EARLIER IDENTIFICATION OF PARKINSON'S DISEASE USING DEEP LEARNING ALGORITHM - AN EVOLUTIONARY APPROACH IN BIOMEDICAL APPLICATION

[1]A S Salma Banu
[1]Assistant professor, Department of Electronics and Communication Engineering, Aalim Muhammed Salegh College of Engineering, AVADI, IAF Chennai

Abstract

Parkinson's disease is a progressive condition characterized by muscle weakness and arm and leg tremors resulting from a central nervous system disorder affecting movement. One approach for detecting Parkinson's disease is through voice analysis, as changes in voice patterns can indicate the presence of symptoms. In this project, a model was developed to detect Parkinson's disease using voice data, achieving an efficiency of 73.8%. The model was trained using a large dataset that included recordings from individuals without Parkinson's disease and those previously diagnosed with the condition. Machine learning algorithms were employed to analyze the data. Sixty per cent of the dataset was used for training, while the remaining 40% was used for testing. By inputting voice data into the model, it could determine whether the person showed symptoms of Parkinson's disease or not. The dataset consisted of 24 columns, each representing symptom values for a patient, except for the "status" column. The "status" column contained values of 0 and 1, where 1 indicated the presence of Parkinson's disease, and 0 indicated normal conditions. The main objectives of this article were to provide an understanding of Parkinson's disease and to detect its early onset. Various machine learning algorithms, such as XG Boost, KNN Algorithm, Support Vector Machines (SVMs), and Random Forest Algorithm, were utilized to achieve these goals. The focus was on evaluating the motor function ability of patients with Parkinson's disease. This project's scope was to demonstrate the high accuracy of detecting Parkinson's disease in its early stages, highlighting the potential for early diagnosis and intervention.

Keywords: Parkinson's disease, machine learning, deep learning, diagnosis, differential diagnosis

Introduction

Parkinson's disease (PD) is a complex neurodegenerative condition predominantly affecting older individuals. It is characterized by a combination of core symptoms, including bradykinesia, rigidity, tremor, and postural instability. The clinical presentation of PD exhibits significant heterogeneity, posing challenges in designing effective clinical trials and providing accurate prognoses to patients. Variations in the age of onset, disease progression rates, treatment complications, and motor and non-motor symptoms contribute to this variability. The diverse phenotypic profiles within the PD population present considerable obstacles in clinical care and trial design. Addressing the individual variability necessitates large-scale, protracted, and costly trials primarily focused on detecting substantial effects [1]. This becomes particularly burdensome during early-stage trials, which hold the greatest potential for effective interventions. Consequently, defining distinct subcategories of PD and predicting disease progression can substantially enhance cohort selection, inform trial design, reduce expenses, and improve the ability to detect treatment effects. Previous efforts to classify PD subtypes have relied on clinical observations based on the age of onset or categorizations centred on prominent features. These classifications have included distinctions such as early-onset versus late-onset, slow-progressing "benign" versus fast-progressing "malignant," PD with or without dementia, or subtypes based on prevalent clinical signs like tremor-dominant or postural instability with gait disorder. However, these dichotomous differentiations fail to fully capture the disease's quantitative, intricate, and interconnected nature. A paradigm shift toward a data-driven, multidimensional approach is warranted to achieve a more precise representation of PD and its disease trajectory. This approach should consider the interplay of multiple features and enable tracking and predicting changes over time. By embracing such an approach, we can gain a more realistic understanding of PD and its progression, leading to improved clinical care, optimized trial design, and enhanced sensitivity in detecting treatment effects. This study presents our research on characterizing and predicting the clinical progression of Parkinson's disease (PD) [2]. Our first step involved constructing a multidimensional space that captures the disease's features and progression rates (velocity). Instead of relying on predefined concepts of symptom differentiation, we employed data dimensionality reduction techniques on complex clinical features observed 60 months after initial diagnosis. This allowed us to create a meaningful spatial representation of each patient's status at that time point. Once the spatial representation was established, we applied unsupervised clustering to identify distinct disease subtypes within this space. This analysis revealed three clinical subtypes corresponding to different groups of patients with varying progression velocities (slow, moderate, and fast progressors). The identified subtypes were validated and replicated in an independent cohort, confirming their robustness. Building upon successfully identifying disease subtypes within the progression space, we developed a baseline predictor capable of accurately predicting an individual patient's clinical group membership after 5 years.Additionally, we explored the predictive potential of biospecimen biomarkers and genetic information in identifying these subtypes. Our findings emphasize the value of machine learning as an additional diagnostic tool for identifying disease subtypes and projecting individualized progression rates based on model predictions. This approach holds promise for improving personalized treatment strategies and our understanding of PD progression.

