IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

Classification of Parkinson's disease using Machine Learning Techniques

Sakshi Jadhav[1], Seema Thorat[2], Sakshi Fokane[3], Rahul Chakre[4]
Students, Department of Computer Engineering[1,2,3]
Assistant Professor, Department of Computer Engineering[4]
Gokhale Education Society's R. H. Sapat College of Engineering, Management Studies and Research, Nashik, India

Abstract: *Biomarkers derived from human speech could be used to investigate neurological illnesses such as paralysis agitans (PD). PD is a neurodegenerative disease that affects about one million people. Clinicians have previously relied on subjective grading systems to determine the severity of Parkinson's disease. Due to difficulties with control, vocalization can be used to detect and diagnose Parkinson's disease. As a result of technology improvements and the widespread usage of audio gathering equipment in everyday life, healthcare providers may profit from less expensive and more accurate diagnosis. Using a medical data dataset obtained from people with and without Parkinson's disease, we provide evidence to support this theory. Decision Tree, Logistic Regression, and Naive Bayes are examples of Machine Learning algorithms, as are Deep Learning algorithms like Recurrent Neural Networks.*

Keywords: Deep Learning, Artificial Neural Network (ANN), Disease Prediction, Parkinson's disease, Parkinson's dataset

I. INTRODUCTION

Dopaminergic neurons in the substantia nigra pars compacta of the midbrain die out in Parkinson's disease (PD). The symptoms of this neurodegenerative disease include issues with coordination, bradykinesia, and medical data changes. Dysarthria, a motor-speech system deficit that affects respiratory, phonatory, articulatory, and prosodic processes, can develop in paralysis agitans (PD) patients. Because of its diverse symptoms and progression, PD is often misdiagnosed for many years. More sensitive diagnostic approaches are required for PD diagnosis due to the fact that PD symptoms worsen as the disease progresses. For example, someone with Parkinson's disease (PD) exhibits a lack of intensity and monotony in pitch and loudness, as well as a decreased stress level (dysphonia). Due to the fact that medical data is captured. The spectrum of medical data-related symptoms appears promising as a viable screening approach because gathering speech data is non-invasive and simple to do with mobile devices. The symptoms of paralysis agitans are subtle at first, making it difficult to detect.

Due to the difficulty in identifying PD at an early stage, researchers have developed screening approaches that employ automated algorithms to distinguish between healthy controls and individuals with PD. The model, on the other hand, lacks a type of differential diagnosis that would allow it to Differentiate PD from a number of illnesses that have PD-like symptoms (eg. Lewy-Body Dementia, Essential Tremor). Delays in diagnosis put the health-care system and its patients under stress. Because Parkinson's disease is notoriously difficult to detect early on, scientists have created automated screening algorithms to distinguish healthy people from those who have the disease. Researchers have devised screening methods that employ automated algorithms to discriminate between healthy controls and those with Parkinson's disease due to the difficulty of recognising the condition early on. On the other hand, the model lacks a type of differential diagnosis that would allow it to distinguish Parkinson's disease from other conditions a variety of disorders with symptoms that are similar to Parkinson's disease (eg. Lewy-Body Dementia, Essential Tremor). Diagnostic delays place a load on the health-care system as well as the people who utilise it. Because early detection of Parkinson's disease is notoriously difficult, scientists have developed automated screening algorithms to identify healthy controls from patients.

DatasetAttribute/Factors	Type
Patno	Binary
Cohort	Binary

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

Subgroup	Binary
Enrlpd	Binary
Enrlprod	Binary
Enrllrrk2	Binary
Enrlgba	Binary
Enrlsnca	Binary
Conpd	Binary
Conprod	Binary
Conlrrk2	Binary
Congba	Binary
Consnca	Binary
Comments	Binary
Condate	Binary

Table 1 Dataset Description

II. LITERATURE REVIEW

This section of the describes the literature survey, starting with an explanation of Parkinson's disease, followed by overviews of machine learning, deep learning, related work and finally PD diagnosis problems.

Gabriel Solana-Lavalle et al. [1] uses the algorithms such as Multilayer Perceptron (MLP), Random Forest (RF), K-Nearest Neighbor (KNN). For the prediction of Parkinson disease, three set of experiences were conducted to obtain the features with highest contribution to PD. This three sets are 1.a population with male and female subjects (balanced), 2.male subjects (balanced and unbalanced), and 3. Female subjects (balanced and unbalanced). In this study, the researchers used acoustic devices to collect speech parameters from 50 persons with Parkinson's disease and fifty healthy people. They employed the k-fold cross validation method for testing and claim that it can deliver 85 percent accuracy.

