***Copyrights reserved***

| Innovating Ai  SVM method of Machine Learning algorithms to recognize Parkinson’s disease  Laraib Ahmad Siddique(Galgotias University)    **Abstract**  Parkinson’s Disease (PD) is the second most prevalent age-related neurological disorder, resulting in a variety of motor and cognitive symptoms.Extensive research has been conducted on treatment and study of PD. Various Machine Learning (ML) and Deep Learning (DL) models have been generated and studied for detection. These symptoms are increasingly manifesting in younger generations, creating an alarming situation. Diagnosing PD is challenging because its symptoms closely resemble those of other conditions, including normal aging and essential tremor.  Rather than ignoring symptoms, early detection and prevention of disease progression need to be made. Noticeable symptoms, such as walking difficulties and communication issues, typically begin to appear around the age of 45-50[1]. For time being there's no cure for PD, certain medications can even accelerate its effects. Patients can manage the complications associated with the disease to sustain stability in their life.Some of the major algorithms that can be used to establish a AI based project that can detect its effectiveness and some of them include[2],Random Forest ,Decision Tree (DT),K-Nearest Neighbor (KNN),Support Vector Machine (SVM),Multi-Layer Perceptron (MLP)  These methods are readily used in the medical industry not only for PD but also to detect and even cure any of the diseases that AI & ML had been helping us for a long time. | D-66,Shaheen bagh  New Delhi,India  **(+91) 97175 47140**  [**labbuxt@gmail.com**](mailto:labbuxt@gmail.com)  Fig .1: Parkinson’s effect on human body |
| --- | --- |
| **Introduction**  Parkinson’s disease is a neurological disorder caused by significant loss in dopamine excretion causing unintended or uncontrollable movements of one's body parts, including shaking, stiffness, and difficulty with balance and to coordinate[3]. While Anyone could be affected by Parkinson’s though research and study tells that male patients are much more affected than women[4]. The reasons for this disparity are not yet clear, but studies are ongoing to identify factors that may tell its gender variance. Studies show that age is a well-established risk factor: although the majority of Parkinson’s cases evolve after age 60, about 5% -10% of cases occur before age 50[5]. Early-onset Parkinson’s is often not readily, but mainly as it could be inherited or has been linked to specific genetic mutations from his/her ancestors.  Parkinson’s disease is marked by the disruption of the dopaminergic cycle in the nerve cells of the substantia nigra. This region is part of the basal ganglia, which is heavily involved in movement control in our body due to secretions such as dopamine, which acts as synapses to control movements in various body parts. The degenerative process begins at the brain's base only , leading to the destruction of olfactory bulbs[6]. It then progresses to the lower brain stem, affecting the substantia nigra and midbrain. Eventually damaging the limbic system and frontal neocortex, thus worsening both physical and mental symptoms.    Fig. 2: *Dopamine synthesis by substantia nigra*  Although genetics is believed to play a role in the transmission of Parkinson’s.But the disease does not typically appear to be familial in most cases it also could be essential tremors. Many researchers now theorize that Parkinson’s arises from a combination of genetic predispositions and environmental influences, such as exposure to toxins and drugs.  **Types of PD and their Symptoms :**  Although there are some noticeable symptoms that could be seen physically but mainly there are mainly 2 types of PD  1. Movement disorder (motor symptoms)  2. Cognitive dysfunction (non-motor symptoms)   * Symptoms related to unintentional body movements. Motor symptoms of Parkinson’s disease also include cardinal symptoms such as bradykinesia, postural instability,cogwheel,etc. Approximately 75% of patients experience a resting tremor, occurring at a frequency of 3-5 Hz, which is known as asymmetrical tremor[7].Bradykinesia, which affects patients' handwriting, is a third type of symptom. In cogwheel rigidity, is when patients feel resistance during joint movements and it's difficult to maintain the strength. Postural instability, where patients lose control over their body balance and become unstable, is also a significant symptom of PD. * Non-motor symptoms include hyposmia, sleep behavior disorder, rapid eye movement (REM) sleep disturbances, constipation, and depression. Many patients also suffer from cognitive dysfunction, ranging from cognitive impairment (PD-MCI ->Mild Cognitive impairment) to (PDD-Parkinson's disease dementia). PD-MCI often occurs at an early stage symptom, whereas PDD develops nearly 20 years after the onset of PD[8]. In PD-MCI, patients exhibit memory and thinking abnormalities beyond what is expected with normal aging, making early diagnosis crucial as it can lead to PDD. PDD symptoms include short-term memory loss, attention impairment, executive dysfunction, visual-spatial deficits, and neuropsychiatric symptoms such as hallucinations, personality changes, and mood alterations.   Fig.3: *Motor and Non-motor impairment*  **Related researches on PD:**  This medical science is a vast field on which AI has planted its roots. The PD is the one on which research through AI and Ml have been conducted and outputs are seen very profoundly. Some of the major researches are :   1. Parvez alam etel.   Researchers have used artificial intelligence techniques to massively accelerate the search for Parkinson’s disease treatments([***University of Cambridge****,****Published: 17 April 2024***](https://www.cam.ac.uk/research/news/ai-speeds-up-drug-design-for-parkinsons-ten-fold#:~:text=The%20researchers%2C%20from%20the%20University,the%20protein%20that%20characterises%20Parkinson's.))To tackle this issue, they proposed a machine learning method to discover small molecule inhibitors of α-synuclein aggregation, a process linked to Parkinson’s disease and other synucleinopathies. Employing structure-based machine learning in an iterative fashion, they first identified and then progressively enhanced secondary nucleation inhibitors. Given that α-synuclein aggregates proliferate via autocatalytic secondary nucleation, they focused on finding compounds that target the catalytic sites on the aggregate surfaces. The findings show that this strategy enables the identification of compounds that are significantly more potent, by two orders of magnitude, than those previously reported.[9]    Fig. 4: Iterative Learning of Bio chemicals   1. Md. Saiful Islam etel.   Online AI-based test for Parkinson’s disease severity shows promising results**(**[*University of rochester*](https://www.rochester.edu/newscenter/ai-test-for-parkinsons-disease-severity-566772/)*,***)** Another significant study utilized an artificial intelligence tool developed by researchers at the University of Rochester to help people with Parkinson’s disease remotely assess the severity of their symptoms within minutes. Published in NPJ Digital Medicine, the study describes a new tool where users tap their fingers 10 times in front of a webcam to evaluate motor performance on a scale of 0–4.The taps could be of different fingers but the main point noted here is its movement speed and stiffness in fingers. The AI model rapidly assesses the severity using the MDS-UPDRS guidelines and automatically generates computational metrics depicting speed, amplitude, frequency, and period within the time bound of raps. These metrics are interpretable, standardized, repeatable, indulging medical guidebooks data, allowing it to classify the severity of tremors effectively[10].      Fig .5:Rigorous tapping of hand with movement detection |  |

