Data Visualization for AMP®-Parkinson's Disease Progression Prediction

Introduction

Parkinson's disease (PD) is a progressive neurological disorder that affects movement. It occurs when the dopamine-producing neurons in the brain, which control movement, become damaged or die. This results in a range of symptoms, including tremors, rigidity, slowness of movement, and difficulty with balance and coordination. As the disease progresses, it can also cause cognitive and behavioral changes, such as depression, anxiety, and dementia.

It is important to predict the occurrence of PD beforehand because early diagnosis and treatment can help manage the symptoms of the disease and improve the quality of life of the patient. Moreover, researchers are still exploring the causes of PD and the development of effective treatments, and early detection can help with the identification of potential risk factors and therapeutic targets. Predicting the occurrence of PD also allows for better planning of healthcare resources and support services for affected individuals and their families.

The goal of the project is to measure and predict the progression of disease in patients. The model will be trained using data on changing protein and peptide levels in patients compared to people who don't develop the disease. Hopefully, the results may contribute to the development of a comprehensive treatment or early diagnosis of the patients.

Data

- *train_peptides.csv* – patients' data on the peptide level:
    - *visit_id, visit_month,* patient_id
    - UniProt - protein ID.
    - Peptide - sequence of amino acids in peptide.
    - PeptideAbundance - frequency of the amino acid.
- *train_proteins.csv* – protein expression frequencies from the peptide level data:
    - *visit_id, visit_month, patient_id, UniProt.*
    - NPX - Normalized Protein eXpression (frequency of the protein's occurrence).
- *train_clinical_data.csv* – general data on decease progression:
    - *visit_id, visit_month, patient_id.*
    - updrs_ [1-4] - Unified Parkinson's Disease Rating Scale Scores.
    - upd23b_clinical_state_on_medication - whether the patient was taking medication during the UPDRS assessment.

Methods

Using the datasets, I have visualized the progression of Parkinson decease in patients expressing various degrees of symptoms intensity, highlighted the distribution of UPDRS scores, patients of and on medications, correlations between symptom types and other numeric values, as well as visualized their dysregulated normalized protein expression. To do this, I used DataSpell JetBrains IDE and such Python libraries as Pandas, NumPy, Seaborn and Matplotlib.

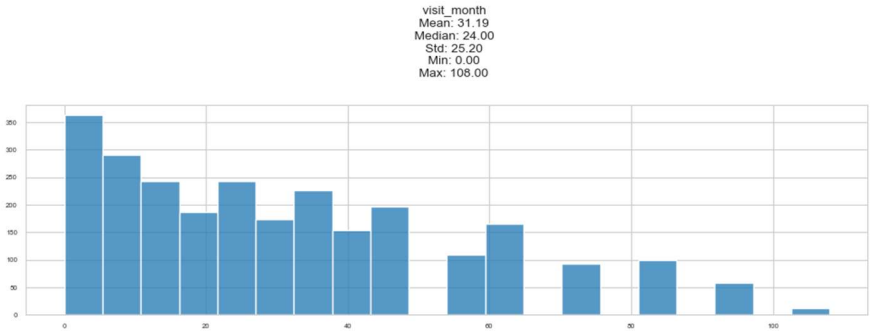
Originally, I intended to try building a prediction model using Time Series Analysis and machine learning algorithms, yet it turned out to be too difficult these topics are not covered by the course. Instead, I have performed the visualization of the provided data sets that yielded meaningful results.

Experimental Results

```
Data columns (total 9 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   visit_id                          2615 non-null   object
 1   patient_id                        2615 non-null   int64
 2   visit_month                       2615 non-null   int64
 3   updrs_1                           2615 non-null   float64
 4   updrs_2                           2615 non-null   float64
 5   updrs_3                           2615 non-null   float64
 6   updrs_4                           2615 non-null   float64
 7   upd23b_clinical_state_on_medication  2615 non-null   object
 8   null_count                        2615 non-null   int64
dtypes: float64(4), int64(3), object(2)
```

