# PARKINSON DISEASE DETECTION USING VARIOUS MACHINE LEARNING ALGORITHMS

**[1]M. Venkata Vinay, [2]G. Ajay Kumar, [3]V. Mani Teja,**
**[4]V. Divya Nandhu, [5]Ms. S. Sarjun Beevi**

[1,2,3,4]Students, [5]Assistant Professor
Department of Computer Science and Engineering,
Bharath Institute of Higher Education and Research,

*Abstract*- **Parkinson disease is a neurodegenerative disorder. It effects on dopamine producing neurons in brain. This disease mostly effects on the organic functions of the human body. It affects the age group above 45, in which out of 31 persons 5 are detected. Rigid muscles, Speech, Amnesia, depression are most likely Symptoms of this disease. The above symptoms will also match with other disorders. Based on the symptoms, it is unpredictable to diagnose the disease. The symptoms include shaking, rigidity, slowness of movement and difficulty with walking, Thinking and behaviour change, Depression and anxiety are also common. The data of any person can be entered in database to check whether the person is affected by Parkinson disease or not.**

## INTRODUCTION:

Parkinson's disease is a disorder of the central nervous system affecting movement and inducing tremors and stiffness a neurodegenerative disorder affecting dopamine neurons in brain. Parkinson's disease is difficult to diagnose. Common diagnostic criteria require the medication before. In this model, the huge data is collected from previously affected person and then by using machine learning algorithm. Parkinson's disease (PD) manifests as the death of dopaminergic neurons in the substantia nigra pars compacta within the mid brain. This neurodegeneration leads to a range of symptoms including coordination issues, bradykinesia, vocal changes, and rigidity. Dysarthria is also observed in PD patients, it is characterized by weakness, paralysis, and lack of coordination in the motor-speech system: affecting respiration, phonation, articulation, and prosody. Since symptoms and the disease course vary, PD is often not diagnosed for many years. Therefore, there is a need for more sensitive diagnostic tools for PD detection because, as the disease progresses, more symptoms arise that make PD harder to treat. The main deficits of PD speech are loss of intensity, monotony of pitch and loudness, reduced stress, inappropriate silences, short rushes of speech, variable rate, imprecise consonant articulation, and harsh and breathy voice (dysphonia). The range of voice related symptoms is promising for a potential detection tool because recording voice data is non-invasive and can be done easily with mobile devices.

PD is difficult to detect early due to the subtle initial symptoms. There is a significant burden to patients and the health care system due to delays in diagnosis.The difficulty in early PD diagnosis has inspired researchers to develop screening tools relying on automated algorithms to differentiate healthy controls from people with PD.This binary diagnosis focuses on the first step of validating digital biomarkers in distinguishing disease from control; it does not offer a form of differential diagnosis where the model may distinguish PD among a variety of disorders that present PD-like symptoms (e.g. Lewy-Body Dementia, Essential Tremor). The current research is a promising first step toward a long-term goal of providing a decision support algorithm for physicians in screening patients for PD. In this paper, we apply several different machine learning models to classify PD from controls using the mPower Voice dataset. The data used for this analysis were collected through mPower, a clinical observational study conducted by Sage Bionetworks using an iPhone app to collect digital biomarkers and health data on participants both with and without PD. To maintain user confidentiality and enable linking across datasets, each participant was uniquely identified with a distinct health code. Typically, the symptoms of PD are attenuated using dopaminergic medications such as levodopa.