**Material and dataset:**

### The dataset used in the research was sourced from the Kaggle Database repository and is available under the title “Parkinson’s Disease Data Set: Detecting Parkinson’s Disease – Python Machine.” This dataset includes various biomedical voice measurements recorded from 31 individuals, 23 of whom having Parkinson's disease (PD) and others with similar symptoms. Each column in the dataset represents a specific voice measure, and each row corresponds to one of the 195 voice recordings from individuals identified in the "name" column. The primary objective of this data is to distinguish healthy individuals from those with Parkinson's disease (PD), indicated by the "status" column, where 0 denotes healthy and 1 denotes PD. The data is provided in ASCII CSV format, with each row representing an instance of one voice recording. There are approximately six recordings per patient, with the patient's name listed in the first column.[11]

### **Attribute Information:**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 the ratio of noise to tonal components in the voice status - The 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

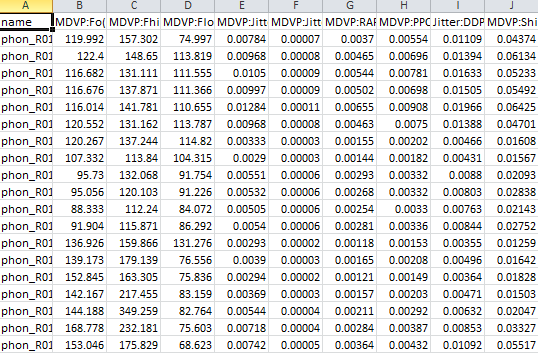


Fig. 6: Dataset look of CSV parkinson’s File

**SVM(Support Vector Machine)**

In the realm of machine learning, support vector machines (SVMs), also known as support vector networks, are supervised max-margin models equipped with associated learning algorithms. These models are utilized for data analysis in both classification and regression tasks. Developed at AT&T Bell Laboratories by Vladimir Vapnik and colleagues (Boser et al., 1992; Guyon et al., 1993; Cortes and Vapnik, 1995; Vapnik et al., 1997)[12], SVMs are among the most extensively studied models. They are grounded in the statistical learning frameworks of VC theory, as proposed by Vapnik (1982, 1995) and Chervonenkis (1974).  Fig . 7: Vladimir Vapnik

SVMs not only handle linear classification efficiently but also excel in non-linear classification tasks through the kernel trick. This technique represents data solely through pairwise similarity comparisons between original observations, effectively transforming the data into higher-dimensional feature spaces. By using the kernel trick, SVMs implicitly map their inputs into these high-dimensional spaces. As max-margin models, SVMs are resilient to noisy data, including misclassified examples. Additionally, SVMs can be adapted for regression tasks by altering their objective function.

The two very reputable Computer scientists namely Hava Siegelmann and Vladimir Vapnik support vector clustering algorithms, extending the principles of SVMs to categorize unlabeled data[13]. This approach, which falls under unsupervised learning, aims to identify natural clustering patterns within data and subsequently mapping the new data points to these clusters.

The widespread adoption of SVMs is due to their theoretical amenability and versatility in handling a wide range of tasks, including structured prediction problems. However, it remains uncertain whether SVMs consistently offer superior predictive performance compared to other linear models, such as logistic regression and linear regression.

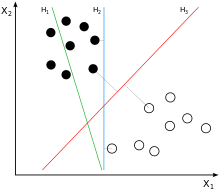
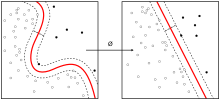
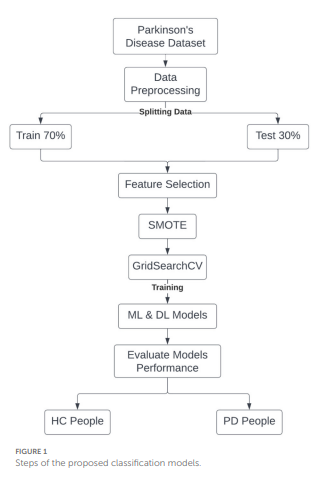
 

Fig. 9:Kernel Machine[12]

Fig. 8: H1 does not separate the classes. H2 does, but only with a small margin. H3 separates them with the maximal margin[11]

**Methodology**



The database was tested with the two datasets trained and tested to differentiate the working algorithm with the new inputs so that it should work relatively and moderately with the same precision when new datasets are input. With the help of sklearn.model\_selection containing train\_test\_split function that splits the datasets for differential actions

Given a rest size of 0.2 and random state of 2.

Through the standardized mean data the database was fitted and transformed to minimize the change fault and make in range respectively.

Created a support variable class to fit it with a linear kernel.Accuracy check :

Finally the was setted to give a accurate data using predict and accuracy\_score libraries from scikit-learn

,that come out to be

**Accuracy score of training data : 0.8846153846153846**

**Accuracy score of test data : 0.8717948717948718**

The accuracy score of Training and test data came out to be nearby almost that is the condition of the model data is nearly correct

Fig. 7 Flowchart of Test : Train ratio[14]

For inputting external data the dataset was applied to check the dataset without the **status** (The status is the final result that will depict the condition of the persons parkinson status if 0 -> means negative but id 1 -> means the person is positive tested )

Thus with the help of numpy the data was converted to an array dataset in the form of tuple and was transformed to standardized and predicted to generate the status value of the person the basis of the specified data.

