Prediction of Impulse Control Disorders in Parkinson's Disease using Machine Learning Algorithms

Venna Rohan[1], Ramesh Ponnala[2]
[1]MCA Student, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad, Telangana State, India
[2]Assistant Professor, Department of MCA, Chaitanya Bharathi Institute of Technology (A),Gandipet, Hyderabad, Telangana State, India

ABSTRACT

Parkinson's disease (PD) is a neurological disorder recognized by non-motor symptoms, such as tremors and bradykinesia. Its cause is unknown, but a combination of both genetic and environmental factors plays a crucial role. Diagnosis relies on clinical and generic evaluation, and treatment involves medication, therapy and lifestyle modifications. Ongoing research aims to improve detection and develop more effective interventions for PD.We conducted the analysis of machine learning algorithms for PD detection using a dataset of clinical and genetic features to improve the accuracy and performance for better prediction of Parkinson's disease (PD) .The Performance metrics such as accuracy, confusion matrix, and classification report were used to assess their effectiveness. The findings revealed the Voting Classifier as the most accurate model, demonstrating its potential as a reliable tool for early PD diagnosis and personalised management. These results contribute to advancing the field of PD detection and hold promise for improving patient care and outcomes.

Keywords *– impulse control disorders, motor symptoms, machine learning algorithms.*

I.INTRODUCTION

Despite the fact that engine side effects are more frequently associated with Parkinson's disease (PD), a number of non-engine side effects have been linked to the condition. The inability to direct one's own motivations and unsuccessful attempts to do so are characteristics of ICDs. In Parkinson's disease, irregularities caused by ICDs are common. After five years of illness, cross-sectional studies reveal a prevalence of 15-20%, annual events of approximately 10%, and total occurrences of more than half. The four most normal ICDs in Parkinson's disease are habitual shopping, neurotic betting, voraciously consuming food, and hypersexuality. Two further ICDs that are normal are beating and hobbyism, and the commonness of each ICD, particularly obsessive betting, varies essentially between societies. ICDs should be treated when possible since they are related to brought down personal satisfaction, stressed relational connections, and a more noteworthy burden on careers. Various contextual analyses show that ICDs disappear when dopamine agonist (DA) drugs are diminished in measurement or halted through and through. ICDs in Parkinson's disease have been interconnection with various variables, including socio-segment, clinical, and hereditary signs. As opposed to ladies, who are more inclined to have dietary issues and over the top shopping, guys are bound to foster obsessive betting and hypersexuality issues. Numerous studies have found a correlation between younger age and ICDs in Parkinson's disease. Anxiety and depression have also been linked to ICDs and abnormal REM rest conduct. Dopamine substitution therapy has been identified as the first risk factor for ICD.

In Parkinson's disease, various elements, including socio-segment, clinical, and hereditary markers, have been related to ICDs [14]. Ladies are more inclined to have dietary issues and urgent shopping than guys are to encounter neurotic betting and hypersexuality problems [15]. In a few studies, ICDs in Parkinson's disease have been linked to younger age [4]. ICDs have likewise been associated with REM rest conduct irregularities, uneasiness, and gloom. The primary gamble factor for ICD has been distinguished as dopamine substitution treatment. ICDs have been related with levodopa and dopamine agonists, with dopamine agonists having a more noteworthy and more grounded affiliation. To combine them, a few single-nucleotide polymorphisms (SNPs) in qualities connected with the dopamine flagging framework have been related with ICDs.

II. LITERATURE SURVEY

Tapan Kumar, Pradyumn Sharma, Nupur Prakash et al [1] in In their study, the author utilised a DT, RF, and hard voting techniques, achieving 100% training accuracy. However, the test accuracy was slightly lower, around 6-7%. The author also observed that reducing the number of estimators in bagging classifiers could enhance accuracy. Overfitting issues were seen in some models, likely attributed to the limited size of the dataset.

