IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF PARKINSON'S DISEASE USING MACHINE LEARNING AND WEB APPLICATION

[1]Naveen G, [2]Paspuleti charan,[3]Lenish, [4]Nishanth,[5]Suraj Kumar Roshan
[1]student, [2]student, [3]Assistent professor3rd ,[4]student,[5]student
[1]Computer science Department,[2]Computer science Department
[1]The oxford college of Engineering, Bangalore, India

Abstract: Parkinson's disease is a degenerative illness of the neurological system defined by tremors, muscular rigidity, and slug gish, inaccurate movement, usually affecting middleaged and older adults. It is connected to brain atrophy in the basal ganglia an d a lack of Dopamine is a neurotransmitter. Cells in one of t movementcontrol regions of the body undergo a process called Unk nown factors that cause the brain to start dying. British physician James Parkinson originally identified PD in the 1800s early. It comprises five phases and affects more than 1 million people in India each year. This condition is more noticeable when it first a ppears, typically in a person's late fifties or early sixties. Symptoms of PD are found in up to 15 percent of people aged 6574, an d over 30 percent. In this paper, in order to improve the present condition, a web application is created in this study to forecast t he sickness. The patient's data must be entered and submitted to make this prediction. Next, the online tool determines whether t he person will get the sickness or not. Different machine learning methods, including logistic regression, KNN, Random Forest, SVM linear kernel, and Decision tree, were applied to increase prediction accuracy.

Index Terms- **Web Application, Logistic Regression, KNN, Random Forest, SVM linear Kernel, Decision Tree.**

I. INTRODUCTION

One of the most prevalent neurodegenerative disorders, Parkinson's disease (PD) affects 12 persons per 1,000 people and has a p revalence rate 1% of people over 60 make people. (2017) Tysnes and Storstein Between 1990 and 2016, the average number of people affected by PD worldwide increased from 2.5 million to 6.1 million due to general rates and an increase in the number of older people. 2018 (Dorsey et al.). According to Jankovic (2008), Parkinson's disease (PD) is a degenerative neurological disord er that affects many aspects of mobility, such as planning, initiation, and execution (Contreras.Vidal and Stlmach, 1995). Prior to mental and behavioral disorders, such as dementia, symptoms related to movement during its development may be observed, inc luding tremors, stiffness, and stiffness (Opara et al., 2012). (2012) Opara et al. Qualitative health (QoL), social interaction, and f amily relationships are all significantly affected by PD, which also has significant financial costs for both the individual and the community (Johnson et al., 2013; Kowal et al., 2013; Yang and Chen, 2017). (Yang and Chen, 2017; Kowal et al., 2013; Johnso n et al., 2013) Historically, diagnoses of PD have been performed using motor signals. Despite the appearance of significant sym ptoms of PD in clinical trials, most of the quantitative measures used to assess disease severity have not been properly researche d and validated (Jankovic, 2008). Despite the fact that many patients experience nonmotorized symptoms prior to the developme nt of Parkinson's disease (PD), such as psychological changes such as attention and planning problems, sleep problems, and abn ormal sensory experiences such as bad breath, these symptoms are usually not. specific. PD diagnosis only (Braak et al., 2003). (Postuma et al., 2015). (Zesiewicz et al., 2006) Nonmotorized symptoms have not yet been identified several conditions, even if s ome have been used as supporting diagnostic criteria.

II.RELATED WORK

Due to the cognitive and neuromuscular activity that underlies them [1], biomarkers extracted from human speech can provide in sight into neurological illnesses like Parkinson's disease (PD). About one million people in the US are thought to have PD, a pro gressive neurodegenerative condition for which 60,000 new clinical diagnoses are made every year. In the past, it has been chall enging to quantify PD, and physicians tend to emphasize some symptoms while neglecting others, depending mostly on subjecti ve rating scales. Voice can be utilized to identify and diagnose PD since the condition is characterized by a loss of motor functio n. The usefulness of employing supervised classification methods, such as deep neural networks, to correctly classify diseasebear ing individuals is examined in this research. With pathological postmortem examination as the gold standard, our peak accuracy of 85% provided by the machine learning models surpasses the average clinical diagnosis accuracy of nonexperts (73.8%) and th e average accuracy of movement disorder specialists (79.6% without follow-up, 83.9% after follow-up).

