Diagnosis Of Parkinson's Disease Through Speech Articulation Using Machine Learning

Fathima G
Professor,
Department of Computer Science and Engineering,
Adhiyamaan College of Engineering,
Hosur, India
fathima.ace@gmail.com

Tharunaa Shoban Babu
Student,
Department of Computer Science and Engineering,
Adhiyamaan College of Engineering,
Hosur, India
tharunaa.shobanbabu@gmail.com

Srimaan A
Student,
Department of Computer Science and Engineering,
Adhiyamaan College of Engineering,
Hosur, India
srimaan200012@gmail.com

Abstract- Parkinson's disease (PD) is one of the major public health diseases in the world which is progressively increasing day by day and has had its effects on many countries. Thus, it is very important to predict it in early age which has been a challenging task among researchers because the symptoms of disease come into existence in either middle or late middle age. In this paper the performance of different models such as adaboost classifier, logistic regression, decision tree classifier and support vector classifier has been evaluated using various metrics i.e. accuracy, precision, specificity. It primarily focuses on the symptoms of speech articulation problems in persons with Parkinson's disease and develops a model using Support Vector Machine. To forecast Parkinson's disease, the Boruta feature selection technique is utilised to determine the most essential features among all the features.

Keywords- Parkinson's Disease, Support Vector Machine, Boruta feature selection, Adaboost classifier, Logistic regression, Decision tree classifier

1. INTRODUCTION

The progressive shredding and loss of neurons in many parts of the nervous system causes neurodegenerative diseases. A unit in the brain that handles the functions are called neurons. Rather than being continuous, they are contiguous. A healthy neuron has extensions called dendrites or axons, a cell body, and a nucleus that houses our DNA, as seen in fig 1. Our genome is DNA, and our entire genome is packaged into it in a hundred billion neurons. When a neuron is sick, it loses its extension and hence its capacity to communicate, which is bad for it, and its metabolism drops, so it starts to amass trash and tries to store it in little packages in little pockets, which is bad for it. When things go worse, the neuron loses its extension, gets spherical, and becomes filled with vacuoles if it's a cell culture.

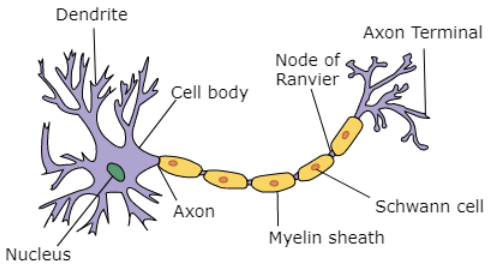

Fig. 1.1 Neurons

This research focuses on the prediction of Parkinson's disease, which is a rapidly growing incurable condition these days. Parkinson's disease is the most common disease, and it was [11] named after James Parkinson, who first defined it as agitation paralysis and then provided his surname as PD. It mostly affects the neurons that control overall body motions. The main molecules that affect the human brain are dopamine and acetylcholine. There are a number of environmental factors [6] that have been linked to Parkinson's disease. Below is a list of the factors that have been linked to Parkinson's disease in individuals.

A. Environmental factors: An individual's environment is a crucial component that affects not just the human brain but also all living organisms in its vicinity. Environmental factors that are rapidly influencing neurodegenerative disorders include: Heavy metals (such as lead and aluminium), pesticide exposure and noise pollution.

B. Air Quality: Pollution causes respiratory problems. Water pollution is caused by the presence of biotic and abiotic pollutants in water, etc. Psychological stress, it raises the level of stress hormone, which impairs neuronal function.

C. Biochemical Factors or Brain Injuries: People suffer brain injuries as a result of certain traumas, which causes some biochemical enzymes to enter the picture, providing stability to neurons and supporting the maintenance of specific chromosomes and genes.

D. Aging factor: Aging is one of the factors that causes Parkinson's disease to develop. According to Indian data [8], Parkinson's disease affects 11,747,102 people out of a total population of 1, 065, 070, 6072.

E. Genetic factors: The size, depth, and influence of individual genes' actions determine the status or severity of neurodegenerative disease, which progresses over time [7].

F. Speech Articulation Factors: Some speech language pathology such as voice, articulation, and swallowing abnormalities are identified due to the condition associated with Parkinson's disease (rigidity and bradykinesia).