Md. Mosharraf Umar,Sameena Naaz et al [2] The study evaluated the Radial Basis Function (RBF) for Parkinson's Disease (PD) prediction using Keras and TensorFlow. RBF-based models gone through in K-fold cross-validation and K-means clustering for improved performance. Compared to a Deep Neural Network (DNN) benchmark, RBF models achieved slightly lower accuracy but demonstrated potential for accurate PD prediction. Early detection and intervention in PD were emphasised for enhanced patient care. Future research should optimise RBF models and explore other machine learning techniques to improve accuracy and facilitate early PD diagnosis. Integration of RBF models holds promise for advancing PD prediction, providing valuable insights, and improving patient outcomes.

Satyabrata Aich,Hee-Cheol Kim, Kim younga, Kueh Lee Hui, Ahmed Abdulhakim Al-Absi and Mangal Sain et al [3] in their study they investigated the use of a supervised machine learning approach with diverse feature selection techniques for predicting Parkinson's Disease (PD) using voice datasets. The objective is to develop a reliable prediction model for early detection and intervention. Various feature selection methods are employed to identify relevant voice features, enabling the training and evaluation of machine learning algorithms. Evaluation metrics to assess model performance. The findings highlight the potential of this approach for accurate PD prediction and improved patient care.

Nagham Mekky,Hassan Soliman,Marwa Helmy, Mohammed Elmogy,Eman Eldaydamony et al[4] the author explores advanced machine learning techniques for enhancing PD gene prediction, including feature engineering, ensemble learning, and deep learning. The study highlights the effectiveness of these enhancements in accurately predicting PD-related genes, advancing PD genetics research.

Pooja Raundale,Chetan Thosar,Shardul Rane et al [5] they had explored the use of algorithms(machine learning and deep learning) for predicting Parkinson's disease and assessing its severity. The study highlights the potential of these algorithms in accurate PD prediction and severity assessment. Feature engineering and data augmentation techniques enhance model performance. Further research can explore advanced algorithms and additional clinical features for improved diagnosis and patient care.

Debasis Patnaik, Mavis Henriques,Ashin Laurel et al [6] in their study they had focused on machine learning techniques for Parkinson's Disease (PD) prediction. It inspects the performance of algorithms such as decision trees, support vector machines, and neural networks.in developing accurate PD prediction models. Preprocessing methods are employed to handle missing values and address class imbalance. The incorporation of enhancements, such as ensemble learning and feature selection, aims to improve prediction accuracy and computational efficiency. The results highlight the potential of machine learning approaches in accurate PD prediction, emphasising the significance of early detection and intervention for effective management of PD.

Ezhilin Freeda, Ezhil Selvan TC, Vishnu Durari RS et al [7] The study explores the use of Decision tree and XGBoost algorithms for accurate Parkinson's disease detection. XGBoost shows high accuracy (92.3%) in early prediction. Future developments and analysis of diverse data types hold potential for improving diagnosis and treatment options.

Valiant Vincent Dmello, Alrich Agnel Kudel, Supriya Kamoji, Dipali Koshti, Nash Rajesh Vaz et al [8] The author conducted a study using different datasets to detect specific symptoms of Parkinson's disease. The freezing of gait dataset achieved 96.06% accuracy in detecting FOG events with the Decision Tree classifier. The clinical speech dataset detected voice irregularities with K-NN achieving 97.43% accuracy. The Spiral and Wave dataset identified arm tremors using a transfer learning CNN model, with the wave dataset providing better results. Early detection of symptoms aids timely treatment, making this system useful in hospitals or for general users to predict disease symptoms conveniently and affordably.

Sahaja Dixit, Akash Gaikwad, Vibha Vyas, Mahesh Shindikar, Ketaki Kamble et al [9] The author conducted tests using machine learning algorithms (CNN, Resnet, VGG-16) on four methods related to human neurological activity, achieving accuracy above 90%. The objective was to detect neurocircuitry diseases that are often interconnected. The research successfully identified neurocircuitry diseases using MRI and CSV frequency data from patients. This study holds significance as it enables early detection and treatment of neurological diseases, contributing to improved patient outcomes.