.

LITERATURE REVIEW

The public transportation system is an important part of urban transportation, which is useful to meet the basic needs of citizens' trips; the basic travel needs of citizens can be met if the mode of public transportation isdesigned and possible to answer the highmobility of the community. With thedevelopment of technology, the publictransportation system refers to an intelligent system, which is proven to have significantly increased the level of public transportation services and the level of user satisfaction, namely by increasing the utilization of transportation resources and reducing passenger travel time [3]. The smart public transportation system utilizing Internet support, IoT, can be applied in various research fields. IoT is concerned with building a network of devices that supportthe internet to develop an intelligent environment, As with the use of mobile devices, which can provide the concept of busmovement in real-time so that it can overcomethe problem of tracking and monitoring buses manually. Bus tracking and monitoring are oneof the main problems in the public transportation sector. Bus tracking andmobile device monitoring can also be designed using IR sensors for passenger count,GPS modules for location tracking and a GSM module for passenger communication. So, the bus arrival time is usually estimated from the passenger ride time at the bus stop, so the lack of data can create difficulties in estimating the bus arrival time [4]. Traffic density, as one of the factors inhibiting bus movement through tracking and monitoring, has also been discussed. With IoT, the bus transportation system can utilize the mobile platform as a sensor to increase network coverageof bus movement information from each bus stop. In addition, a combination of technologies, such as Global Positioning Systems (GPS) and mobile devices, as part of IoT, can help passengers travel on public transportation. A bus transportation tracking system using GPS technology is needed to overcome bus tracking manuallybecause manual bus tracking is prone to errors.Several studies have discussed the problem of using IoT in bus transportation mode in the form of optimization problems using RFID as well as other technologies related to IoT, system performance analysis, use of cable, andsensor networks [5,6]**.**

METHODOLOGY:

In this collection, 31 patients withParkinson's illness provided biological voice measures(PD). Each column in the table represents a voicemeasure. Each row corresponds to one of 195voice recordings from these individuals ("name" column). "status" column is set to 0 for health, and 1, is the primary way to separate healthy persons from those with Parkinson's disease (PD) in the data. ASCIICSV format is used for storing the data in this project. One voice recording is represented in each row of the CSV file; it is estimated that there are between six and eight recordings per subject. Gathering and analyzing information from many different sources is known as data collection. This means the data must be acquired and kept in a way that makes sense for the business challenge. The reference The process of data mapping divides the collected dataset into 80 per cent training data and 20 per cent testing data. The data is split for the modelling dataset into training and testing sets to assign data points to the former and the remaining to the latter. A model is therefore trained using a training set and then applied to a test set. Our application may then be evaluated based on its performance. In the dataset, 8 columns are said to be Control Group (Primary Indicators), and the next 16 columns are said to be Secondary Indicators which act as dependent. Parameters to Primary Indicators; hence this will not be the primary contributors. It is a collection of techniques for giving scores to input features in a predictive model, indicating each item's relative significance in creating a forecast. In this module, the data obtained will be cleansed with data null or not applicable, and the unwanted columns from the dataset will be discarded. Data preparation is a crucial stage of the data mining process. Projects involving Data Mining and machine learning are particularly susceptible to the adage "garbage in, garbage out." Sometimes, data collection techniques are weakly regulated, resulting in unreliable results such as out-of-range or non-existent numbers. An ENTITY RELATIONSHIP(ER) Diagram is a chartthat visually represents the relationship between database entities. It is also known as an entity-relationship diagram. It uses data modelling that can help define business processes and can be used as the foundation for relational databases.ER Diagram modelis an organization's data storage requirements with three main components: entities, attributes, and relationships. Figure 1 shows sequence diagram.

