



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 6, June 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Parkinson Disease Detection using Fine Tree Approach

Isaac Angelo M. Dioses, James Bryan Tababa, Arnel C. Fajardo, Jesusimo L. Dioses Jr.

School of Information Technology, MAPUA University, Makati Campus, Philippines

College of Computing Studies, Information and Communication Technology, Isabela State University, Cauayan
Isabela, Philippines

College of Computing Studies, Information and Communication Technology, Isabela State University, Cauayan
Isabela, Philippines

College of Computing Studies, Information and Communication Technology, Isabela State University, Echague
Isabela, Philippines

ABSTRACT: Parkinson's disease (PD) is a prevalent neurodegenerative disorder characterized by motor dysfunctions such as difficulty in sitting, standing, walking, and maintaining balance, particularly in its advanced stages. Early detection through clinical evaluation and functional testing is crucial for effective intervention and rehabilitation. The disease affects over 1% of people above 60 and is associated with Lewy bodies and degeneration of dopaminergic neurons in the substantia nigra. While most cases are idiopathic, about 10% have genetic origins, usually affecting younger individuals. Diagnosis typically begins with a thorough history and physical examination, often utilizing the Unified Parkinson Disease Rating Scale to assess various aspects of the patient's condition. Electroencephalograms (EEGs) provide valuable insights into brain activity and can support research and healthcare through brain-computer interface (BCI) systems. These systems leverage machine learning algorithms to monitor and analyze cognitive states, offering high temporal resolution measurements of brain electrical activity. In addition to EEG, spectral analysis of voice has proven effective in identifying anomalies in speech characteristics among PD patients, such as decreased fundamental frequency, altered harmonics-to-noise ratio, and increased jitter and shimmer. Given the complexity and high-dimensional nature of voice data, advanced machine learning methods can significantly enhance the accuracy of classifying voice samples for PD detection. This study successfully developed a machine learning model that achieved 90% accuracy in predicting Parkinson's disease based on voice analysis. The model's performance was evaluated using a confusion matrix, revealing 41 true positives, 13 false positives, 12 false negatives, and 183 true negatives. The model's ROC curve showed an Area Under the Curve (AUC) of 86.11% for each class. Local interpretable model-agnostic explanations (LIME) were used to interpret the model's predictions, identifying eight features that negatively impacted accuracy.

KEYWORDS: Parkinsons Disease, Machine Learning, MATLAB, Fine Tree, Confusion Matrix

I. INTRODUCTION

Among the most common neurodegenerative illnesses is Parkinson's disease. In its most severe phases, Parkinson's disease (PD) causes motor dysfunction that makes it difficult to perform fundamental daily tasks like sitting, standing, walking, and balance and it is a common neurodegenerative condition that progresses [1,2]. Healthcare professionals can appropriately intervene in rehabilitation plans when Parkinson's disease (PD) is detected early through clinical evaluation and functional testing [3]. Over the age of 60, at least 1% of people are thought to have Parkinson disease. Lewy bodies and the degeneration of dopaminergic neurons in the substantia nigra are linked to the illness. The majority are idiopathic. Genetic causes account for only 10% of instances, and they are typically young people [4].

Patients with Parkinson's disease (PD) is often evaluated beginning with a history and physical examination designed to identify the indicators mentioned above. The Unified Parkinson Disease Rating Scale can be used by movement disorders clinics to measure a patient's mentation, behavior, mood, activities of daily living, tremors, motor exams, and therapy-related problems [5,6].

An electroencephalogram can be used to show the average activity of cortical neurons throughout the cortex (EEG). EEG signals can be thought of as brain-computer interface systems (BCI) since intelligent BCI based on EEG allows

continuous monitoring of variations in human cognitive states and is useful for research in many domains and healthcare support. Using machine learning algorithms, this system detects how the human brain may connect with the surroundings and cutting-edge technology [7]. EEG signals are the most widely utilized non-invasive modality for both clinical and real-world BCIs, as they provide a high temporal resolution direct measurement of brain electrical activity [8,9]

But there is a Spectral analysis approach to examine the voices of patients with Parkinson Disease [10]. Using spectral analysis, it was possible to identify various anomalies in particular speech characteristics in PD patients, including decreased fundamental frequency and harmonics-to-noise ratio as well as increased jitter and shimmer [11,12,13,14,15,22]. On the other hand, the human voice is a complicated phenomenon with high-dimensional data based on an exponentially large number of features. Therefore, in addition to the independent analysis using spectral analysis of voice features (i.e., fundamental frequency), more sophisticated methods capable of analyzing and dynamically combining high-dimensional datasets of voice features, like machine-learning algorithms, would greatly increase the accuracy of the objective classification of voice samples in Parkinson's disease [16-21]. In fact, machine learning has made it possible to classify voice impairment in a variety of neurologic illnesses objectively and automatically—with hitherto unknown high accuracy.

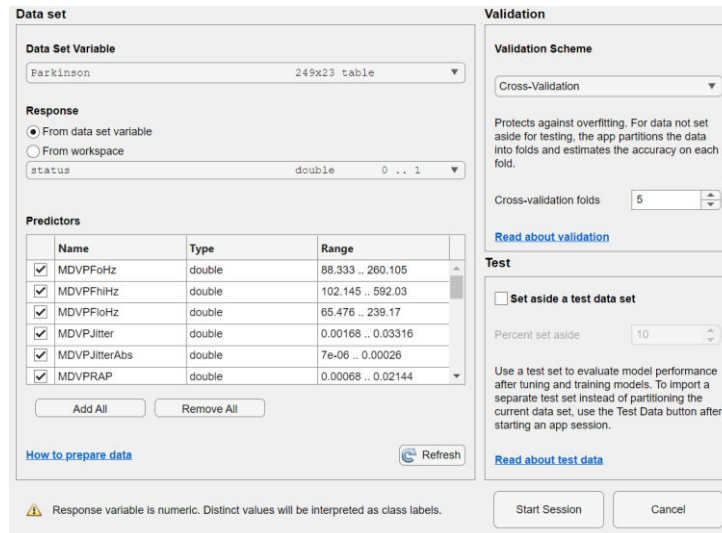
II. METHODOLOGY

The voice dataset was taken from Kaggle with 22 features in CSV files, it consist of voice recording frequencies of healthy people and people with Parkinson disease. These are the features and descriptions: MDVP:Fo(Hz) - Mean fundamental frequency of voice, Maximum vocal fundamental frequency (MDVP:Fhi(Hz)), MDVP:Flo(Hz) - The lowest fundamental frequency of voice, MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:PPQ,Jitter:DDP, and MDVP:RAP - A several measures of variation in fundamental frequency, Multiple amplitude variation measures: MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA NHR and HNR are two measurements of the voice's noise to tone component ratio. status - The subject's health status (zero) - healthy, (one) - Parkinson's Two measurements of nonlinear dynamical complexity are RPDE and D2. DFA: Fractal scaling exponent of a signal Three nonlinear metrics of fundamental frequency variation are spread1, spread2, and PPE. After the data gathering, the data was imported in MATLAB 2023b for classification. In (a), shows some of dataset imported in MATLAB.

Parkinson																					
MDVPFoHz	MDVPFhi...	MDVPFlo...	MDVPJitter	MDVPJitt...	MDVPRAP	MDVPPPQ	JitterDDP	MDVPShi...	MDVPShi...	Shimmer...	Shimmer...	MDVPAPQ	Shimmer...	NHR	HNR	RPDE	DFA	sprea			
Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number			
1	119.992	157.302	74.997	0.00784	0.00007	0.00554	0.01109	0.04374	0.426	0.02182	0.0313	0.02971	0.06545	0.02211	21.033	0.414783	0.815285	-4.8130			
2	119.992	157.302	74.997	0.00784	0.00007	0.00554	0.01109	0.04374	0.426	0.02182	0.0313	0.02971	0.06545	0.02211	21.033	0.414783	0.815285	-4.8130			
3	122.4	148.65	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.262	0.03134	0.04518	0.04368	0.09403	0.01929	19.085	0.458359	0.819521	-4.0751			
4	116.682	131.111	111.555	0.0105	0.00009	0.00544	0.00781	0.01633	0.05233	0.482	0.02757	0.03858	0.0359	0.0827	0.01309	20.651	0.429895	0.825288	-4.4431		
5	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	0.517	0.02924	0.04005	0.03772	0.08771	0.01353	20.644	0.434969	0.819235	-4.1175		
6	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	0.584	0.0349	0.04825	0.04465	0.1047	0.01767	19.649	0.417356	0.823484	-3.7477		
7	120.552	131.162	113.787	0.00968	0.00008	0.00463	0.0075	0.01388	0.04701	0.456	0.02328	0.03243	0.06985	0.01222	21.378	0.417356	0.823484	-3.7477			
8	120.267	137.244	114.82	0.00333	0.00003	0.00155	0.00202	0.00466	0.01608	0.14	0.00779	0.00937	0.01351	0.02337	0.00607	24.886	0.59604	0.764112	-5.6343		
9	107.332	113.84	104.315	0.0029	0.00003	0.00144	0.00182	0.00431	0.01567	0.134	0.00829	0.00946	0.01256	0.02487	0.00344	26.892	0.63742	0.763262	-6.1676		
10	95.73	132.068	91.754	0.00551	0.00006	0.00293	0.00332	0.0088	0.02093	0.191	0.01073	0.01277	0.01717	0.0107	21.812	0.615551	0.773587	-5.4986			
11	95.056	120.103	91.226	0.00532	0.00006	0.00268	0.00332	0.00803	0.02838	0.255	0.01441	0.01725	0.02444	0.04324	0.01022	21.862	0.547037	0.798463	-5.0118		
12	88.333	112.24	84.072	0.00505	0.00006	0.00254	0.0033	0.00763	0.02143	0.197	0.01079	0.01342	0.01892	0.03237	0.01166	21.118	0.611137	0.776156	-5.2497		
13	91.904	115.871	86.292	0.0054	0.00006	0.00281	0.00336	0.00844	0.02752	0.249	0.01424	0.01641	0.02214	0.04272	0.01141	21.414	0.58339	0.79252	-4.9602		
14	136.926	159.866	131.276	0.00293	0.00002	0.00118	0.00153	0.00355	0.01259	0.112	0.00656	0.00717	0.0114	0.01968	0.00581	25.703	0.4606	0.646846	-6.5471		
15	139.173	179.139	76.556	0.0039	0.00003	0.00165	0.00208	0.00496	0.01642	0.154	0.00728	0.00932	0.01797	0.02184	0.01041	24.889	0.430166	0.665833	-5.6602		
16	152.845	163.305	75.836	0.00294	0.00002	0.00121	0.00149	0.00364	0.01828	0.158	0.01064	0.00972	0.01246	0.03191	0.00609	24.922	0.474791	0.654027	-6.1050		
17	142.167	217.455	83.159	0.00369	0.00003	0.00157	0.00203	0.00471	0.01503	0.126	0.00772	0.00888	0.01359	0.02316	0.00839	25.175	0.565924	0.658245	-5.3401		
18	144.188	349.259	82.764	0.00544	0.00004	0.00211	0.00292	0.00632	0.02047	0.192	0.00969	0.012	0.02074	0.02908	0.01859	22.333	0.56738	0.644692	-5.4400		
19	168.778	232.181	75.603	0.00718	0.00004	0.00284	0.00387	0.00853	0.03327	0.348	0.01441	0.01893	0.0343	0.04322	0.02919	20.376	0.631099	0.605417	-2.9310		
20	153.046	175.829	68.623	0.00742	0.00005	0.00364	0.00432	0.01092	0.05517	0.542	0.02471	0.03572	0.05767	0.07413	0.0316	17.28	0.665318	0.719467	-3.9490		
21	156.405	189.398	142.822	0.00768	0.00005	0.00372	0.00399	0.01116	0.03995	0.348	0.01721	0.02374	0.0431	0.05164	0.03365	17.153	0.649554	0.68608	-4.5544		
22	153.848	165.738	65.782	0.0084	0.00005	0.00428	0.0045	0.01285	0.0381	0.328	0.01667	0.02383	0.04055	0.05	0.03871	17.536	0.660125	0.704087	-4.0954		

(a). Parkinson Dataset

After importation of the datasets, the text data were run in MATLAB classification learner for machine learning process. And during the pre-processing, the cross-fold validation of 5 folds was applied for better accuracy validation measures as shown in figure (b). Also, in Figure (c), it shows the model's hyperparameters. The preset is Fine Tree, 100 splits, and Split Criterion is Gini's index without surrogate decision splits. After the pre-processing, the model is ready to train



(b). Dataset Importation

▼ Model Hyperparameters

Preset: Fine Tree

Maximum number of splits: 100

Split criterion: Gini's diversity index

Surrogate decision splits: Off

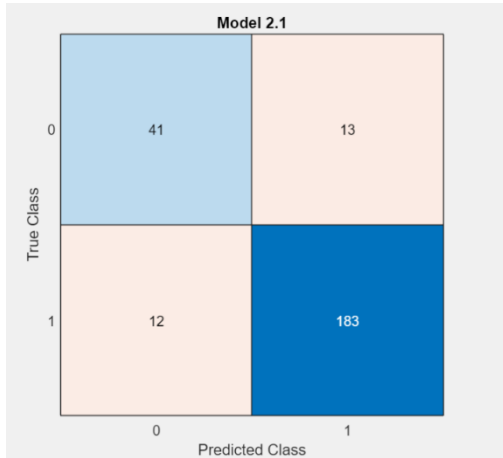
(c) Model Hyperparameters

III. EXPERIMENTAL RESULTS

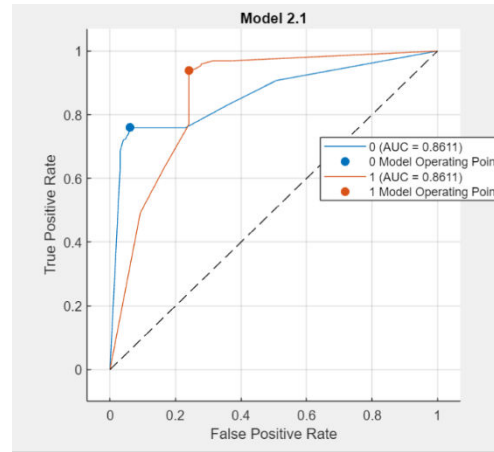
After the training of the model, it appears that the model performs 90% accuracy in predicting Parkinson's Disease through voice. In Figure (d) shows the confusion matrix of the model. Where 41 are True Positive, 13 are false Positive, 12 are False Negatives, and 183 are True Negative achieving its 90% accuracy. In Table 1, displays the accuracy formula of the model. In Figure (e). It shows the AUC or Area Under Curve of the model. The model's ROC curve is shown, illustrating how well it performs at various threshold settings. One of the most important metrics for assessing the model's capacity to discriminate across classes is the Area Under the ROC Curve (AUC). The ROC Curve of this model is 86.11% for each class.

ACCURACY	TRUE POSITIVE + TRUE NEGATIVE
	(TP + TN+ FP + FN)

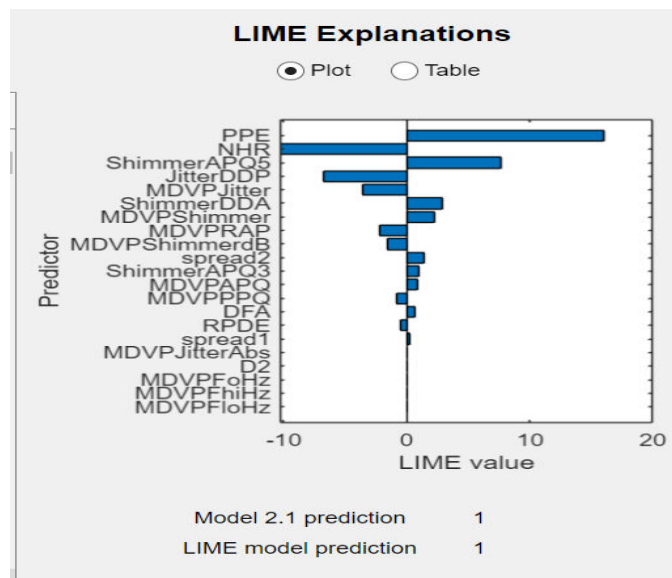
Table 1. Accuracy Formula



(d). Confusion Matrix



(e). ROC AUC Curve



(f). Lime Explanation

By fitting a basic interpretable model to the query point, local interpretable model-agnostic explanations (LIME [1]) interpret a prediction for the query point. Around the query point, the simple model explains model predictions and serves as an approximation for the trained model. In Figure (f) it shows the LIME value and LIME Explanation plot of the model, it shows that there are 8 Features or predictors that has negative values meaning that those features could affect the accuracy of the model. Overall, in this study it is successfully created a model that can detect or predict Parkinson Disease based on voice using Machine Learning Technique.

IV. CONCLUSION

Parkinson's disease is one of the most prevalent neurological diseases. Parkinson's disease (PD) is a common neurological condition that worsens over time and causes motor dysfunction in its most severe phases, making it difficult to perform basic daily tasks like sitting, standing, walking, and balance. Indeed, speech impairment in a range of neurologic diseases may now be objectively and automatically classified—with previously unheard-of high accuracy—thanks to machine learning. In this study, researchers successfully created a Machine Learning Model that can predict Parkinson Disease with 90% accuracy using Fine Tree Model. Also in this study it shows the LIME and

feature importance of each predictors suggesting that in future researches, researchers may remove those features resulting in better accuracy.

REFERENCES

- [1] A. Ascherio and M. A. Schwarzschild, "The epidemiology of Parkinson's disease: risk factors and prevention," **Lancet Neurology**, vol. 15, no. 12, pp. 1257–1272, Nov. 2016, Available: [https://doi.org/10.1016/s1474-4422\(16\)30230-7](https://doi.org/10.1016/s1474-4422(16)30230-7).
- [2] D. K. Simon, C. M. Tanner, and P. Brundin, "Parkinson Disease Epidemiology, Pathology, Genetics, and Pathophysiology," **Clinics in Geriatric Medicine**, vol. 36, no. 1, pp. 1–12, Feb. 2020, Available: <https://doi.org/10.1016/j.cger.2019.08.002>.
- [3] A. Ascherio and M. A. Schwarzschild, "The epidemiology of Parkinson's disease: risk factors and prevention," **Lancet Neurology**, vol. 15, no. 12, pp. 1257–1272, Nov. 2016, Available: [https://doi.org/10.1016/s1474-4422\(16\)30230-7](https://doi.org/10.1016/s1474-4422(16)30230-7).
- [4] S. Zafar and S. S. Yaddanapudi, "Parkinson Disease," **StatPearls - NCBI Bookshelf**, 2023. Available: <https://www.ncbi.nlm.nih.gov/books/NBK470193/>.
- [5] D. Berg et al., "Movement disorder society criteria for clinically established early Parkinson's disease," **Movement Disorders**, vol. 33, no. 10, pp. 1643–1646, Aug. 2018, Available: <https://pubmed.ncbi.nlm.nih.gov/30145841/>.
- [6] M. W. Creaby and M. H. Cole, "Gait characteristics and falls in Parkinson's disease: A systematic review and meta-analysis," **Parkinsonism & Related Disorders**, vol. 57, pp. 1–8, Dec. 2018, Available: <https://pubmed.ncbi.nlm.nih.gov/30041848/>.
- [7] P. L. Nunez, B. M. Wingeier, and R. B. Silberstein, "Spatial-temporal structures of human alpha rhythms: Theory, microcurrent sources, multiscale measurements, and global binding of local networks," **Human Brain Mapping**, vol. 13, no. 3, pp. 125–164, May 2001, Available: <https://doi.org/10.1002/hbm.1030>.
- [8] R. A. Ramadan and A. V. Vasilakos, "Brain computer interface: control signals review," **Neurocomputing**, vol. 223, pp. 26–44, Feb. 2017, Available: <https://doi.org/10.1016/j.neucom.2016.10.024>.
- [9] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A General-Purpose Brain-Computer Interface (BCI) system," **IEEE Transactions on Bio-medical Engineering**, vol. 51, no. 6, pp. 1034–1043, Jun. 2004, Available: <https://doi.org/10.1109/tbme.2004.827072>.
- [10] J. Ruzs, T. Tykalova, L. O. Ramig, and E. Tripoliti, "Guidelines for Speech recording and Acoustic Analyses in Dysarthrias of Movement Disorders," **Movement Disorders**, vol. 36, no. 4, pp. 803–814, Dec. 2020, Available: <https://doi.org/10.1002/mds.28465>.
- [11] J. Ruzs, R. Cmejla, H. Ruzickova, and E. Ruzicka, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease," **Journal of the Acoustical Society of America**, vol. 129, no. 1, pp. 350–367, Jan. 2011, Available: <https://doi.org/10.1121/1.3514381>.
- [12] J. Gamboa et al., "Acoustic voice analysis in patients with Parkinson's disease treated with dopaminergic drugs," **Journal of Voice**, vol. 11, no. 3, pp. 314–320, Sep. 1997, Available: [https://doi.org/10.1016/s0892-1997\(97\)80010-0](https://doi.org/10.1016/s0892-1997(97)80010-0).
- [13] J. Ruzs et al., "Effects of dopaminergic replacement therapy on motor speech disorders in Parkinson's disease: longitudinal follow-up study on previously untreated patients," **Journal of Neural Transmission**, vol. 123, no. 4, pp. 379–387, Feb. 2016, Available: <https://doi.org/10.1007/s00702-016-1515-8>.
- [14] J. Ruzs et al., "Evaluation of speech impairment in early stages of Parkinson's disease: a prospective study with the role of pharmacotherapy," **Journal of Neural Transmission**, vol. 120, no. 2, pp. 319–329, Jul. 2012, Available: <https://doi.org/10.1007/s00702-012-0853-4>.
- [15] Y. Tanaka, M. Nishio, and S. Niimi, "Vocal Acoustic Characteristics of Patients with Parkinson's Disease," **Folia Phoniatica Et Logopaedica**, vol. 63, no. 5, pp. 223–230, Jan. 2011, Available: <https://karger.com/fpl/article-abstract/63/5/223/140688/Vocal-Acoustic-Characteristics-of-Patients-with?redirectedFrom=fulltext>.
- [16] M. A. T. Vu et al., "A shared vision for machine learning in neuroscience," **Journal of Neuroscience**, vol. 38, no. 7, pp. 1601–1607, Jan. 2018, Available: <https://www.jneurosci.org/content/38/7/1601>.
- [17] A. Suppa et al., "Voice Analysis with Machine Learning: One Step Closer to an Objective Diagnosis of Essential Tremor," **Movement Disorders**, vol. 36, no. 6, pp. 1401–1410, Feb. 2021, Available: <https://doi.org/10.1002/mds.28508>.

- [18] S. Hegde, S. Shetty, S. Rai, and T. Dodderi, "A survey on Machine Learning Approaches for Automatic Detection of Voice Disorders," **Journal of Voice**, vol. 33, no. 6, pp. 947.e11-947.e33, Nov. 2019, Available: <https://doi.org/10.1016/j.jvoice.2018.07.014>.
- [19] F. Asci et al., "Machine-Learning Analysis of Voice Samples Recorded through Smartphones: The Combined Effect of Ageing and Gender," **Sensors**, vol. 20, no. 18, p. 5022, Sep. 2020, Available: <https://www.mdpi.com/1424-8220/20/18/5022>.
- [20] F. Asci, G. Costantini, P. Di Leo, G. Saggio, and A. Suppa, "Reply to: "Reproducibility of Voice Analysis with Machine Learning,"" **Movement Disorders**, vol. 36, no. 5, pp. 1283–1284, May 2021, Available: <https://doi.org/10.1002/mds.28601>.
- [21] F. Asci, G. Costantini, G. Saggio, and A. Suppa, "Fostering Voice Objective Analysis in Patients with Movement Disorders," **Movement Disorders**, vol. 36, no. 4, p. 1041, Apr. 2021, Available: <https://doi.org/10.1002/mds.28537>
- [22] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," **BioMedical Engineering Online**, vol. 6, no. 1, p. 23, Jan. 2007, Available: <https://doi.org/10.1186/1475-925x-6-23>.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details