JieMei*et al.* [2]used all basic algorithms of deep learning techniques for the detection of PD.Like SVM, RF, Decision Tree, ANN, KNN, Radial Basis Function Networks (RBF) and Deep Belief Networks (DBN) etc.The early identification of Parkinson's disease is critical. The identification can be performed with the use of a data mining technique. The techniques for detecting PD, such as Naive Bayes, support vector machine, multilayer perceptron neural network, and decision tree, are theoretically explained in this study. This study uses speech input from acoustic devices to predict Parkinson's disease. People from various areas and speech factors are investigated in this article in order to predict Parkinson's disease among patients. The speech dataset was used to recognize Parkinson's illness using Multi - layer Perceptron and Logistic Regression (LR) frameworks.

KaziAmitHasan *et al.* [3] used different classification methods RF, KNN, Decision Tree, Logistic Regression (LR), SVM, and Naïve Bayes for detection of PD. The best result achieved by Decision Tree and Random Forest (RF) classification methods. The data mining techniques may be a more popular in many field of medical, business, railway, education etc. They are most commonly used for medical diagnosis and disease prediction at the early stage. The data mining is employed for healthcare sector in industrial societies.

MosarratRummane*t al.* [4] based on Image Processing and Artificial Neural Network (ANN) classification algorithm According to ANN prediction, if value closer to 1 then suggests PD and value closer to 0 then suggest normal. Parkinson disease is a global public health issue. Machine learning technique would be a best solution to classify individuals and individuals with Parkinson's sickness (PD). This paper gives an entire review for the forecast of Parkinson disease by utilizing the machine learning based methodologies. A concise presentation of varied computational systembased methodologies utilized for the forecast of Parkinson disease are introduced. This paper likewise displays the outline of results acquired by different scientists from accessible information to predict the Parkinson disease.

Yi Xia *et al.* [5]theyhaveconsidered approaches, they include four DL-based models (DCNN, DALSTM, DCLSTM, and CNN-LSTM) and also used two traditional classifications for extraction. In the DL-based model DCNN gives less accuracy than other DL models. Parkinson's disease affects people all around the world. People and people with

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

Parkinson's disease could be classified using machine learning approach. This paper provides a comprehensive overview of machine learning-based approaches for Parkinson disease prediction. A comprehensive overview of various computational system-based techniques for Parkinson disease prediction is presented. This report also includes an overview of the results obtained by several scientists from publicly available data in order to forecast Parkinson's disease.

ShailRaval *et al.* [6] For the detection of PD they include all the aspects such as biological data, chemical data and genetic data. In this paper they mainly focused on the symptoms like rigidity,Tremor at rest, changing medical data etc. The secure data transmission is proposed through authentication check, duplication check and faulty node detection. The proposed method is applicable to long ranges of transmission. It is also supporting a retransmission concept.

Zehra Karapinar Senturk et al. [7] proposed the algorithms to detect PD like support vector machine (SVM), Classification and Regression Tree (CART). It provided about 13% performance improvement for SVM, about 11% for ANN, and about 5% improvement for CART. The result shows that Naive Bayes and decision tree (j48) yield better accuracy when performed upon the discretized PD dataset with cross-validation test mode without applying any attributes selection algorithm.

Timothy J. Wroge *et al.* [8] used Extra Tree and gradient boosted Decision tree classification algorithms are used to detect variations in medical data. LSTMs are a kind of Recurrent Neural Network (RNN) which will learn and memorize long-term dependencies. Recalling past information for long periods is that the default behavior.

RajalakshmiShenbagaMoorthy*et al.* [9] used to novel analytic system for Parkinson's disease Prediction mechanism using Improved Radial Basis Function Neural Network (IRBFNN). RNNs is during a <"one amongst one in every of"> one among the deep learning models that are used for modeling the arbitrary length sequences by applying a transition function to all or any it's hidden states during a recursive manner.