Rhea Mary Josi, R.I. minu et al [10] The author's conclusion emphasises the significance of prediction models for the Parkinson's disease gene in enabling early diagnosis. Identifying the disease gene becomes crucial since the causes and a definitive cure for the disease remain unclear. This analysis aided in identifying more accurate and efficient algorithms to be employed as prediction models, enhancing the overall diagnostic process.

BEHAVIOUR OF PARKINSON'S DISEASE:

The high occurrence of repetitive and reward-seeking behaviours, referred to as Impulse Control Behaviours (ICBs), in Parkinson's disease (PD), could be linked to prolonged dopaminergic replacement therapy (DRT). These behaviours include impulse control disorders (ICDs), such as compulsive gambling, shopping, and dopamine dysregulation syndrome (DDS). Extensive research has been conducted to evaluate the decision, pathophysiology, clinical aspects, risk factors, and management of ICBs. The results indicate that ICBs are common among PD patients, with prevalence rates ranging from 3 to 6 percent for DDS, 0.34 to 4.2 percent for pounding, and 6 to 14 percent for ICDs. DDS is primarily associated with high doses of levodopa, while ICDs are more prevalent in individuals taking dopamine agonists. Several risk factors, including male gender, higher levodopa doses, younger age at PD onset, history of alcohol use, rash, or specific personality traits, are associated with various subtypes of ICBs. The Review for Hurried Hasty Issue in Parkinson's Disease Rating Scale has proven to be an effective tool for gathering relevant data from patients and caregivers. Managing Impulse Control Behaviours (ICBs) continues to be a significant concern, and the primary approach involves adjusting Dopaminergic Replacement Therapy (DRT). Alongside this, psychosocial therapies, atypical antipsychotics, antidepressants, and amantadine are also used to address impulsive episodes resulting from extended DRT. However, it is crucial to carefully consider the effects on motor symptoms and ICBs when making adjustments. For some individuals, deep brain stimulation of the subthalamic nucleus may offer potential benefits. While the specific pathophysiological mechanisms of ICBs in Parkinson's disease are still not fully understood, it is crucial to develop effective treatment options for those currently affected, in addition to gaining a better understanding of the prevalence, characteristics, and risk factors associated with ICBs.

A cross-sectional study conducted on 3,090 Parkinson's disease patients aimed to examine impulse control disorders (ICDs) and their relationship with dopamine replacement therapy and other clinical factors. The study revealed that 13.6 percent of patients had at least one ICD, with gambling, excessive sexual behaviour, compulsive buying, and binge eating being the most prevalent. Drive control issues were more common among patients receiving dopamine agonist therapy, with a 2- to 3.5-fold increased risk compared to those not receiving dopamine agonists. Both pramipexole and ropinirole had similar rates of impulse control disorders. Levodopa use, U.S. citizenship, younger age, single status, smoking, and a family history of gambling were identified as factors associated with ICDs. The findings emphasise the need for further research to better understand the complex relationship between ICDs and other clinical characteristics and to enhance preventive and treatment efforts. Another longitudinal study investigated the long-term associations between PD and ICDs in patients undergoing dopamine replacement therapy. The study found that after five years of follow-up, the prevalence of ICDs increased from 19.7 percent to 32.8 percent. The use of dopamine agonists was significantly associated with ICDs, with higher doses and longer treatment duration showing stronger correlations. Notably, impulse control disorders (ICDs) showed a gradual resolution upon discontinuation of dopamine agonists. The longitudinal analysis revealed that 46 percent of Parkinson's disease (PD) patients undergoing extensive dopamine agonist treatment experienced ICDs, indicating that the dosage and treatment duration contributed to the development of these behaviours.

