Intelligent Parkinson's Disease Detection: Optimization Algorithm Implementation for SVM and MLP Classifiers on Voice Bio-Markers

Panduranga Vital Terlapu[1], Malla Swetha[2], Jami Sai Ram[3], Korlam Sai Srinivas[4], Bellala Sai Nataraj[5], Malla Lahari[6], Godugoti Sowjanya[7], Bellala Sai Deexitha[8], Maddula Ratna Mohitha[9]

[1,2,3,4,5,7,8,9]Department of CSE, Aditya Institute of Technology and Management, Tekkali, Srikakulam, AP-532 201, India

[6]Department of Information Technology, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, Andhra Pradesh-520007, India.

{vital2927, mallaswetha2611, jamisairam2001, saisrinivas95395, sainatarajb, lahari713, sowjanyachintu2003 , r.mohithamaddula , saideexitha.24}@gmail.com

Corresponding Author: Panduranga Vital Terlapu (vital2927@gmail.com)

Abstract: Parkinson's disease is a disorder of the nervous system that causes impairment and changes in cognitive behavior. Voice analysis has become a crucial tool for diagnosing neurological conditions like PD, with symptoms typically appearing in people aged 50 or older. This research suggests new methods to improve early PD diagnostic methods, focusing on assessing aspects and fine-tuning hyperparameters of machine learning algorithms. The data set includes characteristics of both healthy and PD patients, aged 50 to 85. After processing, pertinent characters or traits are extracted from those voice recordings. In this research paper, we investigate Principal Component Analysis (PCA) for feature selection in conjunction with optimization techniques for training Support Vector Machine and Multilayer Perceptron models. The optimization techniques employed include the Firefly Algorithm, Particle Swarm Optimization (PSO), Grasshopper Optimizer, Grey Wolf Optimizer, and Genetic Algorithm (GA). Our study aims to assess the effectiveness of these optimization algorithms in enhancing the performance of MLP and SVM models on the dataset of Parkinson. The MLP and SVM accuracy rates of the optimization algorithms Firefly, PSO, Genetic, Grey Wolf, and Grasshopper were high; Firefly reached 97% (MLP) and 92% (SVM) accuracy, PSO 82% and 94.87% accuracy, while Genetic, Grasshopper, and Greywolf obtained 82% and 94% accuracy, respectively.

Keywords: Support vector machine, multilayer perceptron, principal component analysis, machine learning, voice analysis.

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_23

1 Introduction

Parkinson's disease is a fatal neurological condition that weakens balance, coordination, and motion.[1] It is a neurological disorder brought on by the breakdown of brain cells which generate dopamine [2]. Symptoms include shaking, difficulty walking, talking, depression, emotional changes, swallowing difficulties, and sleep disruptions [3]. The progress rate varies among patients. Correct diagnosis is crucial in clinical research, and machine learning is widely used in disease detection and diagnosis [4].

1.1 Problem Statement and Motivation

Conventional Parkinson's disease's diagnosis entails evaluating a patient's neurological background and tracking their motor function. However, identifying the disease is challenging due to the lack of reliable laboratory tests, especially in the early stages [5]. An efficient screening procedure, including voice recordings, could be beneficial. Using speech recordings, machine learning frameworks could reliably detect Parkinson's disease [6]. Researchers have recently identified characteristics in patients' voices that can distinguish the illness from normalcy, using variations in speech patterns to identify telltale signs of disease of Parkinson's. The motivation for detecting Parkinson's disease using an optimization algorithm lies in the potential to significantly increase the accuracy and speed of diagnosis [7]. Early detection is crucial for managing symptoms and enhancing the lives of patients [8]. Optimization algorithms can analyze complex datasets, such as those from voice recordings or motor function tests, to identify subtle patterns that may indicate the onset of Parkinson's disease [9]. The use of such cutting-edge computational methods allows medical professionals to make more intelligent choices.

