International Conference on Intelligent Technologies & Science - 2022
(ICITS-2022)

Machine Learning based Automated Diagnosis of Parkinson Disease

Aditya Mani Tripathi[1], Dr. C. L. P. Gupta[2]
Computer science and Engineering Department,
Bansal Institute of Engineering and Technology, Lucknow
adityamani33@gmail.com, clpgupta@gmail.com

Abstract: The Parkinson disease prediction has attracted many researchers over a decade. The Parkinson disease dataset is available on the uci machine learning repository website. Researchers have used the various techniques of data mining, machine learning and deep learning to build the prediction system for early detection of Parkinson disease at an early stage so that it can be controlled because the health care for this disease is not affordable for developing and under developed countries. In our proposed work we will pre-process the data based on the exploratory data analysis. And build a prediction system for the Parkinson disease by optimizing machine learning models like kneighbors, logistic regression, decision tree classifier, random forest classifier. All the algorithm is employed to predict the data set and a multi-dimensional result analysis and choose the best fitting algorithm. Thee dataset is generated on study of 31 person out of which 23 patients suffers with the disease.

Keywords: Parkinson's Disease, Classification, Random Forest, Support Vector Machine, Machine Learning.

1. Introduction:

Parkinson's disease (PD) is a chronic, progressive neurodegenerative disease characterized by loss of nigrostriatal dopaminergic pathways [1]. Its prevalence in the general population is 0.1-0.3% [2], showing an increase in individuals aged >65 years [3]. Cardinal findings in PD are tremor, rigidity, akinesia (i.e., bradykinesia, hypokinesia) and postural instability. In addition to motor disturbances, non-motor signs and symptoms are also common in these patients. These symptoms are classified as autonomic, i.e., hyperhidrosis, orthostatic hypotension, sexual-urinary dysfunction, thermoregulation changes, cardiovascular disturbances, peripheral edema, dilated pupillae), sleep disturbances, neuropsychiatric problems, i.e., apathy, fatigue, anhedonia, depression, anxiety, panic attacks, dementia, psychosis, and sensory, i.e., internal tremor, restless leg syndrome, numbness, paresthesia, visual disturbances, and pain [4e7]. Among these sensory symptoms, pain is observed in approximately 30e50% of PD patients; however, the incidence can increase to 68e85% when all types of pain are taken into account [8]. Pain can appear at any time during the disease, and can be present before diagnosis [9]. There is no consensus on the classification and the mechanisms of pain in PD patients. The objective of this review is to review the available data on the possible mechanisms, classification, evaluation and potential risk factors for pain in individuals with PD.

Parkinson's disease (PD) is becoming an important degenerative disease of the central nervous system, affecting the quality of lives of millions of seniors worldwide [1]. Symptoms of PD can progress differently from one person to another because of the variety of the disease. Patients with Parkinson may show symptoms including tremors mainly at rest. Different types of tremors are possible: tremors in hands, limb rigidity, and gait and balance problems. Generally, two types of symptoms of PD can be distinguished: movement-related (i.e., motor) and unrelated to movement (non-motor). In fact, patients showing non-motor symptoms are more affected than whose main symptoms are motor. Non-motor symptoms may include depression, sleep behavior disorders, loss of sense of smell, and cognitive impairment. It has been reported by the Centers for Disease Control and Prevention (CDC) that PD complications are ranked as the 14th leading causes of death in the United States. To date, the cause of PD rests principally unknown. Particularly, the economic burden due to direct and indirect cost of PD covering treatment, social security payments, and lost income is estimated to be approximately $52 billion per year in the United States alone. Actually, the number of people affected by PD has exceeded 10 million worldwide. It should be noted that the timely detection of the PD facilitates rapid treatment and alleviate symptoms significantly as reported in [2]. Therefore, detection of PD at an earlier stage is certainly a key element to slowing down its progression and could give patients the possibility of accessing to disease-modifying therapy, when available.

