International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Parkinson Disease Detection Using XGBoost and SVM

Parag P Chinawale[1], Tanmay Maryapgol[2], Atharv Rayar[3], Jyoti Gaikwad[4]

[1,2,3]Student, Computer Engineering Department, Mumbai University, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India
[4]Professor, Computer Engineering Department, Mumbai University, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

---

Abstract - *Parkinson's disease (PD), is the second most common neurological disease that causes significant disability, lowers the quality of life, and has no cure. The nerve cells in this part of the brain are responsible for producing a chemical called dopamine. Dopamine acts as a link between parts of the brain and the nervous system that help control and coordinate body movements. As dopamine is usually the neurons in the cells begin to enter or complete another simple task. About 90% of people with Parkinson's disease have an early age of about 70 years, and the incidence increases dramatically. However, a small percentage of people with PD have a "first" disease that starts before the age of 50. More than 10 million people worldwide live with PD. There is no cure for PD today, but research is still ongoing and medication or surgery may offer significant improvements in motor symptoms.*

Key Words: Parkinson disease, Artificial Intelligence, Machine Learning

1.INTRODUCTION

Parkinson's disease (PD) is a chronic, debilitating neurological disorder. The main cause of Parkinson's disease is unknown. However, it has been studied that a combination of natural and genetic factors plays a key role in causing PD [1]. To put it bluntly, Parkinson's disease is treated as a disorder of the nervous system that results in the loss of cells in various parts of the brain. These cells include the substantia nigra cells that produce dopamine. Dopamine plays an important role in movement planning. []It acts as a chemical messenger to transmit signals within the brain. As a result of the loss of these cells, patients have difficulty moving.

In many cases, doctors find it difficult to assume that a given patient has already been diagnosed or is expected to have Parkinson's disease [4]. To overcome this, the development of a specific computer model should be performed by examining and summarizing the given patient data and predicting sufficient accuracy when undergoing PD development. Most PD patients are treated with symptoms called voice defects are known as dysphonia. There are several steps associated with dysphonia, in which a voice-related problem can be used to diagnose patients at various stages [14]. This paper is a PD prediction survey using machine learning and in-depth Artificial Intelligence strategies that produce good models and strengths of those algorithms in terms of acquired accuracy, and in terms of the various methods used.

2. LITERATURE SURVEY:

2.1 IMPORTANCE OF VOICE DATA:

Speech or voice data is considered 90% useful. To diagnose a person to detect the presence of a disease. In general, a PD person suffers from speech problems, which is they can be divided into two: hypophonia and dysarthria. Hypophonia indicates a soft, weak voice from the person and dysarthria shows a slow speech or voice, which can never understood simultaneously and this results in damage in the central nervous system. Thus, most therapists PD patients experience dysarthria and try to recover from it certain medications to improve the tone of voice.

3. METHODOLOGY

Upon pre-processing and loading the dataset the first attempt is to perform Exploratory Data Analysis to understand the data available and identify the important features for making accurate models. [3]This is done using Feature Importance analysis, a class of techniques for assigning scores to to find the important features from the given input data, each feature that is used while making a prediction. Feature importance analysis provides insight into the dataset. [2]The relative scores highlight which features is more relevant to the target. The most important scores are calculated by a predictive model that has been fit on the dataset. After preparing the data and gaining valuable insight about important features from the data is implementing various modeling techniques. The modeling techniques are implemented to test the accuracy of models with the training as well as test data. The various modeling techniques proposed for use are as follows-

3.1 SUPPORT VECTOR MACHINE:

Support Vector Machine is a new generation of learning a program based on the latest advances in mathematical learning vision. It is your algorithm for both linear and indirect data. [3]Converts original data to higher magnitude, from where it can find the top flight of data segmentation using important training tuples called support vectors. The