SatyabrataAich*et al.* [10] According to this Random Forest (RF) gives more accuracy. This analysis will help the clinicians to differentiate the PD group from healthy group based on the medical datadata.CNN's, also referred to as ConvNets, contains multiple layers and are mainly used for imageprocessing and object detection. Yann LeCun developed the primary CNN in 1988 when it had been called LeNet. It was used for recognizing characters like ZIP codes and digits.

Rahul Ramesh Chakre *et al*. [11]According to the field of medical diagnosis, bio-inspired computing is also a novel technique. Swarm intelligence and immune computing algorithms, two major subsets of bio-inspired computation, are presented for a wide range of issues. For the feature selection process, swarm intelligence approaches are described, and immune computing techniques are proposed for the classification.

A. Tripati et al. [12] According to the 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.

Rahul R. Chakre et al. [13] According to the hybrid approach, which is a combination of supervised and unsupervised techniques, is also beneficial for classification and feature extraction. Support vector machine is employed as the supervised technique for classification, and ICA is used as unsupervised technique for the feature extraction in multiclass data set.

Prashanth R et al. [14] According to the 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 medical data recordings, to predict the disease with accurate objective diagnosis of Parkinson Disease in its early stages.

Srishti Grover et al. [15] According to this 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.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

WU WANG et al. [16] in this paper they used PPMI dataset to detect PD. They used varies Deep Learning methods to compare accuracy, Specificity, and Sensitivity. First method they used is FNN (forward neural network). Raw data is given into the supervised feed-forward neural network (FNN) addressed in this article, the deep learning algorithm can automatically extract hierarchical representations of the data that are best suited for the underlying learning objective, in our case, classification. They build a FNN with two hidden layers in this tutorial. Twenty neurons make up the first hidden layer, while ten neurons make up the second hidden layer. Forest at random (RF).

Justyna Skibińska et al. [17] According to this 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.

Mohamad Alissa et al. [18] According to this paper Parkinson's disease (PD) is a neurological condition that causes uncontrollable movements and other symptoms. It can be difficult to diagnose Parkinson's disease accurately since the signs and symptoms, especially early on, can be confused with other medical conditions or physiologicalchanges associated with normal ageing. This study aims to aid in the diagnosis of Parkinson's disease by employing a convolutional neural network, a type of deep neural network design, to distinguish healthy controls from PD patients.

They employed the DNN (Deep neural networks) algorithm to diagnose this condition. DNNs are multi-layer network models that convert complex, nonlinear, and unstructured input such as audio, video, pictures, and text into a hierarchical structure of features with multiple levels of abstraction.

Andrés Ortiz et al. [19] According to the research, for detection of PD they used CNN classifier. They used two 3D versions based on well-known architectures were tested in this paper. The first is based on LeNet, and the second is based on the most powerful AlexNet, both of which are fed with pre-processed data obtained through isosurface computing. In this research, they compare classification outcomes in the LeNet-based and AlexNet-based architectures when only a single input volume (isosurfaces) is used. This enables comparison of the performances of both architectures and identifying which isosurfaces provide more meaningful information.

Mahmood Saleh Alzubaidi et al. [20] In this paper CAD (computer-aided diagnosis) is used to diagnoses the PD. They select five different databases PubMed, IEEE, ACM, ScienceDirect, and Google Scholar. According to their research the main goal is to look at the involvement of brain networks in the diagnosis of Parkinson's disease. So that they divided neural networks into five categories: CNNs, RNNs, FNNs, ANNs, and other NNs, which were all used in the investigations.

III. CONCLUSION FROM LITERATURE SURVEY

Prediction of Parkinson's disease is one of the most critical issues that must be identified in the early stages of the disease's onset in order to limit the rate of disease development among individuals. Various studies have been conducted to determine the root cause, with some reaching new heights by proposing a method that uses machine learning techniques to distinguish healthy persons from those suffering from neurodegenerative illnesses. Several feature selection and classification strategies for pre-processing have been implemented and refined during the last few decades. Neutral networks are utilized to diagnose PD in its early stages.