2. Literature Review:

Painwas characterized as a tangible and profound experience related with genuine or likely wounds or portrayed regarding such wounds, by the International Association for the Study of Pain (IASP) [10]. It is realized that few anatomic designs are involved simultaneously in nociception. The interaction prompting torment begins with the feeling of nociceptors. The meagerly myelinated Ad nociceptors reaction to mechanical and warm improvements, while the un-myelinated C-fiber nociceptors (polymodal) typically answer mechanical, warm, or synthetic feeling. The improvements from nociceptors show

up to the dorsal horn neurons of the spinal rope. The lamina II, otherwise called substantia gelatinosa, assumes a significant part in aggravation regulation at the spinal rope [10,11]. This region frames a transitional framework managing the transmission to the tcells on the lamina V, which thus intervenes the transmission of tactile improvements to the mind. The substantia gelatinosa framework goes about as an inhibitory component on the immune system microorganisms. Feeling of the Ad and C strands represses the substantia gelatinosa cells, decreasing the result and their inhibitory activity on the immune system microorganisms. Subsequently, the immune system microorganisms increment their movement. The outcome is a decrease in the limit of tcells to get the upgrades or respond to them. This is, basically, the entryway hypothesis component at the spinal level [12].

Two phylogenetically unmistakable frameworks, the average and horizontal aggravation frameworks, communicate agony to higher focus cerebrum neurons. The average framework is mostly comprised of paleospinothalamic, spinomesencephalic, spinoreticular, spinoparabrachial hypothalamic and spinothalamic lot filaments. These filaments travel in a caudal and rostral bearing to higher focuses by ending in the parabrachial core, the locus cæruleus (reticular development), the periaqueductal dark substance (mesencephalon), intra-laminar and average thalamic cores, thalamic ventral caudal parvocellular core and ventral caudal portae, the insula, parietal operculum, the optional somatosensory cortex, the amygdale and hippocampus.

3. Methodology:

This section deals with the description of the dataset used and the approaches taken to achieve the early prediction of Parkinson's disease in a PD patient .The approaches taken were selected with the aim to distinguish a Parkinson's disease patient from those who are healthy patient. The idea is to do a comparative analysis of different machine learning technique by implementing different models on the selected dataset and finding the best machine learning technique among them by evaluating some performance metrics like accuracy, ROC, AAE, and ARE etc. Further the work is extended by implementing Boruta feature selection technique.

Algorithm:

Step1: Data Gathering
Step 2: Data preparation
Step 3: Model Selection
Step 4: Training
Step 5: Evaluation
Step 6: Prediction

Random Forest

Random Forest is one of the machine learning method which is used for both classification as well as regression tasks. It is a type of ensemble method with which a group of weak model when combines turns into a powerful model. In random forest, multiple tress are created .To classify every tree gives a classification, are supposed to vote for that class. The forest selects the classification having the highest votes.

Random Forest Prediction Pseudo code:

1. Takes the test sets features and make decision trees to predict the outcomes and storesthe predicted outcomes.
2. Calculate the votes for each predicted outcome.
3. Consider the high voted predicted outcome as the final prediction.

4. Result and discussion:

The correlation of data is obtained by using the Pearson coefficient technique. The correlation of each feature in feature set is calculated against to the status label. In the correlation table shown in the figure 4.5 we can clearly see that the features like sex, test time, the features related to Shimmer don't have a good correlation with the status columns hence using these columns would generate a under trained machine learning model. The heat map representing the correlation is shown in the figure 2.

status	1.000000	Shimmer	-0.092105
age	0.258102	Shimmer(dB)	-0.079906
sex	0.005190	Shimmer:APQ3	-0.092579
test_time	0.045205	Shimmer:APQ5	-0.100546
motor_UPDRS	0.428504	Shimmer:APQ11	-0.051410
total_UPDRS	0.412621	Shimmer:DDA	-0.092577
Jitter(%)	-0.045653	NHR	-0.118059
Jitter(Abs)	-0.057245	HNR	0.130205
Jitter:RAP	-0.052936	RPDE	-0.143498
Jitter:PPQ5	-0.073836	DFA	-0.156187
Jitter:DDP	-0.052922	PPE	0.014561

Fig. 1: Correlation table.