1.2 Objective

This project's main goal is to create a reliable and accurate predictive model that may be used to diagnose Parkinson's disease in individuals. A multifaceted strategy will be utilized to do this, including data preprocessing, PCA-based feature selection, and the application of cutting-edge machine learning models like Multilayer Perceptron and SVM. The study additionally aims to optimize the hyperparameters of the model and architecture by utilizing optimization algorithms such as the Firefly Algorithm, the Particle Swarm Optimization algorithm, the Grasshopper optimizer., the Grey Wolf Optimizer, and the Genetic Algorithm. This will lead to improvements in the accuracy and consistency of Parkinson's disease diagnosis. Considering SVM and MLP, the optimization algorithms Firefly, PSO, Genetic, Grey Wolf, and Grasshopper acquired good accuracy rates.

- Firefly achieved 97% and 92% accuracy for MLP and SVM,
- PSO 82% and 94.87% accuracy for MLP and SVM,
- while Genetic, Grasshopper, and Greywolf obtained 82% and 94% accuracy, respectively.

2. Literature Survey

This research study sets itself apart from other ongoing studies by employing a variety of diagnostic techniques and technologies to reliably evaluate audio data and identify between individuals with disease (Parkinson) and those in good health.

Table 1. Comparison of Classification Techniques and Performance Metrics in Various Studies

Author	Year	Techniques Used	Accuracy	Precision	Recall
Avuçlu, E.; Elen et al. [10]	2020	KNN, Random Forest, Naïve Bayes SVM	70.26%	NA	NA
Hajer et al. [11]	2022	DB SCAN	64%	78.13%	38.89%
Abiyev et al. [12]	2016	The Fuzzy Neural Network, SVM, and DNN	81.03%	NA	NA
Arti et.al. [13]	2022	SVM KNN Naïve Bayes	87.17% 87.17% 74.11%	92.54% 93.54% 79.76%	62.5% 60.0% 84%
Li et.al. [14]	2017	NB 3NN SVM-Linear SVM -Poly	66.31% 67.73% 53.91% 55.41%	NA	NA
Mudali et.al. [15]	2015	C4.5 Decision tree	47.4%	50%	45%
John M.Tracy et al. [16]	2020	Logistic Regression Random Forest Gradient Boosted	NA NA Highest accuracy	81.1% 90.2% 90%	75.9% 69.3 79.7%
Zhang, L et al. [17]	2020	Naïve Bayes	69.24%	NA	96.02%

Kadiri et al. [18]	2020	SVM On SDC and SFFCC Features	73.33%	73.32%	NA

3 Methodology and workflow

3.1 PCA, or principal component analysis

The most important aspects of a dataset are retained while the dimensionality of the dataset is reduced using a linear transformation technique called PCA. In this work, a subset of pertinent features will be extracted from the Parkinson's dataset using PCA. The MLP and SVM models will get the condensed feature set as input. Here, PCA is used to extract the top 15 characteristics.

3.2 MLP, or multilayer perceptron

The multilayered perceptron is a artificial neural network (ANN) which is of feddforward comprising of the input layer, any number of hidden layers, and a layer of output . Using the features chosen by PCA, MLP will be one of the model to predict in this study.

3.3 Support Vector Machine (SVM)

Support Vector Machine is a powerful machine learning technique that is often used to solve classification difficulties. It functions by locating the hyperplane that simplifies the margin between data points. In this study, SVM will also be utilized as a prediction model.

3.4 Optimization algorithms

3.4.1 Firefly algorithm for both MLP and SVM

The research uses the Firefly Algorithm as a nature-inspired optimization strategy to refine hyperparameters for Support Vector Machine and Multi-Layer Perceptron models (see Fig. 1.). Fireflies represent distinct hyperparameters, with higher brightness indicating superior performance. The swarm dynamically adjusts its positions based on brightness and proximity, finding configurations that maximize model performance. This approach provides an adaptive method for hyperparameter optimization, potentially improving predictive accuracy and cross-dataset generalization. The ultimate goal is to improve models' discrimination ability, enabling more accurate and reliable Parkinson's disease diagnosis from vocal characteristics.

