iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VI Month of publication: June 2025

DOI: https://doi.org/10.22214/ijraset.2025.72823

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

Exploring Parkinson's Disease Progression Through Python-Based Data Analysis Techniques

Ananya Ravi kumar
Department of Computer Science and Engineering RV University, Bengaluru, Karnataka

***Abstract:** Parkinson's disease is a long-term neurological illness that mainly influences motor functions balance, and speaking abilities. This study aims to present a reproducible Python-based EDA framework to understand the progressive growth of Parkinson's disease using an authentic dataset comprising500affectedpatients. Withtheuseofmajorpython libraries, ExploratoryDataAnalysis(EDA)methodswereappliedtoclean,processandvisualizethedata.Thestudyinvolvedexploring correlation between features like UPDRS score, motor function, speech difficulty and years since diagnosis. Descriptive statistics, correlation heatmaps and visual plots helped in recognizing symptom patterns and perspective-related progression of this condition. This work exhibits how EDAcan be useful in rapid analysis of clinical data, chiefly in emphasizing the role of clinical characteristicsincomplicateddisorderslikeParkinson'sdisease.Ourworkflowidentifiessymptom clusters and temporal trends, offering a template for clinical data analysis.*

***Keywords:** Parkinson'sDisease,ExploratoryDataAnalysis,DataVisualization,UPDRSScore,Python, Clinical Data.*

I. INTRODUCTION

Parkinson's disease is a continuously evolving neurological impairment that largely impacts motor function, speaking ability and equilibrium.With its occurrence increasing noticeably with age, it is oneofthemostfamiliarfunctional movementdisorders. Earlyandaccuratediagnosisofsymptoms are essential for managing and improvement oftreatment planning, and upgrading quality of life. Nevertheless, understanding the trajectory of this disease remains a complicated challenge due to irregularity in patient symptoms and variability. With the rapid growth of medical datasets, data- driven methods such as exploratory data analysis (EDA) are instrumental in extracting interdependencies and meaningful patterns from medical data.

Unlike prior EDAstudies focusing solely on static correlations, our work integrates temporal trend analysis (e.g., UPDRS vs. diagnosis duration) using Python's Seaborn and Matplotlib. This approach enables dynamic symptom tracking, which is critical for early intervention. By using Python-based libraries such asPandas, Seaborn and Matplotlib, researchers can systematically process and visualize complex medical datasets. In this study, a real-world dataset comprising 500 individualsdiagnosedwithParkinson'sdiseasewasexaminedtoidentifysymptomprogressionand correlation across features such as UPDRS score, tremor severity, medication andyears since diagnosis. The main objective of this research is to use Exploratory DataAnalysis (EDA) techniquestounderstandthesymptomtrendsandinterdependenciesofParkinson'sdiseaseusing statistical summaries and visualization. The present study aims to provide insights that can aid practitioners in identifying symptom groupings and presenting early-stage remedies for the diagnosis.

II. METHODOLOGY

Astructured Exploratory DataAnalysis (EDA) workflow is followed in this methodology which is designed to uncover insights from a clinical dataset of 500 Parkinson's patients. With the aid of Pythonlibrarieslike Pandasfordatamanipulation, NumPyfornumericaloperations,Matplotliband Seaborn for visualization and Scikit-learn, the process began with loading the dataset, inspecting its structure and analyzing column types and basic statistical summaries. The dataset consisted of 500 patient records affected by Parkinson's disease. Each record included variables such as age, gender, year since diagnosis, UPDRS Score (Unified Parkinson's Disease Rating Scale), Tremor Severity, Motor Function, Speech Difficulty, Balance Problems, Medications, Exercise Level and the target variable - disease progression.

Inpreprocessingstage,thedatawasexaminedforqualitybyidentifyingmissingvalues,duplicate rows and invalid negative entries. Fortunately, the dataset was clean and required minimal correction in specific columns. In categorical encoding step, Label Encoding and One-Hot Encoding were used to encode categorical variables such as gender and medication types and converted them to numerical format for suitable analysis. Techniques like Min-Max Scaling and Z-score Standardization were applied to make sure the features are uniform depending on their nature for feature scaling.

Outliers were detected using the Interquartile Range (IQR) method to highlightextremevaluesinnumerical columnsandboxplotswereusedto visualizethedistribution anomalies. Outliers beyond 1.5×IQR were removed to ensure robustness, affecting <5% of data points.