In this paper [2], Ioannis G. Tsoulos, Georgia Mitsi, Athanassios Stavrakoudis, and Spyros Papapetropoulos surveyed that insuffi cient symptom monitoring, abnormal access to care, and abnormal communication with health care providers all contribute to or health decisions. and patientrelated outcomes. With the potential to drastically change the nature of PD diagnosis, monitoring, an d treatment, recent advances in digital health strategies have made purposeful and remote monitoring of reduced vehicle perform ance possible. We have recently shown that iMotor, a cloudbased digital platform empowered by artificial intelligence, can discri minate against people with Parkinson's disease (PD) in healthy volunteers using a variety of physiological (HV) testing. This stu dy aims to provide the first evidence that artificial intelligence systems can make it possible to differentiate between PD patients. see the various aspects of the disease. within a cohort of PD participants. Here, data collected through the mobile application (M otor, Apptomics Inc., Wellesley, MA) was divided into two categories: Patient PD and HV, using the newly developed Neural N etwork Construction (NNC) method. This method was tested with various previously collected data, and the results were compar ed with conventional neural network training methods. Individual PD patients can be differentiated from HVs using the NNC alg orithm with 93.11 percent accuracy, and ON vs.76.5 percent accuracy in the OFF position. By generating comprehensive, reliabl e, and sensitive data sets that can be used to make medical decisions both during and between office visits, future use of digital f orums with the potential for intelligent design can improve treatment and clinical research. It is necessary to conduct additional AI-based research in large groups of patients.

Parkinson's disease prognosis has been attempted [3]. Parkinson's disease (PD), a chronic neurological condition, is the second m ost common disease in the world after Alzheimer's. Early illness detection is crucial for providing patients with appropriate treat ment for PD. However, there aren't any medical test(s) that can currently reliably diagnose PD. Computeraided diagnostic (CAD) systems provide a better technique to assist the physician in making the necessary datadriven decisions as a consequence. Man y studies have been conducted to back up the use of CAD to identify PD in its early stages. No comprehensive evaluations have been conducted on the use of AI technology to prevent PD. The study's goal was to examine and research the usage of neural net works in the diagnosis of Parkinson's disease. Methods: The PRISMA Extension for Scoping Reviews was used to conduct this s coping research (PRISMAScR). We examined both technical and medical databases, such as IEEE and PubMed, to discover the relevant publications. Three reviewers independently chose the papers to include and extracted the data from those studies. The d ata that had been retrieved was then combined using the storytelling approach. Results: 91 papers that met the study's inclusion r equirements were chosen from 1061 total submissions. Approximately half of the mentioned studies have employed artificial ne ural networks to diagnose PD. Numerous investigations included focused on the freezing of gait (FoG) Biomedical speech and si gnal datasets were the data types that were most frequently utilized to create and test these models. However, the included studie s also made use of MRI and CT scan pictures. In conclusion, neural networks are crucial to the fight against Parkinson's disease. This review found a lot of potential uses for neural networks, however, most of them are only for academic research.