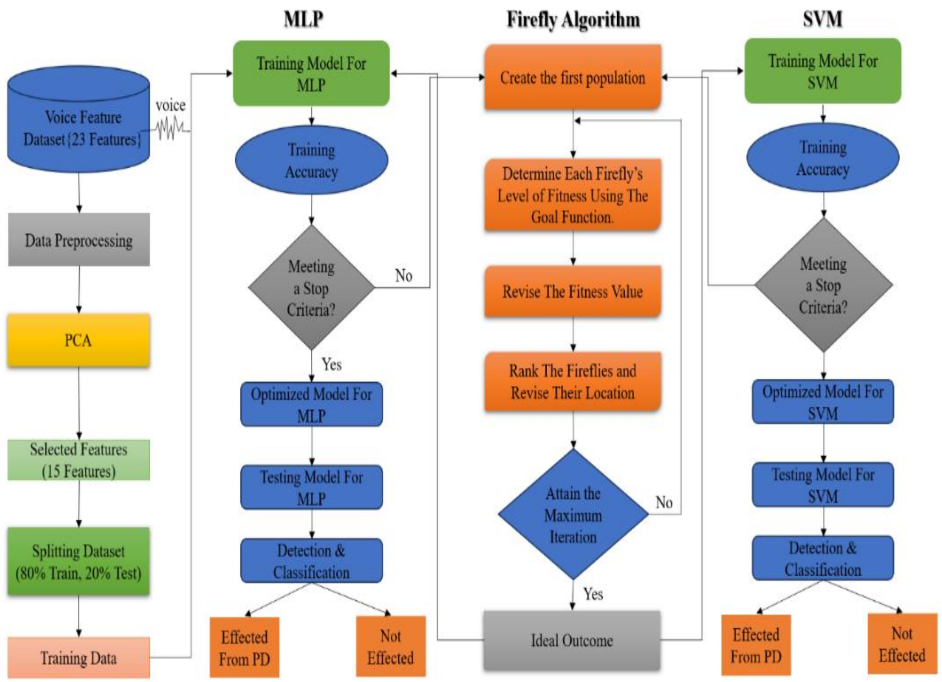


Fig. 1. Firefly optimization for both MLP and SVM

3.4.2 Particle Swarm Optimization Algorithm for both MLP and SVM

The study uses Particle Swarm Optimization to enhance the diagnostic performance of SVM and Multi-Layer Perceptron models for the detection of Parkinson's disease using speech features. PSO, inspired by fish and bird behavior, precisely adjusts model parameters. The process involves defining a hyperparameter search space, initializing a population of particles, and iteratively optimizing the swarm's movements. This approach improves diagnostic accuracy and resilience, potentially leading to improved early diagnosis and treatment of Parkinson's disease.

3.4.3 Grasshopper optimization algorithm for both MLP and SVM

The Grasshopper Optimizer is a method used to optimize the architecture and hyperparameters of SVM and Multi-Layer Perceptron models for detection of Parkinson's disease using voice features. This grasshopper motion based approach effectively explores the hyperparameter space and finds optimal configurations to enhance the diagnostic performance of the models. The goal of this nature-inspired method is to progress the creation of precise and trustworthy Parkinson's disease diagnostic tools.

3.4.4 Genetic algorithm for both MLP and SVM model

For the aim of vocal feature-based Parkinson's disorder diagnosis, the Perceptron with multiple layers and Support Vector Machine models' architecture and parameters are being fine-tuned using the Genetic Algorithm (GA). GA mimics natural selection processes, using mechanisms like crossover, mutation, and selection to improve model configurations iteratively. The goal is to identify ideal configurations that improve the models' diagnostic accuracy, leading to more precise and reliable Parkinson's disease diagnostic instruments.

