International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING MODEL WITH VOICE IMPAIRMENT DATASET

Mrs.M.POORNIMA[1], SAVIREDDY NIKHILA[2], BALARAJU KOWSHIK VARMA[3], SHAIK PREETHI[4], SHAIK MOHAMED SHAFI[5], VUTUKURI VAMSI KRISHNA[6]

[1]Assistant Professor, Dept. of ECE, S V College of Engineering, Tirupati, A.P, India.

[23456]B.Tech Students, Dept. of ECE, S V College of Engineering, Tirupati, A.P, India.

ABSTRACT

Biomarkers derived from human voice can offer in-sight into neurological disorders, such as Parkinson's disease (PD), because of their underlying cognitive and neuromuscular function. PD is a progressive neurodegenerative disorder that affects about six million people across the globe, with approximately sixty thousand new clinical diagnoses made each year. Historically, PD has been difficult to quantity and doctors have tended to focus on some symptoms while ignoring others, relying primarily on subjective rating scales. Due to the decrease in motor control that is the hallmark of the disease, voice can be used as a means to detect and diagnose PD. With advancements in technology and the prevalence of audio collecting devices in daily lives, reliable models that can translate this audio data into a diagnostic tool for healthcare professionals would potentially provide diagnoses that are cheaper and more accurate. We provide evidence to validate this concept here using a voice dataset collected from people with and without PD. This paper explores the effectiveness of using supervised classification algorithm, such as Extreme Gradient Boost(XG Boost) machine learning model, to accurately diagnose individuals with the disease.

Keywords: Machine learning**,** Neurodegenerative, Supervised, Extreme Gradient Boost**.**

1.INTRODUCTION

Parkinson's disease (PD) [1] is a neuropathological disorder which deteriorates the motor functions of the human body. It is the second most common neurological disease seen after Alzheimer's

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

disease and it is estimated that more than one million people are suffering from PD in North America alone. In 1817, PD was termed as shaking palsy by Dr. James Parkinson. Various studies have shown that this number will rise in an ageing population as it is commonly seen in the people whose age is over 60. Parkinson's disease is characterized by the degeneration of certain brain cell clusters that are responsible for producing the neurotransmitters that include dopamine, serotonin and acetylcholine. The loss of dopamine's result in the symptoms like anxiety, depression, weight loss and visual problems. The other symptoms that can be seen in the people with Parkinson's disease are poor balance, voice impairment and tremor. Various research studies have shown that 90% of people who suffer from PD have speech and vocal problems which include dysphonia, monotone and hypophonia. Thus, the degradation of voice is considered to be as the initial symptom of Parkinson's disease. [2] The cause and cure of PD are yet unknown but the availability of various drug therapies offers the significant mitigation of symptoms especially at its earlier stages, thus improving the life quality of patients and also reduces the estimated cost of the Pathology. The analysis of voice measurement is simple and non-invasive. Thus, to track the progression of PD the measurement of voice can be used. For assessing the progression of PD, various vocal tests have been devised which include sustained phonations and running speech texts. The telemonitoring and telediagnosis systems have been widely used as these systems are based on speech signals which are economical and easy to use. Hence, in this paper, there is an attempt to explore a better machine learning based model for an early detection of PD from the voice samples of the subject.

2.LITERATURE REVIEW

Many studies have been conducted on the detection of PD, based on various symptoms like olfactory loss, voice impairment etc. Among the studied symptoms, most of the patients have been reported with vocal impairment and speech problems. [3] Max A. little et al suggested a novel technique for the classification of subjects into Parkinson diseased and control subjects by detecting dysphonia. In their work, pitch period entropy a new robust measure of dysphonia was introduced.Their methodology consisted

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

of three stages; feature calculation, preprocessing and selection of features and finally the classification. For the classification purpose, they used linear kernel support vector machine (SVM). Their proposed model achieved an accuracy of 91.4%.

To separate the healthy subjects from PD subjects, Ipsita Bhattacharya et al [4] used a tool for data mining known as weka. They used SVM, a supervised machine learning algorithm for the classification purpose. Prior to classification, the data preprocessing was done on the dataset. Different kernel values were used to get the best possible accuracy by applying libSVM. The linear kernel SVM produced the best accuracy of 65.2174%, whereas the RBF kernel and polykernel SVM achieved the accuracy of 60. 8696%.

In another work,B.E Sakar et al [5] suggested a model for differentiating the control subjects from the PD subjects. For classification, they used SVM and k-nearest neighbor (k-NN). For cross-validation, they used Summarized Leave-One-Out (s-LOO) and Leave-One-Subject-Out (LOSO). An accuracy of 82.50% was achieved by k-NN and an accuracy of 85% was reported on using SVM classifier.

