ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 9| Issue: 10| October 2023|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2023: 8.224 || ISI Value: 1.188

PARKINSON'S DISEASE DETECTION THROUGH VOICE SIGNALS

Chittiprolu. Saranya[1], Chittibomma. Sai Sindhu[2], Ikkurthi. Chandi Priya[3], Dantu. Swati[4], Anne. Dhatri [5], Alla. Kalavathi [6]

[1,2,3,4,5]*B. Tech Students Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur*

[6] *Professor and HOD, Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur*

Article DOI: https://doi.org/10.36713/epra14758
DOI No: 10.36713/epra14758

ABSTRACT

The primary Parkinson's disease (PD) is a disorder of the central nervous system and about 89% of the people with PD suffering from speech and voice disorders. In this project, we adopted a dynamic feature selection based on fuzzy entropy measures for speech pattern classification of Parkinson's diseases. To investigate the effect of feature selection, XGBoost algorithm was applied to distinguish voice samples between PD patients and health people. The data set of this research is composed of voice signals from 195 people, 147 with Parkinson's disease and 48 healthy people.

The results show that various voice samples need different feature selection. We applied dynamic feature selection can get higher rate of classification accuracy than all features selected.

KEYWORDS: ***Parkinson's disease, linear discriminant analysis, similarity measure, fuzzy***

1. INTRODUCTION

In Parkinson's disease, dopaminergic neurons in the substantia nigra are lost i.e., Compact body of the midbrain. These neurodegenerative changes are numerical and symptoms include communication problems, voice coordination, and allergies. Dysarthria is also observed in PD patients. The miles are marked by weakness, paralysis, lack of strength and communication in the car: Effects on breathing, crying, speech, and prosody. The exact cause of Parkinson's disease is unknown, but it is a complex concept, complex interactions between genetics, biology, and environment.

For many years, this creates uncertainty in the fate of those affected, it also shows therapeutic trials where there are challenges in determining successful endpoints and Further reports on functional changes in the peripheral immune system in Parkinson's disease.

In recent years, evidence has been reported that an intestinal protein called alpha-synuclein enters the intestine. The anxiety machinery and the gut-brain axis are implicated in the pathogenesis of Parkinson's disease. It suggests that Parkinson's disease may also start at the outer edges.

These studies may also provide a way to select markers that anticipate the progression of Parkinson's disease. This usually applies to Parkinson's disease, as the signs and course of the disease can vary. It has not been available for decades. Therefore, more advanced measures may be required a diagnostic device for diagnosing Parkinson's disease according to the progression of the disease, symptoms develop that make Parkinson's disease difficult to treat.

The main disadvantages of PD speech are Loss of depth, tone of voice, pitch, loss of pressure, hallucinations, short speech, Speed, dynamic thrusts, slurred consonants, and a heavy, breathy voice (Speech disorder).

Audio recording is non-invasive and can be easily performed using a mobile device, this research voice recordings are taken from existing voice. The company is also evaluating the effectiveness of using controlled XGBoost Algorithm to better identify people with diseases.

The course of the disease is difficult in young people, and it is more common in older people. Medicine statistics has developed a large amount of data from various clinical fields such as fitness, Nursing care services. There are ways to process these statistics and gain insights from them. To solve this goal, the need for big data evaluation with machine learning various medical and clinical problems.

Although motor symptoms are characteristic, their severity and progression vary, and Parkinson's disease progresses differently in different people. Additionally, non-motor symptoms such as depression, anxiety, and constipation often impact a patient's quality of life more than motor symptoms. Research into the genetic component of Parkinson's disease is ongoing.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 9| Issue: 10| October 2023|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2023: 8.224 || ISI Value: 1.188

Specific genetic mutations associated with familial forms of the disease have been discovered, providing information about possible causes and risk factors.

This is a disease that goes beyond motor symptoms and includes non-motor symptoms and genetic complexity. Comprehensive care approaches, research into the role of the immune system, and personalized medicine promise to improve the lives of people with Parkinson's disease. Recognizing the human and social aspects of the disease is paramount to finding better treatments and support for those affected.