IV. PROPOSED METHODOLOGY

- **Input:** This phase is crucial because the quality and quantity of data you collect will have a direct impact on the size of your prediction model. As a result, we've collected data from the patient's numerous medical data records.We use PPMI (Parkinson’s Progression Markers Initiative) dataset. This is input to the model. This dataset has all the impurities. This dataset contains medical data of persons who are infected by PD and normal persons. The PPMI is a landmark observational clinical study to comprehensively evaluate cohorts of significant interest using advanced imaging, biologic sampling and clinical and behavioural assessments to identify biomarkers of Parkinson’s disease progression

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

- **Data pre-processing**: In this step the info is visualized well to identify the connection between the parameters present within the data so on take the advantage of also as to get the data imbalances. With this, we need to separate the info into two parts. The first part for training the model like in our model we've used 70 percent of knowledge for training and 30 percentage for testing. The first step is to obtain the dataset. The input dataset is then truncated. Following that, all of the necessary libraries are imported. This assists in the pre-processing of the supplied dataset. Numpy, pandas, and matplotlib are the three main libraries used for pre-processing. The next step is to import the PPMI dataset. This is a crucial stage in the data pre-processing process. The following step after importing the dataset is to identify and handle the missing values, which entails finding and accurately processing the missing values. There are two approaches to dealing with deleting a specific row or determining the mean.
- **Feature Selection:** The next step in our workflow is Feature selection. There are various models that have been used till date by researchers and scientist. Some are meant for image processing, some for sequences like text, numbers or patterns. In our case we have defines the PD patient's samples from various patients so we have chosen such models which will classify or differentiates the unhealthy patient with the healthy one. In this step it choose model according to noise present in the data. In this process of decreasing the input variable to your model by just using meaningful data and removing noise. It's the process of selecting appropriate characteristics for your machine learning model based on the sort of problem you're attempting to answer automatically.
- **Training:** Training the dataset is one of the main task of machine learning. we will apply the data to Progressively improve the selected model's ability to predict better i.e. the actual result should be approx. to predict one. Feature extraction produces a model, which is then trained. Model fitting is utilised in the training process. Machine learning algorithms such as Random forest SVM ANN were used to train the model. Check the model's accuracy first, and then load the model for training with the specified dataset. The model will produce results after being trained. Then begin testing the model to see if it is producing the desired results. Follow the same procedure as for training while testing. If the outcomes of testing and training are the same, then our model is working correctly.
- **Model Selection:** first model is Artificial Neural Networks (ANNs) are a type of artificial neural network. ANNs are a type of machine learning algorithm that is designed to mimic the human brain. That is, similar to how neurons in our nervous system can learn from previous data, the ANN can learn from data and produce responses in the form of predictions or classifications. ANNs are nonlinear statistical models that show a positive link between inputs and outputs in order to produce a replacement pattern. Artificial neural networks are used for a number of applications, including image identification, speech recognition, MT, and diagnosis. Second model used to detect the disease is convolutional neural network (CNN). CNNs are deep learning algorithm which takes in an input image. Generally, A Convolutional neural network has three layers. And we understand each layer one by one with the help of an example of the classifier. With it can classify an image of an X and O.
- **Feature Extraction:** The metrics we've calculated are ROC, Accuracy, Specificity, Precision etc. which will highlight the simplest algorithm among all. This is the process of transforming raw data into numerical features that may be handled while keeping the information from the original data collection. This stage takes the model we chose for the prediction as input and converts all of the row data into numeric form. It then calculates precision, specificity, and sensitivity.
- **Prediction:**We eventually get the model ready to detect Parkinson's disease prediction based on the given dataset in this phase. It is the final step in the prediction process. This phase uses the results of all the previous steps and determines whether or not a person has Parkinson's disease.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

