

# A new Diagnosis using a Parkinson's Disease XGBoost and CNN-based classification model Using ML Techniques

Anil Kumar N  
Assistant Professor

Department of Electronics & Instrumentation Engineering  
Sree Vidyanikethan Engineering College, Tirupati, Andhra  
Pradesh,

e-mail: anilkumar.n@vidyanikethan.edu

Ms. P.Anitha,

Assistant Professor,

Department of Information Technology,  
Hindusthan College of Engineering and Technology,  
Coimbatore

e-mail: panitha.rs@gmail.com

Ms. K Komala Devi

Assistant Professor

Department of Computer Applications  
CMR Institute of Technology, Bangalore,  
e-mail: komala.k@cmrit.ac.in

Ms. Bhavini Rajendrakumar Bhatt,  
Assistant Professor,

R. N. G. Patel Institute of Technology-RNGPIT,  
Gujarat

e-mail: brbhattacha@rngpit.ac.in

Mr. Ajay Kumar Yadav

Assistant Professor

Department of Computer Application  
United Institute of Management, Prayagraj Uttar Pradesh,  
India.

e-mail: ajaykumar.uim@gmail.com

Mr. Vivek Chetanbhai Joshi,

Assistant Professor,

R. N. G. Patel Institute of Technology-RNGPIT,  
Gujarat

e-mail: vcjoshi@rngpit.ac.in

**Abstract** - Parkinson's disease (PD) is a neurological condition that affects the brain of the human body and causes difficulty walking, shaking, stiffness, and loss of balance and coordination. Most of the patients suffering from PD face challenges in speaking during the initial stages. In this study, illness has been classified by applying speech features. The standard speech components employed in Parkinson's Disease are Shimmer, Jitter, Harmonic parameters, Frequency parameters, Detrended Fluctuation Analysis (DFA), Recurrence Period Density Entropy (RPDE), and Pitch Period Entropy (PPE) (PD). These features are the baseline features chosen for this study. CNN and XGBoost have been selected to classify the model and recognize Parkinson's Disease in the early stages. From the model feature, the selection was excluded to improve the model.

**Keywords:** Parkinson's Disease, neurodegenerative, XGBoost, CNN, and brain

## I. INTRODUCTION

Learning models are the advanced form of machine learning that focuses on the constraints present in machine learning. Machine learning is used. Parkinson's Disease is classified in this way. Suitable learning criteria are utilized to

anticipate the class label to detect the condition early. Image processing, text, and documents are used for classifying the problems. Various complex data are present globally with different features (Thongsuwan et al., 2021). Currently, deep learning is used as a learning model in multiple fields such as computer vision. Learning models are used for solving the problems of machine learning problems. Vast amounts of information are present on the internet, making models challenging to solve the problem. Therefore, various algorithms recognize the pattern and solve the classification problems. Choosing an effective model to complete the models and deep learning is essential. Currently, deep learning and traditional shallow models are used as a single model for the training. Two hypotheses are used in this article to classify Parkinson's Disease (Nishat et al., 2021). The accuracy for various problems is not met through the capabilities of the single model. The model used in the study has both advantages and disadvantages. We have chosen each model's advantages and rejected the weaknesses by analyzing advantages and disadvantages. In this article, we have used CNN and XGBoost algorithms to attain extreme precision and cutting-edge performance. These two models have been used to solve the classification problem of Parkinson's Disease. Data scientists use XGBoost to avoid overfitting and for providing tree boosting. In machine learning competitions, XGBoost is used to classify the problems.

Convolutional Neural Network has been used for deep learning that contains hierarchical learning differently. Various classification problems are used in the study to show whether XGBoost satisfies the goal of our problem or not (Lamba, Gulati & Jain, 2020). The significant contributions in the paper are Cov XGB which is the combination of CNN and XGBoost used for the classification of the problem. Various stacked convolutional layers consist of the ConvX GB architecture, and XGboost is used as the last layer of the model. CNN used in this study is different from the traditional CNN as in this study, There has been no use of an ultimately linked layer or a pooling layer. A completely linked layer or a pooling layer. These are used to reduce the calculation and make the classification simpler. Problems with image processing and general type are used to measure the performance of the ConvXGB. Auto feature learning is used by convX GB to predict the class labels (Shinde et al., 2019). That has higher accuracy than the two models used in the study.