III. METHODOLOGY

The dataset was initially obtained from an unspecified source and focused on classifying Parkinson's disease. After collecting the data, it underwent a cleaning process to eliminate irrelevant details like the "name" column. Various data visualisation methods were applied, including correlation matrix heatmaps and counterplots, to extract insights from the dataset. To handle missing data, records with incomplete values were removed. The dataset was then divided into training and testing subsets, facilitating the training and evaluation of different machine learning models.

To enhance model performance, feature scaling was employed to normalise input features.Multiple models, such as SVM, Random Forest, Decision Tree, Logistic Regression, and XGBoost, were trained on the training dataset. Accuracy scores for each individual model were computed. By combining their predictions, a voting classifier ensemble was established to assess its accuracy. Additionally, accuracy assessments were conducted on deep learning models, particularly LSTM and GRU. In conclusion, a bar chart was created using Python's seaborn module to compare accuracy scores across all models. This chart served as a tool to evaluate the models' performances. A heatmap generated through seaborn aided in recognizing patterns and associations within the dataset. The heatmap was instrumental in identifying highly correlated features, which in turn led to the exclusion of some variables, revealing potential interdependencies.

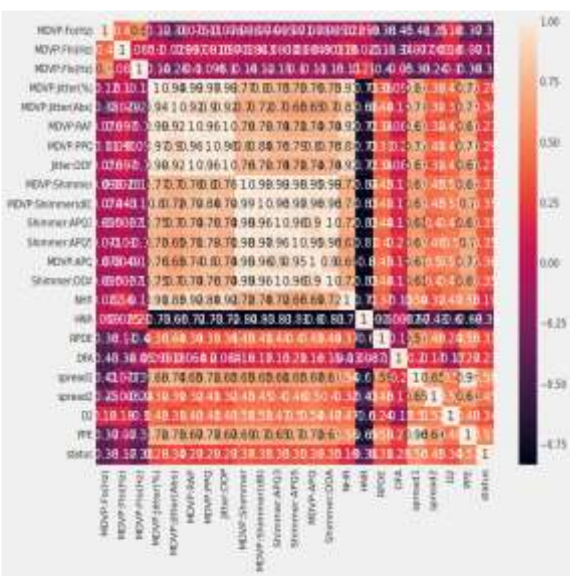

Figure 1: *heatmap for dataset*

MODULES:

In order to conduct the project, we formulated and utilised the following modules:

Information investigation: The analysis on a dataset designed for the classification of Parkinson's disease, encompassing a diverse range of features. The initial stages involve data preprocessing, encompassing the removal of extraneous columns, treatment of missing data, and the encoding of categorical variables. Through the utilisation of data visualisation techniques, including the correlation matrix heatmap and countplot, valuable insights into the relationships between variables and their distributions are obtained.

Subsequently, the dataset is divided, and a process of feature scaling is implemented to ensure optimal performance. Diverse machine learning models, encompassing SVM, Random Forest, Decision Tree, Logistic Regression, and XGBoost, are then trained and evaluated by means of accuracy scores. To consolidate predictions, a voting classifier is employed, and the resultant accuracy is gauged. Delving into deep learning methodologies, the code also engages in the training of LSTM and GRU models utilising the Keras framework. This endeavour culminates in the computation of accuracy scores. In order to provide a comprehensive overview, a bar chart is generated, effectively comparing the accuracy scores across the spectrum of models.

- Handling: We will pursue information for handling, utilising dropping irrelevant columns, handling missing values, and encoding categorical variables. These operations ensure a clean dataset for analysis. The module plays a vital role in data preprocessing, enhancing the accuracy and effectiveness of the predictive models used in Parkinson's disease classification.

- Splitting data into train and test: In this module, we will perform the division of the dataset into training and testing subsets.

- Model generation: In this model generation, We will build models using SVM, RF, DT, LR, XGBoost, Voting classifier, RNN, and GRU. Accuracy computed

- User signup & login:This module facilitates user registration and login.

