International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

Leveraging Machine Learning for the Automated Diagnosis of Parkinson's Disease

Neeraj Kumar Verma, Shashi Verma
Computer science and Engineering Department,
B. N. College of Engineering & Technology, Lucknow
neerajbhanu143@gmail.com

Abstract: The prediction of Parkinson's disease has been a significant area of research for over a decade. Publicly available datasets, such as those found on machine learning repositories, have facilitated extensive exploration in this domain. Researchers have utilized a range of techniques—including data mining, machine learning, and deep learning—to develop predictive systems aimed at the early detection of Parkinson's disease, which is especially important in regions where healthcare costs are high and resources are limited. In our proposed study, we will begin by performing exploratory data analysis (EDA) to preprocess the dataset, followed by the development of a prediction system using optimized machine learning models. These models will include k-nearest neighbors (KNN), logistic regression, decision tree classifier, and random forest classifier. Each algorithm will be evaluated based on its predictive performance, and a comprehensive multidimensional analysis will be conducted to determine the most effective model. The dataset used in this study originates from a research effort involving 31 individuals, of whom 23 have been diagnosed with Parkinson's disease..

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

1. Introduction:

Parkinson's disease (PD) is a chronic and progressive neurodegenerative disorder characterized by the loss of nigrostriatal dopaminergic pathways [1]. Its prevalence in the general population ranges from 0.1% to 0.3%, with a notable increase among individuals aged over 65 years [2]. Key clinical features of PD include tremor, rigidity, akinesia (such as bradykinesia and hypokinesia), and postural instability. Alongside these motor symptoms, PD commonly presents with a spectrum of non-motor signs and symptoms encompassing autonomic dysfunctions (such as hyperhidrosis, orthostatic hypotension, and sexual-urinary dysfunction), sleep disturbances, neuropsychiatric issues (including apathy, fatigue, depression, anxiety, dementia, and psychosis), sensory problems (like internal tremor, restless leg syndrome, numbness, paresthesia, visual disturbances), and pain [4-7]. Pain affects approximately 30% to 50% of PD patients and can increase to 68% to 85% when all types of pain are considered [8]. It can manifest at any stage of the disease, sometimes preceding formal diagnosis [9]. Despite ongoing research, consensus on the classification and underlying mechanisms of pain in PD remains elusive. This review aims to explore existing data on potential mechanisms, classification, assessment, and risk factors associated with pain in individuals with PD.

Parkinson's disease is increasingly recognized as a significant degenerative disorder of the central nervous system, profoundly impacting the quality of life for millions of older adults globally [1]. The progression of PD symptoms varies widely among individuals due to the disease's heterogeneity. Patients may exhibit resting tremors, limb rigidity, and difficulties with gait and balance. PD symptoms are broadly categorized into motor symptoms, which are movement-related, and non-motor symptoms that include depression, sleep disorders, anosmia (loss of sense of smell), and cognitive impairment. Non-motor symptoms often impose greater burden than motor symptoms. According to the Centers for Disease Control and Prevention (CDC), PD complications rank as the 14th leading cause of death in the United States. Despite extensive research, the precise cause of PD remains largely unknown. The economic impact of PD, encompassing direct medical costs, social security expenditures, and lost productivity, is estimated to exceed $52 billion annually in the United States alone. Globally, more than 10 million people are affected by PD. Early detection is crucial as it enables prompt treatment initiation and significant symptom alleviation [2]. Detecting PD early is pivotal in slowing disease progression and potentially enabling access to disease-modifying therapies when available.

2. Literature Review:

Pain, characterized as a profound and tangible experience linked to actual or potential injury, is defined by the International Association for the Study of Pain (IASP) [10]. Nociception, the process initiating pain, involves several anatomical structures. Nociceptors, including thinly myelinated Ad fibers responsive to mechanical and thermal stimuli, and unmyelinated C fibers (polymodal) sensitive to mechanical, thermal, or chemical stimuli, transmit signals to dorsal horn neurons of the spinal cord. Lamina II, known as the substantia gelatinosa, plays a crucial role in regulating these pain signals [10,11]. This region acts as a transitional gateway, modulating transmission to lamina V neurons, which further relay sensory inputs to the brain. The substantia gelatinosa inhibits these neurons, reducing their inhibitory effect on ascending sensory pathways. Consequently, sensory transmission is enhanced, aligning with the gate control theory at the spinal level [12].

Two phylogenetically distinct systems, the medial and lateral pain pathways, relay pain signals to higher brain centers. The medial system comprises fibers such as paleospinothalamic, spinomesencephalic, spinoreticular, spinoparabrachial, hypothalamic, and spinothalamic tracts. These fibers ascend and descend to various brain structures, including the parabrachial nucleus, locus coeruleus (reticular formation), periaqueductal gray (mesencephalon), thalamic nuclei (intralaminar and medial), insula, parietal operculum, secondary somatosensory cortex, amygdala, and hippocampus.

3. Methodology:

This section focuses on detailing the dataset utilized and the methodologies employed for early prediction of Parkinson's disease in patients. The chosen approaches aim to differentiate Parkinson's disease patients from healthy individuals. The strategy involves conducting a comparative analysis of various machine learning techniques by applying different models to the dataset and determining the optimal technique based on performance metrics such as accuracy, ROC (Receiver Operating Characteristic), AAE (Average Absolute Error), and ARE (Average Relative Error). Additionally, the study extends its scope by integrating the 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 a versatile machine learning method utilized for both classification and regression tasks. It belongs to the ensemble methods category, where a collection of weak models collectively forms a robust predictive model. In a random forest, multiple trees are generated, and each tree provides a classification. The final classification is determined by aggregating the votes from all trees in the forest, selecting the class with the highest number of votes.