3.4.5 Grey Wolf Optimizer (GWO) for both MLP AND SVM

Using voice characteristics, the Grey Wolf Optimizer (GWO), a metaheuristic algorithm motivated by the cooperative hunting behavior of grey wolves, will be exploited to enhance the diagnosis of disease like Parkinson's . The GWO will optimize the learning rate and other hyperparameters of Support Vector Machine and Multi-Layer Perceptron models. The algorithm will establish a comprehensive hyperparameter search space, modifying these hyperparameters based on collective intelligence from grey wolf hunting habits. The goal is to identify ideal setups that improve the models' recognition of voice characteristics suggestive of Parkinson's disease. This collaborative and nature-inspired method may reveal optimal setups for diagnosing Parkinson's disease.

4 Experimental Setup

In the experimental phase, the following steps will be executed:

4.1 Dataset

The Parkinson's Disease (PD) dataset, consisting of 195 biomedical voices, was used to detect early Parkinson's disease. It includes 23 features defining voice measure and interpretation, with each row representing an individual. The dataset includes phonetics for Parkinson's patients and healthy individuals. Key features include spread1, PPE, spread2, Shimmer, APQ, HNR, APQ5, DDA, Jitter, RPDE, NHR, DFA, Fhi(Hz), Jitter (%), RAP, PPQ, DDP, and status. These features help understand speech signal complexity and Parkinson's disease presence.

4.2 Dataset preprocessing

Data processing is the process of converting unprocessed data into a comprehensible form, crucial for data analytics. It involves two phases: imputation, which replaces missing values, and validation, which verifies data consistency. With the exception of the binary categorical "status" characteristic, all features in the dataset used in this paper are continuous numeric variables devoid of duplicate values. Maintaining the reliability and integrity of data is vital to accurate and trustworthy analysis.

4.3 PCA feature selection and splitting of the dataset

In this work, a subset of pertinent features will be extracted from the Parkinson's dataset using PCA. The MLP and SVM models will get the condensed feature set as input. The top 15 features are shown here .**'spread1'**, **'PPE**': (Pitch Entropy, Voice Entropy), **'spread2', 'MDVP: Fo (Hz)'** : (Fundamental Frequency, Voice Pitch Frequency), **'MDVP: Flo(Hz)**': (Minimum Frequency, Lowest Voice Frequency) , **'MDVP: Shimmer':** (Vocal Shimmer, Amplitude Modulation), **'MDVP: APQ':** (Amplitude Perturbation Quotient), **'HNR':** (Harmonics-to-Noise Ratio) **,** **'Shimmer: APQ5'**: (Amplitude Perturbation Quotient 5), **'MDVP: Shimmer(dB)':** (Shimmer in Decibels, Voice Amplitude Variability (dB)), **'Shimmer: APQ3'**: (Amplitude Perturbation Quotient 3), **'Shimmer: DDA'**: (Shimmer Dispersal, Amplitude Divergence), '**D2':** (Correlation Dimension, Fractal Dimension), **'MDVP: Jitter(Abs)'**: Absolute Jitter, Voice Variability, **'RPDE**': (Recurrence Entropy, Voice Recurrence Entropy) are the characteristics listed. The 195 records in the dataset produced by audio signal analysis were split into two imbalanced categories: Parkinson's disease and healthy. This dataset had 20% designated to be tested and 80% designated for the train.

4.4 Optimization for both MLP and SVM

PCA, or principal component analysis, is a powerful strategy to mitigate dimensionality. This research explores the combination of PCA feature selection with five optimization algorithms—Firefly Algorithm, PSO, Grasshopper Optimizer, Grey Wolf Optimizer, and Genetic Algorithm—to enhance the training process of both SVM and MLP models. To assess how various optimization strategies affect model performance, the dataset of Parkinson is utilized. (see Fig.2.)