Parkinson's disease (PD) can impact a person in a number of different ways.

(i) The voice becomes more breathy and gentle.

(ii) Smeared speech is possible.

(iii) The person has trouble finding the correct words, causing speech to grow slower.

1.1 PARKINSON'S DISEASE SYMPTOMS

The symptoms of Parkinson's disease can be classified below:

Motor symptoms:
These are symptoms that occur when a person performs a voluntary action. It denotes a movement disorder such as tremor, rigidity, freezing, Bradykinesia, or any other voluntary muscular movement.

Non-motor symptoms:
Non-motor symptoms include mood and affect problems, apathy, cognitive dysfunction, and complicated behavioural disorders, among others. Doctors divide Parkinson's disease into two categories: primary symptom and secondary symptom.

Primary symptoms:
The most essential symptom is primary symptoms. Rigidity, tremor, and slowness of movement are the most common symptoms.

Secondary symptoms:
Parkinson's disease is linked to a wide spectrum of symptoms. These can be motorised or non-motorized. Dysphonia (impaired speech production) and dysarthria (difficulties with speech articulation) [10] are common in parkinson's patients. Micrographia, impaired olfaction, postural instability, digestive system slowdown, constipation, weariness, weakness, and hypotension are some of the symptoms [9].

2. LITERATURE SURVEY

Prediction of Parkinson's disease is one of the most critical issues that must be identified in the early stages of the disease's onset in order to limit the rate of disease development among individuals. Various studies have been conducted to determine the root cause, with some reaching new heights by proposing a system that uses various machine learning approaches to distinguish healthy persons from those suffering from neurodegenerative disorders. In the last few decades, numerous pre-processing, feature selection, and classification algorithms have been adopted and developed. The work done in the prediction of Parkinson's diseases is included below. It is split it into three stages.

(i) Pre-processing Techniques Review

(ii) Classification techniques.

(iii) Computational methods.

Table 2.1: Different technologies used in the prediction of Parkinson's disease

Sno.	Paper Title	Description	Methods
1	Emulation of medical tasks in virtual reality using eye tracking for remote diagnosis of neurodegenerative disease [1]	The paper's author called for a virtual reality system that tracks how an individual's eye movement has reduced medical workload. The step taken to make a virtual reality-based remote diagnosis a reality.	Virtual Reality
2	Application of Deep Machine Learning to the Detection of Aging-Related Preclinical Neurodegenerative Diseases[2]	The paper's author called for a fundamentally new approach: employing AI models to detect preclinical decline using massive datasets obtained from single individuals.	AI stands for artificial intelligence (used in categorising the health states.)
3	A longitudinal study of facial expression recognition in Alzheimer's disease[3]	Face expression traits have been used by the author to distinguish Alzheimer's disease patients. Because of the huge differences in the scenario recognition task, cognition has devised the best way for reading expression in even the most nuanced cases.	Facial /Emotion recognition
4	A case study from the field of biomedical informatics was used to do an exploratory research on big data processing[4].	In biomedical informatics, big data processing is applied. It is thought to be useful in medical imaging and bioinformatics. Big data technology should be used to process the data from these two sectors.	Big Data
5	Simple drawing movements in Parkinson's disease are classified using machine learning[5].	The author distinguishes PD patients from healthy ones using handwriting markers. Because it includes particular muscular movements that were recorded, the patients were instructed to draw a straight line.	Movements recognition

Sahoo et al. (2012) [13] published a study that used data mining approaches to predict Parkinson's illness. Decision stump, Logistic Regression, and Sequential Minimization Optimization were the three approaches employed. According to the findings, the support vector machine model outperforms the others with an accuracy of 76%, a sensitivity of 0.97, and a specificity of 0.62 when compared to two other models.

Bonato et. al (2004) [14] have proposed evidences that data mining and artificial intelligent may help in recognizing the severity of motor fluctuations in PD patient .They collected the data using ACC (accelerometer) and EMG (electromyography) signals which was recorded while execution of standardized sets of motor assessment tasks.

In another study, Saritha.k et al. (2017) [15] utilised a javascript software to record the patient's speech, and then used Praat to transform it into a voice report. Praat accepts input in.wav files and generates a voice report using a script. Among the algorithms used, the decision tree produced the best results, with an accuracy of 100 percent without feature selection and 94 percent with feature selection.