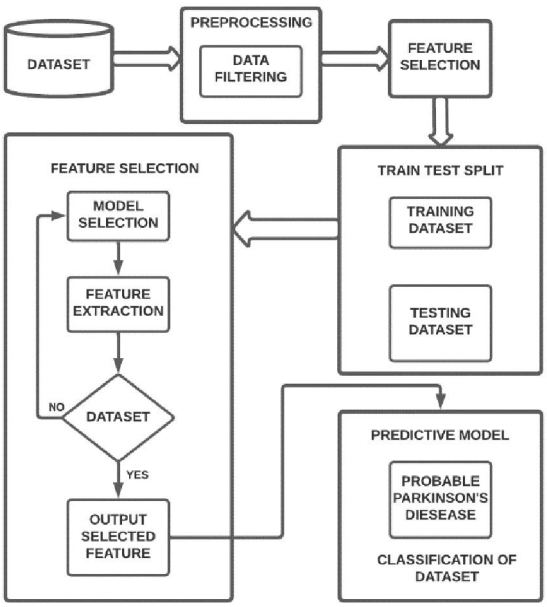

Fig 1. Block Diagram of Proposed Methodology

V. ALGORITHM DETAILS

5.1 Algorithm 1: ANN (Artificial Neural Network)

Artificial Neural Networks are a form of machine learning algorithm inspired by the human brain. That is, much as neurons in our nervous system can learn from previous data, the ANN may learn from data and respond in the form of predictions or classifications. Nonlinear statistical models, or ANNs, are nonlinear statistical models that demonstrate a complex relationship between inputs and outputs in order to uncover a new pattern. Artificial neural networks are used for a range of tasks, including image recognition, speech recognition, machine translation, and medical diagnosis. In the field of data mining, neural networks have a lot of applications. For example, pattern recognition in economics, forensics, and other fields.After thorough training, it can also be utilized for data classification in vast amounts of data. The following three layers can be found in a neural network:

- **Input layer:** The input layer's job is to take in the values of each observation's explanatory characteristics as input. An input layer's number of input nodes is usually equal to the number of explanatory variables. The patterns are presented to the network through the 'input layer,' which communicates with one or more 'hidden layers.' The input layer's nodes are passive, which means they don't affect the data. They take a single value from their input and duplicate it across all of their outputs. It repeats each value from the input layer and sends it to all of the hidden nodes.
- **The Hidden layer:** inside the network, apply provided changes to the input values Incoming arcs from other hidden nodes or input nodes connected to each node are included in this. It connects to output nodes or other hidden nodes through outgoing arcs. The real processing is done in the hidden layer via a system of weighted 'connections.' There could be one or more layers buried beneath the surface. Weights, a set of predetermined numbers contained in the programme, are multiplied by the values entering a hidden node. After that, the weighted inputs are summed together to form a single number
- **Output layer:** After then, the concealed layers are linked to a 'output layer.' Connections are received by the output layer from hidden layers or the input layer. It returns a value that corresponds to the response variable's prediction. There is frequently only one output node in classification issues. To produce the output values, the active nodes of the output layer integrate and alter the data.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

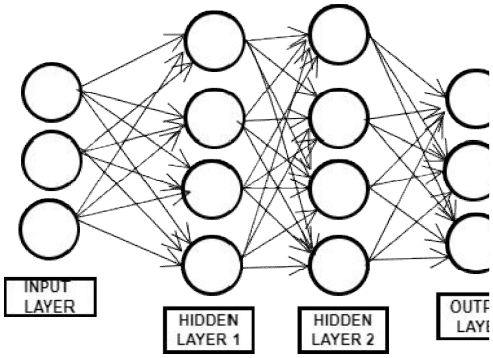

Fig 1. Architecture of Artificial Neural Network

5.2 Algorithm 2: CNN (Convolutional Neural Network)

Convolutional Neural Networks (CNN or ConvNet) are complicated feed forward neural networks used in machine learning. Because of its great accuracy, CNNs are utilized for picture categorization and recognition. Yann LeCun, a computer scientist, proposed it in the late 1990s, after being inspired by human visual perception of object recognition. The CNN uses a hierarchical model that builds a network, similar to a funnel, and then outputs a fully-connected layer in which all neurons are connected to each other and the output is processed. A Convolutional neural network has three layers in general. And we'll go over each layer one by one with the help of a classifier example It can be used to categories an image of an X and an O. So, given the case, we'll be able to comprehend all four layers. The layers of Convolutional Neural Networks are as follows:

- **Convolutional:** This is the initial stage in obtaining useful information from a photograph. The convolution action is performed by many filters in a convolution layer. Every image is viewed as a pixel value matrix.
- **ReLU Layer:** The rectified linear unit is abbreviated as ReLU. After the feature maps have been removed, they must be moved to a ReLU layer. ReLU performs an elementwise operation, converting all negative pixels to 0 pixels. It causes the network to become non-linear, and the result is a rectified feature map.
- **Pooling:** Pooling is a downsampling process that decreases the feature map's dimensionality. To create a pooled feature map, the rectified feature map is now sent via a pooling layer. The pooling layer employs a variety of filters to recognize various aspects of the image, including edges, corners, the body, feathers, eyes, and the beak.