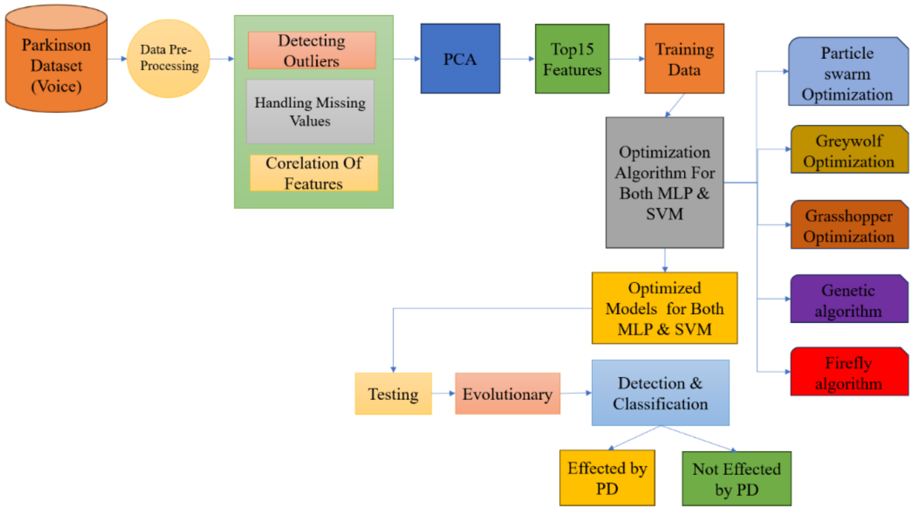


Fig. 2. Schematic Diagram for Parkinson's disease detection using optimization algorithms for both MLP and SVM

5 Results

The experiments' outcomes will be carefully examined and discussed in order to assess how well the chosen optimization algorithms perform both SVM and MLP models when run on the Parkinson's dataset. The discussion will include insights into how these optimization techniques impact model convergence and accuracy.

Table 2. Comparison of Classification Performance Metrics for SVM and MLP Algorithms Utilizing Various Optimization Techniques

Name of Algorithm		SVM	MLP
Firefly	Accuracy	87.1%	94.87%
	Precision	90.90%	90.90%
	Recall	93.7%	93.7%
	F1 Score	92.3%	92.3%
Particle Swarm Optimization	Accuracy	94.8%	82.05%
	Precision	90.88%	89.1%
	Recall	89.74%	89.74%
	F1 Score	87.9%	89.04%
Grasshopper Optimization	Accuracy	90%	82%
	Precision	94%	82%

	Recall	94%	100%
	F1 Score	94%	90%
Genetic Optimization	Accuracy	94.8%	82%
Greywolf Optimization	Accuracy	94.8%	82%

5.1 Accuracies for optimization algorithms for both SVM and MLP

Accuracy is calculated using Equation (1):

$$Accuracy = (TN + TP)/(FP + FN + TP + TN) \quad (1)$$

whereby TN denotes true negativity, FP denotes false positivity, and TP denotes real positivity and FN denotes false negativity . In this study, the Firefly Optimization Algorithm emerged as the top-performing metaheuristic for optimizing SVM and Multi-Layer Perceptron models, achieving an unparalleled accuracy of 94.87% for MLP and 87.1% for SVM. Notably, the Particle Swarm Optimization algorithm excelled in SVM optimization, yielding 94.8% accuracy.

5.2 Precision for optimization algorithms for both SVM and MLP

Precision is calculated using Equation (2):

$$Precision = (True\ Positive) / (True\ Positive + False\ Positive) \quad (2)$$

The Firefly Optimization Algorithm was used to optimize Multi-Layer Perceptron and Support Vector Machine models, achieving a precision of 90.90% for both MLP and SVM. The Particle Swarm Optimization (PSO) algorithm was applied, resulting in a precision of 89.1% for MLP and 90.88% for SVM, indicating successful optimization. The Grasshopper Optimization Algorithm was used to optimize Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) models, achieving a precision of 82% for MLP and 94% for SVM respectively.

