International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072

Detection of Parkinson's Disease Using Machine Learning and Deep Learning Algorithms

Shrihari K Kulkarni[1], K R Sumana[2]

[1]PG Student, Department of MCA, The National Institute of Engineering, Mysuru, Karnataka, India
[2]Assistant Professor, The National Institute of Engineering, Mysuru, Karnataka, India

---

Abstract - Neurological diseases, like as Parkinson's disease (PD), may be studied using biomarkers obtained from human speech. PD is a progressive neurodegenerative illness that affects around one million people. In the past, clinicians have relied on subjective grading systems to gauge the severity of Parkinson's disease. Difficulties with motor control make it possible to detect and diagnose PD via vocalization. Healthcare professionals could benefit from cheaper and more accurate diagnoses as a result of technological advancements and the widespread use of audio collecting devices in everyday life. We provide evidence to validate this concept here using a voice dataset collected from people with and without PD using Machine Learning algorithms: Decision Tree, Logistic Regression, and Naive Bayes and Deep Learning algorithm like Recurrent Neural Networks (RNN) by predicting with accuracy rate and performance comparison of all Machine Learning and Deep Learning algorithms.

Key Words: PD, Machine Learning (ML), Deep Learning (DL), Decision Tree (DT), Logistic Regression (LR), Naive Bayes (NB), Recurrent Neural Networks (RNN).

1. INTRODUCTION

Parkinson's disease is characterised by the death of dopaminergic neurons in the substantia nigra pars compacta of the midbrain (PD). Coordination problems, bradykinesia and voice alterations are among the signs of this neurodegenerative disease. Parkinson's disease (PD) patients can also develop dysarthria, an impairment of the motor-speech system that affects respiratory, phonatory, articulatory and prosodic functions. PD is often not identified for many years due to its variable symptoms and progression. Due to the fact that PD symptoms increase as the disease advances, more sensitive diagnostic techniques are needed for PD diagnosis. For example, a person with Parkinson's disease (PD) has a lack of intensity and monotony in pitch and loudness as well as a lowered level of stress (dysphonia). Because capturing speech data is non-invasive and easy to accomplish with mobile devices, the spectrum of voice-related symptoms appears promise as a possible screening technique. Early signs of Parkinson's disease are mild, making it difficult to identify.

Researchers have developed screening methods that use automated algorithms to distinguish between healthy controls and patients with PD because of the difficulties in diagnosing PD at an early stage. However, the model does not provide a type of differential diagnosis that would allow it to identify PD from a number of diseases that exhibit PD-like symptoms (e.g. Lewy-Body Dementia, Essential Tremor). The health care system and its patients are burdened by delays in diagnosis. PD is notoriously difficult to diagnose at an early stage, which has prompted scientists to create automated algorithms for screening to distinguish healthy controls from those with PD. These biomarkers are used to distinguish illness from control, however they are not used in differential diagnosis to determine whether PD is present among a range of diseases that exhibit PD-like symptoms. As a first step toward a long-term objective of providing clinicians with a decision support algorithm for screening patients for Parkinson's disease, the current research seems encouraging. In this research, we use the Power Voice dataset to distinguish PD from controls using a variety of machine-learning models and algorithms. Participants with and without Parkinson's disease were enrolled in Sage Bionetworks' Power clinical observational research, which collected digital biomarkers and health data via an iPhone app. Each participant was assigned a unique health code in order to protect user confidentiality and facilitate data linkage between datasets.

2. LITERATURE REVIEW

As a result of medical attention and clinical indicators, including the description of motor symptoms, Parkinson's disease (PD) is usually diagnosed. Due to the fact that they rely on the evaluation of motions that are often subtle to human sight and hence difficult to identify, traditional diagnostic methods may be vulnerable to subjectivity and misclassification.

As a result, these indications are typically ignored, making the early diagnosis of Parkinson's disease difficult to achieve. Machine Learning approaches have been used for the categorization of PD and healthy controls or patients with comparable clinical presentations in order to overcome these problems and enhance the diagnostic and assessment processes of PD. The major goal of the proposed system is to overcome the limitations of the present system and create a system that is accurate enough to diagnose Parkinson disease in its earliest stages. [1].