## II. MATERIALS AND METHOD

### A. Convolutional Neural Network (CNN)

CNN formulation has been done in these sections, and the mathematical theory behind the CNN formulation has been done.  $W \times H$  has been assumed as input, When the intensity of the pixel gets  $x(m,n)$  at  $(m, n)$ , the convolution produced by the map is represented by the  $K$ . The discrete convolution is defined as

$$(I \otimes K)_{m,n} = \sum_k W_k u = -\sum_k W_k \sum_h h_k v = -\sum_h h_k \sum_u v_{m+u, n+j+v},$$

An additive bias and a convolution operation have been indexed in each convolutional layer for feature map indeed by  $f$  belonging to  $\{1 \dots f(l)\}$ . The results of the preceding layers feeds into the  $l$ th layer's output  $y(l)$ , which is denoted by

$$Y(l)_i = \phi(B(l)_i + \sum_j fX(l-1)_j = \sum_k K(l)_i \cdot Y(l-1)_j),$$

Layer output is modified through the pooling layer, which helps to downsample and avoid overfitting output. The pooling layer replaces the average value present in the rectangular neighbourhood.  $P(.)$  will act as a pooling function on the previous output  $(Y(l)_{m,n})$ . Average pooling and max pooling are In the pooling layer, and there are two forms of pooling (Ma et al., 2021). When the maximum value of each window is utilised to build a pool, it is known as max pooling.  $P(Y(l)_i)_{m,n} = \max(Y(l)_i)_{m,n}$ , as the output of the max-pooling function. Here the max process is applied to the max-pooling window as

$$dP(Y(l)_i)_{m,n} = ((W - w_k)/S_p + 1) \times ((H - h_k)/S_p + 1),$$

The final CNN architecture is the Fully Connected (FC) layer. This step takes the output from the preceding pooling layer and stretches it into a single column vector. The last layer's neurons are all coupled to the other connected layer. Well-known equations are used in the fully connected coatings for the multilayer perceptrons (Pranmanik et al., 2021). If the number of totally connected layers is  $L$ , then The size of the number of feature maps is  $f_1$ , and the size of the feature map is  $f_2 \times f_3$ . It is computed as

$$\begin{aligned} (Y(l)_i)_{m,n} &= \phi(f_1 X(l-1)_p = f_2 X(l-1)_q \\ &= f_3 X(l-1)_r \\ &= \sum_i w(l)_i, p, q, r(Y(l-1)_j)q, r). \end{aligned}$$

### B. Extreme Gradient boosting (XGBoost)

XGBoost is an end-to-end, highly scalable boosting system that may be used to learn how to classify regression cases. For each  $K$ , that corresponds to 1.  $K$  nodes, the XGBoost employs a set of  $k$  classification and regression trees in an ensemble. The sum of each tree's prediction scores yields the final projection.

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in F,$$

$X_i$  is considered a training set member in the above equation, and  $y$  denotes the corresponding class labels. Leaf scores are represented by  $f$ , and  $F$  is viewed as an asset for all  $K$  scores. The final results are improved by applying regularization.

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

The differential loss function is represented by the first term  $l$  to measure the difference between target  $y$  and prediction  $\hat{y}_i$ . Overfitting is avoided by the second term, where the complexity of the model is represented through  $\Omega$ .

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_j w_j^2$$

Regularization degree is controlled through the  $\gamma$ ,  $\lambda$  constant, several leaves in the tree are represented through  $T$ , and the weight of each leaf is represented through  $w$ . In classification problems and regression problems, Gradient boosting is very effective. The loss function is used with the Gradient Boosting, and it is extended through the second-order Taylor expansion to produce simplified objectives.

$$\begin{aligned} \tilde{L}(t)' \sum_i x_i &= \sum_i [g_i f(x_i) + \frac{1}{2} h_i f^2(x_i)] + \Omega(f_t) \\ &= \sum_i x_i = \sum_i [g_i f(x_i) + \frac{1}{2} h_i f^2(x_i)] + \gamma T + \frac{1}{2} \lambda \sum_j w_j^2 \end{aligned}$$

$$= X^T j = 1 [(X_i \in I_j g_i) w_j + 1/2 (X_i \in I_j h_i + \lambda) w_j^2] + \gamma T$$