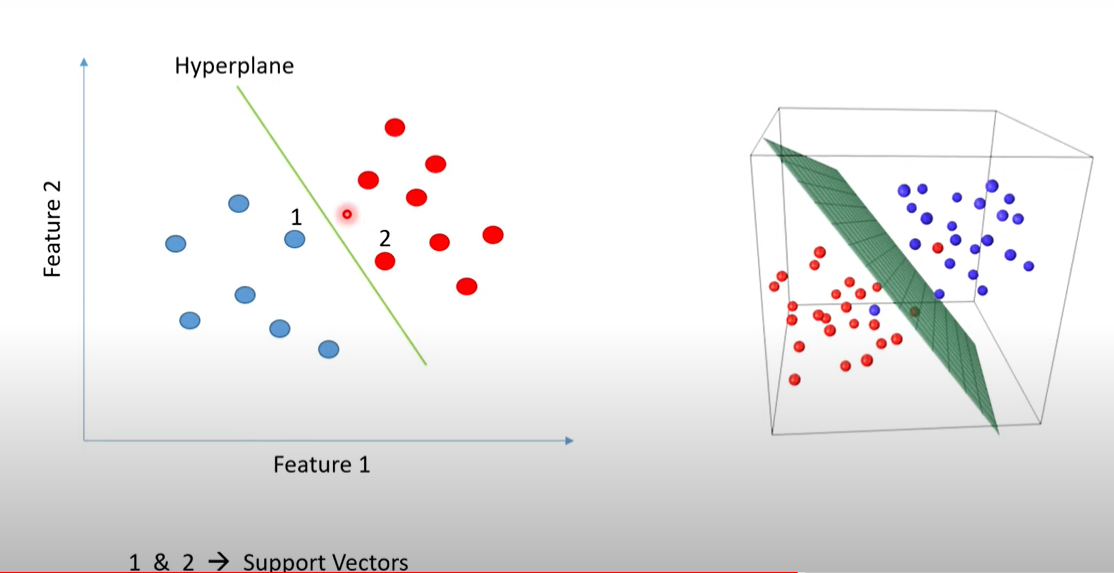


Fig. 8: Hyperplane Depiction of a SVM dataset

**Result :**

This paper aims to develop an efficient method for detecting Parkinson's disease (PD) using voice samples. I utilized a dataset from Kaggle, which contains 196 records of voice signal features collected from 147 individuals with PD and 48 healthy controls (HC). While I focused on the SVM model, other traditional machine learning and deep learning algorithms, such as k-nearest neighbors (KNN), decision tree (DT), random forest (RF), and multilayer perceptron (MLP), are also available.

To demonstrate the performance of the proposed strategy using the dataset, I split the sample into two groups. The first group consists of the training samples, which comprise 80% of the total samples. The remaining samples are used for testing, validating, and checking the system's accuracy. Our dataset appears to be unbalanced, with a significantly higher number of PD cases than HC cases.

Additionally, when we fed our SVM model with external outputs, the results were compared to the original tested results and were found to be correct.