Many studies have already shown that ML knowledge exists algorithm showed significantly improved overall performance in classifying major medical problems. However, supervised ML strategies are among the most practical. Practical programs in techniques and clinical areas for the research community. This work is to improve detection and analysis strategies for Parkinson's disease.

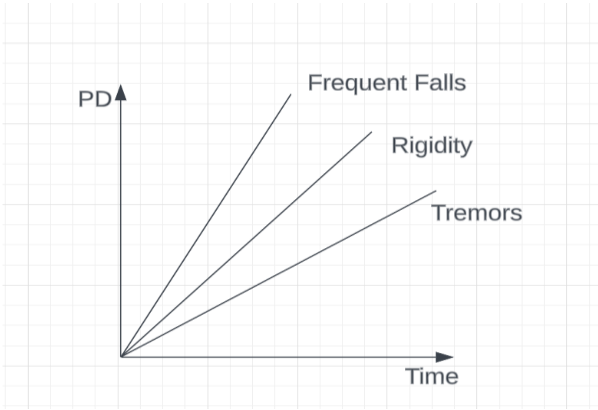
There is no cure for Parkinson's disease. However, medicines may help dramatically manipulate signs and symptoms on a regular basis. Therefore, if it is recognized within in the early stages, drug costs are low. Therefore, our observations can play an important role in Parkinson's disease detection using machine learning algorithm.

2.LITERATURE REVIEW

James Parkinson (1755–1824) was a British pharmacist, geologist, paleontologist, and political activist. He spent his radical youth in political activity, and later devoted himself to medical work and interests in medicine and earth sciences. His 1817 Essay on Trembling Paralysis combined elements of resting tremor, rigidity, and the characteristic posture and gait of the disease, which he called paralytic agitans.

Fragments of the disease can be found in ancient Chinese, Indian, Babylonian, and Greek texts, but Jean-Martin Charcot added bradykinesia to the cardinal symptoms in his 1872 lecture and coined the name Parkinson's disease. Charcot's school advocated the use of anticholinergic drugs. French neurologist Edouard Brisseau described a characteristic pathology of the substantia nigra, and this finding was confirmed by Greenfield and Bosanquet in 1953. Although Berger and Ewens synthesized dopamine in 1910, it was not until the late 1950s that neurochemists linked dopamine deficiency in the substantia nigra to James Parkinson's disease.

Birkmeyer and Hornykiewicz's 1961 injection of the dopamine precursor L-dihydrophenylalanine (levodopa) into Parkinson's disease patients produced immediate but short-term improvement.



The subsequent development of peripherally acting dopamine blockers increased the uptake of levodopa into the central nervous system and decreased the incidence of side effects. Invasive techniques such as destruction of the globus pallidus and placement of deep brain electrodes in the midbrain are modern treatments with unclear indications for use. James Parkinson's disease has become a model for combined chemical and surgical attacks on the globally prevalent degenerative neurological disease.

Generally, PD is defined as a progressive and chronic neurodegenerative disease and its effects. According to a report published by Parkinson's disease patients in 2015, the Disease Foundation approximately 10 million people worldwide suffer from Parkinson's disease. There are about 1 million of them in the United States, another report says. According to the Parkinson's Disease Society, one in every 500 people in the UK suffer from Parkinson's disease. The disease worsens over time and affects people between the ages of 50 and 70.

3.METHODOLOGY

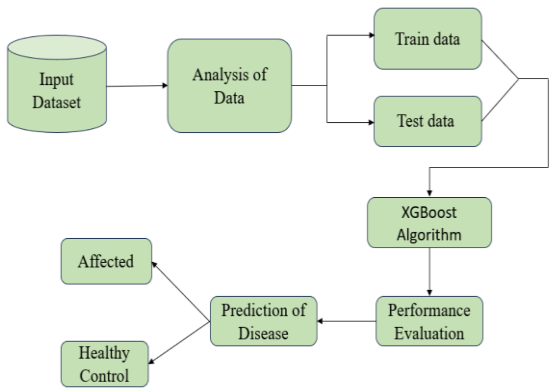
The methodology employed in this research involved the retrieval and analysis of a 7-year dataset comprising global publications related to the detection of Parkinson's disease through voice signals. Using specific search parameters. The study focused on extracting publications from the period between 2014 and 2021, utilizing keywords associated with Parkinson's disease and voice signal analysis.

