques. He uses deep learning with multiple hidden

layers so as to provide a meaningful and higher level of abstraction. Fig illustrates the approach in three steps.

- (i) Started by pre-processing raw data to overcome main issues such as missing values, outliers, and data quality.
- (ii) Second step is to apply unsupervised deep learning for producing a higher level of abstraction of input data.
- (iii) Finally supervised learning method is implemented for predicting the target value and model evaluation.

2.3 Review on different methodologies used in neurodegenerative disorder’s prediction

Bioinformatics is emerging day by day and lots of researchers are now inclined towards this branch of science as Bioinformatics deals with the biological aspects of individuals like health, nutrition, environment. One of the most trending disorders is neurological disorders which have shown a tremendous increase in recent years. So by analyzing all the disorders, we found that different researchers have used different technologies to distinguish the ND patient from the healthy one. The technologies used are big data processing, Virtual reality, facial and emotion recognition, handwriting recognition, and artificial intelligence.

All these techniques are summarized in following table-

Sr. No.	Author	Paper name	Method	Accuracy (in %)
1	Kamal Nayan Reddy Challa	An improved approach for prediction of Parkinson's disease using machine learning techniques	Boosted Logistic Regression	97.159
2	Geeta Yadav	Predication of Parkinson's disease using data mining methods: a comparative analysis of tree, statistical, and support vector machine classifiers	tree classifier, statistical classifier, and support vector machine classifier	97
3	Aarushi Agarwal	Prediction of Parkinson's disease using speech signal with Extreme Learning Machine	Extreme Learning Machine	90.76
4	Marius Ene et al,	Neural network-based approach to discriminate healthy people from those with Parkinson’s disease	PNN	79-81
5	David Gil A, Magnus Johnson B	Diagnosing Parkinson by using Artificial Neural Networks and Support Vector Machines	ANNs and SVMs	90
6	Resul Das et al	A comparison of multiple classification methods for diagnosis of Parkinson disease	Neural Networks, DMneural, Decision Tree and Regression	93
7	Leandro A. Passos	Parkinson Disease Identification using Residual Networks and Optimum-Path Forest	ResNet-50 , Optimum-Path Forest (OPF) classifier	96
8	Salama A. Mostafa	Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson’s disease	Decision Tree, Random Forests, Naïve Bayes, Support Vector Machine and Neural Network.	accuracy improved in the raange 5-15%

Table 1 : Comparative study of different Parkinson's disease detection techniques

III. CONCLUSION

Machine learning techniques play a fundamental role in data analysis, allowing one to obtain patterns and relationships between different classes automatically and efficiently. In Parkinson's disease detection different machine learning as well as deep learning based techniques shows better prediction as compared to other existing diagnostic methods. This paper presents a brief review on ML techniques like- Logistic regression, Decision Tree, Random Forests, Naïve Bayes, Support Vector Machine and Neural Network etc.