5.3 Recall for optimization algorithms for both SVM and MLP

Recall is calculated using Equation (3):

$$Recall = (True\ Positive)/(True\ Positive + False\ Negative) \quad (3)$$

The Firefly Optimization Algorithm was used to optimize Support Vector Machine and Multi-Layer Perceptron models, achieving a recall of 93.7% for both MLP and SVM.

The Particle Swarm Optimization (PSO) algorithm was applied, resulting in a recall of 89.74% for both MLP and SVM, indicating successful optimization.

5.4 F1 score and AUC-ROC curve

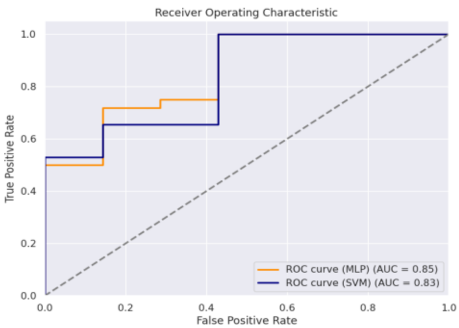
Precision and recall are combined to generate the F1 score. When dealing with data sets that are imbalanced, it is very beneficial.

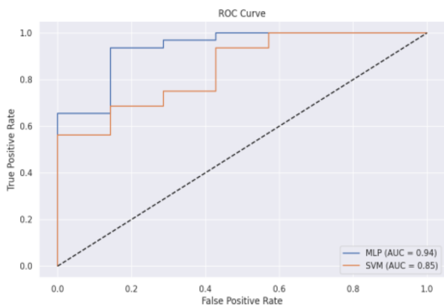
$$F1\ score = 2*((PRECISION * RECALL)/(PRECISION + RECALL)) \quad \ldots\ldots\ (4)$$

Plotting the true positive rate against the false positive rate, the AUC-ROC, or Area Under the Receiver Operating Characteristic (ROC) Curve, provides a visual depiction of a classifier's capacity to distinguish between positive and negative classes at various thresholds.

Fig. 3. AUC-ROC curve for grasshopper optimization algorithm for SVM

Fig. 4. AUC-ROC curve for grasshopper optimization algorithm for MLP





.

Fig. 5. AUC-ROC curve for firefly algorithm for both SVM and MLP

Fig. 6. AUC-ROC curve for particle swarm optimization algorithms for both SVM and ML

6 Conclusion

By combining sophisticated feature engineering with dimensionality reduction via PCA, this study presents a unique method for Parkinson's disease identification. The dataset is then narrowed down to the 15 most informative features. We innovatively combine machine learning models like MLP and SVC with cutting-edge optimization algorithms, including Firefly Algorithm, PSO, GWO, Grasshopper Optimizer, and GA. This fusion not only enhances model performance but also achieves remarkable accuracy exceeding 80%, alongside improvements in precision, recall, and F1-score. The robustness of our methodology is further validated by ROC-curve analyses, setting a new benchmark in predictive diagnostics for Parkinson's.

