e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

A SURVEY ON PARKINSION DISEASE PREDICTION USING MACHINE LEARNING AND DEEP LEARNING

Kanagaraj S[*1], Harini KR[*2], Shahid Majeed Sheikh[*3], Sriram S[*4]

[*1]Assistant Professor, Department Of Information Technology, Kumaraguru College Of Technology, Tamil Nadu, India.

[*2,3,4]Students, Department Of Information Technology, Kumaraguru College Of Technology, Tamil Nadu, India.

DOI : https://www.doi.org/10.56726/IRJMETS32779

ABSTRACT

Parkinson's disease is a progressive, long-term neurological condition. The central nervous system of the human brain is disturbed by Parkinson's disease (PD), which can negatively impact a patient's quality of life. Basic cognitive abilities such as speaking, writing, walking, and others become increasingly difficult as dopamine-producing neurons in some areas of the brain are damaged or die. Patients' symptoms worsen with time, which causes them to become more severe. In this essay, we examined three key characteristics of PD: speech, gait signal measurements made by PD patients. The algorithms Deep Neural Network, K-Nearest Neighbour, Bagger Classification, Support Vector Machine, Learn Without Forgetting Approach, and Extreme Gradient Boosting Algorithm have all been compared here.

I. INTRODUCTION

Dopaminergic neurons in the substantia nigra pars compacta of the midbrain die out in Parkinson's disease (PD). issues with coordination and bradykinesia among the symptoms of this neurodegenerative illness are voice changes. Patients with Parkinson's disease (PD) may also experience dysarthria, a motor-speech system dysfunction that impairs respiratory, phonatory, articulatory, and prosodic processes. Due to its diverse symptoms and rate of development, PD is sometimes not discovered for many years. More sensitive diagnostic methods are required for PD diagnosis since PD symptoms worsen as the disease progresses. For instance, a person with Parkinson's disease (PD) has lessened stress and a lack of intensity and monotony in pitch and loudness (dysphonia). The healthcare industry is increasingly using machine learning techniques. Machine learning, as the name suggests, enables a computer programme to learn from data in a semi-automatic fashion and derive meaningful representations from it.

II. LITERATURE SURVEY

[1] The paper is about predicting the severity of Parkinson's disease Using Deep Learning. They have used voice data set of patients and used TensorFlow deep learning of python of neural networks to anticipate security.

[2] This writing is about Cloud-Based Diagnosis of Parkinson's using Machine Learning. This suggested method consists of two steps. To create a model that may be utilised for PD early therapy, two machine learning techniques, Random Forest (RF) and Long-Short-Term Memory (LSTM), are performed in the first step. A feature selection strategy is employed in this step to choose the smallest subset of the best characteristics that may be used subsequently to produce the classification model. The most effective diagnosis model is implemented in cloud computing at the second step. In comparison to the RF model, the LTSM model offers greater accuracy with 95% of the F-score.

[3] This talks about the predictive and detection analytics of Parkinson's disease with tremor analysis using Deep Learning Algorithm. The purpose of the study is to supplement clinical research by using a variety of standardised tremor rating measures created by neurological patients. They employed Random Forest and Convolution neural networks, and the CNN algorithm's accuracy was close to 96.97%.

[4] This proposal informs us about Deep Learning for the Diagnosis of Parkinson's Disease. This paper uses voice difficulties from the patient as dataset, It uses Random Forest Classifier for the maximum accuracy of 83.12%

[5] Here the papers tell us about Deep Learning-Based Parkinson's Disease Diagnosis. Here, the frequency, amplitude, and two voice components—NHR and HNR—from the voice data set are employed. Discriminating between healthy and PD-affected individuals is the goal. When using the test dataset, the accuracy was 95%.

[6] The paper suggests that Parkinson's diagnosis employing machine learning. The dataset was used from the PubMed and IEEE Xplore databased, with including an examination of their objectives, data sources, data kinds, machine learning techniques, and related results.