Achraf Benba et al [6] aimed to separate the people with PD from the control subjects. In their work, the data comprised of 34 sustained vowels, which was collected from 34 people of which 17 were PD subjects. From each subject, 1 to 20 Mel-frequency cepstral coefficients (MFCC) were obtained. SVM with different kernel types was used for classification. LOSO was used as a cross-validation technique. The best accuracy of 91.17% was reported by linear kernel SVM on taking the top 12 MFCC coefficients.

For PD detection, the different speech signal processing algorithms were compared by C.O Sakar et al [7]. In their work, a new feature was introduced called as tunable Q-factor wavelet transform (TQWT). The effectiveness of TQWT outperformed the state-of-the-art speech signal processing methods that were used for the extraction of features in PD detection. On different feature subsets, different classifiers were used and using the ensemble techniques the prediction of the classifiers were combined. It was found that MFCCs and TQWT achieved the highest accuracies and thus are

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

considered as important features in the problem of PD classification. Also, the minimum redundancy- maximum relevance feature selection technique was used as a data preprocessing step. The highest accuracy of 86% was reported by RBF kernel SVM on all feature subsets.

Richa Mathur et al [8] suggested a method for predicting the PD. They used a weka tool for implementing the algorithms to perform preprocessing of data, classification and the result analysis on the given dataset. They used k-NN along with Adaboost.M1, bagging, and MLP. It was observed that k-NN + Adaboost.M1 yielded the best classification accuracy of 91.28%.

From the review above, it may be observed that various ML techniques have been applied in recent research works over Voice based PD detection. But it may be observed that in none of these works, the Ensemble based ML approaches like the Extreme Gradient Boost(XGBoost) were used for model construction, which now have been used in this work. The success of proposed machine learning model was also evaluated using various performance metrics like accuracy, precision, recall. These results were also compared with results obtained from various other ML models which were used in the recently reviewed works to establish the model's efficiency.

3. EXISTING METHOD

In The existing system, Parkinson's Disease is detected in a secondary stage which leads to medical challenges. Thus, mental disorders are been poorly characterized and have many health complications like thinking difficulties, depression and emotional changes, and sleep disorders. Parkinson's Disease is generally diagnosed with the following clinical methods such as MRI or CT scan, PET scan, and SPECT scan.

3.1. DISADVANTAGES

1. Low accuracy.
2. Disease prediction is not good.
3. Performance is low.

4.PROPOSED METHOD

This paper aims to develop a machine learning model containing a voice impairment classifier implementing the XG Boost algorithm for a cheaper and to get the accurate objective diagnosis of Parkinson's disease in its early stages. It is based on voice impairment. The dataset used in this paper has been taken from the UCI machine learning repository. This paper provides

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

insight into all the decisive features and takes some of these features as input to the implemented model which is used to predict the disease. The success of implemented machine learning model(XG Boost) was also evaluated by comparing these accurate results with results obtained from other ML models (Logistic regression and SVMClassifiers).

4.1. BLOCK DIAGRAM

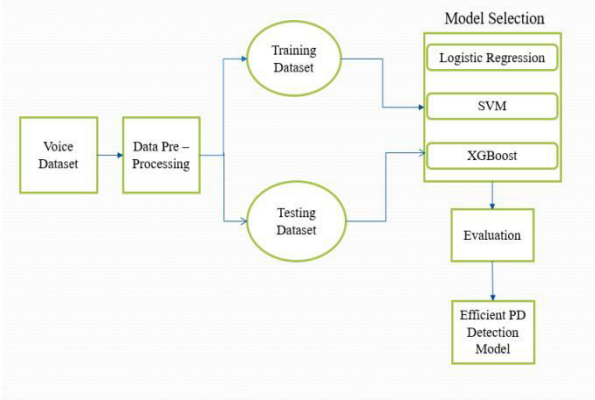


Figure-1: Block diagram for Parkinson's disease detection.