SEQUENCE DIAGRAM:

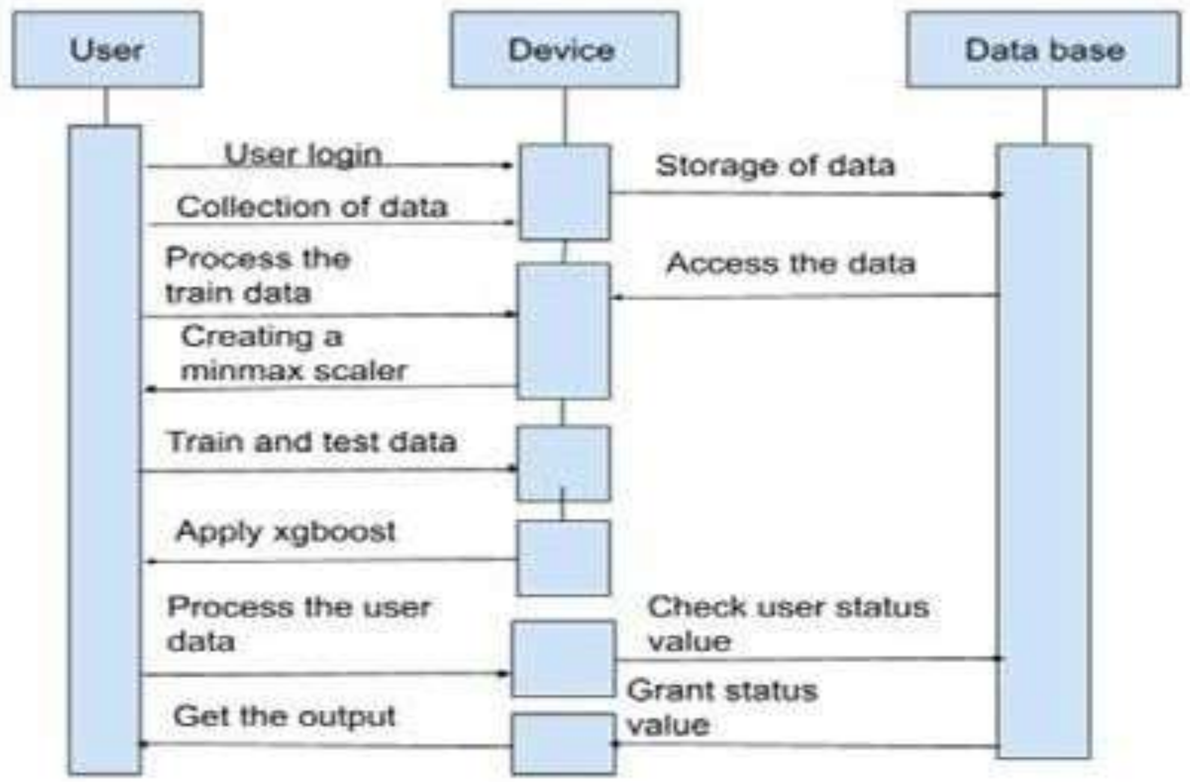

Figure 1: Sequence Diagram

Using data mapping, the collected dataset is separated into two parts: 80 per cent training data and 20 per cent testing data. The data has been separated into training and testing sets to allocate data points to the former and the latter in the modelling dataset. A model is therefore trained using a training set and then applied to a test set. Our application may be evaluated in this manner. The model is ready to detect Parkinson's Disease and predict it based on the given dataset. The data features obtained from the test are compared. Machine learning algorithms can only be fairly compared if assessed on the same data. We may force algorithms to be assessed on a uniform test harness when testing algorithms.XGBoost is an algorithm. That has recently been dominating applied gadget learning. XGBoost set of rules is an implementation of gradients boosted choice timber. That changed into the design for pace andoverall performance. It is a machine learning algorithm where the data is continuously split according to certain parameters.The two main functions are nodes and leaves, as we have imported. It is a machine learning algorithm based on the Bayes theorem for solving classification problems and making predictions. There are three types of models. Here we used Gaussian for prediction. So this is also imported from the sklearn library. Another algorithm for classification and regression analysis is the support vector machine. It isa supervised machine algorithm used. Image classification and hand-written recognition where the support vector machine comes to hand used. It sorts the data into one of two categories and displays theoutput with the margin between them as far as possible. Random forests are an ensemble version of many choice bushes, wherein each tree will specialize its focus on a specific feature while maintaining a top-level view of all capabilities. Each tree within therandom wooded area will do its own random train/check breakup of the information, referred to as bootstrap aggregation and the samples no longer covered are called the 'out-of-bag samples. Moreover, every tree will do characteristic bagging at every node-branch split to lessen the results of a characteristic mostly correlated with the response. While an individual tree is probably touchy to outliers,the ensemble version will no longer be the same. It is the level of software testing where individual unitsand components are tested. In the proposed project,the data of a person is taken and tested. The accuracy is 100% when tested with a single person's data. It may be a level of software testing where individual units are combined and tested as a gaggle. In the proposed project, all the data is combined and tested. The accuracy level is 94.87%. This testing will test thewhole project at a time. It reduces the time complexityin integration testing. Functional testing may be a sort of software testing thatvalidates the software against the functional requirements/specifications. This testing detects Parkinson’s will based on a machine learning algorithm.ML algorithm will boost up the speed. Functional testing typically involves the following steps: Identifying the functions the software is expected to perform. Create input data based on the function's specifications. It determines the output based on the function'sspecifications. Execute the test case. Compare the actual and expected outputs.