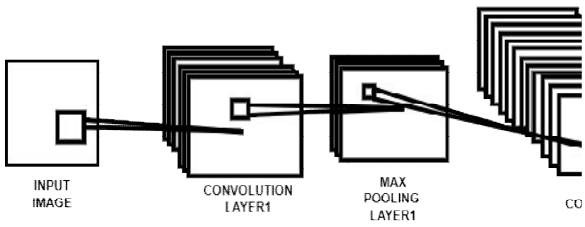

Fig 3. Architecture of CNN

5.3 Algorithm 3: SVM (Support Vector Machine)

The Support Vector Machine, or SVM, is a popular Supervised Learning technique that may be used to solve both classification and regression issues. However, it is mostly utilized in Machine Learning for Classification difficulties. The SVM algorithm's purpose is to find the optimum line or decision boundary for categorizing n-dimensional space into classes so that additional data points can be readily placed in the correct category in the future. A hyperplane is the optimal choice boundary. There are two types of SVM:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

- **Linear SVM**: Linear SVM is used for linearly separable data, which means that if a dataset can be classified into two classes using only a single straight line, it is called linearly separable data, and the classifier employed is called Linear SVM.
- **Non-linear SVM**: Non-Linear SVM is used to classify non-linearly separated data, which implies that if a dataset cannot be classified using a straight line, it is classified as non-linear data, and the classifier employed is the Non-Linear SVM classifier. The followings are important concepts in SVM
  - Support Vectors: Support vectors are data points that are nearest to the hyperplane. These data points will be used to define a separating line.
  - Hyperplane: As seen in the diagram above, it is a decision plane or space that is partitioned between a group of objects of various types.
  - Margin: It's the distance between two lines on a closet data point of different classifications. The perpendicular distance between the line and the support vectors can be computed. A large margin is seen as a good margin, whereas a small margin is regarded as a poor margin.

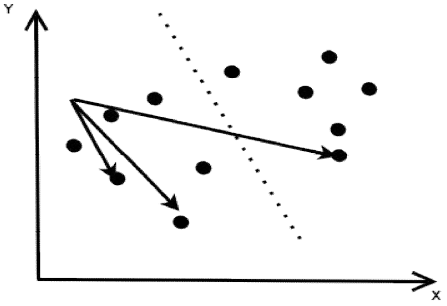

Fig 4. Working of Support Vector Machine

5.4 Algorithm 4: Random Forest

Random Forest is a well-known machine learning algorithm that uses the supervised learning method. In machine learning, it can be utilized for both classification and regression issues. It is based on ensemble learning, which is a method of integrating numerous classifiers to solve a complex problem and increase the model's performance. "Random Forest is a classifier that contains a number of decision trees on various subsets of a given dataset and takes the average to enhance the predicted accuracy of that dataset," according to the name. Instead than relying on a single decision tree, the random forest collects the forecasts from each tree and predicts the final output based on the majority votes of predictionsThe greater the number of trees in the forest, the more accurate it is and the problem of overfitting is avoided. The first part of Random Forest is to generate the random forest by mixing N decision trees, and the second phase is to make predictions for each tree created in the first phase.A random forest algorithm's building components are decision trees. A decision tree is a decision-making tool with a tree-like structure. A basic understanding of decision trees will aid our understanding of random forest algorithms. There are three parts to a decision tree: decision nodes, leaf nodes, and a root node. A decision tree method separates a training dataset into branches, each of which is further divided into branches. This pattern repeats until a leaf node is reached. There is no way to separate the leaf node any farther. The attributes utilized to forecast the outcome are represented by the nodes in the decision tree. The leaves are connected to the decision nodes. The three types of nodes in a decision tree are depicted in the diagram below.

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that winsthe majority votes.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

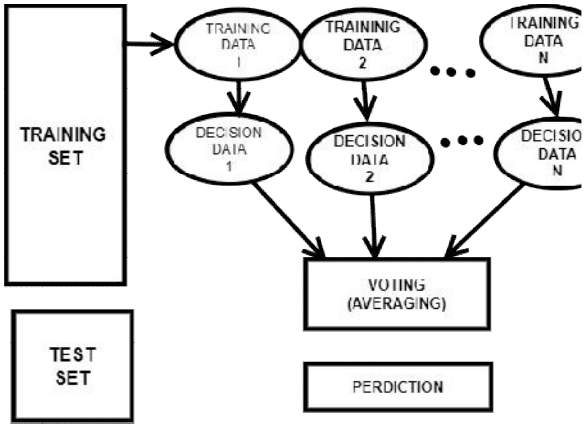

Fig 5. Working of Random Forrest

VI. DATASET INFORMATION

We use PPMI dataset for predication of Parkinson's disease. In the PPMI csv dataset there are 195 row and 24 columns. It contain some medical data. Columns are: Matrix column entries (attributes): name - ASCII subject name and recording number

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 x 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

Three nonlinear fundamental frequency variation measurements We used ANN, Random Forest, and SVM classifiers to train the model on this dataset. For prediction in image processing, we use an MRI dataset. We employed the CNN classifier for image processing.