Random Forest Prediction Pseudo code:

Here's a revised version:

1. It uses the features of the test set to build decision trees for predicting outcomes, saving these predictions.
2. Votes are computed for each predicted outcome.
3. The final prediction is based on the highest-voted outcome among all predictions.

4. Result and discussion:

The data correlation is determined using the Pearson coefficient technique. Each feature in the feature set is correlated with the status label. In the correlation table depicted in Figure 4.5, it is evident that features such as sex, test time, and those related to Shimmer do not exhibit strong correlations with the status column. Utilizing these columns could result in an undertrained machine learning model. Figure 2 displays a heatmap illustrating these correlations.

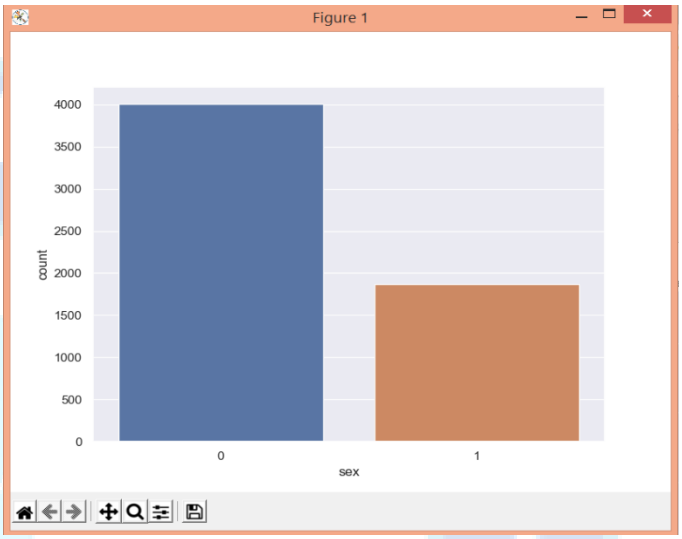

Fig 1: Total amount of masculine and feminine patients

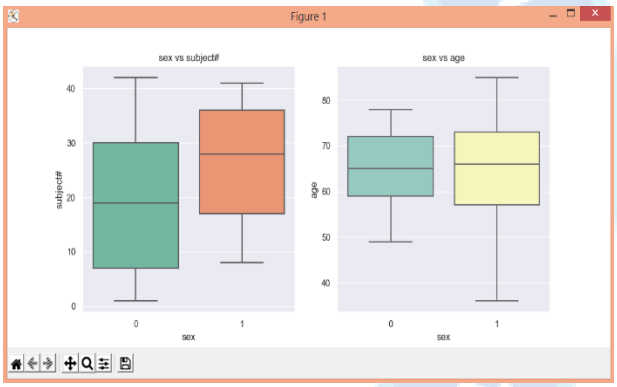

Fig. 2: The box plot of sex versus subject

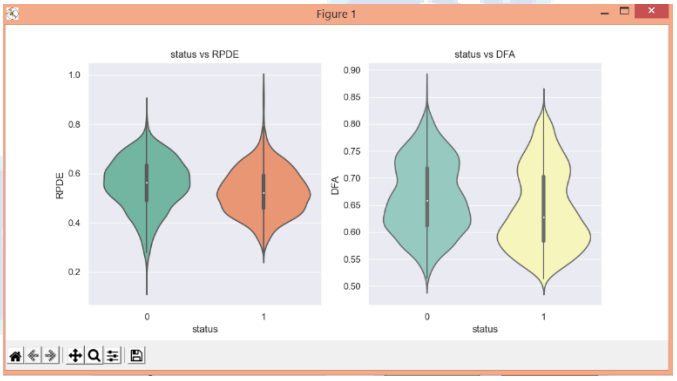

Figure 4.3: the violin plot of status versus RPDE and DFA.

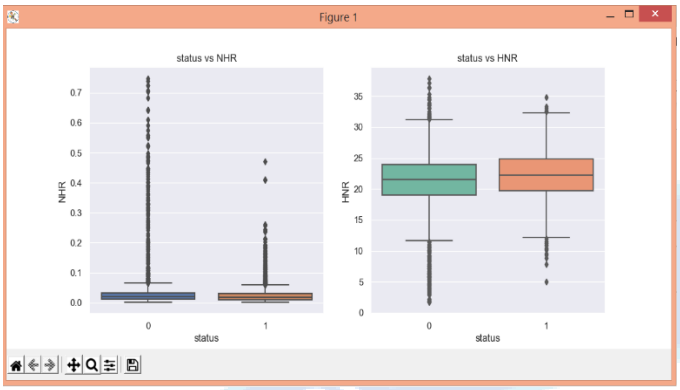

Figure 4.4: the box plot of status versus NHR and HNR.

5. Conclusion:

Parkinson's disease directly impacts the central nervous system and primarily affects patients' motor functions. The disease is primarily caused by a reduction in dopamine levels due to damage to nerve cells. Key symptoms include stiffness, impaired movement, and tremors. It is a progressive disease with no cure, managed through ongoing medical care. Machine learning models were applied to a dataset after performing exploratory data analysis and preprocessing steps such as scaling and feature extraction. The models underwent hyperparameter tuning using the gridsearchcv module. Based on the results, the random forest model emerged as the most effective for predicting Parkinson's disease.

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.
[20] Anurag et. al., "Load Forecasting by using ANFIS", International Journal of Research and Development in Applied Science and Engineering, Volume 20, Issue 1, 2020
[21] Raghawend, Anurag, "Detect Skin Defects by Modern Image Segmentation Approach, Volume 20, Issue 1, 2020