Support Vector Machine is a discriminatory category officially defined by separating the hyper plane. [5]In another names, given labeled training Vector vector support builds a high-rise aircraft or a set of high-altitude planes at high altitudes or an infinite-dimensional space, which can be used separation, retreat, or other activities. Intuitively, well the separation is achieved by a hyper plane with very large distance to the nearest training data center of any class called a functional margin, as it is usually undermines the separator generalization error The useful created data is then applied on the pre loaded SVM algorithm.

```
"from sklearn import svm\n",
"\n",
"#Create a svm Classifier\n",
"model2= svm.SVC(kernel='linear') # Linear Kernel\n",
"\n",
"#Train the model using the training sets\n",
"model2.fit(xtrain,ytrain)\n",
"\n",
"#Predict the response for test dataset\n",
"y_pred = model2.predict(xtest)"
]
```

Fig 1: SVM Code

3.2 XGBOOST:

XGBoost is another use of gradient boosting (GB) algorithm, based on the tree of determination as divides into categories. Used for its speed, efficiency, and proportions. By type, GB and XGBoost can be explained. [3]If we have D = [x, y] it represents data sets that contain n observation, i.e. x feature (independent variables) and y dependent variations. In GB, imagine that there is a value of k development, rather than having the B function of predicting the effect using Ÿ as a predictor of i-th sample b-th boost, means the formation of a tree q, which has a j-shaped leaf. [1]Then for a given sample, the final prediction can be determined by summarizing the points further all leaves, this is shown in Eq.1

$$\hat{y}_i = \sum_{b=1}^{B} f_b(x_i)$$

Fig -2: Formula for XGBoost

```
"source": [
 "from xgboost import XGBClassifier\n",
 "\n",
 "\n",
 "model=XGBClassifier()\n",
 "model.fit(xtrain,ytrain)\n",
 "predict=model.predict(xtest)\n",
 "\n"
]
},
{
 "cell_type": "code",
 "execution_count": 33,
 "metadata": {},
 "outputs": [],
```

Fig -3: Code for XGBoost

4. ANALYSIS:

The front end of the system is created using stremlit and the interface is kept simple for better understanding. The front end consists of various voice and speech related parameters which user needs to enter. The parameters are then processed at the backend and the output is displayed on the frontend. The accuracy of the system is 94.87%.

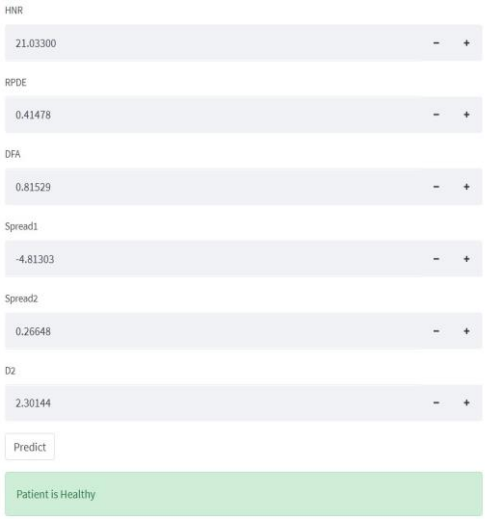

Fig 4: Frontend Output

Fig 5: Accuracy Level

5. CONCLUSION

The paper is an effort for providing the broader view of Parkinson Disease and the methods such as Machine Learning and Artificial Intelligence use to detect it. In the Methodology in above section the algorithms used provide a good amount of accuracy using the voice and speech parameters.[4] By the previous research it has been identifies that the new machine learning and artificial intelligence technologies can maximize the output if combined together.

REFERENCES

[1] Implementation of Xgboost for classification of parkonsion's disease: G Abdurrahman, M Sintawati – 2020.

[2] Detection of Parkinson's Disease by Employing Boosting Algorithms - Mirza Muntasir Nishat , Tasnimul Hasan , Sarker Mohammad Nasrullah , Fahim Faisal – 2021.

[3] Vocal Feature Extraction-Based Artificial Intelligent Model for Parkinson's Disease Detection – Muntasir Hoq , Mohammed Nazim Uddin , Seung-Bo Park – 2021.

[4] Akshaya Dinesh and Jennifer He, "Using Machine Learning to Diagnose Parkinson's Disease from Voice Recording", in Proceedings of the IEEE MIT Undergraduate Research Technology Conference (URTC), pp. 1-4, 2017.

[5] Ipsita Bhattacharya et al, SVM Classification to Distinguish Parkinson Disease Patients, Conference: Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India, September 16-17, 2010