[7] Here the paper tells us about An Enhanced Machine Learning Approach for Parkinson's Disease Prediction. The cost of diagnosing PD patients using voice recognition and machine learning technology may be lower. We used the voice recorded dataset from the UCI machine learning repository for this study.In this research, they've put out an improved method of data pre-processing that improves prediction precision for PD diagnosis. With improved sensitivity, specificity, precision, F1 score, and kappa, we achieve 97.4% prediction accuracy, by utilising AdaBoost.

[8] The paper suggests about Machine Learning and Deep Learning for Parkinson's Disease Detection. This aims to identifies disease via vocalisation, Parkinson's disease (PD) can be detected and diagnosed due to difficulties with motor control. The algorithms used ere are Recurrent Neural Network, by all Machine Learning and Deep Learning algorithms are compared in terms of performance and accuracy rate for prediction.

[9] The paper Using deep learning to estimate the severity of Parkinson's illness. They suggest a deep learning strategy that makes use of sensors for measuring ground reaction force. Features are gathered from sensor data and supplied into a hybrid deep learning model. Convolutional neural networks and locally weighted random forest are both used in this model. The accuracy was about 98.7%.

[10 Parkinson's disease prediction from sound analysis using ensemble machine learning. In this paper, using machine learning classifiers, they compare the voice measurement aspects of the patient dataset in this study to determine if a patient has Parkinson's disease (PD) or not. Decision Trees, Logistic Regression, and K-Nearest Neighbours have been used as base classifiers, and their performance has been compared to that of Ensemble learning classifiers Bagging, Random Forest, and Boosting. We have examined the accuracy (%) of the classifiers and talked about whether one is better at predicting how the sickness would turn out. Additionally, they identified the classification's most significant properties and ordered them according to feature relevance.

[11] Machine Learning for Parkinson Disease Detection. This study investigates the efficacy of applying XG boost algorithm and other extreme gradient boosting techniques to precisely diagnose patients with the condition. This phrase is a fundamental description of a PD, according to researchers. more recent. Using a speech dataset gathered from individuals with and without Parkinson's disease, this effort gives proof to support this idea. The comparison was conducted primarily to clarify the findings and demonstrate that the XG boost algorithm offers more accuracy when compared to other algorithms

VOICE DATA SET:

D ATA COLLECTION:

Measurements of the biomedical voice make up the dataset. The data is in ASCII CSV format.

DATA PREPROCESSING:

The dataset was normalised using min-max normalisation to have a range of 0 to 1. Equation was used to do the column-wise normalisation.

$$\text{Normalized value of } x = x\text{-min}(x) / \max(x) \text{-min}(x)$$

ALGORITHMS USED:

DNN (DEEP NEURAL NETWORK):

The motor-UPDRS score for the dataset ranges from a minimum of 5.0377 to a high of 39.511, while the total-UPDRS score for the dataset spans from a minimum of 5.0377 to a maximum of 54.992. The train and test datasets were created by dividing the normalised dataset in two equal parts: 80% for the train dataset and 20% for the test dataset. The approach takes a dataset as input, creates an input pipeline around it, and then specifies iterators over it; these variables help in data set scanning. The suggested method also provides the option to shuffle the dataset in order to provide randomness. After building the input pipeline, the next step is to utilise a lambda function to feed the input data into the training model. The model conducts training, evaluation, and prediction after receiving data. By defining arrays of hidden layers with initialised weights, the processing system's hidden layer—which creates and stores models—is trained. Evaluate the generated DNN classifier last but not least. The DNN Classifier was built with TensorFlow, while Keras served as the backend. THE neural network's input layer comprises 16 units, whilst its three hidden layers each include 10 neurons, 20 neurons, and 10 neurons. Following that, the network received 1000 and 2000 further training steps, respectively.