Visualizationswereextensivelyusedtointerpretpatterns.Multiplestatisticalgraphsweregeneratedtointerpretfeatureinterdependenciesand trends.Histogramsandboxplotswereusedtounderstand the distribution of continuous variables and the bar charts provided comparisons among categorical features. Correlation heatmaps were used to reveallinear associations among clinical variables and multivariaterelationshipswereexaminedthroughpairplots.Groupedbarchartsandlineplotsaided in comparison of disease progression across patient subgroups. This EDApipeline enabled the extraction of interpretable insights about Parkinson’s disease feature interdependencies and progression of the disorder.

III. RESULTS AND DISCUSSION

The results obtained from the exploratory data analysis performed on the Parkinson’s disease datasetarepresentedinthissection. Severalstatisticaltechniquesandvisualizationtoolswereused to identify patterns, distributions and interrelationships among clinical features. By visualizing symptoms, identifying feature trends and analyzing relationships across affected groups, this study aims to convert raw data into perceptive observations. This studynot only points symptoms that support to disease advancement but also assist in building a clear understanding through structured data interpretation.

A. Feature Distributions

Histograms and box-plots were used for acomprehensive view of the distribution of key numeric featuressuchas*UPDRSScore*, *TremorSeverity*, *MotorFunction*and*SpeechDifficulty*.withalarge number of patients falling within mid-range severity scores, most of these variables exhibited moderately skewed distributions.Aright-skewed distribution was observed by *Exercise Level* variablewhichindicatedthatconsiderableportionofaffectedindividualsmaintainedlowerphysical activity levels.

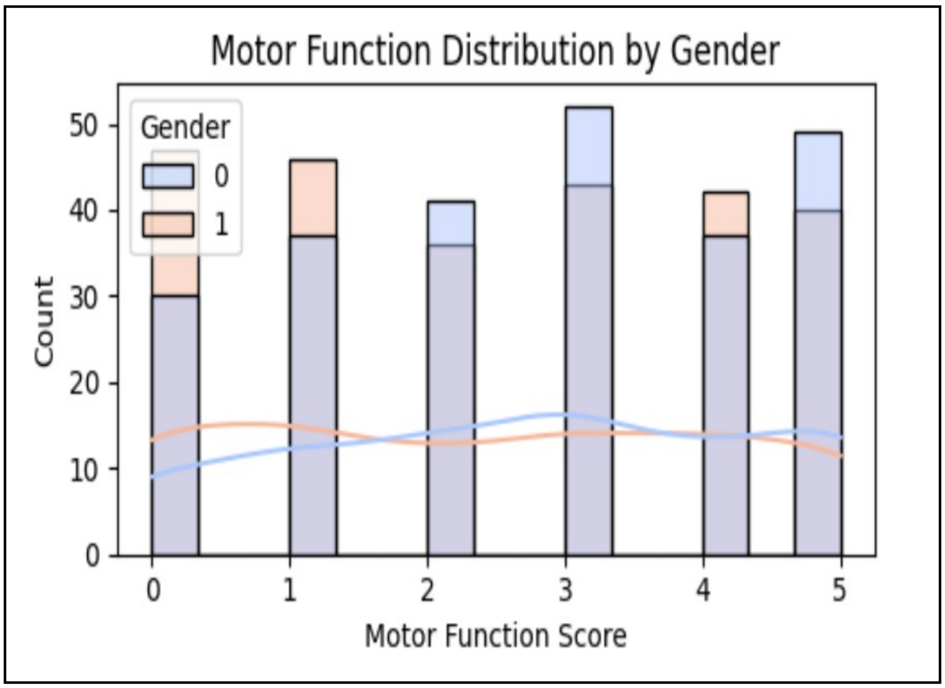


Fig1:VisualizingFeatureDistributionswith Histograms

B. Categorical observations

Barchartsaidedusinexaminingcategoricalvariablessuchas*Gender*and*Medications*.Asaresult, thedatasetshowedanear-equalgender distribution. Thisindicatesabalancedsampleofdata.When visualized against severity-related variables, grouped bar charts revealed a slightly higher average tremor and motor difficulty in male patients. Medication categories varied, with some combinations more common among patients with higher disease progression.