4.2. METHODOLOGY

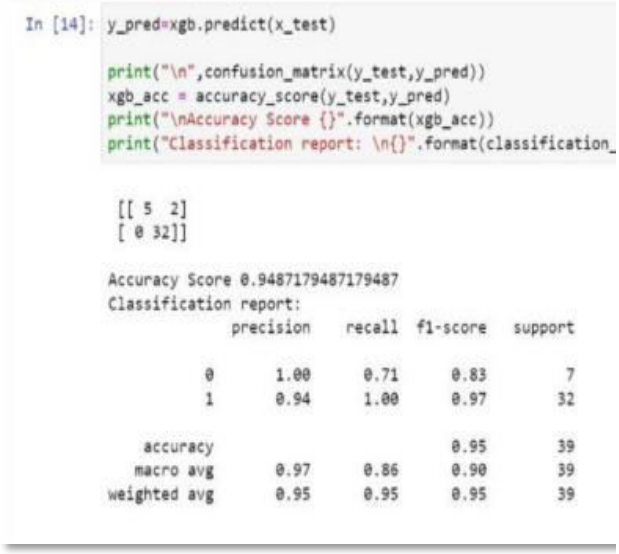
XG BOOST CLASSIFIER

In this Python machine learning paper, using the Python libraries scikit-learn, numpy, pandas, and XG Boost, we will build a model using an XGB Classifier. The XGBoost model for classification is called XGB Classifier. Initialize an XGB Classifier and train the model. This classifies using Extreme Gradient Boosting- using gradient boosting algorithms for modern data science problems. It falls under the category of Ensemble Learning in ML, where we train and predict using many models to produce one superior output. We can create and fit it to our training dataset. Models are fit using the scikit-learn API and the model.fit() function. To detect the presence of Parkinson's Disease in individuals using various factors. We used an XGB Classifier for this and made use of the scikit-learn library to prepare the dataset.

The paper will be made by a new machine learning algorithm called the XG Boost. XG Boost is a new Machine Learning algorithm designed with speed and performance in mind. XG Boost stands for Extreme Gradient Boosting and is based on decision trees. In this paper, we will import the XGB Classifier from the XG boost library; this is an implementation of the scikit-learn API for XG Boost classification. To build a model to accurately detect the presence of Parkinson's disease in an individual.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

4.3.SIMULATION RESULT



```
In [14]: y_pred=xgb.predict(x_test)

         print("\n",confusion_matrix(y_test,y_pred))
         xgb_acc = accuracy_score(y_test,y_pred)
         print("\nAccuracy Score {}".format(xgb_acc))
         print("Classification report: \n{}".format(classification_

          [[ 5  2]
          [ 0 32]]

         Accuracy Score 0.9487179487179487
         Classification report:
                       precision    recall  f1-score   support

                    0       1.00      0.71      0.83         7
                    1       0.94      1.00      0.97        32

             accuracy                           0.95        39
            macro avg       0.97      0.86      0.90        39
         weighted avg       0.95      0.95      0.95        39
```

Figure-2: Output Screenshot of Proposed Model.

The values of accuracy of Voice Impairment Classifier for detailed features of speech were evaluated with different epochs.The implemented method Extreme Gradient Boost(XG Boost) gives better accuracy 94.87% as compared to other machine learning algorithms such as Support Vector Machine(82.05%), Logistic Regression(89.7%).

Table-1: Comparative Analysis of various models for Parkinson's disease detection

Sl. No	ModuleName	DataSplitting(Training–Testing)	Training Epochs	Accuracy Comparisons		
				LR	SVM	XGBoost
1	VoiceImpairment Classifier	156-39	35	89.7%	82.05%	94.87%

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

4.4.ADVANTAGES

1. High accuracy.
2. Disease prediction is good, decreasing misdiagnosis rate.
3. Performance is High.

5.CONCLUSION

The analysis of which algorithm provide the high accuracy of prediction for the Parkinson's disease dataset, here the classification accuracy was studied and compared, with good performance and fast implementation XGBoost with multiple fold data achieved a high accuracy with 94.8%. This system provides the comparison between machine learning classifiers of Logistic Regression(LR), Support Vector Machine(SVM) and Extreme Gradient Boost(XGBoost) in PD disease diagnosis with high dimensional data.

6. FUTURE SCOPE

In future work, we can focus on different techniques to predict the Parkinson disease using different datasets. In this research, we using binary attribute (1- diseased patients, 0-non-diseased patients) for patient's classification. In the future we will use different types of attributes for the classification of patients and also identify the different stages of Parkinson's disease.

REFERENCES

[1]. Van Den Eeden, S. K., Tanner, C. M., Bernstein, A. L., Fross, R. D., Leimpeter, A., Bloch, D. A., & Nelson, L. M. (2019). Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. American journal of epidemiology, 157(11), 1015-1022.

[2]. Singh, N., Pillay, V., & Choonara, Y. E. (2020). Advances in the treatment of Parkinson's disease. Progress in neurobiology, 81(1), 29-44.

[3]. M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," IEEE Trans. Biomed. Eng., vol. 56, no. 4, pp. 1010–1022, 2009.

[4]. Bhattacharya, I., & Bhatia, M. P. S. (2010, September). SVM classification to distinguish Parkinson disease patients. In Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India (p. 14). ACM.

[5]. Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., ... & Kursun, O. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics, 17(4), 828-834.

[6]. Benba, A., Jilbab, A., Hammouch, A., & Sandabad, S. (2015, March). Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease. In 2015 International conference on electrical and information technologies (ICEIT) (pp. 300-304). IEEE.

[7]. Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., ... & Apaydin, H. (2019). A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. Applied Soft Computing, 74, 255-263.

[8]. Mathur, R., Pathak, V., & Bandil, D. (2019). Parkinson Disease Prediction Using Machine Learning Algorithm. In Emerging Trends in Expert Applications and Security (pp. 357-363). Springer, Singapore.