The relevance of non-motor systems has been emphasised by Nayan reddy challa et al (2016) [16], which has been overlooked by many clinicians in favour of motor systems. Rapid eye movement (REM), sleep behaviour distortion, and smell loss were all evaluated in the study, and predictions were made using four machine learning techniques: Multilayer Perceptron, Bayes Net, RF, and Boosted Logistic Regression. Boosted logistic regression, which has an accuracy of 97.159 percent and a 98.9% area under the ROC curve, is considered to be a better method.

Chandrayan et al. (2016) [17] proposed extreme learning machines to predict PD..Using ELM they have done a comparative analysis and inferred that unlike conventional Neural Network elm doesn't require iterative variation of hidden neurons. So the simple architecture make elm a reliable method than others for prediction.

Among all voice recording variables, Jennifer He et al. (2017) [18] discovered that fundamental frequency is the best factor for predicting Parkinson's illness. They used Microsoft Azure Machine Learning Studio to test a variety of machine learning algorithms, including Boosted decision trees, Decision jungle, Locally Deep SVM, Logistic regression,Neural Networks, and SVM, with the best being Two-class Boosted decision trees, which is an ensemble technique.

3. PROPOSED MODEL

In this research, Kaggle is used to collect the Parkinson's Dataset, which is made up of a variety of biological voice measurements from 31 persons, 23 of whom have Parkinson's disease (PD). Each row in the table corresponds to one of the 195 voice recordings, and each column contains a specific voice measure. From the dataset, must split trained data and testing data where 80% of the code is taken by training data and the remaining for test data, which is further used to evaluate the model. Before choosing the appropriate model, comparison with other models such as AdaBoost classifier, Decision tree classifier, Logistics regression and Support vector classifier is done in terms of the effective model. After which, the trained Support Vector Machine model shall predict if the Person is affected by Parkinson's Disease or Not. Furthermore deployment will be done using Flask, that can be available for the users to use.

3.1 ARCHITECTURE DESIGN

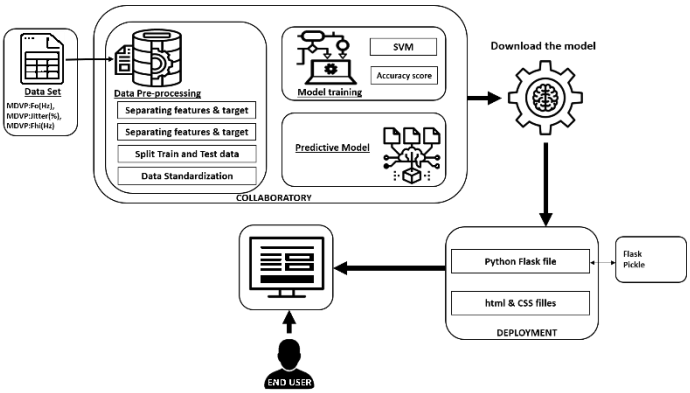

Fig. 3.1. Architecture diagram

Firstly, obtain the Parkinson's dataset from Kaggle website. After which, import the dataset file with .csv extension using Pandas into the Google Colaboratory. Reduce the amount of input variables in training data using Dimensionality Reduction Techniques. This is used to deal with high dimensional data, it is often useful to reduce the dimensionality by projecting the data to a lower dimensional subspace which captures the "essence" of the data.

Splitting the features and target variables, the target variable of a dataset is the feature of a dataset that you want to learn more about. A supervised machine learning algorithm learns patterns and uncovers links between other features in the dataset and the target using previous data. The model focuses on the status values i.e. 0 and 1 to produce an output. Where 0 indicates that the person does not have Parkinson's disease and 1 indicates that the person has Parkinson's disease. The serial number and the patient's name are the only attributes in the dataset that are considered features.

Splitting the dataset into two parts, one is training and the other one is testing. The train-test split procedure is used to evaluate the performance of machine learning algorithms that are used to generate predictions on data that was not used to train the model. It can be used for any supervised learning technique and can be utilized for classification or regression tasks. Taking a dataset and separating it into two subgroups is the technique. The training dataset is the first subset, which is used to fit the model. The second subset is not used to train the model; instead, the dataset's input element is given to the model, which then makes predictions and compares them to the predicted values. The test dataset is the name given to the second dataset.