- User input: This module is responsible for providing input data for prediction.

- Prediction: The final predicted value will be shown by using this module.

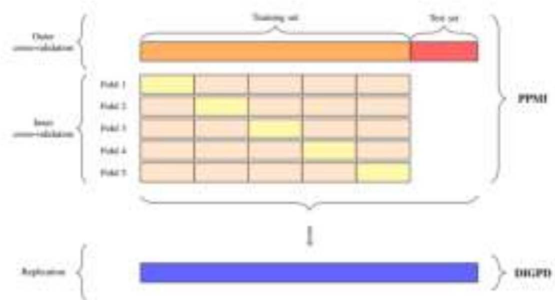

Figure 2: *System architecture*

IV.IMPLEMENTATION

In this project, various machine learning models are trained and evaluated using a Parkinson's disease dataset. The models include SVM, Random Forest, Decision Tree, Logistic Regression, XGBoost, and Voting Classifier. The dataset is split into input features and the target variable. The fit method is called on the training data for each model, and predictions are made on the same data. Accuracy scores, confusion matrices, and classification reports are calculated to assess model performance. XGBoost, an ensemble learning method that combines weak prediction models using gradient boosting techniques, achieves the highest accuracy score among the individual models. The Voting Classifier, which combines predictions using majority voting, has a lower accuracy score compared to XGBoost in this implementation. Visualisations such as histograms and heatmaps are used to understand the data and model performance. The trained models, including the Voting Classifier, are saved for future use. Finally, an attempt is made to implement a saved XGBoost model for making predictions on new input data. The project aims to accurately predict Parkinson's disease and highlights XGBoost as the best-performing model in this specific implementation.

ALGORITHMS:

SVM: SVM is a directed ML procedure that might be utilised for both order and relapse. However we call them relapse issues, they are the most appropriate for order. The SVM's calculation will likely recognize a hyperplane in a N-layered space that obviously orders the info focuses.

RF: The Random Forest algorithm stands out as a potent machine learning technique that leverages an ensemble of decision trees to achieve precise predictions. Through the strategic utilisation of random feature subsets and data samples for individual trees, it adeptly mitigates the risk of overfitting while enhancing the ability to generalise. The ultimate prediction stems from the amalgamation of predictions originating from each tree, thereby yielding a sturdy and dependable model primed for informed decision-making.

DT: A decision tree is a widely used algorithm that imitates human decision-making processes. It employs a flowchart-like structure with nodes to represent features and possible outcomes. By recursively dividing data based on these features, decision trees establish an interpretable hierarchy. To make predictions, the tree is traversed from the root to leaf nodes, utilising feature values along the way. Decision trees are versatile and can handle both classification and regression tasks with categorical and numerical data. They are appreciated for their simplicity, interpretability, and ability to handle intricate relationships in data.

LR: Logistic regression, a statistical technique, is implemented for binary classification tasks to model the connection between input features and the probability of a binary outcome. It estimates feature coefficients to build a linear equation, subsequently transformed using the logistic function to yield a probability ranging from 0 to 1. A threshold is set to determine predictions based on whether the probability surpasses the threshold. Renowned for its simplicity, interpretability, and effectiveness in binary outcome prediction, logistic regression remains widely adopted in various applications.

XGBoost:XGBoost, or (Extreme Gradient Boosting), is a powerful machine learning algorithm that excels in various tasks like classification, regression, and ranking. By combining predictions from multiple weak learners, such as decision trees, it constructs a robust predictive model. XGBoost's gradient boosting framework sequentially creates new models to rectify previous model errors, employing regularisation techniques to prevent overfitting. With efficient gradient descent optimization, it can handle large datasets with high-dimensional features. This algorithm is widely acclaimed for its speed, scalability, and wide adoption across industry and academia.