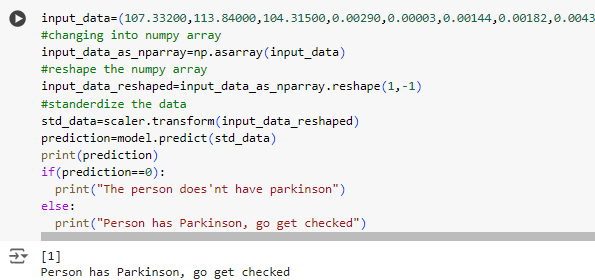


Fig. 9: Result output

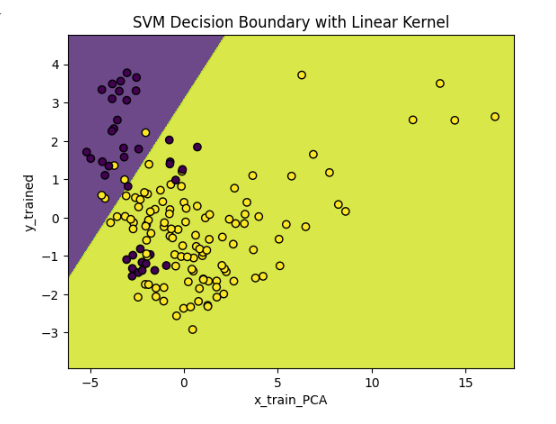


Fig. 10:SVM datagraph representation of trained and test variable

**Motivation:**

Parkinson's disease (PD) is a chronic and progressive neurological disorder that affects millions of people globally. Despite advancements in medical science, the early diagnosis of Parkinson's disease still remains a significant challenge for us. Traditional diagnostic methods mainly rely on clinical treatments and symptomatic observations and behavioral changes , which can be subjective and often detected when one is fully affected by it . This delay in diagnosis can limit the effectiveness of treatments and adversely affect the quality of life of patients.

1.1 Why Early Detection?

Early detection of Parkinson's disease is crucial for several reasons:

* Improved Patient Outcomes: Early diagnosis allows for early diagnosis , which can slow its effects through medication, improve patient outcomes, and smoothe their life.
* Better Disease Management: Patients diagnosed early can benefit from lifestyle adjustments and treatments that can mitigate symptoms and delay further degeneration.
* Research Advancements: Early detection provides a larger window for researchers to study it, potentially leading to new treatments and a better understanding of Parkinson's disease.

1.2 Limitations of Current Diagnostic Methods

Current diagnostic techniques for Parkinson's disease include clinical evaluations, imaging tests, and various motor function assessments. However, these methods have significant limitations:

* Subjectivity: Clinical evaluations are often subjective and dependent on the experience and judgment of the doctorate.
* Late-stage Detection: Many current methods detect PD only after significant motor symptoms have appeared, missing the early symptoms that were crucial.
* ResourceConsumption: Advanced diagnostic tests, such as imaging, can be expensive and resource-intensive, making them inaccessible to the financially weak section of society.

1.3 The Potential of AI and Machine Learning

Artificial Intelligence (AI) and Machine Learning (ML) offer promising solutions to overcome these challenges. By analyzing large datasets and identifying patterns that are not apparent through traditional methods, AI and ML can enhance the accuracy and speed of Parkinson's disease diagnosis. Key motivations for applying AI and ML to PD detection include:

* High Accuracy: Machine learning algorithms can process complex datasets and identify subtle markers of Parkinson's disease with high precision.
* Non-Invasive Methods: AI models can utilize non-invasive data, such as voice recordings and other biomarkers, to detect PD, making the diagnostic process less burdensome for patients.
* Scalability: AI-driven diagnostic tools can be scaled and deployed widely, making advanced diagnostic capabilities accessible even in resource-limited settings.

1.4 Impact and Future Directions

The application of AI and ML in Parkinson's disease recognition has the potential to transform the landscape of medical diagnostics. By providing a reliable, non-invasive, and scalable diagnostic tool, this project aims to:

* Enhance Early Diagnosis: Enable healthcare providers to detect Parkinson's disease at an earlier stage, improving patient prognosis.
* Reduce Diagnostic Costs: Offer a cost-effective alternative to expensive diagnostic tests, broadening access to advanced medical diagnostics.
* Advance Medical Research: Contribute valuable insights and data to the field of Parkinson's disease research, fostering further advancements and innovations.

In summary, this project is driven by the critical need for early and accurate diagnosis of Parkinson's disease, the limitations of existing methods, and the transformative potential of AI and ML technologies. By addressing these challenges, we aim to make significant contributions to patient care, medical research, and the broader healthcare community.