Table 2. Comparison with different methods

Sr. No.	Method	Accuracy	Specificity	Sensitivity
1	Artificial Neural Network	97%	94%	96%
2	Random Forest	90%	96%	99%
3	Support Vector Machine	92%	94%	99%
4	Convolutional Neural Network	95%	97%	90%

VII. RESULT ANALYSIS

It is based on how accurate each algorithm is in detecting the disease that results in the final results. Below Fig-2 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. To evaluate the performance of machine method for discriminating Parkinson patients, we employ the following criterion:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, August 2022
Impact Factor: 6.252

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

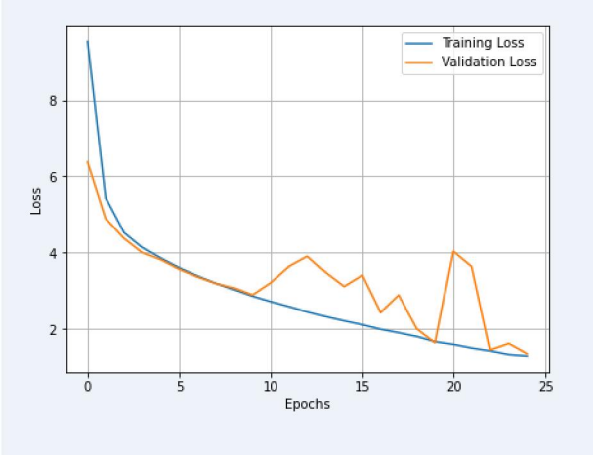

Figure: Accuracy Comparison Char to fall Algorithms

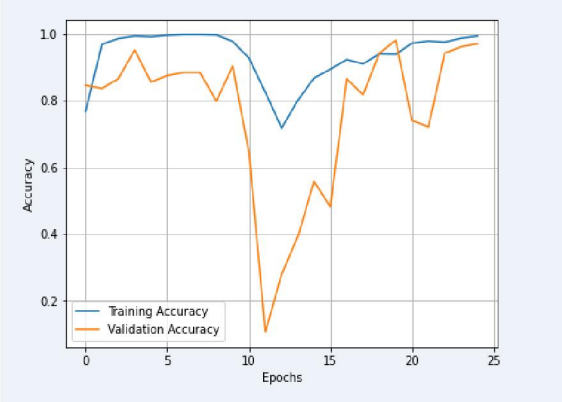

Figure: Confusion matrix

- Accuracy of ANN is 97%Specificity 94% and Sensitivity 96%
- Accuracy of Random Forest is 90%Specificity 96% and Sensitivity 99%
- Accuracy of SVM is 92%Specificity 94% and Sensitivity 90.4%
- Accuracy of CNN is 95%, Specificity 90% and Sensitivity 97%

VIII. CONCLUSION

We reviewed the ideas of Deep Learning and their application in Parkinson prediction in this paper. Deep Learning approaches have a considerable impact on early Parkinson's identification with a high accuracy rate, according to the studies included.However, the majority of the proposed approaches are still in the early stages of development and have yet to be evaluated in a clinical environment. Our research focuses on improving predictive models to attain 95 percent accuracy, 90% specificity, and 97 percent sensitivity in forecasting valid disease outcomes using deep learning methods. Convolutional Neural Network (CNN)-based prediction. Deep Learning approaches were proposed in this research for the early-stage prediction of Parkinson Disease.

REFERENCES

[1]. Gabriel Solana-Lavalle, “Analysis of medical data as an assisting tool for detection of Parkinson’s disease and its subsequent clinical interpretation”, ELSEVIER, Biomedical Signal Processing and Control, 2021.

[2]. Jie Mei, Christian Desrosiers and Johannes Frasnelli, “Machine Learning for the Diagnosis of Parkinson’s Disease” Front Aging Neuroscience, 2021.