The instance set of leaf  $t$  is denoted through  $I_3$  and

$$g_i = \partial l(\hat{y}_i(t-1), y_i) / \partial \hat{y}(t-1)_i \quad (12)$$

$$h_i = \partial^2 l(\hat{y}_i(t-1), y_i) / \partial (\hat{y}(t-1)_i)^2$$

### III. CONVXGB ARCHITECTURE

In figure 1, the architecture of the ConvXGB has been shown. The architecture has six layers, including An input layer, a data preparation layer, convolutional, a reshaping layer, a classification performance layer, and a network output are all included in the algorithm. All these layers are assigned to perform different roles and responsibilities. These layers play an essential role in the success of each model. Feature learning and predicting the class labels are the two parts of each layer. Feature Layers include input data preprocessing convolutional layers (Mounika & Rao, 2021). Effective feature learning determines prediction accuracy. The input layer is the first layer in the feature learning portion, and it is used to feed data into the models. We've used X as a training data set in this investigation. The second layer enhances the system's flexibility, a data preprocessing layer. It's used to keep track of data from several sources. The convolutional layer is the architecture's fundamental layer, and It's in charge of applying convolution to the input data and adding additive bias. According to the requirements, the number of computation times increases. In the convolutional layers, feature learning predicts class labels from the training data. Before entering the prediction portion, the input should be in vector format. The class prediction layer's primary purpose is to forecast the class using XGBoost. XGBoost uses the tree structure. As a criterion, the quality of the tree structure is used.

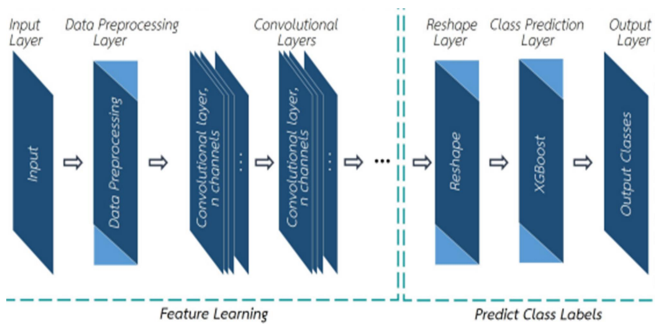


Figure 1: The ConvXGB model's architecture

### IV. RESULT AND DISCUSSION

The mode performance for classification problems is done by the data set gathered from the machine learning data sets repository.

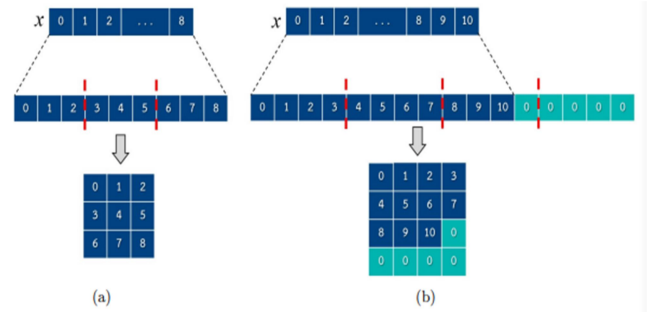


Figure 2: Example of data conversion into standard criterion

These datasets were selected from various sources, including data sets containing a variety of cases. Table 1 demonstrates that the number of attributes exceeds the number of issues. To train and evaluate the models, the researchers utilized three-fold cross-validation. In three disjoint subsets, each dataset was classified. As a training set, two subsets were used, whereas other subsets were used as testing sets. Three This method was performed several times. As a testing set, each subset is utilized only once. (Rajnoha et al., 2018). The result obtained from the testing setting set was average and helped calculate the standard deviation. Experimental parameters have been set carefully that helps resource balancing and help in getting good performance. Experimental parameters have been balancing guided through the complexity of the model. A variety of convolutional layers or maps were initially observed. d. in painting, when  $L=2$  and the output of the convolution layer were investigated as in  $z-2n$ , The value of  $n$  is the natural number starting from 1 to 10 (Bikias et al., 2021). (Bikias et al., 2021). When we choose a limited number of maps  $z$ , it's difficult to describe the data's features, and if we use a high number of maps  $z$ , the data will be overfitted from the computation.

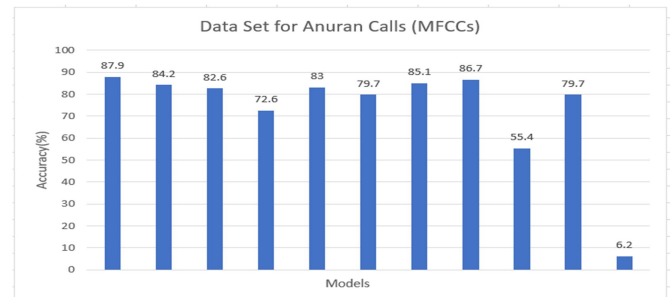


Figure 3: ConvXGB vs. all models for Anuran calls

Table 1: The ConvXGB model contains multiple layers, each with its own set of parameters.