KNN (K-Nearest Neighbour)

K-Nearest Neighbour, one of the simplest machine learning algorithms, is based on supervised learning techniques and may be used to solve classification or regression issues. It is frequently used to address classification problems. K is a key variable in the construction of a KNN classifier. The KNN approach inserts the new data in the category that resembles the current categories the most on the premise that the new data and the old data are comparable. Based on similarity, the KNN algorithm categorises new data points and preserves all previously collected data. This suggests that new data may be swiftly categorised into a fit category using the KNN approach. Since the KNN approach is robust to noisy training data, it can be more effective with large training data sets.

STEPS INVOLVED:

- Put the data in.
- Set up the value of k.
- KNN algorithm adaptation to the training
- To anticipate a data Measure the separation between each row of training data and the test data. Euclidean distance is applied here.
- Based on the distance value, sort the computed distance.
- Check the result's accuracy.
- Displaying the test-set outcome.

BAGGER CLASSIFIER:

The bagger classifier is additionally referred to as bootstrap aggregating. It is an ensemble meta-algorithm for machine learning that aims to boost the precision and stability of machine learning algorithms used in statistical classification and regression. This classifier is nearly identical to Random Forest, but it differs in that it takes into account all the features that can be split at each node. In Random Forest, only a subset of the total number of features is randomly chosen, and the best split feature from the subset is used to split each node in the tree.

XGBOOST CLASSIFIER:

The XGBOOST technique is often used in machine learning, whether the problem is one of classification or regression. It is recognised for its high performance when measured against all other machine learning algorithms. Extreme gradient boosting algorithm is the name of a method that uses decision trees. In comparison to existing gradient booster techniques, it speeds up the process by around ten times and has the highest option for accuracy.

SUPPORT VECTOR MACHINE (SVM):

A popular classification technique is the support vector machine, or SVM. Classification or regression issues can be solved using supervised machine learning techniques like the support vector machine (SVM). It generally deals with classification problems. Because SVM provides appreciable accuracy while consuming minimal computing power, it is widely used. SVM's main objective is to locate a hyper plane in an N-dimensional space with N features that clearly categorises the input points. The SVM classifier is the frontier that most effectively discriminates between the two classes (hyper plane/line). To obtain the largest marginal hyper plane, divide the dataset into classes using SVM (MMH). Its high degree of generalizability allows for effective use across a range of classification fields.

GAIT SIGNAL MEASUREMENTS:

DATASET:

This study employed the "Physionet Gait in Parkinson's Disease" public access dataset, which provides gait data. The dataset consists of gait signal measurements for PD patients and a control group. Patients with PD are 66.3 years old on average, and 37% of them are female. The control group's average age is 63.7 years, and 45% of its members are female. The experiments involve 93 PD patients and 73 control volunteers. Measurements of the gait signal were conducted while the individuals walked for two minutes. Gait signal measurements are taken when walking while doing a dual task and while walking normally. The subjects were told to proceed 20 metres on flat ground at their normal walking pace before turning around and going back to the beginning. Respondents in a dual tasking approach were directed to continuously subtract the number 7 from a specified number (for instance, 300, 293, 286, 279) while moving along the same path at their normal walking speed. Each foot has eight sensors connected to measure the gait signals. A function of time is used to derive the force value (in Newton) for each sensor. The Gait dataset contains a number of demographic information for each individual. The Unified Parkinson's Disease Rating Scale (UPDRS), a measure used to quantify numerical symptoms, is also included. It provides severity of PD ratings.