Voting classifier: The concept of a Voting classifier involves an advanced machine learning algorithm that merges the predictions made by several distinct individual classifiers in order to arrive at a conclusive decision. This mechanism functions by permitting each classifier to submit its prediction, and the ultimate outcome is established through a consensus reached among these classifiers. This procedure mirrors a collaborative decision-making process, akin to a group discussion, where the diverse viewpoints and proficiency of individual classifiers collaborate to attain a prediction that is both precise and resilient. Voting classifiers prove to be exceptionally advantageous when disparate classifiers furnish distinct insights relevant to the task, consequently leading to heightened accuracy and dependability in prediction outcomes.

RNN: RNN, which stands for Recurrent Neural Network, is a specialised type of neural network used for processing sequences of data like time series or text. Unlike regular neural networks that analyse data in a linear manner, RNNs have a built-in memory that allows them to consider previous information while processing the current input. This memory feature enables RNNs to capture temporal patterns and relationships.Imagine processing a sentence word by word and comprehending its meaning based on the words read so far. This is similar to how Recurrent Neural Networks (RNNs) operate, as they excel in tasks like language modelling, speech recognition, and machine translation by effectively analysing sequential data and understanding its context.

GRU:GRU, which stands for Gated Recurrent Unit, is a specialised neural network renowned for its proficiency in handling sequential data. Just like other recurrent neural networks, GRU processes information step by step. However, GRU has a clever design that enables it to selectively remember and use important details from before the steps. This makes GRU especially good at capturing long-term relationships in the data, which is valuable in tasks like understanding language, recognizing speech, and analysing time series data. In simpler terms, you can think of GRU as a smart system that learns from the past to make better predictions about what comes next in a sequence.

LSTM:LSTM, or Long Short-Term Memory, represents an advanced version of the recurrent neural network (RNN) tailored to process and interpret sequential data. Unlike conventional RNNs, LSTM effectively captures long-term relationships by incorporating a memory cell and three gates: input, forget, and output. These gates regulate information flow, selectively retaining or discarding relevant data at each step. With its ability to handle long-term dependencies, LSTM finds extensive application in natural language processing, speech recognition, and time series analysis. In essence, LSTM can be visualised as an intelligent system with enhanced memory, proficiently retaining and utilising vital information from the past to make precise predictions about the future.

V.METHODOLOGY

Among the evaluated machine learning algorithms, the Voting Classifier demonstrated exceptional performance and emerged as the most accurate model for Parkinson's disease (PD) detection. It achieved a high accuracy score on the entire dataset, indicating its ability to correctly classify PD cases. The analysis of the confusion matrix associated with the Voting Classifier further provided valuable insights into the distribution of “true positive”, “true negative”, “false positive”, and “false negative” predictions. These results highlight the Voting Classifier's potential as a robust and reliable algorithm for the detection of PD.

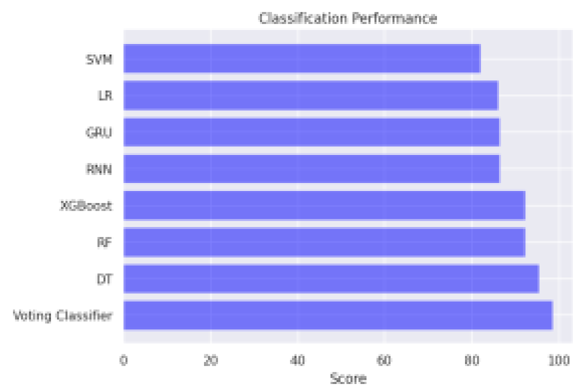

Figure 3: *Accuracy score on different model*

VI. CONCLUSION

According to my findings, this project demonstrates the potential benefits of utilising machine learning algorithms and integrating socio-segment, clinical, and genetic indicators to predict Impulse Control Disorders (ICDs) in Parkinson's disease. The project's focus on enhancing prediction accuracy, enabling personalised healthcare, facilitating early identification, and supporting proactive patient management shows promising results for improving patient outcomes and informing clinical decision-making. The gained insights into ICD risk factors contribute to a deeper understanding of the condition and provide valuable guidance for future research. Although challenges such as data limitations, ethical considerations, and the need for external validation exist, the project offers valuable insights and holds potential for future applications in precision medicine.