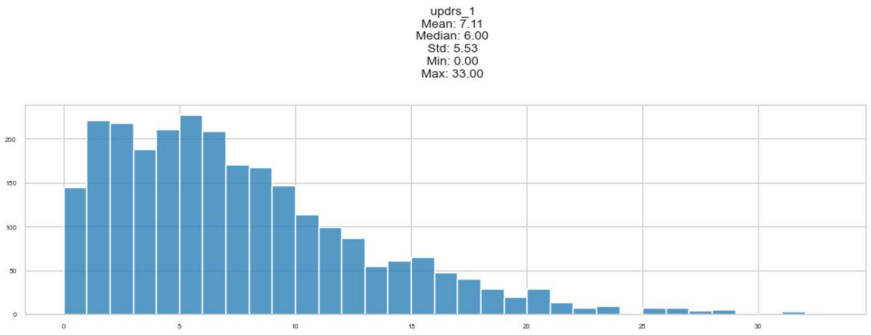
Dataset Info:

The yielded data above showcases that in the provided data set, there are 2615 entries of each category. Yet, it doesn't tell that the values are unique as, for example, the visiting identifiers are being repeated for each vising month.



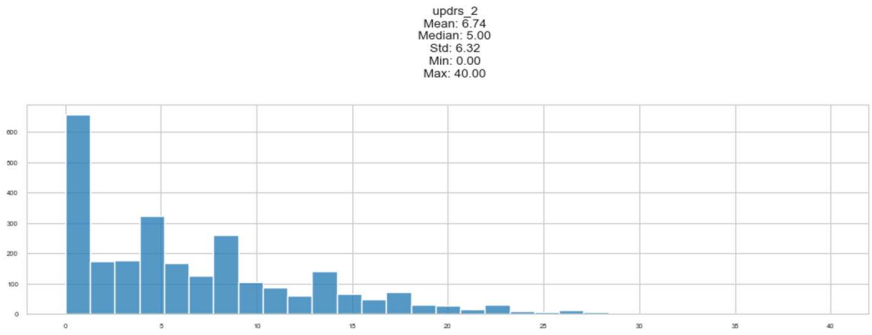
Visiting Data

Most patients visited treatment facilities for less than 40 months since diagnosis. This, however, might indicate that patient death during this period was not a significant factor, considering the average lifespan of PD patients is approximately 16 years after diagnosis.



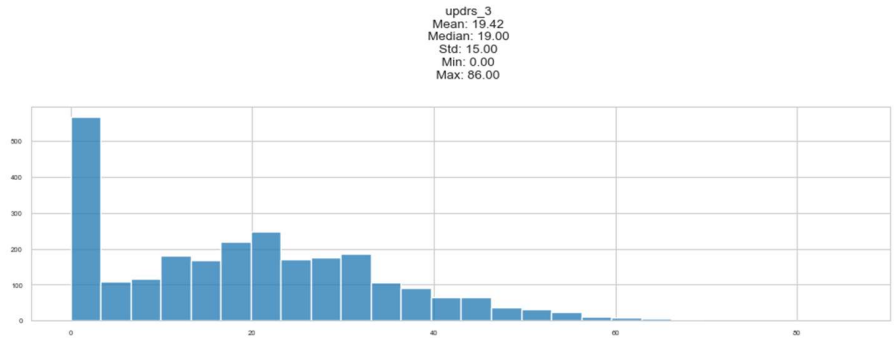
MDS-UPDRS Part I: Nonmotor Experiences of Daily Living

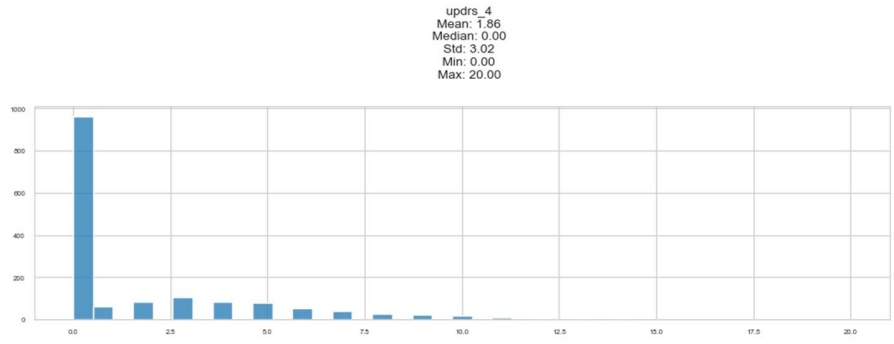
Most patients exhibited mild impairments in non-motor experiences, such as cognitive impairment, hallucinations and psychosis, depressed mood, anxiety, apathy, and dopamine dysregulation syndrome. This can be attributed to the recent diagnosis of most patients, as severe symptoms typically take longer to develop.