Data sets	Area	Instances	Features	Following are the features that show after reshaping.	Depth of output
Anuran Calls (MFCCs) are a type of anuran call.	Life	7.193	21	799	31
privacy	Computer	605	6.399	50.199	7
Breast Cancer Wisconsin Original	Life	695	8	510	30
Parkinsons	Life	195	22	795	30
QSAR Biodegradation	N/A	1055	40	780	15
Waveform Database Generator (ver. 2)	Physical	5,000	35	750	30
Waveform Database Generator (ver .2)	Physical	5,000	35	12,540	250
Sensorless Drive Diagnosis	Computer	58,499	18	1,507	30

Table 2: The ConvXGB model comprises several layers, each with its own set of parameters.

Layer	Type	Output
1	Input	$\sqrt{N} \times \sqrt{N}$
2	Pre-processing of Data	$\sqrt{N}, \sqrt{N}, 1$
3	Convolution	$\sqrt{N}, \sqrt{N}, z(l-1)$
4	Reshape	$\sqrt{N} \times \sqrt{N} \times z(l)$
5	Predictions for each class	The number of classes available
6	Output	

The overfitted data use up the computer's memory. Table 2 shows the exact value used. The filter size is  $K=2 \times 2$ , and the field stride is  $S_k=1$ . The results of the ConvXGB model were compared to Convolutional neural network (CNN) and eXtreme Gradient Boosting (XGBoost) models, multilayer perceptron, and decision tree classifier models illustrated in Tables 3 and 4. (Hoq, Uddin & Park, 2021).

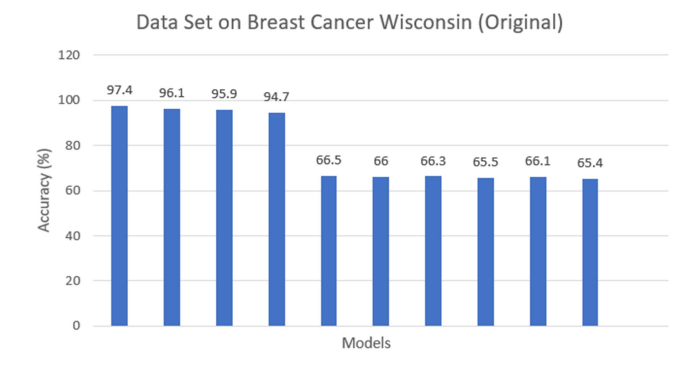


Figure4: ConvXGB vs. all models from Parkinsons data set

Table 3: CNN, XGBoost, and DTC were compared to our model.

Data set	ConveX GB Acc. (%)	CNN Impv. (percentage) Acc.	XGBoost Impv. (percentage) Acc.	DTC Impv. (percentage) Acc.
Anuran Screams (MFCCs)	87.9 ± 1.3	84.2 ± 0.93.7%	83.6±1.34.3%	72.6±1.0 15.3%
DrivFace	97.4 ± 0.7	96.1±0.91.3%	95.9±0.21.5%	94.7±0.92.7%
Breast Cancer Wisconsin (Original)	93.7 ± 1.9 9	91.9±2.61.8%	91.4±0.62.3%	86.6±2.37.1%
QSAR Biodegradation	96.4 ± 0.7	90.8±1.35.6%	94.9±3.61.5%	92.3±6.74.1%
Parkinsons	88.4 ± 0.6	86.5±0.51.9%	87.6±0.60.8%	81.2±1.07.2%
Generator of Waveform Databases (ver. 1)	99.6±0.0	87.9±2.2 11.7%	99.0±0.10.6%	98.3±0. 1.3%
Generator of Waveform Databases (ver. 2)	86.6 ± 0.4	86.1±0.50.5%	84.8±0.91.8%	74.8±0.1 11.8%
Sensorless Drive Diagnosis	86.7 ± 1.0	82.7±0.84.0%	85.0±0.41.7%	73.8±1.3 12.9%

Table 4: Compared to other models, our model takes longer to run.