When the dataset available is small, the train-test procedure is not appropriate. If the dataset is split into train and test sets, there wont be enough data in the training dataset for the model to learn an acceptable mapping of inputs to outputs. There will also be insufficient data in the test set to evaluate the model's performance appropriately. But it is appropriate to partition the dataset for this model. The remaining 80% of the data will be considered train data, with 20% of it being considered test data.

The process of rescaling one or more attributes to have a mean value of 0 and a standard deviation of 1 is known as data standardisation. The data is assumed to have a Gaussian (bell curve) distribution during standardisation. Although this is not required, the strategy is more effective when the attribute distribution is Gaussian.

After which the preceding step is model training. All that training a model involves is deterministic learning as good values for all the weights and the bias from descriptive samples. A machine learning algorithm generates a model in supervised learning by studying numerous examples and tries to find a model that minimises loss; this process is known as empirical risk minimization.

Support Vector Machine (SVM) was used to train the model. Although SVM is considered a classification strategy, it may be used to solve both classification and regression problems. We investigated several algorithms before deciding on the most effective method to use. DecisionTreeClassifier, Adaboost Classifier, and Logistic regression, as well as support vector machine, are some examples. In the a statistical method used to estimate the skill of these machine learning models found to have the Support vector machine or classifier to be the highest. After which the built our model using SVM.

Using Support Vector Classifier, must created a model, which will predict if the person is having parkinson's disease or not. Model works perfectly fine in Google colaboratory, but it cannot be used by the public here. In order to make this product available for the users the code has been deployed using Flask as Web app Application.

From colaboratory two files are been exported that are modelForPrediction.sav and standardScalar.sav. These files are to be imported to the python file in Flask for Web app application. These operations are done with the help of pickle library. After which will have the final product that will be available for the users to use.

4. IMPLEMENTATION

4.1 DATA SET

This set of data includes biological voice measurements from 31 people, 23 of whom have Parkinson's disease (PD).Each row in the table corresponds to one of the 195 voice recordings produced by these people, and each column relates to a specific voice measure ("name" column). The primary goal of the data is to distinguish healthy persons from those with Parkinson's disease using the "status" column, which is set to 0 for healthy people and 1 for those with PD.

The information is stored in ASCII CSV format. Each row in the CSV file corresponds to a single voice recording instance. There are around six recordings per patient, with the patient's name in the first column.

Matrix column entries (attributes):
name - ASCII subject name and recording number
MDVP:Fo(Hz) - Average vocal fundamental frequency
MDVP:Fhi(Hz) - Maximum vocal fundamental frequency
MDVP:Flo(Hz) - Minimum vocal fundamental frequency
MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP,MDVP:PPQ,Jitter:DDP - Several measures of variation in fundamental frequency
MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA - Several measures of variation in amplitude
NHR,HNR - Two measures of ratio of noise to tonal components in the voice
status - Health status of the subject (one) - Parkinson's, (zero) - healthy
RPDE,D2 - Two nonlinear dynamical complexity measures
DFA - Signal fractal scaling exponent
spread1,spread2,PPE - Three nonlinear measures of fundamental frequency variation[12].

Table 4.1: Dataset description

Data Set Characteristics:	Multivariate	**Number of Instances:**	197	**Area:**	Life
Attribute Characteristics:	Real	**Number of Attributes:**	23	**Date Donated**	2008-06-26
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	319213

Table 4.2: Features extracted from voice recordings

Feature	Group
Shimmer (dda) Shimmer (local) Shimmer (apq3) Shimmer (apq11) Shimmer (apq5) Shimmer (local,dB)	Amplitude Parameters
Number Of pulses Mean period Number Of periods Standard deviation Of period	Pulse Parameters
Jitter (ddp) Jitter (local) Jitter (rap) Jitter (local, absolute) Jitter (ppq5)	Frequency Parameters

Number of voice breaks Fraction of locally unvoiced frames Degree of voice breaks	Voicing Parameters
Mean pitch Median pitch Standard Deviation Maximum pitch Minimum itch	Pitch Parameters
Harmonic-to-Noise Noise-to-Harmonic Autocorrelation	Harmonicity Parameters

4.2 DIMENSIONALITY REDUCTION TECHNIQUE

Dimension reduction technique is used in such a situation when the attributes n has to be reduced in the data into the subset of k attributes i.e k<<n. It can be used to retain the most relevant attributes, leaving behind the least relevant attributes or finding smaller group of variables to combine with other variable without changing the actual meaning of the data value.