MDS-UPDRS Part II: Motor Aspects of Experiences of Daily Living

Most patients showed little to no impairment in various motor aspects, including speech, saliva and drooling, chewing and swallowing, eating tasks, dressing, hygiene, handwriting, and other activities. This aligns with the recent diagnosis of most patients.



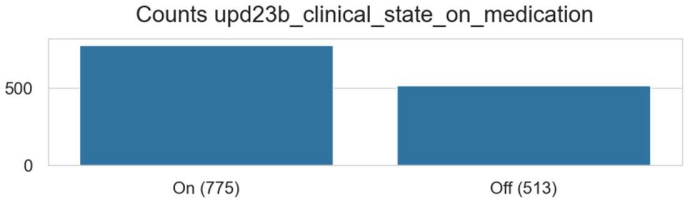


MDS-UPDRS Part III: Motor Examinations

The results from speech, facial expression, rigidity, finger tapping, hand movement, pronation-supination movements of hands, toe-tapping, leg agility, etc. follow the same pattern as Part II.

MDS-UPDRS Part IV: Motor Complications

The overwhelming majority of patients showed no motor complications, such as dyskinesias (involuntary movements), motor fluctuations, or dystonia (unintentional sustained muscle contractions leading to abnormal postures).

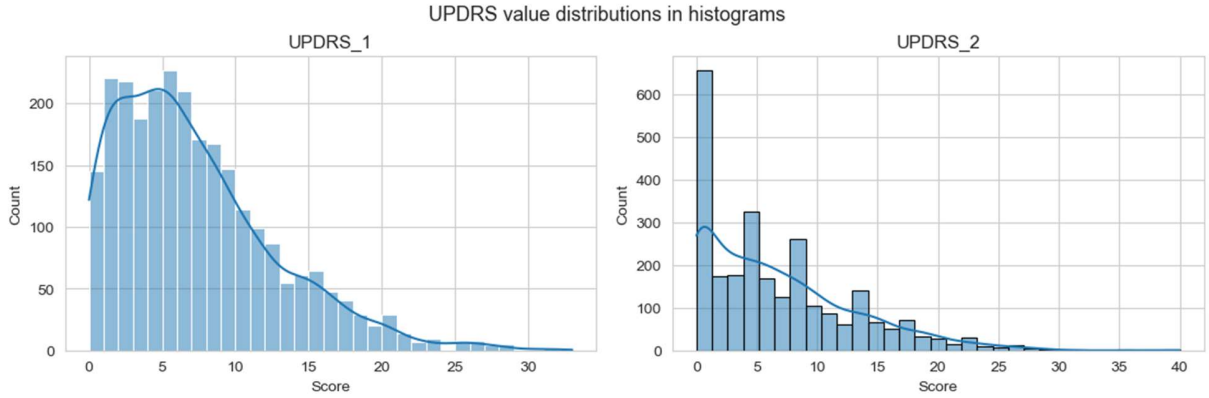


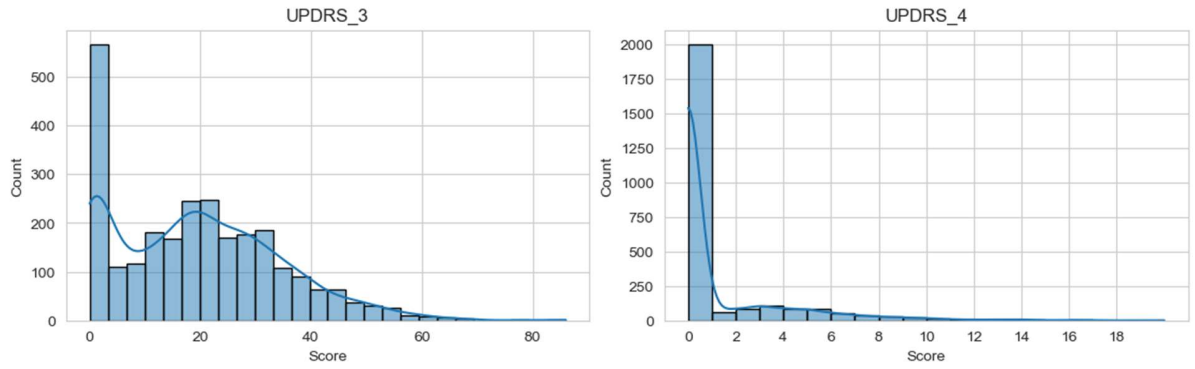
Medication Data

Out of the 1318 medication recordings, representing the total number of months patients were on medication (cumulatively), it was found that 775 recordings (58.8%) indicated medication usage, while 513 recordings (41.2%) indicated the absence of medication usage.