An extensive empirical evaluation of CNNs (Convolutional Neural Networks) has been implemented on large-scale image classification of gait signals converted to spectrogram images and deep dense ANNs (Artificial Neural Networks) on the voice recordings, to predict the disease with accurate objective diagnosis of Parkinson Disease in its early stages. [2]. We use the Patient Questionnaire (PQ) portion from the widely used Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) to develop prediction models that can classify early PD from healthy normal using machine learning techniques that are becoming popular in biomedicine: Logistic Regression, Random Forests, Boosted Trees and Support Vector Machine.

We carried out both subject wise and record-wise validation for evaluating the machine learning techniques. We observe that these techniques perform with high accuracy and high area under the ROC curve (both >95%) in classifying early PD and healthy normal. The logistic model demonstrated statistically significant fit to the data indicating its usefulness as a predictive model. It is inferred that these prediction models have the potential to aid clinicians in the diagnostic process by joining the items of a questionnaire through machine learning [3].

We presented the novel attitude of detection of Parkinson's disease based on facial features expression. Firstly, we elaborated on the features, which could differentiate healthy control group versus Parkinsonians.

The features describing differences in expressing fear during the time were the most significant from the statistical point of view. The XGBoost classifier outperformed other classifiers and achieved 0.69 balanced accuracy.

3. PROPOSED SYSTEM

In the below Fig-1, at the initial stage user will gather the dataset required to make a diagnosis of Parkinson Disease. Once the data has been loaded, system will preprocess the data and extraction of features will be done. Once the features required for the prediction have been extracted system will compare the features with model and prediction will be given as final result in the end. Graphical visualization were made to compare the performance parameters of all Machine Learning and Deep Learning algorithms considered in our work.

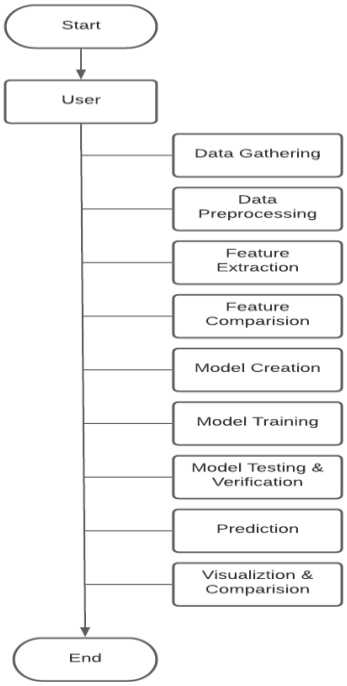

Fig-1 Block Diagram of the Proposed Work

4. IMPLEMENTATION

4.1 Data Collection

In this collection, 31 patients with Parkinson's illness provided biological voice measures (PD). Each column in the table represents a voice measure, and each row corresponds to one of 195 voice recordings from these individuals ("name" column). "status" column, which 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. ASCII CSV 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.

4.2 Data Loading

Gathering and analyzing information from many different sources is known as data collection. This means that the data we collect must be acquired and kept in a way that makes sense for the business challenge at hand. The reference process of data mapping divides the collected dataset into 80 percent training data and 20 percent testing data. The data is split for the modeling dataset into training and testing sets is to assign data points to the former and the remaining to the latter. A model is therefore trained using a training set, then applied to a test set. Our application may then be evaluated based on its performance.

In the dataset 8 columns is said to be Control Group (Primary Indicators) and the next 16 columns is said to be Secondary Indicators where it acts as a 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, which indicate the relative significance of each item in creating a forecast. Using a prediction model that has been fitted to the dataset, the most important scores are identified.

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 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.

4.2 Data Preprocessing

In this module, the data obtained will be cleansed with data null or not applicability and the unwanted colums from the dataset will be discarded. As part of the data mining process, data preparation is a crucial stage. 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.