To find the missing value in the dataset, .isnull().sum() is used to calculate it. As a result shown in figure given below, there isn't any missing values with a percentage of 0%.

```
[ ]  parkinsons_data.isnull().sum()

name                0
MDVP:Fo(Hz)         0
MDVP:Fhi(Hz)        0
MDVP:Flo(Hz)        0
MDVP:Jitter(%)      0
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
DFA                 0
spread1             0
spread2             0
D2                  0
PPE                 0
dtype: int64
```

Fig. 4.1: Result of Dimensionality Reduction Technique

4.3 MODEL ACCURACY SCORE

In this paper Decision Tree Classifier, SVM, Adaboost Classifier and Logistic Regression are used to check the efficiency using accuracy score. From the figure below is evident that SVM has obtained 86% accuracy score being the highest, while other algorithms such as Decision Tree Classifier has obtained 82%, Adaboost Classifier has obtained 84% and Logistics regression has obtained 83%. Since SVM has high efficiency rate, it has been implemented as the main algorithm to build the predictive model.

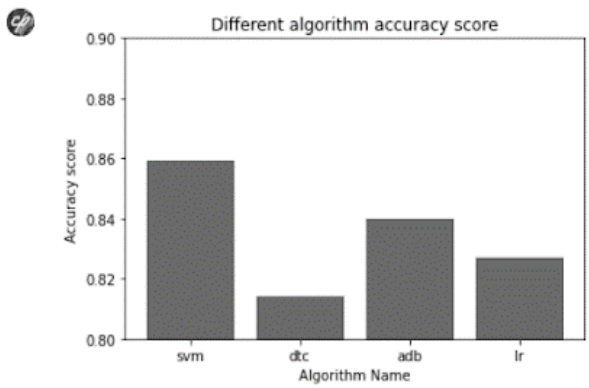

Fig. 4.2. Different algorithm accuracy score

Evaluating the model using accuracy score is important. Having said that if the score lies between 70% to 80% the model is good enough, if the score lies between 80% to 90% the model is said to be an excellent model, if the score lies between 90% to 100% the model is said to be overfitting.

From the figure below we have obtained the accuracy score for trained data as 88% and for the test data as 87%. There is no much of a difference between the trained and test accuracy score. Thus, making the model efficient to predict if the person is having Parkinson's disease or not.

```
# accuracy score on training data
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(Y_train, X_train_prediction)

print('Accuracy score of training data : ', training_data_accuracy)

Accuracy score of training data :  0.8846153846153846

# accuracy score on test data
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(Y_test, X_test_prediction)

print('Accuracy score of test data : ', test_data_accuracy)

Accuracy score of test data :  0.8717948717948718
```

Fig. 4.3. Accuracy score on train and test data

4.4 BUILDING A DIAGNOSIS SYSTEM

Passing the kernel as Linear for SVM algorithm. Once the model has been trained, pass the input units for different attributes. Changing the inputs into the array form using numpy. After which reshaping the numpy array and standardization.

```
# train dataset
input_data = (197.07600,206.89600,192.05500,0.00289,0.00001,0.00166,0.00168,0.00498,0.01098,0.09700,0.8

# changing input data to a numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the numpy array
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

# standardize the data
std_data = scaler.transform(input_data_reshaped)

prediction = model.predict(std_data)
print(prediction)


if (prediction[0] == 0):
  print("The Person does not have Parkinsons Disease")

else:
  print("The Person has Parkinsons")

[0]
The Person does not have Parkinsons Disease
```

Fig. 4.4. Diagnosis system

5. EXPERIMENTAL RESULTS

In terms of the total number of input samples divided by the number of valid predictions. It only works when there are an equal number of samples in each class. The accuracy score of the Training data and Test data in the SVM model alone is 0.88 and 0.87, respectively.

Also in this paper, to determine which attribute is contributing the most in prediction i.e among the 23 attributes which attribute/s plays a major role in determining if the person is having the parkinson's disease or not. From the figure given below we see that spread1 plays the most important role in finding out the result.