Data set	Run time (sec. )					
	Conv XGB	CNN	XGBoost	DTC	MLP	SVC
Anuran Calls (MFCCs)	42s	1.6×106 s	5.4 s	13 s	17 s	23 s
DrivFace	2.7 s	5.2×103 s	8.1×10−1	9.3×10−1 s	9.8×10−1 s	2.1×10−2 s
Breast Cancer Wisconsin (Original)	2.2×102 s	2.4×104 s	7.4×10−1 s	3.1 s	4.7 s	1.4×10−2 s
QSAR Biodegradation	1.2s	1.5×103 s	4.5×10−1 s	3.6×10−1 s	6.3×10−1 s	4.7×10−1 s
Parkinsons	6.2 s	7.9 ×103	2.1s	2.3 s	3 s	2.7 s
Waveform Database Generator (ver. 1)	6.7 x 102s	2.4×106 s	3.8×102 s	1.3×102 s	3.3×102 s	1.2×104 s
Waveform Database Generator (ver. 2)	57 s	5.6×104 s	8.8 s	9.7 s	16 s	13 s
Sensorless Drive Diagnosis	4.5 x 104s	6.1×104	14 s	11 s	21 s	18 s

The parameters in the convolutional layers of the CNN model have been determined with the exception that, according to the model because the convolutional layer parameters have not been established adding to the CNN model the pooling layer of the CNN was of the size 2x2 pool and the number of the neuron present in the FC was 2n where the value of n was from 6 to 10. Similarly, the value of the XGBoost model was set, and the accuracy was carried out based on the resource. The greatest depth of the tree was described using the DTC model. The DTC model explains the maximum depth of the dress. All the leaves contain less than 2 sample splits. Minimum samples are needed to split the internal node (Pramod et al., 2019). Gini impurity was used to determine the quality of the split. The number of neurons in 2n was set at n=6 to 10 for

the MLP model, and the pace of learning was set to 0.00-1. MLP1, MLP2, MLP3, and ReLU were the four variations used for distinct activation functions (MLP4). The study used four alternative SvC models with different RBF (SVC1), Linear (SVC2), Polynomial (SVC3), and Sigmoid (SVC4) are the kernel functions (SVC4). The other models were tested with Python (3.6.4) and TensorFlow functions. To finish the prediction step, convolutional operations were done in CNN, and functions from the TensorFlow library were integrated with the XGBoost Python library (Nissar, Mir & Shaikh, 2021). Another machine learning library that has been developed is Scikit Learn.

## V. CONCLUSION

The purpose of this post was to create a deep learning framework model for parkinson disorder by using CNN and XGBoost. Feature learning and prediction of class labels are the two significant parts of the study that helps in classifying the problems. ConvXGB model has been used as a classification based on image data and used for preprocessing modules. The number of needed parameters has been reduced to simplify the ConveX GB. Depending on the data, the number of convolution layers increases. From the result of our study, it has been clear that the ConveX GB model provides better results than individually using CNN and XGBoost. For some data sets, such as the Breast cancer data set, CNN fared better. XGBoost, on the other hand, performed better across the board for all of the parkinson's data. Therefore, ConvXGB was better for all datasets than XGBoost and CNN.