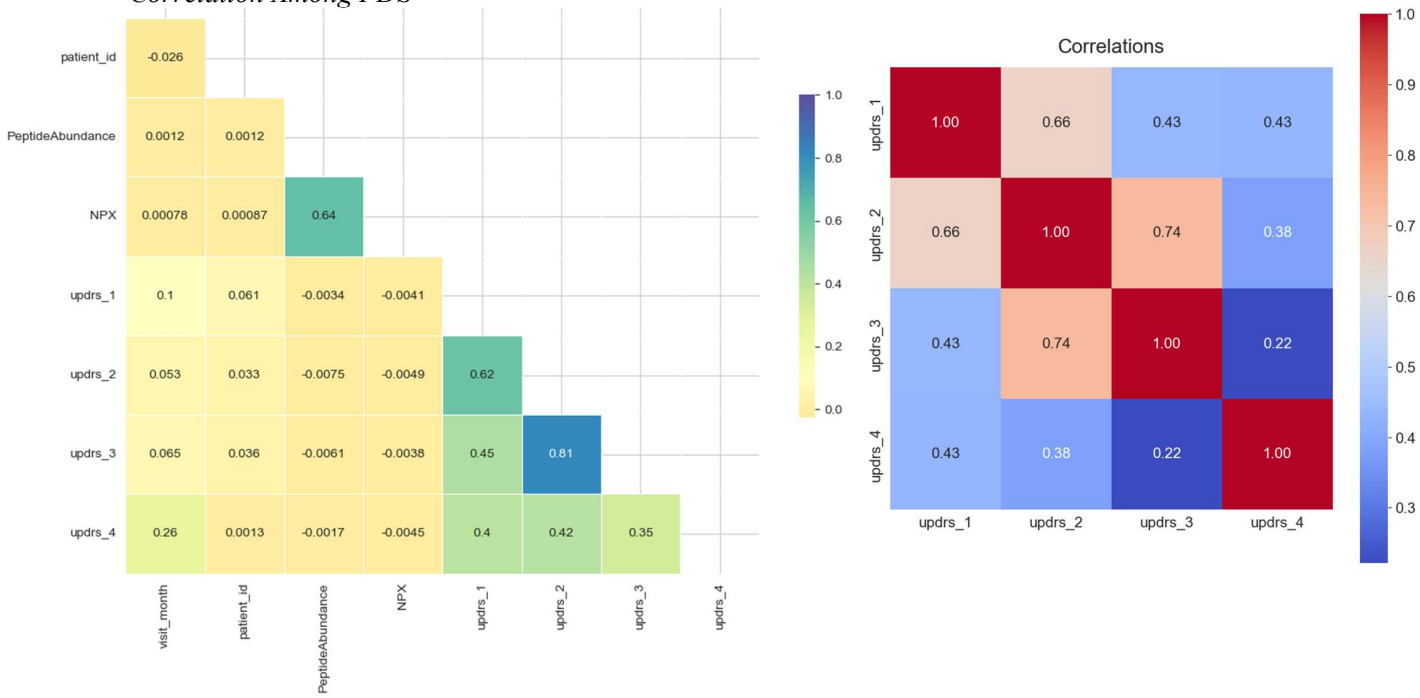
Total MDS-UPDRS Distribution

All symptoms displayed a downward trend, indicating that most patients either exhibited mild symptoms or had no severe impairments.