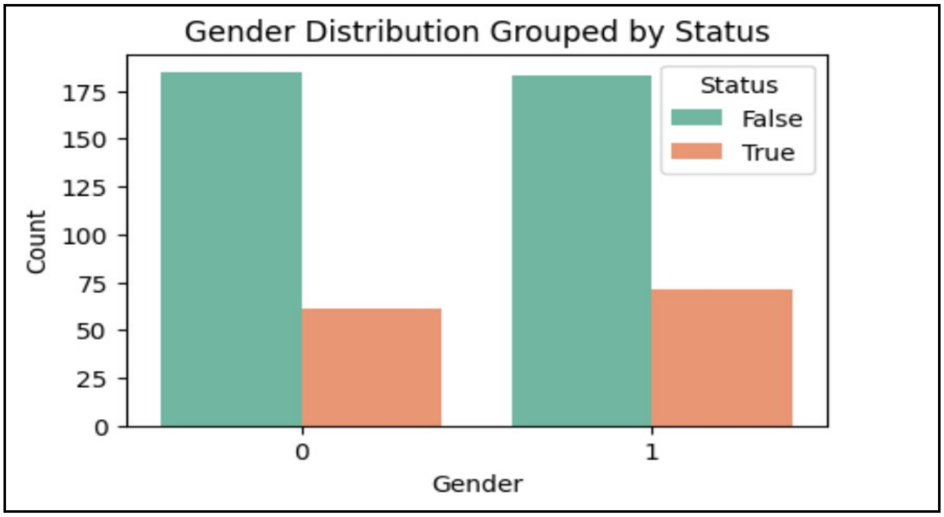


Fig 2: Gender Distribution

C. *Correlation Analysis*

Acorrelation heat-map was generated to present linear relationships among numerical features. UPDRS Score and Motor Function showed a strong positive correlation, confirming motor symptomsasakeyprogressionmarker. *YearsSince Diagnosis*alsohadamoderatecorrelationwith disease progression which was the target variable, corresponding with clinical expectations of a gradual worsening over time.