REFERENCES

- [1] Kamal Nayan Reddy, Challa, Venkata Sasank Pagolu and Ganapati Panda, "An Improved Approach for Prediction of Parkinson's Disease using Machine Learning Techniques", in Proceedings of the International conference on Signal Processing, Communication, Power and Embedded System (SCOPE)-2016, pp. 1446-145, 2016.
- [2] Geeta Yadav, Yugal Kumar and G. Sahoo, "Prediction of Parkinson's disease using Data Mining Methods: a comparative analysis of tree, statistical and support vector machine classifiers", in Proceedings of the National Conference on Computing and Communication Systems (NCCCS), pp. 1-4, 2012.
- [3] Paolo Bonato, Delsey M. Sherrill, David G. Standaert, Sara S. Salles and Metin Akay, "Data Mining Techniques to Detect Motor Fluctuations in Parkinson's Disease", in Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4766-4769, 2004.
- [4] Sonu S. R., Vivek Prakash and Ravi Ranjan, "Prediction of Parkinson's Disease using Data Mining", in Proceedings of the International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 1082-1085, 2017.
- [5] Aarushi Agarwal, Spriha Chandrayan and Sitanshu S Sahu, "Prediction of Parkinson's Disease using Speech Signal with Extreme Learning Machine", in Proceedings of the International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 1-4, 2016.
- [6] Akshaya Dinesh and Jennifer He, "Using Machine Learning to Diagnose Parkinson's Disease from Voice Recording", in Proceedings of the IEEE MIT Undergraduate Research Technology Conference (URTC), pp. 1-4, 2017.
- [7] Giulia Fiscon, Emanuel Weitschek, Giovanni Felici and Paola Bertolazzi, "Alzheimer's disease patients classification through EEG signals processing", in Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM), pp 1-4, 2014.
- [8] Pedro Miguel Rodrigues, Diamantino Freitas and Joao Paulo Teixeira, "Alzheimer electroencephalogram temporal events detection by K-means", in Proceedings of the International Conference on Health and Social Care Information Systems and Technologies HCIST, pp. 859 – 864, 2012.
- [9] Elva Maria Novoa-del-Toro, Juan Fernandez-Ruiz, Hector Gabriel Acosta-Mesa and Nicandro Cruz-Ramirez, "Applied Machine Learning to Identify Alzheimer's Disease through the Analysis of Magnetic Resonance Imaging", in Proceedings of the International Conference on Computational Science and Computational Intelligence, pp. 577-582, 2015.
- [10] Daniel Johnstone¹, Elizabeth A. Milward¹, Regina Berretta¹ and Pablo Moscatol¹, "Multivariate Protein Signatures of Pre-Clinical Alzheimer's Disease in the Alzheimer's Disease Neuroimaging Initiative (ADNI) Plasma Proteome Dataset", in Proceedings of the Disease Neuroimaging Initiative, vol-7, pp. 1-17, 2017.
- [11] Jason Orlosky, Yuta Itoh, Maud Ranchet, Kiyoshi Kiyokawa, John Morgan, and Hannes Devos, "Emulation of Physician Tasks in Eye-tracked Virtual Reality for Remote Diagnosis of Neurodegenerative Disease", in Proceedings of the IEEE Transactions on Visualization and Computer Graphics, vol. 23, pp. 1302 – 1311, 2017.
- [12] Mathew J. Summers, Vienna, Austria, Alessandro E. Vercelli, Georg Aumayr, Doris M. Bleier and Ludovico Ciferri, "Deep Machine Learning Application to the Detection of Preclinical Neurodegenerative Diseases of Aging", in Proceedings of the Scientific Journal on Digital Cultures, vol. 2, pp. 9-24, 2017.
- [13] Bianca Torres, Raquel Luiza Santos, Maria Fernanda Barroso de Sousa, Jose Pedro Simoes Neto, Marcela Moreira Lima Nogueira, Tatiana T. Belfort¹, Rachel Dias¹, Marcia and Cristina Nascimento Dourado, "Facial expression recognition in Alzheimer's disease: a longitudinal study", pp. 383-389, 2014.
- [14] Smitha Sunil and Kumaran Nair, "An exploratory study on Big data processing: a case study from a biomedical informatics", 3rd MEC International Conference on Big Data and Smart City, pp. 1-4, 2016.
- [15] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis and M. Arnaoutoglou, "Machine learning-based classification of simple drawing movements in Parkinson's disease", in Proceedings of the Biomedical Signal

Processing and Control, pp. 174–180, 2017.

[16] Santosh S. Rathore and Sandeep Kumar, “An empirical study of some software fault prediction techniques for the number of faults prediction”, in Proceedings of the Soft Computing, vol. 21, pp 7417–7434, 2017.

[17] Arvind Kumar Tiwari, “Machine Learning Based Approaches for Prediction of Parkinson’s Disease”, in Proceedings of the Machine Learning and Applications: An International Journal (MLAIJ), vol.3, pp. 33-39, 2016.

[18] Polina Mamoshina, Armando Vieira, Evgeny Putin and Alex Zhavoronkov, “Applications of Deep Learning in Biomedicine”, in Proceedings of the American Chemical Society Mol. Pharmaceutics, pp. 1445–1454, 2016.

[19] Alexis Elbaz, James H. Bower, Brett J. Peterson, Demetrius M. Maraganore, Shannon K. McDonnell, J. Eric Ahlskog, Daniel J. Schaid, Walter A. Rocca, “Survival Study of Parkinson Disease in Olmsted County, Minnesota”, Arch Neurol. Vol. 60 pp. 91-96, 2003.

[20] Tanner CM, Ross GW, Jewell SA, ” Occupation and risk of Parkinsonism: a multicentre case- control study” Arch Neurol,66(9):1106–1113,2009.

[21] Marius Ene et al, Neural network-based approach to discriminate healthy people from those with Parkinson’s disease, Annals of the University of Craiova, Math. Comp. Sci. Ser. Volume 35, 2008, Pages 112–116 ISSN: 1223-6934

[22] DAVID GIL A, MAGNUS JOHNSON B, Diagnosing Parkinson by using Artificial Neural Networks and Support Vector Machines, Global Journal of Computer Science and Technology, 2009

[23] Resul Das et al, A comparison of multiple classification methods for diagnosis of Parkinson disease, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 13, NO. 2, MARCH 2002.

[24] Leandro A. Passos, Parkinson Disease Identification using Residual Networks and Optimum-Path Forest, SACI 2018, IEEE 12th International Symposium on Applied Computational Intelligence and Informatics, May 17-19, Timiúoara, Romania

[25] S. A. Mostafa, A. Mustapha, M. A. Mohammed et al., Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson’s disease, Cognitive Systems Research , 2018.