A comprehensive search string, incorporating relevant field tags such as "Keyword" and "Title of Paper," was formulated to filter the dataset. Further refinement was implemented by limiting the search to specific categories related to Parkinson's disease and voice signal detection techniques. This refined search string enabled the collection of data on the distribution of global research output concerning Parkinson's disease detection, encompassing aspects such as the collaboration between countries, author contributions, affiliations, and the dissemination of findings through various academic journals.

Additionally, citation data for the retrieved publications was collected from the date of publication until the 5th of July 2021, facilitating a comprehensive analysis of the impact and influence of the research within the academic community and beyond. The methodology aimed to provide an insightful

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 9| Issue: 10| October 2023|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2023: 8.224 || ISI Value: 1.188

overview of the evolution and current landscape of research endeavors pertaining to the use of voice signals for the detection and analysis of Parkinson's disease.



3.1 PD Diagnosis

For diagnosing the PD, it is important to differentiate between Parkinsonism and PD. Parkinsonism is the clinical representation of the symptoms that are attributed to PD like rigidity, tremors, and bradykinesias. Whereas, PD is a form of parkinsonism and hence, not all the patients with parkinsonism will have PD as it might be due to other health issue.

It is difficult to differentiate PD from other form of parkinsonism based on motor symptoms and hence it becomes necessary to monitor other symptoms as well,
to diagnose the disease.

with different health conditions.

The research seeks acoustic parameters dataset. Information regarding parameters such as the average vocal fundamental frequency (Fo), maximum and minimum vocal fundamental frequencies (Fhi and Flo), along with various measures of variation in frequency (Jitter (%), Jitter (Abs), RAP, PPQ, and DDP), and measures of variation in amplitude (Shimmer, Shimmer (dB), Shimmer:APQ3, Shimmer:APQ5, APQ, and Shimmer: DDA), were extracted from voice recordings. Furthermore, parameters including the noise-to-harmonics ratio (NHR), harmonics-to-noise ratio (HNR), recurrence period density entropy (RPDE), D2, and signal fractal scaling exponent (DFA) were calculated to characterize the vocal properties. Careful attention was given to ensure the uniformity of measurement protocols and data extraction methods across all sources, preventing any potential statistical disparities in the acoustic parameter analysis.

The voice recordings utilized in this study were collected from a diverse range of individuals, encompassing various age groups, genders, and geographical backgrounds. Each recording accurately depicted the distinctive fundamental frequencies, pitch changes, and amplitude fluctuations of the individuals. The dataset included many vocal samples, ensuring a thorough representation of the acoustic properties of the human voice in various settings. The study aimed to provide a thorough understanding of the complex interaction between acoustic parameters and individual vocal attributes by incorporating this wide range of voice recordings, which contributed to a nuanced exploration of voice analysis and its potential applications in more general research and clinical settings.

index	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	MDVP:Shimmer(dB)	Shimmer:APQ3
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	7e-05	0.0037	0.00554	0.01109	0.04374	0.426	0.02182
1	phon_R01_S01_2	122.4	148.65	113.819	0.00968	8e-05	0.00465	0.00696	0.01394	0.06134	0.626	0.03134
2	phon_R01_S01_3	116.682	131.111	111.555	0.0105	9e-05	0.00544	0.00781	0.01633	0.05233	0.482	0.02757
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	9e-05	0.00502	0.00698	0.01505	0.05492	0.517	0.02924
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	0.584	0.0349

index	Shimmer:APQ5	MDVP:APQ	Shimmer:DDA	NHR	HNR	status	RPDE	DFA
0	0.0313	0.02971	0.06545	0.02211	21.033	1	0.414783	0.815285
1	0.04518	0.04368	0.09403	0.01929	19.085	1	0.458359	0.819521
2	0.03858	0.0359	0.0827	0.01309	20.651	1	0.429895	0.825288
3	0.04005	0.03772	0.08771	0.01353	20.644	1	0.434969	0.819235
4	0.04825	0.04465	0.1047	0.01767	19.649	1	0.417356	0.823484