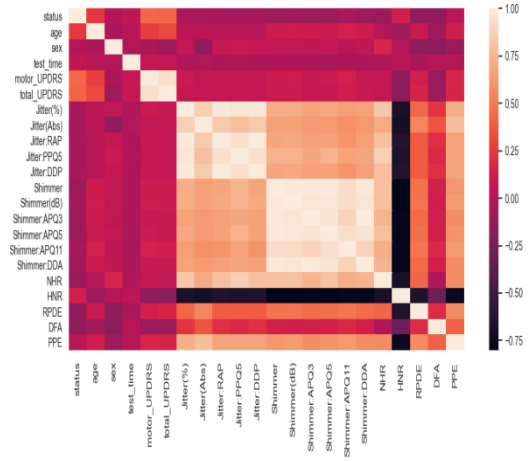

Fig. 2: heat map of the Correlation.

The data set is divided into the feature set and the label. Then the data set is splitted into 80:20 ratio as training and testing set. The dataset is scaled by using the standard scalers which uses the standard deviation of the data and scales the values in the data. The training data is used to train various machine learning models like Kneighbors, Logistic regression, Decision tree classifier, Random forest classifier. The models are hyper tuned for various parameters before training the models by using the Grid searchCv hyper tunening model. The KNN classifier is hyper tuned for the n-neighbors parameter which is by default 5. the hyper tuning resulted the maximum accuracy with the value of 7. The reasulting prediction accuracy of both the tuned and non-tuned models are shown in the figure 3.

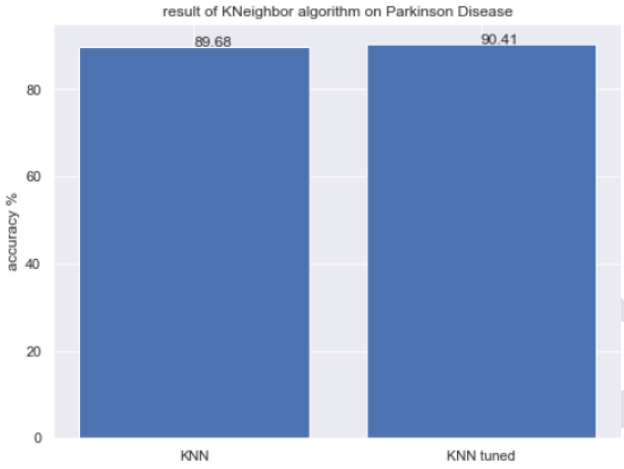

Fig 3: the prediction accuracy of KNN.

The logistic regression model is hyper tuned for the learning rate (C), penalty, maximum iteration to find the best theta which plays an important role in building the logistic regression models. The values of parameters after hyper tuning are c=1000, penalty is l2, maximum iteration is120000.

5. Conclusion:

Parkinson disease directly affects the central nervous system and mainly affects the motion of the patients. The main cause of the Parkinson disease is the drop of dopamine due to the damages in the nerve cell. The primary symptoms of the disease are stiffness, impaired movement, tremors etc. it is a non-curable disease only can be controlled by proper medical observation. Various machine learning models were used on the dataset. Before using the machine learning on the dataset we have performed the exploratory data analysis and done the pre-processing which includes scaling and feature extraction. The models were hyper tuned using the gridsearchcv module. All the models were trained using the Parkinson disease and the results shows that random forest is the best model for the prediction.