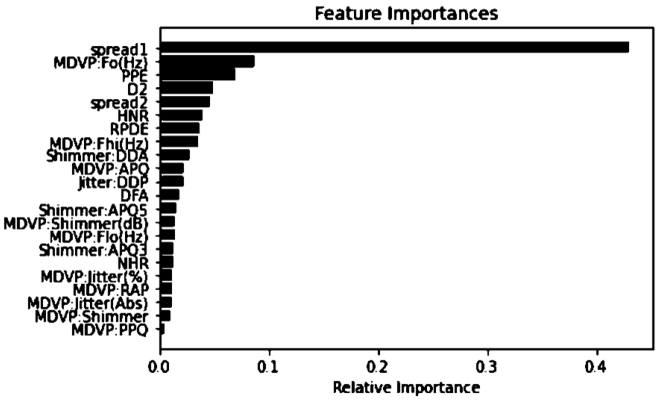

Fig. 5.1. Feature importance

The Webapp application is an user interface that allows the user to enter the values correctly; after submitting the answers, the outcome will show whether or not the person has Parkinson's disease.

The below figure 5.2. shows the webapp interface before entering the values. The values are obtained by physical examinations as well as with the help of the person's speech. From a person's speech, various values can be extracted like maximun fundamental vocal frequency, if some is having jitters or not and many more. With those values that has been obtained, either the doctors or the patient can enter the values in our application to find if he/she has Parkinson's disease or not.

Fig. 5.2. Website – before entering the values

The below figure 5.3. shows the sample values entered in our application. Every attribute that is involved in contributing the prediction has its range where the values lie. Unless the values entered by the user is in the range it can predict if he/she has parkinson's disease or not.

Fig. 5.3. Website – after entering the values

The below figure shows the resultant value of the sample values entered in figure 5.3. The model has predicted that the person has parkinson's disease and should consult a specialist.

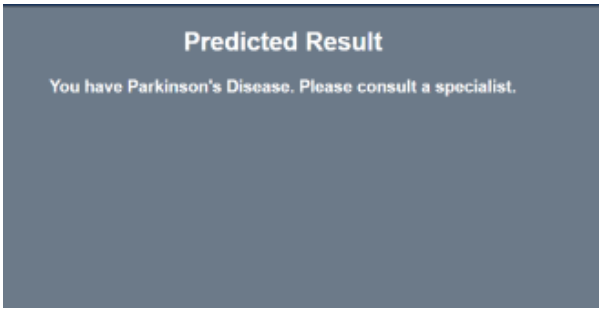

Fig. 5.4. Website – Resultant page

6. CONCLUSION

In this paper the machine learning model is being built using a Support Vector Classifier for Diagnosing the disease so that at the earlier stages the symptoms can be reduced and try to lead a normal life. While comparing the SVM model with other algorithms such as AdaBoost classifier, Logistic regression and Decision tree classifier, we found SVM more accurate. After building the predictive model, we have deployed the model using Flask to be available to the end-users. The Webapp application can be used by doctors to ease their prediction terminologies.

7. FUTURE SCOPE

Machine learning techniques are utilized in this paper, however there has been relatively little research on deep learning approaches. The work can be expanded in the future by employing autoencoders to minimise the number of characteristics and extract the most significant ones. Also, because the dataset utilised in this study is not particularly complex, the autoencoder did not learn well from it; nonetheless, a more complex dataset would almost certainly yield better results.

REFERENCE

[1] Jason Orlosky, Yuta Itoh, Maud Ranchet, Kiyoshi Kiyokawa, John Morgan, and Hannes Devos, "Emulation of Physician Tasks in Eye-tracked Virtual Reality for Remote Diagnosis of Neurodegenerative Disease", in *Proceedings of the IEEE Transactions on Visualization and Computer Graphics*, vol. 23, pp. 1302 – 1311, 2017.

[2] Mathew J. Summers, Vienna, Austria, Alessandro E. Vercelli, Georg Aumayr, Doris M. Bleier and Ludovico Ciferri,"Deep Machine Learning Application to the Detection of Preclinical Neurodegenerative Diseases of Aging", in *Proceedings of the Scientific Journal on Digital Cultures*, vol. 2, pp. 9-24, 2017.