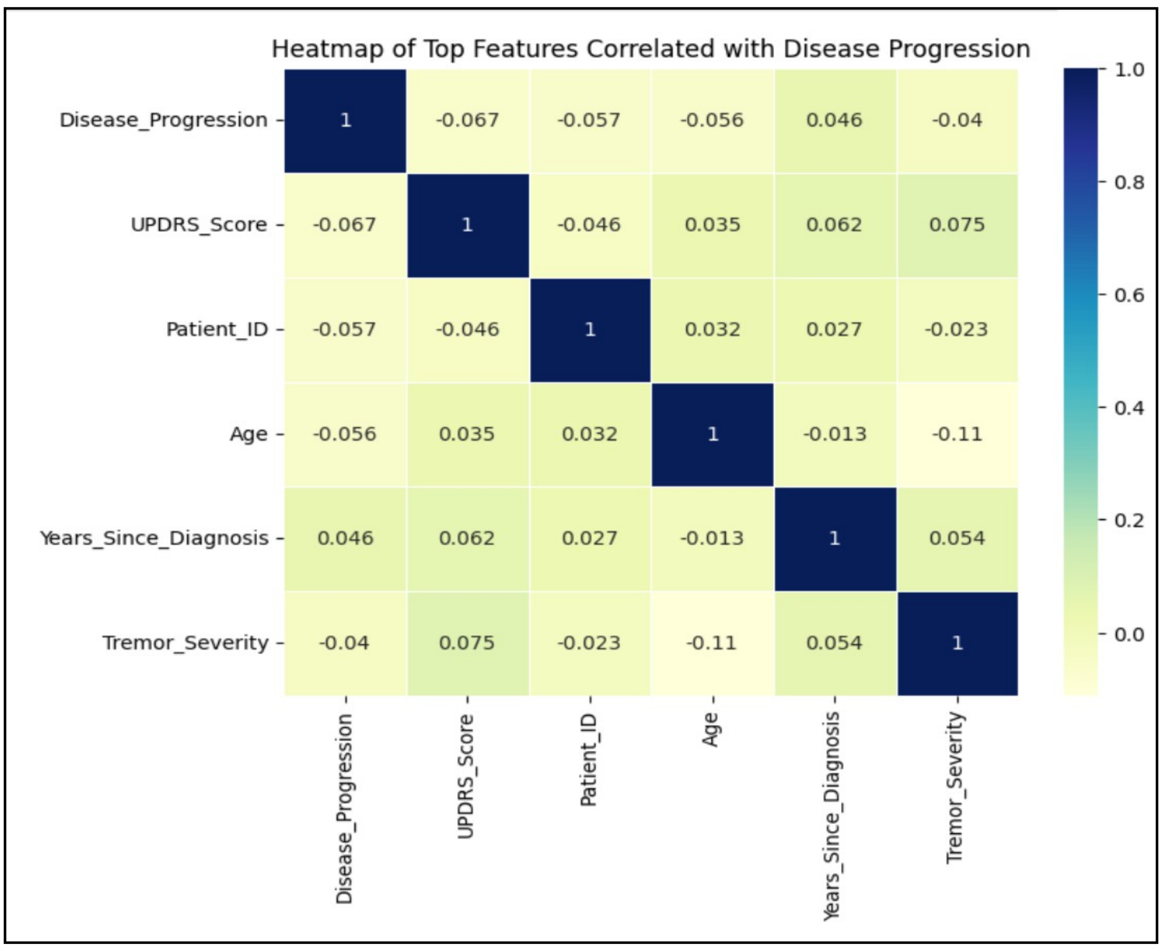


Fig 3 : Correlation heat map

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

D. Multivariate Visualizations

Pairplotswerecreatedtoexploretheinteractionsbetweenmultipleclinicalfeatures.Overlapping clusters between *Motor Function*, *Speech Difficulty*and *Balance Problems*representedcooccurringsymptomprogression. Scatterplotsshowedaspread of*UPDRS Scores*acrossvarying levels of symptom severity, indicating heterogeneous disease expression across patients.