## VI. REFERENCES

- [1]. Bikias, T., Iakovakis, D., Hadjidimitriou, S., Charisis, V., & Hadjileontiadis, L. J. (2021). DeepFoG: an IMU-based detection of freezing of gait episodes in Parkinson's disease patients via deep learning. *Frontiers in Robotics and AI*, 8.
- [2]. Goyal, J., Khandnor, P., & Aseri, T. C. (2020). Classification, Prediction, and Monitoring of Parkinson's Disease Using Computer-Assisted Technologies: A Comparative Analysis. *Engineering Applications of Artificial Intelligence*, 96, 103955.
- [3]. Hoq, M., Uddin, M. N., & Park, S. B. (2021). Vocal Feature Extraction-Based Artificial Intelligent Model for Parkinson's Disease Detection. *Diagnostics*, 11(6), 1076.
- [4]. Johri, A., & Tripathi, A. (2019, August). Parkinson's disease detection using deep neural networks. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-4). IEEE.
- [5]. Lamba, R., Gulati, T., & Jain, A. (2022). A Hybrid Feature Selection Approach for Parkinson's Detection Based on Mutual Information Gain and Recursive Feature Elimination. *Arabian Journal for Science and Engineering*, 1-14.
- [6]. Lamba, R., Gulati, T., Al-Dhlan, K. A., & Jain, A. (2021). A systematic approach to diagnose Parkinson's Disease through kinematic features extracted from handwritten drawings. *Journal of Reliable Intelligent Environments*, 7(3), 253-262.
- [7]. Ma, Y. W., Chen, J. L., Chen, Y. J., & Lai, Y. H. (2021). Explainable deep learning architecture for early diagnosis of Parkinson's Disease. *Soft Computing*, 1-10.
- [8]. Mounika, P., & Rao, S. G. (2021, November). Machine Learning and Deep Learning Models for Diagnosis of Parkinson's Disease: A Performance Analysis. In 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC) (pp. 381-388). IEEE.
- [9]. Nahar, N., Ara, F., Nelay, M., Istiek, A., Biswas, A., Hossain, M. S., & Andersson, K. (2021, September). Feature Selection Based Machine Learning to Improve Prediction of Parkinson Disease. In International Conference on Brain Informatics (pp. 496-508). Springer, Cham.
- [10]. Nilashi, M., Abumalloh, R. A., Minaei-Bidgoli, B., Samad, S., Yousoof Ismail, M., Alhargan, A., & Abdu Zogaan, W. (2022). Predicting Parkinson's Disease Progression: Evaluation of Ensemble Methods in Machine Learning. *Journal of Healthcare Engineering*, 2022.
- [11]. R. Lokeshkumar, E. Maruthavani, and A. Bharathi, A new perspective for decision makers to improve efficiency in social business intelligence systems for sustainable development, *International Journal of Environment and Sustainable Development* 2018 17:4, 404-416.
- [12]. S. Balakrishnan, V. K. and M. S. S. Hameed, "An Embarking User Friendly Palmprint Biometric Recognition System with Topnotch Security," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1028-1032, doi: 10.1109/ICICCS51141.2021.9432230.
- [13]. V. Arulkumar A Survey on Multimedia Analytics in Security Systems of Cyber Physical Systems and IoT, 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC) pp 1556-1561, Nov 2021. 10.1109/ICOSEC51865.2021.9591754
- [14]. Ravi Kumar Poluru & Lokesh Kumar R (2021) An Improved Fruit Fly Optimization (IFFOA) based Cluster Head Selection Algorithm for Internet of Things, *International Journal of Computers and Applications*, 43:7, 623-631, DOI: 10.1080/1206212X.2019.1600831
- [15]. Selvan, C. S. R. Balasundaram, (2020), Data Analysis in Context Based Statistical Modeling in Predictive Analytics, IGI Global, Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics, pages 98-114.
- [16]. V. Arulkumar An Intelligent Face Detection by Corner Detection using Special Morphological Masking System and Fast Algorithm, 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC) pp 1556-1561, Nov 2021.
- [17]. Nishat, M. M., Hasan, T., Nasrullah, S. M., Faisal, F., Asif, M. A. A. R., & Hoque, M. A. (2021, August). Detection of Parkinson's Disease by Employing Boosting Algorithms. In 2021 Joint 10th International Conference on Informatics, Electronics & Vision (ICIEV) and 2021 5th International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (pp. 1-7). IEEE.
- [18]. Nissar, I., Mir, W. A., & Shaikh, T. A. (2021, March). Machine Learning Approaches for Detection and Diagnosis of Parkinson's Disease-A Review. In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 898-905). IEEE.
- [19]. Pramanik, M., Pradhan, R., Nandy, P., Bhoi, A. K., & Barsocchi, P. (2021). Machine learning methods with decision forests for Parkinson's detection. *Applied Sciences*, 11(2), 581.
- [20]. Pramod, N. S., Sajitha, L., Mohanlal, S., Thameem, K., & Anzar, S. M. (2021, November). Detection of Parkinson's Disease Using Vocal Features An Eigen Approach. In 2021 Fourth International Conference on Microelectronics, Signals & Systems (ICMSS) (pp. 1-6). IEEE.
- [21]. Rajnoha, M., Mekyska, J., Burget, R., Eliasova, I., Kostalova, M., & Rektorova, I. (2018, November). Towards the identification of hypomimia in Parkinson's Disease based on face recognition methods. In 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT) (pp. 1-4). IEEE.
- [22]. Shinde, S., Prasad, S., Saboo, Y., Kaushick, R., Saini, J., Pal, P. K., & Ingahlalikar, M. (2019). Predictive markers for Parkinson's Disease using deep neural nets on neuromelanin sensitive MRI. *NeuroImage: Clinical*, 22, 101748.
- [23]. Thongsuwan, S., Jaiyen, S., Padcharoen, A., & Agarwal, P. (2021). ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost. *Nuclear Engineering and Technology*, 53(2), 522 – 53.