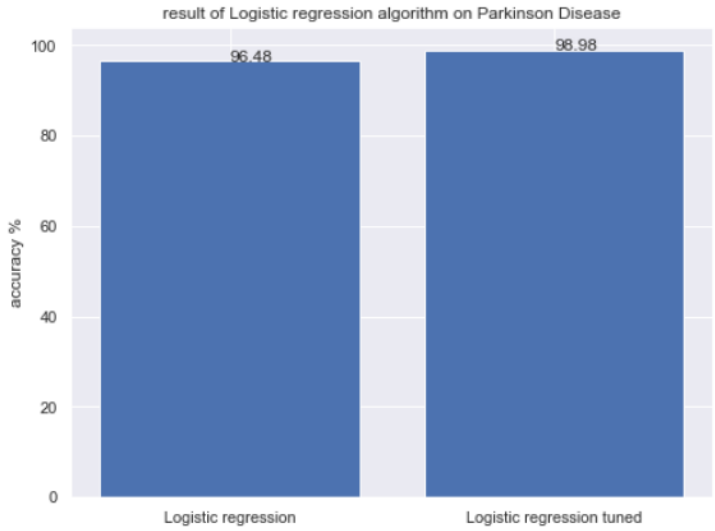

Fig. 4: the prediction accuracy of Logistic regression.

References:

[1] McNamara P, Stavitsky K, Harris E, Szent-Imrey O, Durso R. Mood, side of motor symptom onset and pain complaints in Parkinson's disease. Int J Geriatr Psychiatry 2010;25:519e24.
[2] Weintraub D, Comella CL, Horn S. Parkinson's disease-Part 1: pathophysiology, symptoms, burden, diagnosis, and assessment. Am J Manag Care 2008; 14:40e8.
[3] Alves G, Forsaa EB, Pedersen KF, Dreetz-Gjerstad M, Larsen JP. Epidemiology of Parkinson's disease. J Neurol 2008;255:18e32.
[4] Park A, Stacy M. Non-motor symptoms in Parkinson's disease. J Neurol 2009; 256:293e8.
[5] Simuni T, Sethi K. Nonmotor manifestations of Parkinson's disease. Ann Neurol 2008;64:S65e80.
[6] Bayulkem K, Lopez G. Non-motor fluctuations in Parkinson's disease: clinical spectrum and classification. J Neurol Sci 2010;289:89e92.
[7] Wolters ECH. Non-motor extranigral signs and symptoms in Parkinson's disease. Parkinsonism Relat Disord 2009;3:S6e12.
[8] Beiske AG, Loge JH, Rønningen A, Svensson E. Pain in Parkinson's disease: prevalence and characteristics. Pain 2009;141:173e7.
[9] Tolosa E, Compta Y, Gaig C. The premotor phase of Parkinson's disease. Parkinsonism Relat Disord 2007;13:S2e7.
[10] Almeida TF, Roizenblatt S, Tufik S. Afferent pain pathways: a neuroanatomical review. Brain Res 2004;1000:40e56.
[11] Millan MJ. The induction of pain: an integrative review. Prog Neurobiol 1999; 57:1e164.
[12] Wall PD, Melzack R. Textbook of pain. Churchill London: Livingstone; 1994.
[13] Scherder E, Wolters E, Polman C, Sergeant J, Swaab D. Pain in Parkinson's disease and multiple sclerosis: its relation

to the medial and lateral pain systems. Neurosci Biobehav Rev 2005;29:1047e56.
[14] Willis WD, Westlund KN. Neuroanatomy of the pain system and of the pathways that modulate pain. J Clin Neurophysiol 1997;14:2e31.
[15] Nolano M, Provitera V, Estraneo A, Selim MM, Caporaso G, Stancanelli A, et al. Sensory deficit in Parkinson's disease: evidence of a cutaneous denervation. Brain 2008;131:1903e11.
[16] Braak H, Sastre M, Bohl JR, de Vos RA, Del Tredici K. Parkinson's disease: lesions in dorsal horn layer I, involvement of parasympathetic and sympathetic pre- and postganglionic neurons. Acta Neuropathol 2007;113: 421e9.
[17] Braak H, Del Tredici K, Rub U, de Vos RA, Jansen Steur EN, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging 2003;24:197e211.
[18] Braak H, Rub U, Jansen-Steur EN, Del Tredici K, de Vos RA. Cognitive status correlates with neuropathologic stage in Parkinson disease. Neurology 2005; 64:1404e10.
[19] Chudler EH, Dong WK. The role of the basal ganglia in nociception and pain. Pain 1995;60:3e38.