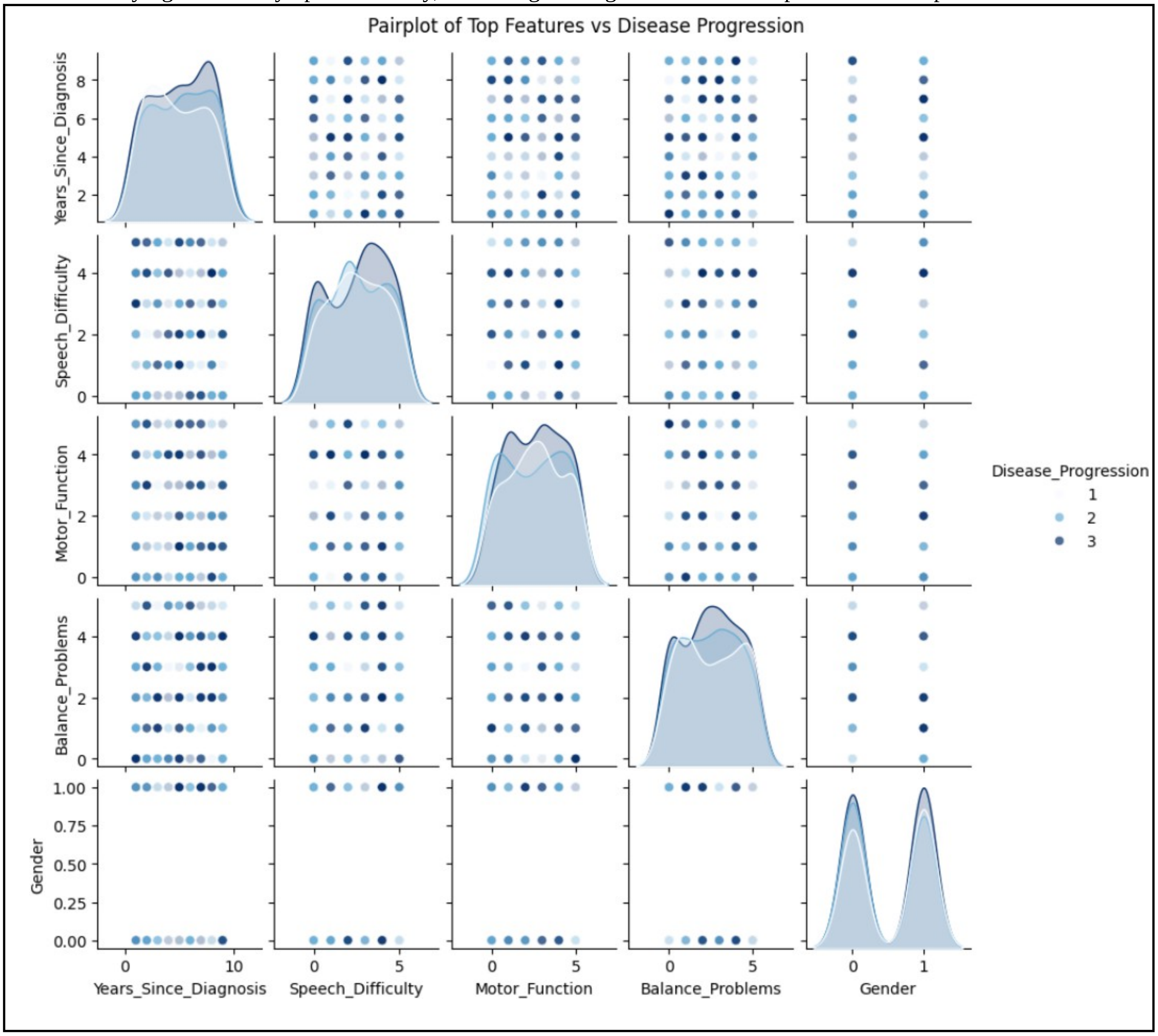


Fig4: Pair-plotoftop Features

E. Temporal Trends

Aconsistentupwardtrendwasclearlyillustratedbetween*YearsSinceDiagnosis*and*UPDRSScore* in a line plot. Significantly higher symptom scores were observed with the patients with a longer disease duration. This emphasizes the progressive nature of Parkinson's disease and confirms that diagnosis duration is a key predictor of clinical severity. The symmetry of the violins indicates a roughly normal distribution for both groups, with few outliers at the lower end of progression scores.