ISSN: 2278-4632
Vol-13, Issue-11, November 2023

Overall, this project represents a significant step towards enhancing the identification, management, and prevention of ICDs in Parkinson's disease, with the aim of improving patient care and well-being.

REFERENCES

[1] Tapan Kumar, Pradyumn Sharma, Nupur Prakash | Comparison of Machine learning models for Parkinson's Disease predictionDOI 978-1-7281-9656-5/20/$31.00 ©2020 IEEE.

[2] Md. Mosharraf Umar,Sameena Naaz | Classification Using Radial Basis Function for Prediction of Parkinson's Disease| DOI: 10.1109/GCAT55367.2022.9971828| 978-1-6654-6855-8/22/$31.00 ©2022 IEEE.

[3] Satyabrata Aich,Hee-Cheol Kim, Kim younga, Kueh Lee Hui, Ahmed Abdulhakim Al-Absi and Mangal Sain | A Supervised Machine Learning Approach using Different Feature Selection Techniques on VoiceDatasets for Prediction of Parkinson's Disease | DOI ISBN 979-11-88428-02-1

[4] Marwa Helmy,Nagham Mekky,Hassan Soliman,Mohammed Elmogy,Eman Eldaydamony |Enhanced Parkinson's Disease Genes Prediction | DOI 978-1-6654-5895-5/22/$31.00 ©2022 IEEE

[5] Pooja Raundale,Chetan Thosar,Shardul Rane|Prediction of Parkinson's disease and severity of the disease using Machine Learning and Deep Learning algorithm | DOI 978-1-7281-7029-9/21/$31.00 ©2021 IEEE.

[6] Debasis Patnaik, Mavis Henriques,Ashin Laurel|Prediction of Parkinson's Disorder: A Machine Learning Approach | DOI 978-1-6654-7886-1/22/$31.00 ©2022 IEEE.

[7] Ezhilin Freeda, Ezhil Selvan TC, Vishnu Durari RS | Prediction of Parkinson's disease using XGBoost | DOI 978-1-6654-0816-5/22/$31.00 ©2022 IEEE

[8] Sahaja Dixit, Akash Gaikwad, Vibha Vyas, Mahesh Shindikar, Ketaki Kamble | United Neurological study of disorders: Alzheimer's Disease, Parkinson's disease detection, Anxiety Detection, and Stress detection using variousMachine learning Algorithms | DOI 978-1-7281-6885-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICONSIP49665.2022.10007434.

[9]Rhea Mary Josi, R.I. minu |Review of computational approaches to Parkinson's disease gene prediction | DOI 978-1-7281-7089-3/20/$31.00 ©2020 IEEE

[10] A. H. Erga, G. Alves, O. B. Tysnes, and K. F. Pedersen, "Impulsive and compulsive behaviours in Parkinson's disease: Impact on quality of and satisfaction with life, and caregiver burden," Parkinsonism Related Disorder., vol. 78, pp. 27–30, 2020.

[11] E. Mamikonyan et al., "Long-term follow-up of impulse control disorders in Parkinson's disease," Movement Disorder., vol. 23, no. 1, pp. 75–80, 2008.

[12] M. J. Nirenberg and C. Waters, "Compulsive eating and weight gain related to dopamine agonist use," Movement Disorder., vol. 21, no. 4, pp. 524–529, 2006.

[13] M. Grall-Bronnec et al., "Dopamine agonists and impulse control disorders: A complex association," Drug Saf., vol. 41, no. 1, pp. 19–75, Jan. 2018.

[14] D. Weintraub and D. O. Claassen, "Impulse control and related disorders in Parkinson's disease," Int. Rev. Neurobiol., vol. 133, pp. 679–717, 2017