3.2 Data and Variables

The input data for this research comprises a diverse collection of voice signals obtained from both healthy individuals and those affected by specific health conditions, including Parkinson's disease. These voice signals were acquired using standardized recording procedures to ensure consistency and reliability across the dataset. The recorded audio data were subsequently converted into numerical form using the Multidimensional Voice Program (MDVP), facilitating the extraction of various acoustic parameters. These parameters include fundamental frequency (Fo), jitter, shimmer, noise-to-harmonics ratio (NHR), harmonics-to-noise ratio (HNR), and other relevant measures. The use of MDVP enabled the comprehensive characterization of the vocal properties of both healthy and diseased subjects, thereby establishing a robust foundation for the subsequent analysis and comparison of the acoustic features associated

3.3 XGBOOST

XGBoost or Extreme Gradient Boost is designed for high mobility and mechanical use. Gradient learning algorithms solve data science problems quickly and the exact way. This is an extension of gradient boosted decision trees and is much faster than GBM. It Supports regularized learning, increasing, and decreasing gradient trees, Column subsampling, Much faster than other algorithms. The performance of the model is also much better than other models used for surveillance and monitoring. forecast.

The algorithm works based on decision trees and creates a diagram after checking various if statements. The algorithm adds more and more ifs to the tree to obtain a more powerful model construction. XGBoost is an ensemble learning method based on boosting. This cheering up the trees, each subsequent

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 9| Issue: 10| October 2023|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2023: 8.224 || ISI Value: 1.188

tree is arranged in a way that aims to alleviate the shortcomings of the previous tree wood. Each tree learns from its predecessors and checks for any remaining errors. Therefore, the next tree to grow will learn from a revised version of the fossil.

3.4 Train Test Split

For effective model calculation in machine learning, training and create algorithms to help predict data. The data provided is typically categorized and converted into a dataset and reused for training and testing purposes. This method is used to measure the overall performance of ML. Algorithms can be used to infer unspecified facts and teach models. It is a quick and easy way to test ML performance with results, adapt algorithms to the complexity of predictive modeling. Easy to use and translate even if you have a mild illness, you may not be able to receive treatment at this time. Databases and situations where additional configuration is required even when used classes and databases are heterogeneous.

The model was first included in the training data. The model was trained using supervised learning methods. The current model or model is the development used with a set of training data and results are generated based on the results. You can predict whether your model can predict prices correctly. Embedded models help predict validation datasets that provide: Evaluate the model unbiasedly at the end of the dataset, providing unbiased data.

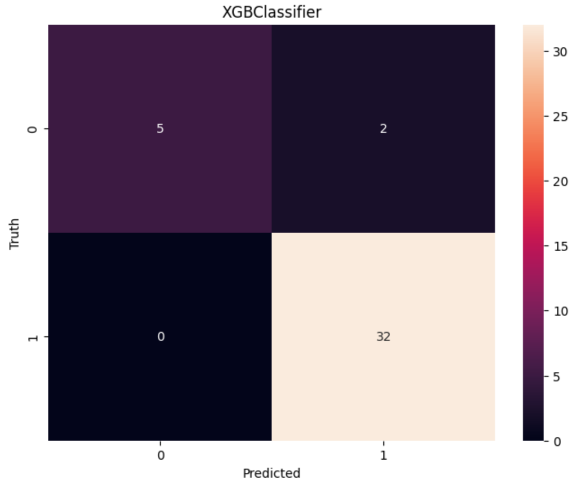
Evaluation of the final model on the training data set. The training model and test data set are used for testing purposes. The test is 80% tensile and 20% inspection.

4.RESULTS

Using XG booster for that first, installed XG booster then imported it and we have fitted X train and Y train values in our model after that we have predicted values and calculated need accuracy with XG booster we have got an accuracy of 94.8%.

```
Accuracy 94.87179487179486
Precision 94.11764705882352
Recall 100.0
F1 96.96969696969697
```

After that we have constructed a confusion matrix for the same model mentioned above.



5.CONCLUSION

This research work developed a Parkinson's Predictor. The work Initially started with a database of human voice parameters of 195 different people; 147 persons were suffering from the disease. The system develops a model with the highest accuracy and with greater true positives and true negatives.