## LITERATURE SURVEY:

### Detection of Parkinson Disease Using Clinical Voice Data Mining:

For all 195 phonations, features are extracted. For feature extraction only first half of the recordings are considered, because the second half of the recording is influenced by reduced lung pressure. Various traditional measures and nonstandard measures are extracted. A set of 22 features is prepared. Feature set consists of Fo(Hz), Fhi(Hz), Flo(Hz), Jitter(%), Jitter(Abs), RAP, PPQ, Jitter:DDP, Shimmer, Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, APQ , Shimmer:DDA, NHR, HNR, RPDE, DFA, spread1, spread2, D2, PPE. The vocal fold vibration frequency is known as fundamental frequency. The perturbation in the frequency and amplitude in successive vocal fold cycles is termed as jitter and shimmer respectively. The noise-to harmonic and harmonic-to-noise measures are measured using estimates of signal to noise by calculation of autocorrelation of each cycle. D2 is the correlation dimension between the signal and its first time delay embedded signal whereas the RPDE(recurrence period density) is the measure of periodicity of the reconstructed signal after embedding time delay. DFA (Detrended fluctuation analysis) is the log-log plot of the time scales L and amplitude variation F(L). Non-linear measure of fundamental frequency variation is defined in terms of spread 1 and spread 2. The logarithmic scale of pitch sequence is explained as semitone pitch p(t) where t is the time. The entropy of relative semitone variation is known as pitch period entropy (PPE). All these parameters show variation for the healthy and parkinson's case. Next, features are selected among these to get best classification among the two groups. Feature are selected which have more separable values than others and a new feature data subset is prepared which contains 15 features as shown in Table 1 . Support vector machine classifier is used and their performance is evaluated. Classifier used is supervised classifiers and

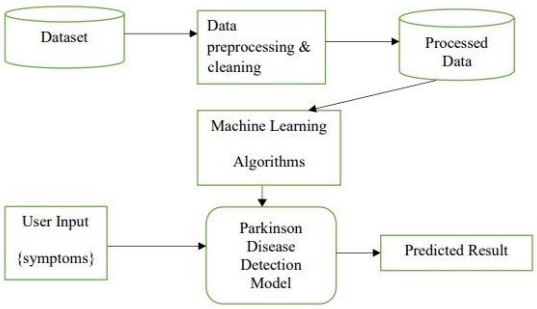
therefore dataset is divided into training and testing datasets.75% of the data is used for training purpose and rest 25% is for testing. Out of 195 observations, 146 are used for training (110 Parkinson +36 healthy) and 49 are used for testing

**Chirag Mittal, Amrit Anshu Sharma, "Parkinson's Disease Detection Using Different Machine Learning Algorithms'':** Parkinson's Disease is a widespread degenerative syndrome that affects the nervous system. Its early symptoms include tremor, rigidity, and vocal impairment (dysphonia). This paper proposes performance of Machine Learning Methods in Diagnosing Parkinson's Disease. Several machine-learning techniques were considered and trained with the same data set to classify healthy and Parkinson's Disease patients. Methods:-We have used various machine learning-based techniques for Parkinson's disease (PD) diagnosis. These ma[1]chine learning techniques includes K Nearest Neighbours' (k[1]NN), Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), Random Forest and XGBoost. We have used these algorithms to check the best algorithm for detection of Parkinson's Disease. Result:-After applying these algorithms we found our accuracies of these algorithms are as follows: - Naive Bayes(71.79%), SGD(84.61%), Decision Tree(84.61%), KNN(97.43%), Random Forest(89.74%), SVM(87.18%) and XGBoost(94.87%). After considering all algorithms and analysing their accuracies we found out that KNN is the best of all the algorithms used by us for detection of Parkinson Disease with accuracy of 97.43 percent.
early prediction and accurate medicinal diagnosis to the affected people.

## PROPOSED SYSTEM:

The proposed system of Parkinson Disease Detection using machine learning is that we have used many techniques and algorithms and all other various tools to build a system which detects the Parkinson Disease of the patient using the symptoms and by taking those symptoms we are comparing with the system's dataset that is previously available. By taking those datasets and comparing with the patient's symptoms we will detect the accurate percentage Parkinson Disease of the patient. The dataset and symptoms go to the prediction model of the system where the data is pre-processed for the future references and then the user will enter the various symptoms. Then the classification of those data is done with the help of various algorithms and techniques such as Decision Tree, KNN, Random Forest and SVM. This system takes symptoms from the user and detects the disease accordingly. Based on the symptoms that it takes and from the dataset. And then it calculates the accuracy that the user may have disease or not in the future and shows the result.