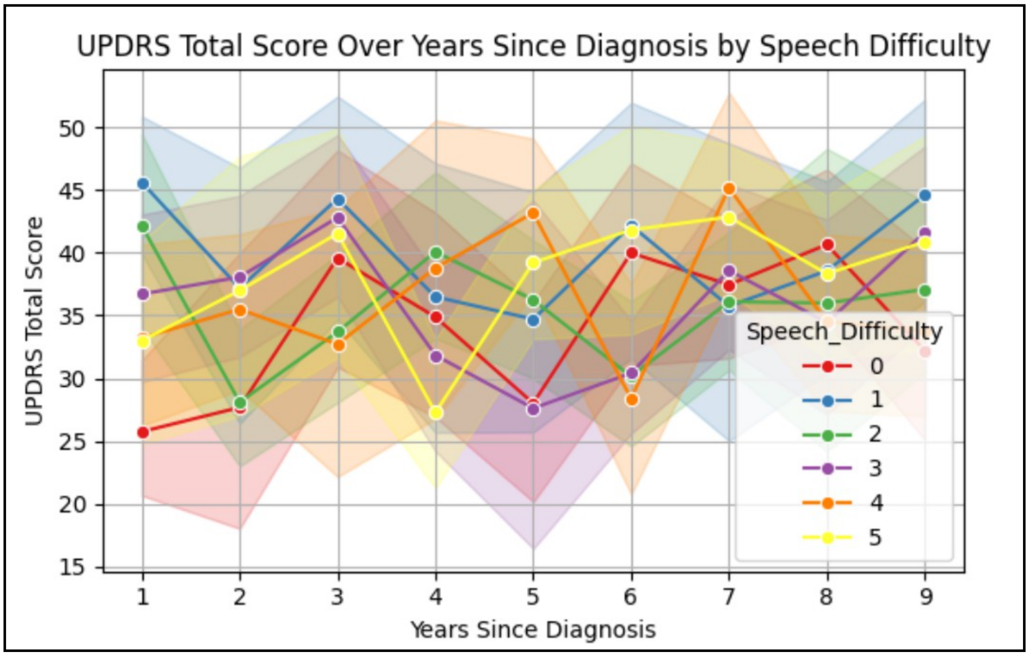


Fig5:years SinceDiagnosis &UPDRS Scoreline plot

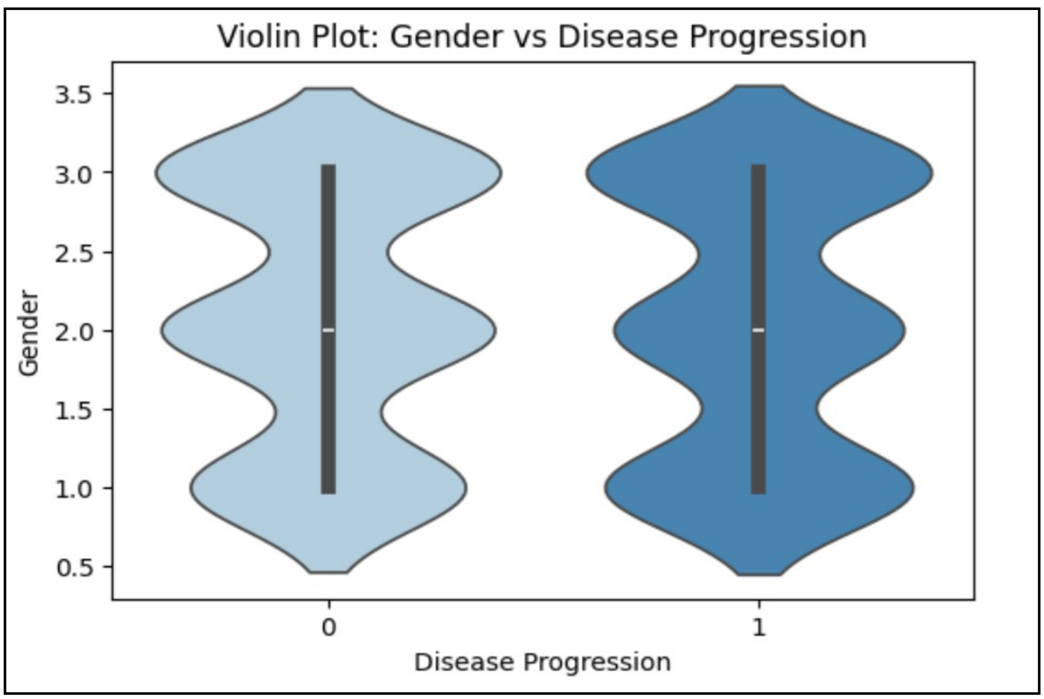


Fig6:Violin Plot

F. Summaryof Findings

The data visualizationperformed through this Exploratory DataAnalysis project acknowledge severalmeaningfulobservations. Theoveralldiseaseprogressionwasstronglyaffectedbyfeatures such as*Motor Function*and *Speech Difficulty*. Gender and exercisepatterns provided to variability in symptom progression. Thus, the combination of statistical and visual methods successfully highlighted key symptom collections and temporal trends, proposing valuable input for further clinical research or predictive modelling.

G. ImplicationsforClinicalTools

Our correlation heatmaps and temporal trends could inform the design of a lightweight Python dashboard for clinicians to track symptom progression. For example, integrating thesevisualizationswithwearablesensordata(e.g.,tremorfrequency)couldenablereal-timemonitoring.