[3] Bianca Torres, Raquel Luiza Santos, Maria Fernanda Barroso de Sousa, Jose Pedro Simoes Neto, Marcela Moreira Lima Nogueira, Tatiana T. Belfort1, Rachel Dias1, Marcia and Cristina Nascimento Dourado, "Facial expression recognition in Alzheimer's disease: a longitudinal study", pp. 383-389, 2014.

[4] Smitha Sunil and Kumaran Nair, "An exploratory study on Big data processing: a case study from a biomedical informatics", 3rd MEC International Conference on Big Data and Smart City, pp. 1-4, 2016.

[5] C. Kotsavasilogloua, N. Kostikis, D. Hristu-Varsakelis and M. Arnaoutoglouc, "Machine learning-based classification of simple drawing movements in Parkinson's disease", in *Proceedings of the Biomedical Signal Processing and Control*, pp. 174–180, 2017.

[6] Tanner CM, Ross GW, Jewell SA ," Occupation and risk of Parkinsonism: a multicenter case- control study" Arch Neurol,66(9):1106–1113,2009.

[7] V. A. Sukhanov, I. D. Ionov, and L. A. Piruzyan, "Neurodegenerative Disorders: The Role of Genetic Factors in Their Origin and the Efficiency of Treatment" in *Proceedings of the Human Physiology US National Library of Medicine National Institutes of Health*, vol. 31, pp. 472–482, 2005.

[8] Marras C, Tanner C."Epidemiology of Parkinson's Disease", Movement Disorders: Neurologic Principles and Practice, 2nd ed.2004, Watts, RL, Koller, WC (Eds). The McGraw-Hill Companies:New York, pp. 177.

[9] Cnockaert, L., Schoentgen, J., Auzou, P., Ozsancak, C.,Defebvre, L., & Grenez, F., "Low frequency vocal modulations in vowels produced by Parkinsonian subjects", Speech Communications, vol 50, pp. 288-300, 2008.

[10] Kenneth Revett, Florin Gorunescu and Abdel-Badeeh Mohamed Salem, "Feature Selection in Parkinson's disease: A Rough Sets Approach", Proceedings of the International Multi onference on Computer Science and Information Technology, pp. 425 – 428,2004, ISBN 978-83-60810- 22-4.

[11] Alexis Elbaz,James H. Bower, Brett J. Peterson, Demetrius M. Maraganore, Shannon K. McDonnell, J. Eric Ahlskog, Daniel J. Schaid, Walter A. Rocca, "Survival Study of Parkinson Disease in Olmsted County, Minnesota ", Arch Neurol. Vol. 60 pp. 91-96, 2003.

[12] https://www.kaggle.com/nidaguler/parkinsons-data-set?select=parkinsons.names

[13] Geeta Yadav, Yugal Kumar and G. Sahoo, "Predication of Parkinson's disease using Data Mining Methods: a comparative analysis of tree, statistical and support vector machine classifiers", in *Procedings of the National Conference on Computing and Communication Systems (NCCCS)*, pp. 1-4, 2012.

[14] Paolo Bonato, Delsey M. Sherrill, David G. Standaert, Sara S. Salles and Metin Akay, "Data Mining Techniques to Detect Motor Fluctuations in Parkinson's Disease", in *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4766-4769, 2004.

[15] Sonu S. R., Vivek Prakash and Ravi Ranjan, "Prediction of Parkinson's Disease using Data Mining", in *Proceedings of the International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 1082-1085, 2017.

[16] Kamal Nayan Reddy, Challa, Venkata Sasank Pagolu and Ganapati Panda, "An Improved Approach for Prediction of Parkinson's Disease using Machine Learning Techniques", in *Procedings of the International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016*, pp. 1446-145, 2016.

[17] Aarushi Agarwal, Spriha Chandrayan and Sitanshu S Sahu, "Prediction of Parkinson's Disease using Speech Signal with Extreme Learning Machine", in *Proceedings of the International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pp. 1-4, 2016.

[18] Akshaya Dinesh and Jennifer He, "Using Machine Learning to Diagnose Parkinson's Disease from Voice Recording", in *Proceedings of the IEEE MIT Undergraduate Research Technology Conference (URTC)*, pp. 1-4, 2017.