4.3 Model Creation

In this module, the custer based on transaction and geolocations are done to identify the fradulent activity. This is a machine learning approach that divides unlabeled data into categories. A definition of it would be that it is "Data clustering is the process of arranging related data points into separate clusters. The objects with probable similarities stay in a group that has fewer or no similarities to another group." Shape, size, colour, behaviour, etc., are all examples of comparable patterns seen in the unlabeled dataset that are used to classify the data into groups based on the existence or absence of these patterns.

4.5 Training And Testing

With the use of data mapping, the collected dataset is separated into two parts: 80 percent training data, and 20 percent testing data. In order to allocate data points to the former and the latter in the modelling dataset, the data has been separated into training and testing sets. A model is therefore trained using a training set, then applied to a test set. Our application may be evaluated in this manner.

4.6 Prediction and Comparison

The model ready to detect Parkinson's Disease and predict based on the given dataset. The data features obtained from test is compared. Machine learning algorithms can only be fairly compared if they are assessed on the same data. When testing algorithms, we may force them to be assessed on a uniform test harness.

4.7 Visualization

The graphical visual representation created gives a user-friendly way to explore and comprehend data trends, outliers, and patterns in the data.

5. RESULT ANALYSIS

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

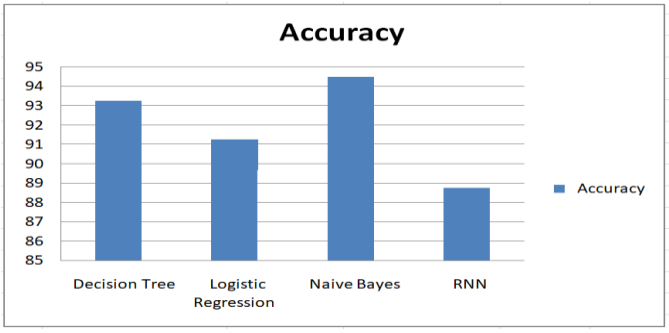

Fig-2 Accuracy Comparison Chart of all Algorithms

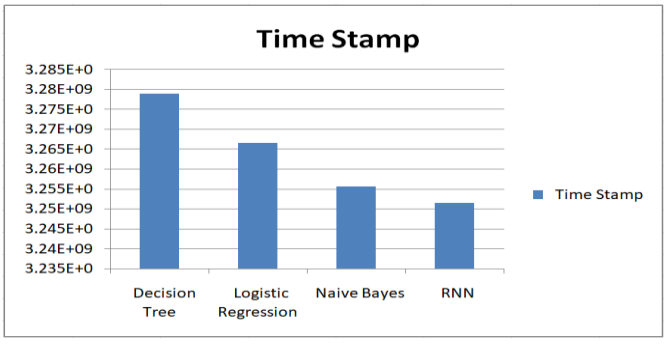

Fig-3 Performance Comparison Chart of all Algorithms

7. CONCLUSIONS

To think of an effective way for the discovery of Parkinson's Disease, distinctive exploration papers were considered are dependent on Parkinson's Detection utilizing different Machine Learning and Deep Learning calculations to group subjects into the class of typical and dubious dependent on different manifestations. The outcomes got by different investigations were looked at utilizing changed methods and reasoned that it is ideal to execute 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] "Parkinson Disease Detection Using Deep Neural Networks", Shivangi, Anubhav Johri and Ashish Tripathi, Department of Computer Science Jaypee Institute of Information Technology Noida, India.

[2] Prashanth R, Dutta Roy S, Early Detection of Parkinson's Disease through Patient Questionnaire and Predictive Modelling, International Journal of Medical Informatics (2018), https://doi.org/10.1016/j.ijmedinf.2018.09.008

[3] Justyna Skibińska, Radim Burget, "Parkinson's Disease Detection based on Changes of Emotions during Speech", Brno University of Technology Brno, Czech Republic, 2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)

[4] Srishti Grover, Saloni Bhartia, Akshama, Abhilasha Yadav, Seeja 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) 1788–1794.

[5] https://www.kaggle.com/margot234/parkinsons-disease-dataset

[6] https://archive.ics.uci.edu/ml/datasets/parkinsons