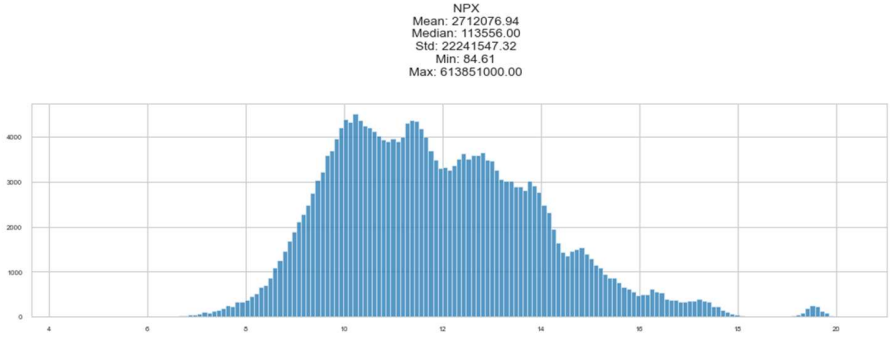




Correlation Among PDS



Strong correlations were observed between Part I and Part II, as well as between Part II and Part III of the UPDRS. Moderate correlations were found between Part I and Part III, as well as Part I and Part IV. The correlation between Part III and Part IV was relatively weak. A strong correlation was observed between Normalized Protein eXpression (NPX) and protein abundance. Additionally, the number of visiting months positively correlated with Part IV of the UPDRS symptoms.



Protein Distribution

Since its total protein distribution, the results might vary if proteins are explored individually. However, as seen here, the overall protein distribution demonstrated a rightward shift, indicating dysregulation in net protein expression.

This dysregulation is expected and is associated with non-motor experiences, which are often expressed as mood swings.

Discussion and Conclusions

In this project, I conducted data analysis using various visualization techniques to examine the progression of Parkinson's disease and its associated factors. I utilized datasets containing information on peptide and protein levels, as well as general clinical data, such as visit details and UPDRS scores. Using Python libraries, I visualized the data to gain insights into the disease progression and related factors.

Key Discoveries:

- *Disease Progression Patterns*. Patients exhibited mild impairments in non-motor experiences, such as cognitive impairment, hallucinations, depressed mood, anxiety, apathy, and dopamine dysregulation syndrome. This suggests that most patients were recently diagnosed without severe symptoms. Research has shown that during the early phases, non-motor symptoms are more common than severe motor impairments.
- *Motor Impairments*. Patients showed little to no impairment in various motor aspects of daily living, such as speech, saliva and drooling, chewing and swallowing, eating tasks, dressing, hygiene, handwriting, and other activities. This aligns with the recent diagnosis of most patients.
- *UPDRS Scores*. The total distribution of UPDRS scores exhibited a downward trend, indicating that most patients either had mild symptoms or no severe impairments and is in line with research that suggests UPDRS scores are typically lower in the early stages of Parkinson's disease. As the disease progresses, the scores tend to increase, reflecting more significant motor and non-motor symptoms.
- *Correlations*. Strong correlations were observed between Part I and Part II, as well as between Part II and Part III of the UPDRS, indicating interdependencies among different symptom categories. Moderate correlations were found between Part I and Part III, as well as Part I and Part IV of the UPDRS. The correlation between Part III and Part IV was relatively weak. The finding reflects the interconnections between non-motor symptoms and motor impairments in Parkinson's disease.
- *Protein Dysregulation*. The distribution of protein levels demonstrated a rightward shift, indicating dysregulation in net protein expression. This dysregulation is expected and is associated with non-motor experiences, often manifested as mood swings. Studies have identified dysregulated protein levels, particularly in relation to dopamine-related pathways, contributing to the pathogenesis and progression of the disease.

References

Kirsch, L., Dane, S., Adam, S., & Dardov, V. (2023). AMP®-Parkinson's Disease Progression Prediction. Kaggle. https://www.kaggle.com/competitions/amp-parkinsons-disease-progression-prediction/

Erro, R., Picillo, M., Vitale, C., Amboni, M., Moccia, M., Longo, K., Cozzolino, A., Giordano, F., De Rosa, A., De Michele, G., Pellecchia, M. T., & Barone, P. (2010). Non-motor symptoms in early Parkinson's disease: a 2-year follow-up study on previously untreated patients. Journal of Neurology, Neurosurgery & Psychiatry, 84(1), 14–17. https://doi.org/10.1136/jnnp-2012-303419

Martinez-Martin, P., Chaudhuri, K. R., Rojo-Abuin, J. M., Rodriguez-Blazquez, C., Alvarez-Sanchez, M., Arakaki, T., ... & Goetz, C. G. (2015). Assessing the non-motor symptoms of Parkinson's disease: MDS-UPDRS and NMS Scale. European Journal of Neurology, 22(1), 37-43.