***References:***

[1] [*John L. Lehr*](https://www.ncoa.org/author/john-l-lehr) *”****Parkinson's Disease: Early Signs, Symptoms, and What to Do When Diagnosed****”Apr 19, 2022*

[2] *Raya Alshammri, Ghaida Alharbi, Ebtisam Alharbi and Ibrahim Almubark “****ML approach to identify parkinson’s disease using voice signal feature****”Original Research PUBLISHED 28 March 2023 DOI 0.3389/frai.2023.1084001*

[3]”***Bias Investigation in Artificial Intelligence Systems for Early Detection of Parkinson’s Disease: A Narrative Review Sudip****” Paul , Mahesh Rao Maindarkar , Sanjay Saxena , Luca Saba , Monika Turk , Manudeep Kalra , Padukode R. Krishnan and Jasjit S. Suri Review. Diagnostics 2022, 12, 166.*

[4]*Raya Alshammri, Ghaida Alharbi, Ebtisam Alharbi and Ibrahim Almubark “****ML approach to identify parkinson’s disease using voice signal feature****”Original Research PUBLISHED 28 March 2023 DOI 0.3389/frai.2023.1084001*

[5]**”*What to Know About Parkinson’s Symptoms in Adults”****Medically Reviewed by C. Nicole Swiner, MD on April 07, 2023 Written by WebMD Editorial Contributors*

[6]”*Olfaction in Neurologic and Neurodegenerative Diseases: A Literature Review” Maria Dantas Costa Lima Godoy,1 Richard Louis Voegels,1 Fábio de Rezende Pinna,1 Rui Imamura,1 and José Marcelo Farfel. PMCID: PMC4399182 PMID: 25992176*

[7]”***The clinical symptoms of Parkinson's disease”,****Sigurlaug SveinbjornsdottirFirst published: 11 July 2016https://doi.org/10.1111/jnc.13691Citations: 796.This article is part of a special issue on Parkinson disease.*

[8]”***The clinical symptoms of Parkinson's disease”,****Sigurlaug SveinbjornsdottirFirst published: 11 July 2016https://doi.org/10.1111/jnc.13691Citations: 796.This article is part of a special issue on Parkinson disease.*

[9]”**Discovery of potent inhibitors of α-synuclein aggregation using structure-based iterative learning”**

Robert I. Horne, Ewa A. Andrzejewska, Parvez Alam, Z. Faidon Brotzakis, Ankit Srivastava, Alice Aubert, Magdalena Nowinska, Rebecca C. Gregory, Roxine Staats, Andrea Possenti, Sean Chia, Pietro Sormani, Bernardino Ghetti, Byron Caughey, Thomas P. J. Knowles & Michele Vendruscolo *Nature Chemical Biology* 20, 634–645 (2024)

[10]*”****Using AI to measure Parkinson’s disease severity at home”****Md Saiful Islam supported by the U.S. Defense Advanced Research Projects Agency (DARPA) under grant W911NF19-1-0029, National Science Foundation Award IIS-1750380, National Institute of Neurological Disorders and Stroke of the National Institutes of Health under award number P50NS108676, and Gordon and Betty Moore Foundation.*

[11]*Max A. Little, Patrick E. McSharry, Eric J. Hunter, Lorraine O. Ramig (2008),* ***'Suitability of dysphonia measurements for telemonitoring of Parkinson's disease'****, IEEE Transactions on Biomedical Engineering (to appear).*

[12]*Vladimir Vapnik Wikipedia available under the Creative Commons Attribution-ShareAlike License 4.0*

[13]*Vladimir Vapnik Wikipedia available under the Creative Commons Attribution-ShareAlike License 4.0*

[14]*Raya Alshammri, Ghaida Alharbi, Ebtisam Alharbi and Ibrahim Almubark “****ML approach to identify parkinson’s disease using voice signal feature****”Original Research PUBLISHED 28 March 2023 DOI 0.3389/frai.2023.1084001*