In this review [4], they examined subjects who used mechanical learning (ML) methods to diagnose PD using motor function or test restingstate electroencephalography (EEG). Methods: Preferred Reporting Materials for Systematic Reviews and MetaAnaly sis (PRISMA) standards followed during the review. Prior to May 2020, all articles were included, and their key features and res ults were evaluated and recorded. Results: There were nine subjects in total. Two engine EEGs were used and seven EEGs in the rest circuit. In 83.3 percent of the studies, subbrand models were employed. The accuracy of the PD class was 62 percent. The p arameters downloaded to the EEG were very different and there was no uniform cleaning process. However, spectral features too k precedence. Conclusions: For the model to perform well in predicting classification, both the included features and model desi gn were required. On the other hand, EEG purification procedures differed significantly between investigations and had no effect on the results. A new and growing topic is the use of ML methods in EEG to differentiate neurodegenerative diseases.

III PROPOSED SYSTEM

The issue can be resolved with a low mistake rate by utilizing machine learning techniques. The input is the Parkinson's disease speech dataset from the UCI Machine learning library. By combining the spiral drawing inputs from healthy and Parkinson's dise ase subjects, our suggested approach also yields correct findings. We provide a method that accurately analyses patient data from spiral drawings and voice recordings. By comparing the two data, the doctor may determine if anything is normal or aberrant an d then give the appropriate medication for the afflicted stage.To identify PD, we employ the algorithms KNN, SVM, Decision Tr ee, Random Forest, and Logistic

IV METHODOLOGY

The following algorithms are being implemented in this project. SVM, kNearest Neighbors, Logistic Regression, Decision Treen dand Random Forest.

4.1 SVM (Support Vector Machine)

Support vector machines (SVMs) are potent but adaptable supervised machine learning algorithms used for both regression and c lassification.they are typically employed in categorization issues. Compared to other machine learning algorithms, SVMs are im plemented in a different method. Because they can manage numerous continuous and categorical variables, they have recently be come very popular. An SVM model is just a hyperplane in a multidimensional space that represents several classes. SVM will co nstruct the hyperplane in an iterative manner to reduce error. SVM seeks to discover a maximum marginal hyperplane by classif ying the datasets (MMH). Important SVM principles include the following: Support Vectors are the data points that are closest t o the hyperplane. These data points will be used to define the separating line. As seen in the picture above, a hyperplane is a dec ision plane or space that is partitioned between a collection of objects of various kinds. The distance between two lines on the ne arest data points of various classes is known as the margin. It may be computed as the perpendicular distance from the line to the support vectors. A large margin is seen favorably, whereas a small margin is unfavorable.

4.2 KNN (K-Nearest Neighbors)

A supervised machine learning (ML) technique known as KNearest Neighbors (KNN) may be applied to classification and regre ssion predicting issues.it is mostly employed in the industry for categorization and forecasting issues. The next two characteristic s would accurately describe KNN. KNN is a lazy learning algorithm since it uses all the data for training while classifying rathe than having a separate training phase. KNN is another example of a nonparametric learning algorithm since it makes no assumpt ions about the underlying data.

KNN Algorithm Operation

A new data point will be given a value depending on how closely it resembles the points in the training set according to the K-Nearest Neighbors (KNN)which employs "featuresimilarity" to forecast the values of new data points. With the aid of the proces ses below, we may comprehend how it functions-
Step 1: A dataset is necessary before we can build any algorithm.at the first stage of KNN both the training data's rows and the t est data.
Step 2: The next step is to select the K value or the closest data points. Any integer can be K.
Step 3: Complete the steps below for each point in the test data and calculate the distance using either the Euclidean, or Hamming distances between both the training data's rows and the test data. Euclidean is the approach that is most frequently u sed to compute distance. Sort them now in ascending order depending on the distance value. The top K rows from the sorted arra y will then be selected from this point on, the test point will be assigned a class based on the most prevalent class of these rows
Step 4: Finish