The system used XGBooster algorithm, and achieved 94% accuracy, 94% precision, 100% recall and 96% f1 score. The confusion matrix values are as follows:
TP=5, TN=32, FN=0, FP=2.

6.FUTURE ENHANCEMENT

While achieving a 94% prediction accuracy in our research project, potential for enhanced efficiency and accuracy when considering the possibility of a larger and more complex dataset. To further advance the project, increasing the dataset size can pave the way for the implementation of various advanced deep learning techniques, with the goal of achieving 100% accuracy.

To achieve this objective, it is crucial to engage in several case studies that explore innovative strategies for data handling and model creation. These case studies serve as essential tools to understand how the data set works and what are the liabilities, for better predictions.

7.REFERENCES

1. *Vallejo, M., Jamieson, S., Cosgrove, J., Smith, S.L., Lones, M.A., Alty, J.E. and Corne, D.W., 2016, December. Exploring diagnostic models of Parkinson's disease with multi-objective regression. In Computational Intelligence (SSCI), 2016 IEEE Symposium Series on (pp. 1-8). IEEE*
2. *Abujrida H., Agu E., Pahlavan K. (2017). Smartphone-based gait assessment to infer Parkinson's disease severity using crowdsourced data, in 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT) (Bethesda, MD:), 208–211. 10.1109/HIC.2017.8227621 [CrossRef] [Google Scholar]*
3. *Alam M. N., Garg A., Munia T. T. K., Fazel-Rezai R., Tavakolian K. (2017). Vertical ground reaction force marker for Parkinson's disease. PLoS ONE 12: e0175951. 10.1371/journal.pone.0175951 [PMC free article]*

[PubMed] [CrossRef] [Google Scholar].

4. *Pereira, C.R., Pereira, D.R., Papa, J.P., Rosa, G.H. and Yang, X.S., 2017. Convolutional neural networks applied for Parkinson's disease identification. In Machine Learning for Health Informatics (pp. 377-390). Springer, Cham.*
5. *Lones, M.A., Alty, J.E., Cosgrove, J., Duggan-Carter, P., Jamieson, S., Naylor, R.F., Turner, A.J. and Smith, S.L., 2017. A New Evolutionary Algorithm-Based Home Monitoring Device for Parkinson's Dyskinesia. Journal of medical systems, 41(11), p.176.*
6. *Alaskar H., Hussain A. (2018). Prediction of Parkinson disease using gait signals, in 2018 11th International Conference on Developments in eSystems Engineering (DeSE) (Cambridge:), 23–26. 10.1109/DeSE.2018.00011 [CrossRef] [Google Scholar]*
7. *Adeli E., Thung K.-H., An L., Wu G., Shi F., Wang T., et al. (2019). Semi-supervised discriminative classification robust to sample-outliers and feature-noises. IEEE Trans. Pattern Anal. Mach. Intell. 41, 515–522. 10.1109/TPAMI.2018.2794470 [PMC free article] [PubMed] [CrossRef] [Google Scholar]*
8. *Aich S., Kim H., younga K., Hui K. L., Al-Absi A. A., Sain M. (2019). A supervised machine learning approach using different feature selection techniques on voice 70 datasets for prediction of Parkinson's disease, in 2019 21st International Conference on Advanced Communication Technology (ICACT) (PyeongChang:), 1116–1121.10.23919/ICACT.2019.8701961 [CrossRef] [Google Scholar]*
9. *Alharthi A. S., Ozanyan K. B. (2019). Deep learning for ground reaction force data analysis: application to wide-area floor sensing, in 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE) (Vancouver, BC:),, 1401–1406. 10.1109/ISIE.2019.8781511 [CrossRef] [Google Scholar]*
10. *Ali L., Khan S. U., Arshad M., Ali S., Anwar M. (2019a). A multi-model framework for evaluating type of speech samples having complementary information about Parkinson's disease, in 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (Swat:), 1–5. 10.1109/ICECCE47252.2019.8940696[CrossRef] [Google Scholar]*
11. *Z. Karapinar Senturk, "Early diagnosis of Parkinson's disease using machine 68 learning algorithms," Medical Hypotheses, vol. 138, May 2020, doi: 10.1016/j.mehy.2020.109603.*