[3]. Kazi Amit Hasan, “Classification of Parkinson’s Disease by Analyzing Multiple Vocal Features Sets”, Department of Computer Science & Engineering Rajshahi University of Engineering & Technology Rajshahi, Bangladesh, 2021, pp.758-761.

[4]. Mosarrat Rumman, “Early detection of Parkinson’s disease using image processing and artificial neural network”, Department of Computer Science & Engineering, BRAC University, Bangladesh, 2018.

[5]. Yi Xia, “A dual- modal attention-enhanced deep learning network for quantification of Parkinson's disease characteristics”, IEEE Transactions on Neural Systems and Rehabilitation Engineering, pp.1-12, 2019.

[6]. ShailRaval,“A Comparative Study of Early Detection of Parkinson’s Disease using Machine Learning Techniques”, Dept. of Information Technology Vishwakarma Government Engineering College Ahmedabad, India, pp. 509-516, 2020.

[7]. Zehra Karapinar Senturk, “Early Diagnosis of Parkinson’s Disease Using Machine Learning Algorithms”

Duzce University, Faculty of Engineering, Computer Engineering Department, Duzce, Turkey, ELSEVIER, pp.1-20,2020.

[8]. Timothy J. Wroge,"Parkinson's Disease Diagnosis Using Machine Learning and Medical data", Department of Bioengineering, University of Pittsburgh, Pittsburgh, Pennsylvania, USA 2018.

[9]. Rajalakshmi Shenbaga Moorthy, "Prediction of Parkinson's Disease Using Improved Radial Basis Function Neural Network", Madras Institute of Technology, Chennai 2021.

[10]. Satyabrata Aich, "A Nonlinear Decision Tree based Classification Approach to Predict the Parkinson's disease using Different Feature Sets Medical Data." International Conference on Advanced Communications Technology (ICACT). ISBN 979-11- 88428-01-4. ICACT 2018 February 11-14, 2018.

[11]. Rahul Ramesh Chakre, Dipak Patil "A Survey on Bio-inspired Computing on Medical Image Analysis" in THE JOURNAL OF ORIENTAL RESEARCH MADRAS, ISSN: 0022-3301, MARCH 2021.

[12]. Shivangi, Anubhav Johri, Ashish Tripathi "Parkinson Disease Detection Using Deep Neural Networks", Shivangi, Anubhav Johri and Ashish Tripathi, Department of Computer Science Jaypee Institute of Information Technology Noida, India.

[13]. Rahul R. Chakre, Dr. Radhakrishna Naik, "Performance Analysis of Hybrid (supervised and unsupervised) method for multiclass data set" IOSR Journal of Computer Engineering, Volume 16, Issue 4, Ver. III (Jul – Aug. 2014), PP 93-99.

[14]. 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

[15]. 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.

[16]. WU WANG, JUNHO LEE, FOUZI HARROU, (Member, IEEE), AND YING SUN "Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning" Statistics Program, Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia August 12, 2020

[17]. 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)

[18]. Mohamad Alissa1 • Michael A. Lones2 • Jeremy Cosgrove3 • Jane E. Alty4 • Stuart Jamieson3 • Stephen L. Smith5 • Marta Vallejo6 "Parkinson's disease diagnosis using convolutional neural networks and figure-copying tasks" College of Science and Engineering, Hamad Bin Khalifa University, Doha 53 September 2021

[19]. Andrés Ortiz, Jorge Munilla1, Manuel Martínez-Ibañez, Juan M. Górriz, Javier Ramírez and Diego Salas-Gonzalez "Parkinson's Disease Detection Using Isosurfaces-Based Features and Convolutional Neural Networks" Department of Communications Engineering, Universidad de Málaga, Malaga, Spain, Department of Signal Theory, Networking and Communications, University of Granada, Granada, Spain july 2019.

[20]. Mahmood Saleh Alzubaidi, Uzair Shah, Haider Dhia Zubaydi, Khalid Dolaat, Alaa A. Abd-Alrazaq , Arfan Ahmed 1 and Mowafa Househ "The Role of Neural Network for the Detection of Parkinson's Disease: A Scoping Review", College of Science and Engineering, Hamad Bin Khalifa University, Doha 53, Qatar; National Advanced IPv6 Centre, Universiti Sains Malaysia, Gelugor 11800, Malaysia, 16 June 2021