AMP PARKISION DISEASE PROGRESSION PREDICTION

Dr. S Mohana*1, Dheeraj Prakash S*2, Michael Jones J*3,
Mukesh Manikandan M*4, Muthuraaj KJ*5

*1Associate Professor, Dept Of Computer Science And Engineering, Saranathan College Of Engineering, Tiruchirappalli, Tamilnadu, India.
*2,3,4,5Student, Dept Of Computer Science And Engineering, Saranathan College Of Engineering, Tiruchirappalli, Tamil Nadu, India.

DOI: https://www.doi.org/10.56726/IRJMETS43893

ABSTRACT
Parkinson's disease (PD) is a neurodegenerative disorder affecting 60% of people over the age of 50 years. Patients with Parkinson's (PWP) face mobility challenges and speech difficulties, making physical visits for treatment and monitoring a hurdle. PD can be treated through early detection, thus enabling patients to lead a normal life. The rise of an aging population over the world emphasizes the need to detect PD early, remotely and accurately. This paper highlights the use of machine learning techniques in telemedicine to detect PD in its early stages.

Research has been carried out on the MDVP audio data of 30 PWP and healthy people during training of 4 ML models. Comparison of results of classification by Support Vector Machine (SVM), Random Forest, Lasso Regression, Light Gradient Boosting and XGBoosting models, yield Random Forest classifier as the ideal Machine Learning (ML) technique for detection of PD.

Keywords: Support Vector Machine (SVM), Random Forest, Lasso Regression, Light Gradient Boosting And Xgboosting Models, Yield Random Forest Classifier.

I. INTRODUCTION

1.1 PARKINSON DISEASE:
Parkinson’s disease (PD) is a common neurological disorder impacting muscle movement in the body. It affects mobility, speech and posture leading to tremors, muscle rigidity and bradykinesia. It occurs due the death of neurons, resulting in a decrease of dopamine levels in the brain. Low levels of dopamine hamper communication between synapses, causing ineffective motor functions. While the progress of symptoms may vary from patient to patient, balance problems and tremors are the most prevalent side-effects of dopaminergic neuron death. Unfortunately, there is no cure for PD, hence Patients with Parkinson’s (PWP) rely on early detection and tailored treatments to slow the progress of the disease.

1.2 PROGRESSION STAGES:
PD has 5 stages of progression and 90% PWP display signs of vocal cord injuries as a symptom in stage 0. Vocal impairment is not only easy to measure, but also falls under the category of telemedicine or remote medicine. Patients need not travel physically to a doctor instead; they can record audio using phones and perform a simple test at home. Common voice modulation symptoms include dysphonia and dysarthria. Patients can be asked to hold a single vowel's pitch for as long as possible, also known as sustained phonation or running speech tests can be administered, as a realistic test of impairment. These phonation tests can be used for diagnosis of Parkinson’s at stage 0.

Following early detection, doctors can cater therapeutic solutions or deep brain simulation to reactivate the dopamine producing neurons in the brain, thereby slowing the progress of PD. Owing to its complex nature, there is no cure for Parkinson’s till date. However, early identification followed by the right medication can reduce the tremors and imbalance symptoms in patients, enabling them to lead a normal life. We are using several Regressor ML algorithms and SVR algorithm. To achieve efficient prediction results.

II. LITERATURE SURVEY

Previous studies to predict PD have been implemented on MRI scans, gait and genetic data, but research on audio impairment for early detection is minimal. For instance, Bilal et. al. studied genetic data to predict the onset of PD in senior patients with SVM model. They trained an SVM model to reach an accuracy of 0.889, while
this research paper describes an improved SVM model with an accuracy of 0.9183. These results also corroborate the merits of classification of PD based on audio data, over genetic data. Raundale, Thosar and Rane used keystroke data from UCI telemonitoring dataset to train a Random Forest classifier to predict the severity of PD in older patients. Cordella et al. use audio data to classify PWP, however their models are heavily reliant on MATLAB. Our research uses open-source models trained in Python, that are faster and memory efficient.

Majority of research done emphasizes the use of deep learning in PD detection, such as, Aditi Govindu et al who explains the process of predicting the progress of PD by using e KNN, logistic regression, random forest regression and SVM models and got 91.83% of success rate. Ali et al. who explains the use of ensemble deep learning models applied to phonation data, to predict the progress of Parkinson’s disease. Their work lacked the use of feature selection that would improve Deep learning model (DNN) performance. Hence, this paper implements PCA on 22 attributes to select 7 major voice modalities in PD detection. Huang et al. aim to reduce PD diagnosis dependence on wearable equipment by training a traditional decision tree on 12 complex speech features of the MDVR-KCL dataset. Wodzinski et al. trained a ResNet model on images of audio data, instead of training the model on the nuances of the frequency of audio. Wroge et al. aimed to remove subjectivity of doctors in prediction of PD using an unbiased ML model, however their results achieved peak accuracy of 85% only.

Wang et al. implemented 12 machine learning models on 401 voice biomarkers dataset to classify patients as PD or not. They built a custom deep learning model (DEEP) with a classification accuracy of 96.45%, however the model was expensive due to large memory requirements. Alkhatib et al. implemented a linear classification model with 95% accuracy to characterize shuffling movement of PD patients. Their study focused on the gait of patient and future work encouraged the use of audio and sleep data to improve the results. Ricciardi et al. performed spatial-temporal analysis of brain MRI scans. They implemented decision trees, random forest and KNN to detect Mild Cognitive Impairment (MCI) in PWP. However, the dataset was small and artificial data augmentation was needed.

III. PROPOSED SYSTEM

3.1 LOADING DATASETS

Mass spectrometry data at the peptide level. Peptides are the component subunits of proteins

- visit_id - ID code for the visit.
- visit_month - The month of the visit, relative to the first visit by the patient.
- patient_id - An ID code for the patient.
- UniProt - The UniProt ID code for the associated protein. There are often several peptides per protein.
- Peptide - The sequence of amino acids included in the peptide. See this table for the relevant codes. Some rare annotations may not be included in the table. The test set may include peptides not found in the train set.
- Peptide Abundance - The frequency of the U. Haq and colleagues implemented L1-support SVM, without feature identification on vowel phonation dataset for neurological disorder patients. Their paper focused on the patient age group of 46-85 years, without considering healthy individuals in a lower age bracket.

Based on our literary review, we have implemented a PD detection algorithm by combining several algorithms. The algorithms are SVR, Lasso Regression, Decision Tree Regression, XGB Regression, Random Forest Regression, LGBM Regression.

3.2 FEATURE ENGINEERING MODULE

The feature engineering algorithm includes filling in any missing values, followed by grouping the data into peptide, clinical, and protein datasets. Aggregate functions are then applied to each group to summarize the data. Next, the datasets are merged into a single dataset for training the module. Finally, Min-Max normalization is performed on the merged dataset to ensure all features are on a similar scale between 0 and 1

3.2.1 TRAIN_PEPTIDE.CSV
3.2.2 TRAIN_PROTEINS.CSV
3.2.3 TRAIN_CLINICAL.CSV
3.3 GROUPING AND APPLYING THE AGGREGATE FUNCTIONS FOR THE NEEDED FEATURES (FOR ALL 3 DATASETS) AND MERGING AND NORMALIZATION

**IV. CONCLUSION**

Through the utilization of UPDRS scores, we have come to the conclusion that they can be a reliable evaluation tool in determining the severity of an individual’s condition. We can predict it in the initial stage itself, so the death rate or severity may be reduced. This project has the potential to transform how Parkinson’s disease is treated and improve the lives of millions of patients worldwide.

**V. REFERENCES**