4.3 Logisticiregression

The probability of a target variable is predicted using the supervised learning classification technique known as logistic regressio n. There are only two viable classes since the goal or dependent variable is dichotomous in nature. Simply said, the dependent va riable is a binary variable, with data recorded as either 1 (which represents success/yes) or 0 (which represents failure/no). A log istic regression model makes mathematical predictions about P(Y=1) as a function of X. One of the most basic machine learning algorithms, it may be used for a number of categorization issues, including spam identification, diabetes prediction, cancer diagn osis, etc.
How logistic regression works: Binary or binomial logistic regression, in which the objective or dependent variable can only ha ve one of two potential types 1 or 0 is the most basic type of logistic regression. We can model the link between several predicto r variables and binary or binomial target variables using this method. The linear function is essentially utilized as an input to ano ther function, such as g in the following relation, in logistic regression.

$$g(z) = \frac{1}{1+e^{-z}} where\ z = \emptyset^T x$$

The logistic or sigmoid function, g, in this case, can be expressed as follows. − The graph below can be used to illustrate the sig moid curve. As can be seen, the y-axis values range from 0 to 1 and cross the axis at 0.5.

4.4 RandomiForest

The supervised machine learning method known as random forest is a wellliked approach for problems with classification and re gression. It builds decision trees from different samples, using their average for categorization and majority vote for regression. One of the most important features of the Random Forest Algorithm is its capacity to handle data sets including both continuous variables, as in regression, and categorical variables, as in classification. It offers better outcomes for classification problems. Th e steps of the random forest algorithm are as follows.
Step 1: From a data collection with k records, n random records are selected at random and used in the Random Forest algorithm
Step 2: For each sample, a unique decision tree is built.
Step 3: An output will be produced by each decision tree.
Step 4: For classification and regression, the final result is evaluated using a majority vote or an average.

4.5 Decision Tree

The supervised learning algorithms family includes the decision tree algorithm. The decision tree approach may be used to resol ve classification and regression issues as well, unlike some learning algorithms are monitored. By studying the specific decision rules found in the previous data, the Decision Tree is used to create a training model that can be used to predict class or volume variables (trainingdata). In the decision trees, we start at the root of the tree while waiting for the record class label. We compare the root values of the attribute to that of the attribute in the record. We follow the branch associated with that number and move on to the next node based on the comparison.Definitions of key Decision Tree terms:
Origin Node: The complete population or sample is represented by it, and this is then partitioned into two or more homogenous s ets.A node can be split into two or more sub-nodes by the process of splitting.
Decision Node: A sub-node is referred to as a decision node when it divides into more subnodes.
Terminal Node / Leaf: Leaf or Terminal nodes are nodes that do not divide.
Pruning: Pruning is the process of removing subnodes from a decision node. The process of separating in the opposite direction.
Branches and Sub-Trees: A branch or sub-tree is a section of the main tree.

System Architecture:

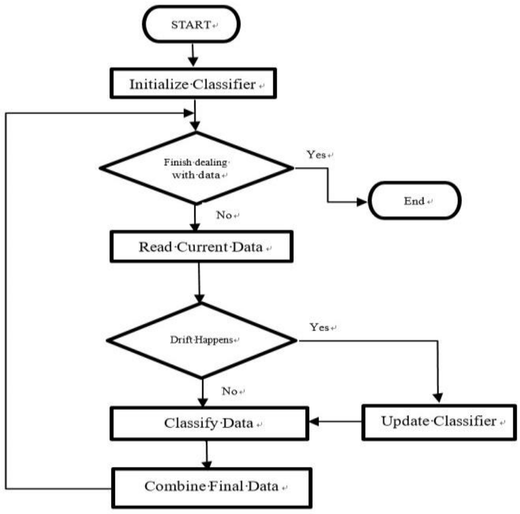

Fig 1: System Architecture

V RESULTS

Result Analysis of Algorithms:

The algorithms were successfully implemented on the data set we collected. The MSE, MAE, RSquared Parameter, RMSE, and Accuracy Values for each algorithm are calculated and displayed in the form of a graph