PROPOSED FRAME WORK:

The proposed framework is a hybrid deep learning approach that uses CNN and LWRF architectures. For LWRF, a locally weighted learning mechanism is required. Local connections used by the LWRF model only utilise a portion of the dataset. The locally weighted learning procedure creates a local model nearest to the query point rather than a global model for functional space. Each data point has been given a weighted value. Weight values are taken into consideration when computing a target value using a predictor approach. The nearest data points to the query have weighted values that are greater than the furthest ones. This approach is improved for the Random Forest model by locally weighting each data point while generating split points and utilising bootstrap samples for decision trees. The CNN is used to extract deep local features from stacked sensor signal channels. In order to extract deep local relationship information from CNN, this is done rather than manually creating signal characteristics. Our method intends to extract the key components of GRF signals by merging Convolutional layers with a collection of locally weighted random trees that employ relevant local descriptors that characterise data. Each gait sample yields a sixteen-channel sample. Constructed multichannel samples include 23 values along the x axis, which indicate seven frequency and sixteen time-domain features, and 16 channels, which correspond to the gait sensors on each foot. All samples have been converted into multi-channel samples and are now ready to be sent to a hybrid model. This hybrid model uses a layer of input, convolution layer blocks, and the LWRF technique to forecast the UPDRS values of the recovered local deep features. The model's CNN component has four convolutional layers and 40 filters altogether. These filters are available in 16 and 71 filter sizes, among others. These filter sizes are used by this model to extract the local vertical and horizontal gait characteristics.

While horizontal local features make use of relationships between time and frequency domain data, vertical local features make use of relationships between numerous GRF sensors. As a consequence, integrating these helpful local characteristics might produce a better representation than relying just on manually generated time and frequency domain features. A normalisation layer and a max pooling layer are components of the CNN design. The deep characteristics of the final convolution layer are sent into the LWRF model.

III. CONCLUSION

From the algorithms above, the test accuracy rate is achieved as follows, for KNN the rate is 0.95, for XGBOOST the rate is 0.95, for Bagging the arrived test rate id 0.92. But comparing the suggested DNN model to other methods already in use, accuracy was improved.

Experiments showed that when predicting the UPDRS value of samples, learnt deep features include more representative information than hand-crafted ones. Another finding from our research is that the LWRF model, with its locally weighted structure, is ideally suited to learning the connections between the numerous GRF sensors that CNN layers use.

IV. REFERENCE

[1] G.Priyadharshini, T.Gowtham, M.Harshavardhan Bhoopathi, M.Reshma, V.Tamilarasi, P.Nandhini Professor, Detection of Parkinson Disease Using Machine Learning Assistant Under Graduate Students, Department of Computer Science and Engineering, Coimbatore Institute of Technology ,Coimbatore ,India

[2] Shikha Singh, Nikita Shingade, Priti Sarote, Deepti Yelale and Nihar Ranjan parking's disease prediction using machine learing Information Technology Department, JSPM's RSCOE, Pune

[3] Krishna Prasad S J , Mohana Kumar S , Puttabore Gowda G,Swati Swami Early Prediction of Parkinson Disease Using Machine Learning and Deep Learning Approaches

[4] Mousumy Kundu An Optimized Machine Learning Approach for Predicting Parkinson's Disease Department of Computer Science and Engineering, Jashore University of Science and Technology, Jashore, 7408, Bangladesh

[5] Shrihari K Kulkarni1, K R Sumana of Parkinson's Disease Using Machine Learning and Deep Learning Algorithms PG Student, Department of MCA, The National Institute of Engineering, Mysuru, Karnataka, India Assistant Professor, The National Institute of Engineering, Mysuru, Karnataka, India

[6] Amit Kumar Patra et Prediction of Parkinson's disease using Ensemble Machine Learning classification from acoustic analysis To cite this article: al 2019 J. Phys.: Conf. Ser. 1372 012041

[7] Srishti Grover, Saloni Bhartia, Akshama, Abhilasha Yadav, Seeja K. R.Predicting Severity Of Parkinson's Disease Using Deep Learning Department of Computer Science and Engineering Indira Gandhi Delhi Technical University for Women, Kashmere Gate, Delhi, 110006, India

[8] Jie Mei, Christian Desrosiers and Johannes Frasnelli Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature

[9] Tunç Aşuroğlul Hasan Oğul2A deep learning approach for parkinson's disease severity assessment Received: 2 July 2022 / Accepted: 20 August 2022 © The Author(s) 2022