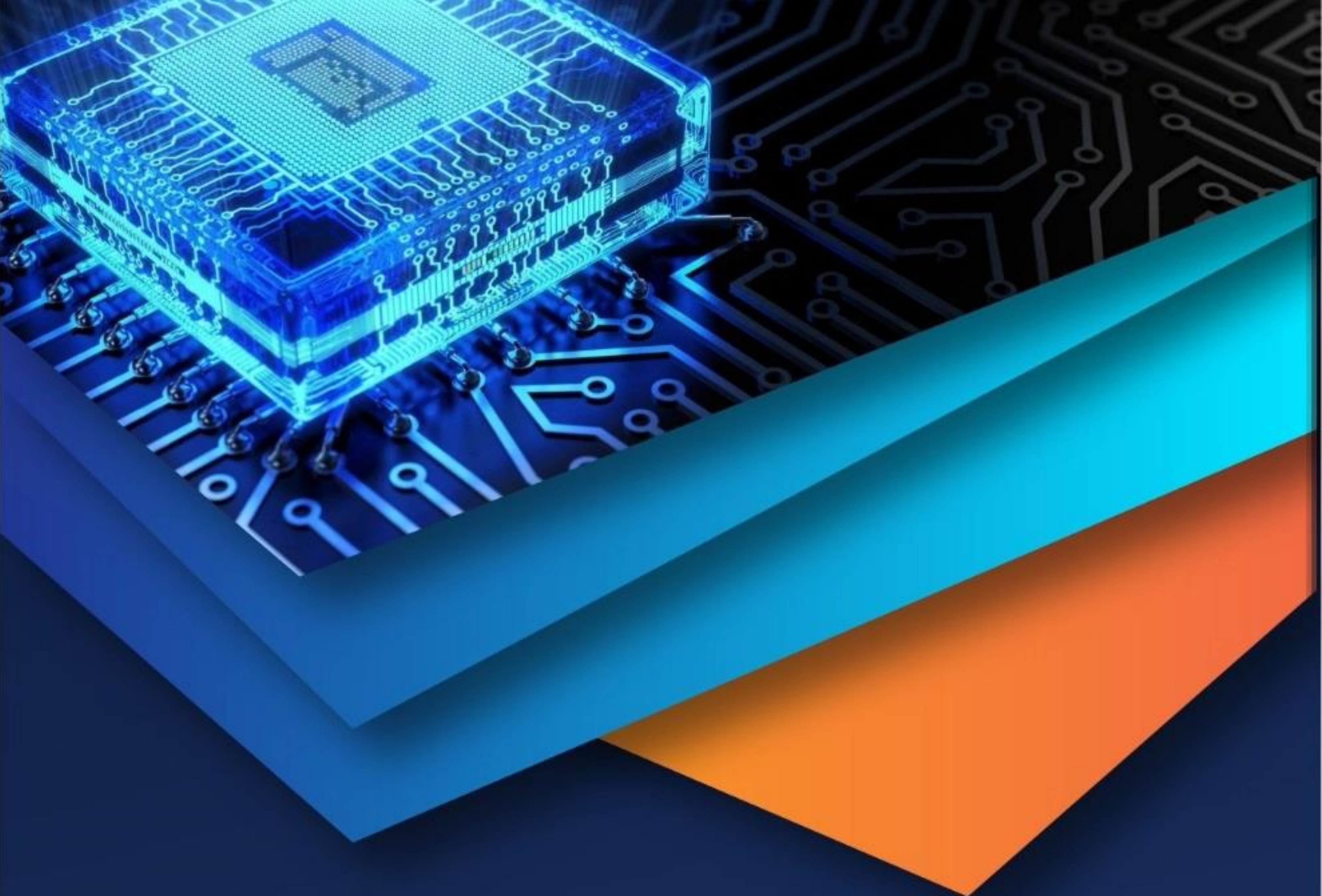
IV. CONCLUSION

ThiscurrentstudypresentedanExploratoryDataAnalysisofParkinson'sDiseaseprogressionusing a medical dataset comprising 500 patients. By applying Python-based data analysis methods, we were able to clean, preprocess and visualize the dataset to discover important trends and symptom Interdependencies.Astrong correlation was exhibited between features such as UPDRS Score, Motor Function and Balance Problems, revealing how these symptoms worsen parallelly as the disease progresses. Visual tools such as histograms, bar plots, correlation heatmaps, pair plots and line graphs showed valuable insights into how various features interact over time. Clinical patterns of Parkinson's Disease were further supported by gender-based comparisons and trend analysis.

While this study was restricted to exploratory analysis of the data and did not include predictive modeling, the results demonstrate the effectiveness of EDAin understanding symptom progression of Parkinson's disease. In future work, machine learning models can be applied to predict disease progressionandclassifyseveritylevelsbasedonpatientdata.Theoutcomeofthisstudyservesasa helpful basis for more advanced data-driven approaches in Parkinson's research and healthcare planning.

REFERENCES

[1] UCIMachineLearningRepository:Parkinson'sDiseaseProgressionDataset.Availableat:https://archive.ics.uci.edu/
[2] Smithetal.(2023).'Open-SourceToolsforClinicalDataVisualization'.J.Med.Syst.,47(3).
[3] Chenetal.(2023).'MachineLearningforParkinson'sDiseaseProgressionPrediction'.IEEE Journal of Biomedical Health Informatics, 27(4), 1234–1245.
[4] Rahmanetal.(2022).'IoT-BasedWearablesforParkinson'sSymptomMonitoring'.Sensors,22(9),3401.
[5] Automatic and Early Detection of Parkinson's Disease Available at: https://www.mdpi.com/2075-4418/13/11/1924
[6] WorldHealthOrganization.Parkinson'sDiseaseFacts.Availableat:https://www.who.int/news-room/fact-sheets/detail/parkinson-disease
[7] Parkinson's EDA Code Repository. GitHub. Available: https://github.com/ananyaravikumar548/Parkinson-s-Disease-Progression.git



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY