Parkinson's Disease Detection By MachineLearning Using SVM

Doneti Sowmya[1], Dodla Kavya[2], J. Rashmitha[3], V. Satheesh kumar[4], Preethi Jeevan[5]

[4,5] Professor, Dept. of Computer Science and Engineering,SNIST,Hyderabad-501301,India
[1,2,3] B.TECH Scholars, Dept. of Computer Science and Engineering Hyderabad-501301,India

---***---

Abstract:-

Parkinson's disease (PD) is a neuro degenerative movement disorder in which the signs initially appear as a mild tremor in one hand and a general sense of stiffness, but gradually get worse. Over 6 million people are impacted globally. Currently, non-specialist clinicians have not been able to definitively diagnose this disease, especially in the early stages of the illness when it is very challenging to identify the symptoms. The issue can be resolved with a low mistake rate by utilising machine learning techniques. As a result, data mining offers a prediction method for systematically identifying Parkinson's disease. Parkinsonism cannot be diagnosed using a traditional test, hence we suggest a statistical method based on the most prevalent PD symptoms, Given that there is no established test to identify Parkinsonism, we suggest a statistical method based on the three most prevalent PD symptoms—gait, tremors, and micrographia. In order to determine the classification method that provides the maximum accuracy in diagnosing PD patients, this involves studying the correlation between the symptoms and classifying the obtained data using several classification algorithms. By combining inputs from Parkinson's sufferers and healthy individuals, our suggested method produces reliable results for the data sets obtained as the input. Our work will demonstrate how early disease identification can extend a patient's life and lead to a serene existence through appropriate medical care and medication.

Keywords: *Machine learning; SVM; Parkinson's disease; Gait analysis.*

INTRODUCTION:

Parkinson's disease is characterized by the death of dopaminergic neurons in the substantia nigra pars compacta of the midbrain. Coordination problems, bradykinesia and voice alterations are among the signs of this neurodegenerative disease. Parkinson's disease (PD) patients can also develop dysarthria, an impairment of the motor-speech system that affects respiratory, phonatory, articulatory and prosodic functions. Parkinson's disease symptoms can be different for everyone. Early signs are mild that goes unnoticed. Symptoms usually begin on one side of your body and gets worsen on that side, afterwards it affects both the sides. Parkinson's symptoms may include:

- Tremor
- Slowed movement
- Rigid muscles.
- Impaired posture and balance.
- Loss of automatic movements
- Speech changes
- Writing changes

The cause of Parkinson's disease is still a question mark,but several factors appear to play a role, including:

- Genes
- Environmental
- Triggers

BACKGROUND STUDY (LITERATURE)

Parkinson's illness is generally incurable, however Medication can frequently help control the symptoms. dramatically. Surgery may be required in some more severe situations. Additionally, yourhealthcare physician could counsel lifestyle adjustments, particularly consistent aerobic activity. In a few instances, physical treatment that emphasises balance and Exercising is crucial. A therapist for speech-language disorders help with speech issues.

Medicines may be able to assist you control tremor, mobility, and walking issues. Dopamine-boosting or dopamine- substituting drugs. those with Brain dopamine levels are low in those with Parkinson's disease. However, because it cannot enter the body directly, dopamine the mind. Your performance could significantly improve. symptoms following the start of treatment for Parkinson's disease. However, over time, the advantages of medications frequently decreasing or losing consistency .Usually, you still have good control over your symptoms. The following are current methods for identifying Parkinson's disease:

- PET scans are utilised to evaluate the function and activity of the brain's movement-related areas.

- SPECT scans are able to spot alterations in brain chemistry, including a drop in dopamine.

- MRI or CT scan - Conventional MRI is unable to identify Parkinson's disease's early symptoms.

- Parkinson's disease can be detected using voice and speech data.

These procedures result in unwelcome bias (up to 25%), mistakes, high medical expenses, and late PD diagnosis, which may have an impact on the patient's quality of life.

METHODOLOGY:

By using machine learning techniques, the problem can be solved with minimal error rate. Parkinson's disease detection using gait, tremors and handwriting samples as the dataset, in order to increase the accuracy by finding the co-relation between these symptoms. Since individual analysis of every symptom has some drawback attached to it, for examplehandwriting is a complex activity where other factors can influence motor movement, in speech recognition additional steps such as noise removal and speech segmentation are required, using breath samples has been proved to fail to meet clinically relevant results. Thus, in order to avoid the above problems, we have included multiple symptoms rather than relying on oneof them.

Our proposed system provides accurate results by integrating the input data of healthy and Pd affected patients. Thus, by the results, doctor can conclude normally or abnormality and prescribe the medicine based on the affected stage. It saves time and helps in early detection. No need to spend a lot over hospitals. This study represents the experimental process (figure 3.1) of the experiment, including machine learning techniques. Parkinson's Disease data sets have been considered in this work. Firstly, wefocused on preparing and combining data from the main datasets. 30 characteristics were also taken out of the Parkinson datasets.

Next, we looked at the co-related and missing values. Second, a key task in this machine learning-based industry is data set separation. We were unable to locate split and test datasets in this dataset. The Parkinson data set has been divided into train set and test sets, as shown in Figure 3.1. Three supervised- based classifiers then went into operation. After these algorithms were successfully implemented, SVM showed the best performance.

Keep meticulous records of attendance. If no match is detected, the system will simply resume looking for the next visitor, suggesting that the person's photo has not yet been entered into the system.

Until the system is shut off, this continues. By visiting the configured home page with proper credentialsAdmin can view the attendance and details of employee or student. We update the attendance in real time into a local excel sheet that can be read by admin.

ALGORITHMS:

4.1 Support Vector Algorithm:

Support vector machines have been first introduced by Vladimir Vapnik and Alexey Chervonenkis (Chervonenk is, 2013)(Vapnik, Guyon, Learn, & 1995, n.d.). SVM is a method of machine learning that can solve both linear and nonlinear problems. It provides good performance to solve both regression and classification problems. The SVM classification technique inspects for the optimal separable hyperplane to classify the dataset between two classes (Smola & Schölkopf, 2004). Finally, the model can estimate noisy data problems for new cases.

A Support Vector Machine is a supervised learning algorithm. An SVM models the data into k categories, performing classification and forming an N- dimensional hyperplane. These models are very similar to neural networks. Consider a dataset of N dimensions. The SVM plots the training data into an N-dimensioned space. Following that, using hyper- planes with n various dimensions, the training data points are split into k distinct regions based on their labels. The test points are shown in the same N- dimensional plane following the testing phase. The points are correctly classified in the respective regionin which they are placed. Divided into a train and test set, our dataset was then fitted to the SVM model as seen below.

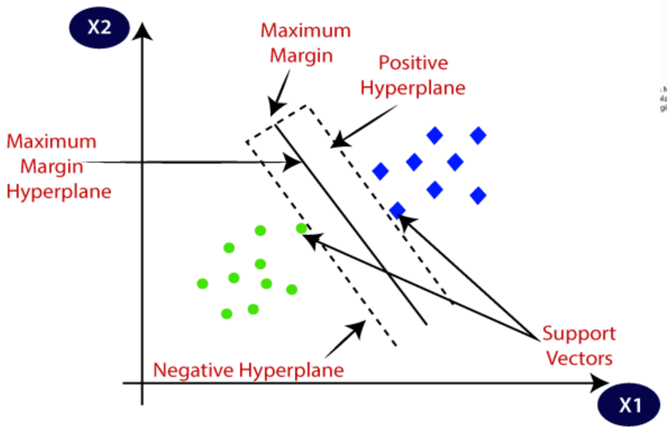
X_train X_test, Y_train, Y_test = train_test_split(X_all, y_all) clf = svm.LinearSVC() clf.fit(X_train,Y_train) pred = clf.predict(X_test)

Result is stored in csv file by using below code result2=open("Output/resultSVM.csv","w") result2.write("ID,Predicted Value" + "\n")

for j in range(len(pred)): result2.write(str(j+1) + ","

+str(pred[j]) + "\n") result2.close();

Support vector machines (SVMs) are regarded as effective learning techniques and are frequently used to issues in biomedical and health informatics [49]. An SVM model's output after training is an ideal hyperplane that can increase the distance between any class and the closest training data points. The following are the main factors that drive machine learning researchers to utilise SVM for their issues. (1) The first justification is that SVMs are very good at generalising to new data. (2) SVMs' reliance on a relatively small set of hyperparameters is the second factor.



IMPLIMENTATION:

Any project's implementation phase is a real showcase for the turning points that determine whether it will succeed or fail. The installation and operationalization of the system or system modifications in a production environment is referred to as the implementation step. The Support Vector Machine (SVM) technique is employed to create the prediction system, and Python is the language used.

Libraries / Algorithms Used**:**

➢ NUMPY: which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays.

➢ PANDAS: is a Python library. Pandas is used to analyze data, Learning by Reading.

➢ SK-learn: Scikit-learn is a free machine learning library for Python. It features various algorithms like support vector machine, random forests, and k-neighbours, and it also supports Python numerical and scientific libraries like NumPy and SciPy .

DATASET:

Number of Instances: 197 Number of Attributes: 23 Missing Values? N/A

Source: Max Little of the University of Oxford generated the dataset in association with the National Centre for Voice and Voice in Denver, Colorado, which captured the speech signals. The feature extraction techniques for general voice problems were published in the original paper.

Data Set Information: This dataset includes various biological voice measurements taken from 31individuals, 23 of whom have Parkinson's disease (PD). Each row in the table corresponds to one of the

195 voice recordings from these people, and each column in the table represents a specific voice measure ("name" column). According to the "status" column, which is set to 0 for healthy and 1 for PD, the maingoal of the data is to distinguish between healthy individuals and those with PD. The information is in CSV ASCII format. One instance per voice recordingis present in each row of the CSV file. Each patient has about six recordings, and the first column lists thepatient's name.

MDVP:Shi	Shimmer:/	Shimmer:/	MDVP:AP	Shimmer:I	NHR	HNR	status	RPDE	DFA	spread1	spread2	D2	PPE
0.426	0.02182	0.0313	0.02971	0.06545	0.02211	21.033	1	0.41478	0.81529	-4.81303	0.26648	2.30144	0.28465
0.626	0.03134	0.04518	0.04368	0.09403	0.01929	19.085	1	0.45836	0.81952	-4.07519	0.33559	2.48686	0.36867
0.482	0.02757	0.03858	0.0359	0.0827	0.01309	20.651	1	0.4299	0.82529	-4.44318	0.31117	2.34226	0.33263
0.517	0.02924	0.04005	0.03772	0.08771	0.01353	20.644	1	0.43497	0.81924	-4.1175	0.33415	2.40555	0.36898
0.584	0.0349	0.04825	0.04465	0.1047	0.01767	19.649	1	0.41736	0.82348	-3.74779	0.23451	2.33218	0.41034
0.456	0.02328	0.03526	0.03243	0.06985	0.01222	21.378	1	0.41556	0.82507	-4.24287	0.29911	2.18756	0.35778
0.14	0.00779	0.00937	0.01351	0.02337	0.00607	24.886	1	0.59604	0.76411	-5.63432	0.25768	1.85479	0.21176
0.134	0.00829	0.00946	0.01256	0.02487	0.00344	26.892	1	0.63742	0.76326	-6.1676	0.18372	2.06469	0.16376
0.191	0.01073	0.01277	0.01717	0.03218	0.0107	21.812	1	0.61555	0.77359	-5.49868	0.32777	2.32251	0.23157

PROCEDURE/ALGORITHM:

➢ Start

➢ CREATING A MACHINE LEARNINGMODEL{

IMPORTING DEPENDENCIES;

DATA COLLECTION AND ANALYSIS {

loading the data from csv file to a Pandas DataFrame;

Data Normalization(removing inconsistentvalues);

}

DATA PREPROCESSING{

Separating the features & Target;

Splitting the data to training data & Test data;

}

Data Standardization;Model Training; Model Evaluation{

calculating accuracy of training dataset;calculating accuracy of test dataset;

}

}

➢ BUILDING A PREDICTIVE SYSTEM{

user_input_data(parameters/attributes);

user_input_data to numpy.array; standardization of data:

std_data = scaler.transform(input_data);

#predicting

prediction = predict(std_data)print(prediction);

if prediction is '0' then do

print("The Person does not have Parkinsons Disease");

else do

print("The Person has Parkinsons Disease");

}

➢ End

Results (Accuracy) / Output Screens

1	name	MDVP:Fo(	MDVP:Fhi	MDVP:Flo	MDVP:Jitt	MDVP:Jitt	MDVP:RA	MDVP:PP(	Jitter:DDP	MDVP:Shi
2	phon_R01	119.992	157.302	74.997	0.00784	0.00007	0.0037	0.00554	0.01109	0.04374
3	phon_R01	122.4	148.65	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134
4	phon_R01	116.682	131.111	111.555	0.0105	0.00009	0.00544	0.00781	0.01633	0.05233
5	phon_R01	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492
6	phon_R01	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425
7	phon_R01	120.552	131.162	113.787	0.00968	0.00008	0.00463	0.0075	0.01388	0.04701
8	phon_R01	120.267	137.244	114.82	0.00333	0.00003	0.00155	0.00202	0.00466	0.01608
9	phon_R01	107.332	113.84	104.315	0.0029	0.00003	0.00144	0.00182	0.00431	0.01567
10	phon_R01	95.73	132.068	91.754	0.00551	0.00006	0.00293	0.00332	0.0088	0.02093

```
[1]
diseased
```

```
[0]
No diseased
```

```
[[ 0.63239631 -0.02731081 -0.87985049 ... -0.97586547 -0.55160318
   0.07769494]
 [-1.05512719 -0.83337041 -0.9284778  ...  0.3981808  -0.61014073
   0.39291782]
 [ 0.02996187 -0.29531068 -1.12211107 ... -0.43937044 -0.62849605
  -0.50948408]
 ...
 [-0.9096785  -0.6637302  -0.160638   ...  1.22001022 -0.47404629
  -0.2159482 ]
 [-0.35977689  0.19731822 -0.79063679 ... -0.17896029 -0.47272835
   0.28181221]
 [ 1.01957066  0.19922317 -0.61914972 ... -0.716232    1.23632066
  -0.05829386]]
```

Standardized trained data

```
     MDVP:Fo(Hz)  MDVP:Fhi(Hz)  MDVP:Flo(Hz)  ...   spread2        D2       PPE
0        119.992       157.302        74.997  ...  0.266482  2.301442  0.284654
1        122.400       148.650       113.819  ...  0.335590  2.486855  0.368674
2        116.682       131.111       111.555  ...  0.311173  2.342259  0.332634
3        116.676       137.871       111.366  ...  0.334147  2.405554  0.368975
4        116.014       141.781       110.655  ...  0.234513  2.332180  0.410335
..           ...           ...           ...  ...       ...       ...       ...
190      174.188       230.978        94.261  ...  0.121952  2.657476  0.133050
191      209.516       253.017        89.488  ...  0.129303  2.784312  0.168895
192      174.688       240.005        74.287  ...  0.158453  2.679772  0.131728
193      198.764       396.961        74.904  ...  0.207454  2.138608  0.123306
194      214.289       260.277        77.973  ...  0.190667  2.555477  0.148569

[195 rows x 22 columns]
```

```
MDVP:Jitter(Abs)    0
MDVP:RAP            0
MDVP:PPQ            0
Jitter:DDP          0
MDVP:Shimmer        0
MDVP:Shimmer(dB)    0
Shimmer:APQ3        0
Shimmer:APQ5        0
MDVP:APQ            0
Shimmer:DDA         0
NHR                 0
HNR                 0
status              0
RPDE                0
```

Normalized pd data attributes

Featured data

```
0      1
1      1
2      1
3      1
4      1
      ..
190    0
191    0
192    0
193    0
194    0
Name: status, Length: 195, dtype: int64
```

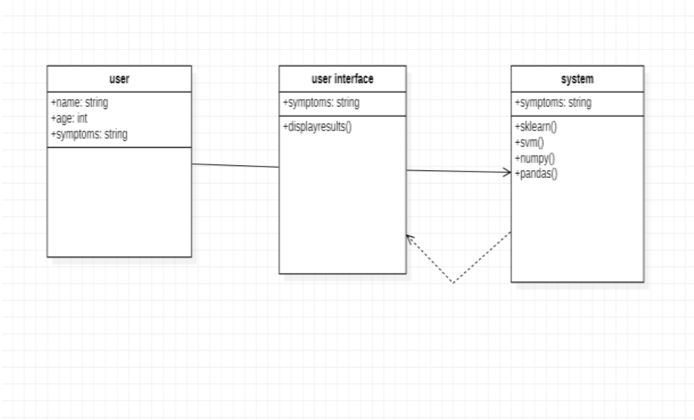
status	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	MDVP:Shimmer(dB)	Shimmer:APQ3	Shimmer:APQ5	MDVP:APQ
0	181.937771	223.636750	145.207292	0.003866	0.000023	0.001925	0.002056	0.005776	0.017615	0.162958	0.009504	0.010509	0.013305
1	145.180762	188.441463	106.893558	0.006989	0.000051	0.003757	0.003900	0.011273	0.033658	0.321204	0.017676	0.020285	0.027600

Target data

UML DIAGRAMS

Class Diagram

The visual depiction of a class object in a model system, arranged according to class types, is called a class diagram. With three compartments for the class name, attributes and operations each class type is shown as a rectangle. Ovals are used to represent objects, and inside each oval are compartments with class names.



Sequence Diagram

The project's sequence diagram for disease prediction using machine learning includes all the different elements that a typical sequence diagram needs. This flow chart demonstrates how the model progresses from the initial step to the last.

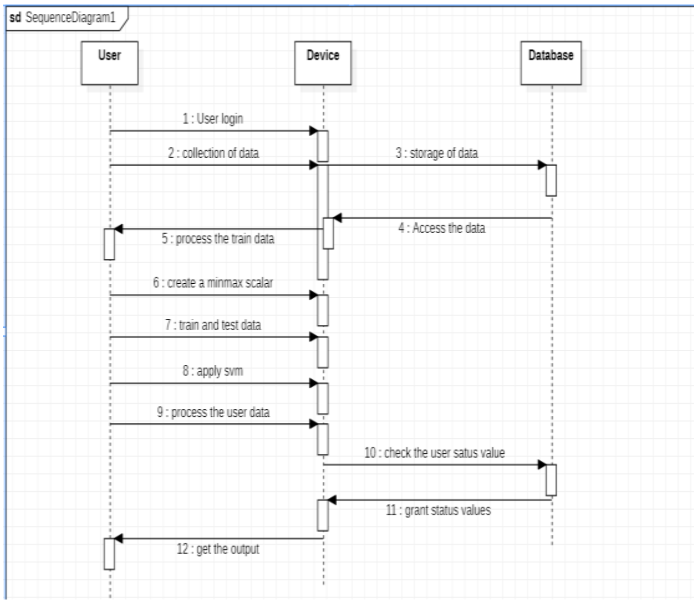


Fig.17SequenceDiagram

Use case diagram

The Use Case diagram of the project disease prediction using machine learning consist of all the various aspects a normal use case diagram requires. This use case diagram shows how from starting the model flows from one step to another.

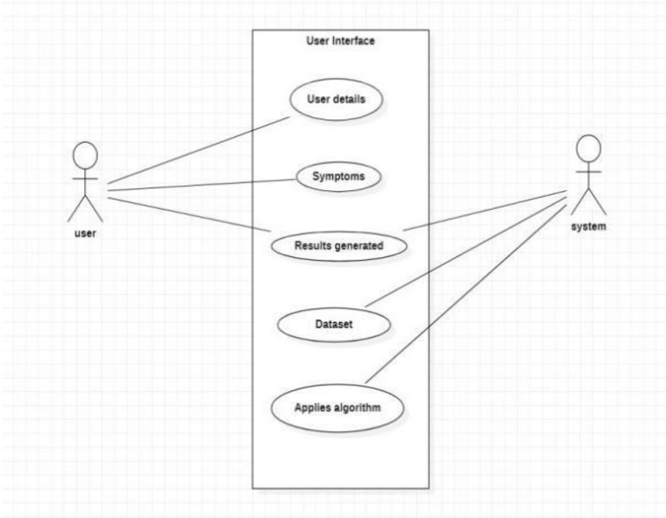


Fig.Use Case Diagram

ADVANTAGES OF THE SYSTEM:

The suggested system uses a much more straight forward and effective method. The usage of an easy-to use Frame work makes the system simpler. It has fewer complicated database setups and a more effective algorithm. Due to its platform independence, the system is more effective.

CONCLUSION

Parkinson's disease, which affects the brain's CNS, is incurable unless it is caught early. Lack of treatment and life loss result from late discovery. Therefore, it is important to diagnose it early. We used the machine learning technique Support Vector Machine for early illness identification.

SVM is the best method that delivers the best accuracy(up to 86%) comparison to other algorithms to forecast the commencement of the disease, allowing for early treatment and perhaps saving a life. We checked our Parkinson disease data and discovered this.

FUTURE ENHANCEMENT

Future research can concentrate on various methods for anticipating Parkinson's disease utilizing various datasets. In this study, we classify patients using a binary attribute (1- diseased patients, 0- non-diseased patients). The classification of patients and the distinct stages of Parkinson's disease will be done in the future using various types of features.

ACKNOWLDEMENT

For his insightful suggestion, knowledgeable counsel, and moral support during the writing of this work, we would like to express our gratitude to Mr. V. Satheesh Kumar.

REFERENCES

[1] DraganaMiljkovic et al, "Machine Learningand Data Mining Methods for Managing Parkinson's Disease" LNAI 9605, pp 209-220, 2016.

[2] Arvind Kumar Tiwari, "Machine Learning based Approaches for Prediction of Parkinson's Disease," Machine Learning and Applications- An International Journal (MLAU) vol. 3, June 2016.

[3] Dr. Anupam Bhatia and RaunakSulekh, "Predictive Model for Parkinson's Disease through Naive Bayes Classification" International Journal of Computer Science & Communication vol. 9, March 2018.

[4] M. Abdar and M. Zomorodi-Moghadam, "Impact of Patients' Gender on Parkinson's disease using Classification Algorithms" Journal of AI and Data Mining, vol. 6, 2018.

[5] Md. Redone Hassan et al, "A Knowledge Base Data Mining based on Parkinson's Disease" International Conference on System Modelling& Advancement in Research Trends, 2019

[6] I. Cantürk and F. Karabiber, ''A machine learning system for the diagnosis of Parkinson's disease from speech signals and its application to multiple speech signal types,'' Arabian J. Sci. Eng., vol. 41, pp. 5049–5059, Dec. 2016

[7] Ho, A.K., Iansek, R., Marigliani, C., Bradshaw, J.L. and Gates, S. (1998) Speech Impairment in a Large Sample of Patients with Parkinson's Disease. Behaviour Neurology, 11, 131-137. http://dx.doi.org/10.1155/1999/327643.

[8] Gil, D. and Johnson, M. (2009) Diagnosing Parkinson by Using Artificial Neural Networks and Support Vector Machines. Global Journal of Compute Science and Technology, 9, 63-71

[9] Logemann, J.A., Fisher, H.B., Boshses, B. and Blonsky, E.R. (1978) Frequency and Co-Occurrence of Vocal-Tract Dysfunctions in Speech of a Large Sample of Parkinson Patients. Journal of Speech Hearing Disorder, 43, 47-57.

[10] L. M. de Lau and M. M. Breteler, "Epidemiologyof Parkinson's disease," *The Lancet Neurology*, vol. 5, no. 6, pp. 525–535, 2006.

[11] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Earlydiagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Systems with Applications*, vol. 137, pp. 22–28, 2019.

[12] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–10, 2019.

[13] J. Rusz, M. Novotny, J. Hlavnivcka, T. Tykalova, and E. Ruuvzivcka, "High-accuracy voice-based classification between patients with Parkinson's disease and other neurological diseases may be an easy task with inappropriate experimental design," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 8, pp. 1319–1321, 2016.

[14] Z. A. Dastgheib, B. Lithgow, and Z. Moussavi, "Diagnosis of Parkinson's disease using electrovestibulography," *Medical & Biological Engineering & Computing*, vol. 50, no. 5, pp. 483– 491, 2012.

[15] L. Naranjo, C. J. Pérez, and J. Martín, "Addressing voice recording replications for tracking Parkinson's disease progression," *Medical & Biological Engineering & Computing*, vol. 55, no. 3, pp. 365–373, 2017.