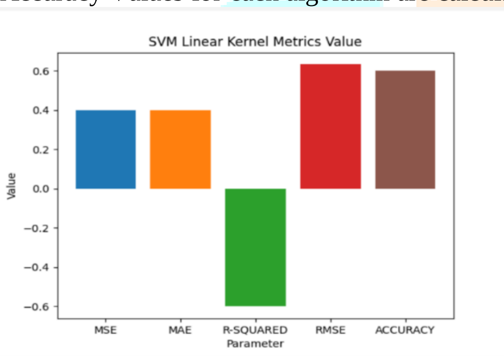

FIG:2 SVM LINEAR KERNEL METRICS VALUE

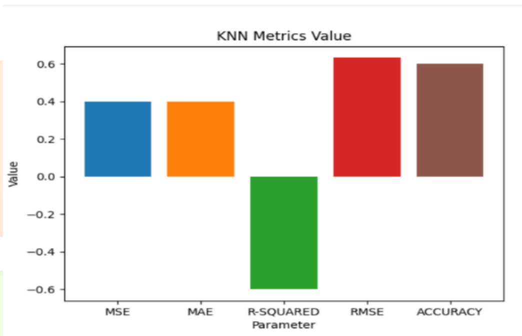

FIG 3: KNN METRICS VALUE

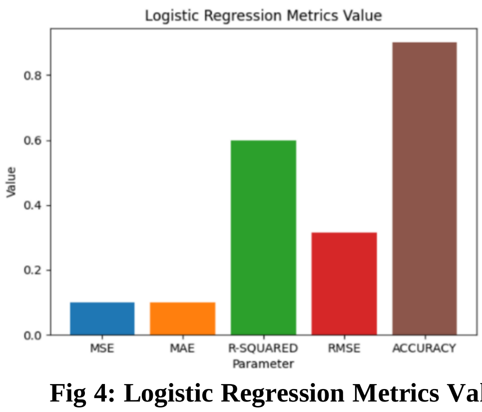

Fig 4: Logistic Regression Metrics Value

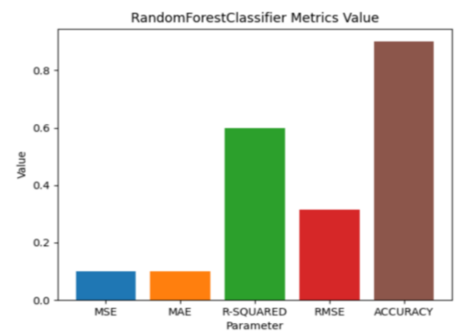

Fig 5: Random Forest Classifier Metrics Value

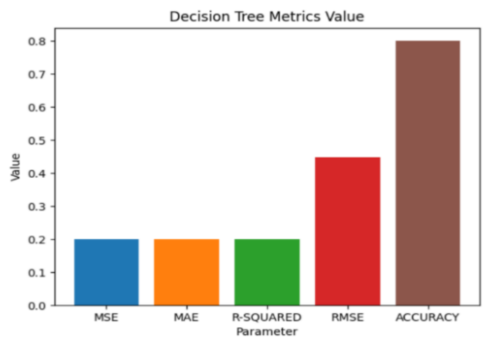

Fig 6: Decision Tree Metrics Value

SL.NO	Name of the Algorithm	MSE	MAE	R-Squared Parameter	RMSE	Accuracy
1	SVM	0.4	0.4	-0.60	0.6324	0.60
2	CNN	0.4	0.4	-0.60	0.63245	0.60
3	Logistic Regression	0.100	0.100	0.60	0.316228	0.90
4	Random Forest	0.0	0.0	1.00	0.0	1.00
5	Decision tree	0.20	0.20	0.20	0.447214	0.80

Fig: Metric Values of the Algorithms

As shown in the above table, the Error Values are compared with each other and the fewest error value algorithm with more Acc uracy is selected.In this prediction, Random Forest Classifier gives 100% Accuracy with fewer errors. So, we can use this algorit hm for the Web Application to predict Parkinson's Disease with 100% Accuracy.