RESULT:

It is based on how accurate each algorithm is in detecting the disease that results in the results. Below Fig-2 and Fig-3 show the comparison chart of all ML and DL algorithms used in our project, which shows the slight difference between the algorithms in their accuracy and time stamp. Accuracy and performance comparison is shown in figure 2 and 3.

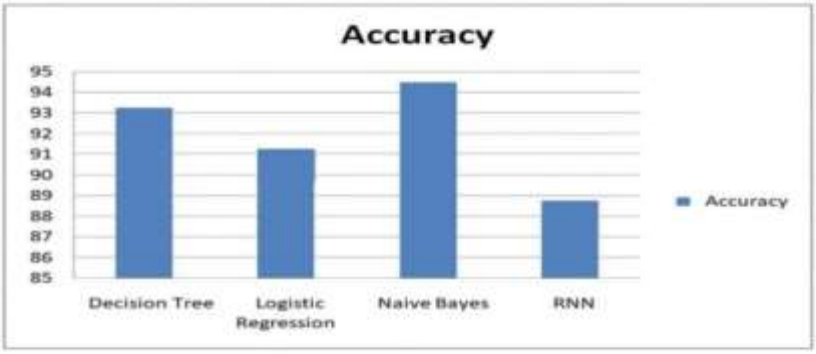

Figure 2: Accuracy comparison of all algorithms

PERFORMANCE COMPARISON:

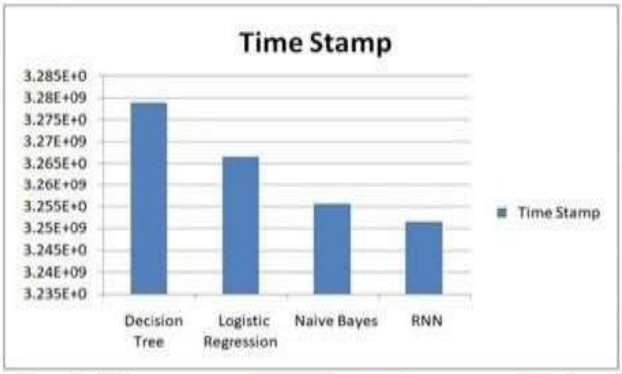

Figure 3: A performance comparison chart of all algorithm

CONCLUSION:

To think of an effective way to discover Parkinson's Disease, distinctive exploration papers were considered dependent on Parkinson's Detectionutilizing different Machine Learning and Deep Learning calculations to group subjects into the class of typical and dubious depending on different manifestations. The outcomes obtained from different investigations were looked at utilizing changed methods and reasoned that it is ideal for executing Deep Learning for discourse weakness and discovered the proposed model is more proficient and returns better precision. The results for algorithms based on accuracy are like this: Decision Tree 93.25%, Logistic Regression 91.25%, Naive Bayes 94.5%, and RNN 88.75%.

REFERENCES

[1] Prashanth R, Dutta Roy S, Early Detection Of Parkinson's Disease through Patient Questionnaire and Predictive Modeling, International Journal of Medical Informatics (2018),

[2] Srishti Grover,Saloni Bhartia, Akshama, Abhilasha Yadav, Seja K. R. "Predicting Severity Of Parkinson's Disease Using Deep Learning". International Conference on Computational Intelligence and Data Science (ICCIDS 2018). Procedia Computer Science 132 (2018)

[3] J. Farlow, N. D. Pankratz, J. Wojcieszek, and T. Foroud, "Parkinson disease overview," 2014.

[4] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic features identify Alzheimer's disease in narrative speech," Journal of Alzheimer's Disease, vol. 49, no. 2, pp. 407– 422, 2016.

[5] D. M. Huse, K. Schulman, L. Orsini, J. Castelli-Haley, S. Kennedy, and G. Lenhart, "Burden of illness in Parkinson's disease," Movement Disorders, vol. 20, no. 11, pp. 1449– 1454, 2005.

[6] Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms forhigh accuracy classification of Parkinson's disease,"IEEE Transactions on biomedical engineering, vol. 59, no. 5, pp. 1264–1271, 2012.