## SYSTEM ARCHITECTURE:




## MODULES:

**SVM:**

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

**KNN:**

K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much like the new data.

**XGBOOST:**

XGBoost or the Extreme Gradient boost is a machine learning algorithm that is used for the implementation of gradient boosting decision trees. Why decision trees? When we talk about unstructured data like the images, unstructured text data, etc., the ANN models (Artificial neural network) seems to reside at the top when we try to predict. While when we talk about structured/semi-structured data, decision trees are currently the best. XGBoost was basically designed for improving the speed and performance of machine learning models greatly, and it served the purpose very well.

**Random Forest:**

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. The fundamental concept behind random forest is a simple but powerful one. The wisdom of crowds. In data science speak, the reason that the random forest model works so well is many relatively uncorrelated models (trees) operating as committee will outperform any of the individual constituent models.

## CONCLUSION:

So, Finally We conclude by saying that this project Parkinson Disease Detection using machine learning is very much useful in everyone's day to day life and it is mainly more important for the healthcare sector, because they are the one that daily uses these systems to detect the Parkinson disease of the patients based on their general. information and their symptoms that they are been through. Now a day's health industry plays major role in curing the diseases of the patients so this is also some kind of help for the health industry to tell the user and also it is useful for the user in case he/she doesn't want to go to the hospital or any other clinics, so just by entering the symptoms or information the user can get to know whether he/she having Parkinson disease or not.

## REFERENCES:

[1] D. Heisters, "Parkinson's: symptoms, treatments and research," British Journal of Nursing where the paper published, vol. 20, no. 9, pp. 548– 554, 2011.
[2] L., Naranjo, C.J., Pérez, J., Martín, Y., Campos-Roca, "A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications", COMPUT METH PROG BIO, vol. 142, pp.147- 156, 2017.
[3] H., Lee, L., Guan, I., Lee,"Video analysis of human gait and posture to determine neurological disorders", EURASIP J. Image Video PROC., vol. 1, pp. 380867, 2008.
[4] Buckley, C.; Alcock, L.; McArdle, R.; Ur Rehman, R.Z.; Del Din, S.; Mazzà, C.; Yarnall, A.J.; Rochester, L. The Role of Movement Analysis in Diagnosing and Monitoring Neurodegenerative Conditions: Insights from Gait and Postural Control. Brain Sci. 2019, 9, 34.
[5] Espay, A.J.; Bonato, P.; Nahab, F.B.; Maetzler, W.; Dean, J.M.; Klucken, J.; Eskofier, B.M.; Merola, A.; Horak, F.; Lang, A.E.; et al. Technology in Parkinson's Disease: Challenges and Opportunities. Mov. Disord. 2016, 31, 1272–1282.
[6] Siragy, T.; Nantel, J. Quantifying Dynamic Balance in Young, Elderly and Parkinson's Individuals: A Systematic Review. Front. Aging Neurosci. 2018, 10, 387
[7] 7. Castiglia, S.F.; Trabassi, D.; Tatarelli, A.; Ranavolo, A.; Varrecchia, T.; Fiori, L.; Di Lenola, D.; Cioffi, E.; Raju, M.; Coppola, G.; et al. Identification of Gait Unbalance and Fallers Among Subjects with Cerebellar Ataxia by a Set of Trunk AccelerationDerived Indices of Gait. Cerebellum.
[8] Phinyomark, A.; Petri, G.; Ibáñez-Marcelo, E.; Osis, S.T.; Ferber, R. Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. J. Med. Biol. Eng. 2018, 38, 244–260.
[9] Sharma, P.; Pahuja, S.; Veer, K. A Systematic Review of Machine Learning Based Gait Characteristics in Parkinson's Disease. Mini Rev. Med. Chem. 2021, 21, 1197–1210.
[10] Mirelman, A.; Ben or Frank, M.; Melamed, M.; Granovsky, L.; Nieuwboer, A.; Rochester, L.; Del Din, S.; Avanzino, L.; Pelosin, E.; Bloem, B.R.; et al. Detecting Sensitive Mobility Features.