VI CONCLUSION

To the best of our knowledge, this study is the first to include data from all studies that used machine learning techniques to diag nose Parkinson's disease. Here, we have provided access to information, including machine learning techniques used in the diagn osis of PD and related results from research included in the overall highquality view. In conclusion, although the adoption of ne w biomarkers may result in easier access to PD diagnoses in the past, the implementation of machineassisted PD diagnostics offe rs great promise for a highly structured clinical decisionmaking process. Therefore, machine learning strategies have the potentia l to give doctors many tools to diagnose, diagnose, or diagnose PD.

REFERENCES

[1].Parkinson's Disease Diagnosis Using Machine Learning and VoiceAuthor: Timothy J. Wroge; Yasin Özkanca; Cenk Demiro glu; Dong Si; David C. Atkins; Reza Hosseini Ghomi.
[2].Application of Machine Learning in a Parkinson's Disease Digital Biomarker Dataset Using Neural Network Construction (N NC) Methodology Discriminates Patient Motor StatusIoannisG.Tsoulos,GeorgiaMitsi,Athanassios Stavrakoudis and Spyros Papa petropoulos.
[3][The Role of Neural Network for the Detection of Parkinson's Disease: A Scoping ReviewTimothy Mahmood Saleh Alzubaid i, Uzair Shah , Haider Dhia Zubaydi , Khalid Dolaat ,Alaa A. Abd-Alrazaq ,Arfan Ahmed and Mowafa Househ.
[4].Machine Learning Approaches for Detecting Parkinson's Disease from EEG Analysis: A Systematic Review María Maitín , Alvaro José García-Tejedor and Juan Pablo Romero Muñoz.
[5]. Agarwal, Aarushi; Chandrayan, Spriha; Sahu, Sitanshu S (2016). *[IEEE 2016 International Conference on Electrical, Electr onics, and Optimization Techniques (ICEEOT) - Chennai, India (2016.3.3-2016.3.5)] 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - Prediction of Parkinson's disease using speech signal with Extreme Learning Machine. , (), 3776–3779.*
[6].Pooja Raundale;Chetan Thosar;Shardul Rane; (2021). Prediction of Parkinson's disease and severity of the disease using Mac hine Learning and Deep Learning algorithm 2021 2nd International Conference for Emerging Technology (INCET),
[7].Kolte, Ashish; Mahitha, Bodireddy; Raju, N V Ganapathi (2019). [IEEE 2019 International Conference on Emerging Trend s in Science and Engineering (ICESE) Hyderabad, India (2019.9.182019.9.19)]2019InternationalConference on Emerging Trend s in Science and Engineering (ICESE) Stratification of Parkinson Disease using python scikitlearn ML library.4.doi:10.1109/ice se46178.2019.9194627.
[8]. Aich, Satyabrata; Sain, Mangal; Park, Jinse; Choi, Ki-Won; Kim, Hee-Cheol (2017). *[IEEE 2017 International Conference on Inventive Computing and Informatics (ICICI) - COIMBATORE, India (2017.11.232017.11.24)] 2017 International Conference on Inventive Computing and Informatics (ICICI) A mixed classification approach for the prediction of Parkinson's disease using nonlinear feature selection technique based on t he voice recording.959962.* doi:10.1109/ICICI.2017.8365279.
[9].iAdaptiveiNeuroFuzzyiInferenceiSystemiAsiNewiReaTimeiApproachiForiParkinsoniSeizuresiPredictioniAlaDaher;iSallyiYas sin;iHadiiAlsamra;iHassaniAbouiAliiJanuaryi2022.
[10].iAccuracyibasediPerformanceiAnalysisiofiClassificationiAlgorithmsiusingiParkinsoniDiseaseiDatasetiJ.iDhinakaran;iT.iVel muruganMay2022Doi:https://doi.org/10.1109/ICAAIC53929.2022.9792981