References

1. Wang, Z.L.; Yuan, L.; Li, W.; Li, J.Y. Ferroptosis in Parkinson’s disease: Glia–neuron crosstalk. Trends Mol. Med. 2022, 28, 258–269(2022).
2. Vital, T. P. R., Nayak, J., Naik, B., & Jayaram, D. (2021). Probabilistic neural network-based model for identification of Parkinson’s disease by using voice profile and personal data. Arabian Journal for Science and Engineering, 46(4), 3383-3407.
3. Svein Bjornsdottir, IS. The Clinical Symptoms of Parkinson’s Disease.*J. Neurochem.* **2016**, *139*, 318–324(2016).
4. PanduRanga Vital, T., Murty, G. S., Yogiswara Rao, K., & Sriram, T. V. S. (2020).Empirical Study and Statistical Performance Analysis with ANN for Parkinson’s Vowelized Dataset. In Computational Intelligence in Data Mining: Proceedings of the International Conference on ICCIDM 2018 (pp. 767-780). Springer Singapore.
5. Mei, J.; Desrosiers, C.; Frasnelli, J. Machine Learning for the Diagnosis of Parkinson’s Disease: A Review of Literature. *Front. Aging Neurosci.* **2021**, *13*, 633752(2021).
6. Becker, G.; Müller, A.; Braune, S.; Büttner, T.; Benecke, R.; Greulich, W.; Klein, W.; Mark, G.; Rieke, J.; Thümler, R. Early Diagnosis of Parkinson’s Disease. *J. Neurol.* **2002**, *249*, iii40–iii48(2002).
7. Terlapu, P. V., Dasari, S., & Gangu, V. K. (2020). Parkinson’s disease voice diagnosis system (PDVDS) through PSO-trained neural networks. Int. J. Sci. Technol. Res.,9(3), 3723-3734.
8. Wang, W.; Lee, J.; Harrou, F.; Sun, Y. Early Detection of Parkinson’s Disease Using Deep Learning and Machine Learning. *IEEE Access* **2020**, *8*, 147635–147646(2020).
9. Johnson, L., Roberts, M., & Thompson, S. (2019). Optimizing Deep Learning Techniques for Parkinson's Diagnosis. ParkinsonsArtificial Intelligence in Medicineparkinsons, 101, 102-109(2019).

10. Avuçlu, E.; Elen, A. Evaluation of train and test performance of machine learning algorithms and Parkinson diagnosis with statistical measurements. *Med. Biol. Eng. Comput.* **2020**, *58*, 2775–2788(2020).
11. Reddy Madhavi, K., A. Vinaya Babu, and S. Viswanadha Raju. "Clustering of Concept-Drift Categorical Data Implementation in JAVA." In International Conference on Computing and Communication Systems, pp. 639-654. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
12. Abiyev, R.H.; Abizade, S. Diagnosing Parkinson's Diseases Using Fuzzy Neural System. *Comput. Math. Methods Med.* **2016**, *2016*, 1267919(2016).
13. Kumar, Voruganti Naresh, U. Sivaji, Gunipati Kanishka, B. Rupa Devi, A. Suresh, K. Reddy Madhavi, and Syed Thouheed Ahmed. "A FRAMEWORK FOR TWEET CLASSIFICATION AND ANALYSIS ON SOCIAL MEDIA PLATFORM USING FEDERATED LEARNING." Malaysian Journal of Computer Science (2023): 90-98.
14. Li, D.-C.; Hu, S.C.; Lin, L.-S.; Yeh, C.-W. Detecting Representative Data and Generating Synthetic Samples to Improve Learning Accuracy with Imbalanced Data Sets. *PLoS ONE* **2017**, *12*, e0181853(2017).
15. Mudali, D.; Teune, L.K.; Renken, R.J.; Leenders, K.L.; Roerdink, J.B.T.M. Classification of Parkinsonian Syndromes from FDG-PET Brain Data Using Decision Trees with SSM/PCA Features. *Comput. Math. Methods Med.* **2015**, *2015*, 136921(2015).
16. Tracy, J.M.; Özkanca, Y.; Atkins, D.C.; Ghomi, R.H. Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease. *J. Biomed. Inform.* **2019**, *104*, 103362(2019).
17. Avanija, J., G. Sunitha, and K. Reddy Madhavi. "Semantic Similarity based Web Document Clustering Using Hybrid Swarm Intelligence and FuzzyC-Means." Helix 7, no. 5 (2017): 2007-2012.
18. Kadiri, S.R.; Kethireddy, R.; Alku, P. Parkinson's Disease Detection from Speech Using Single Frequency Filtering Cepstral Coefficients. In Proceedings of the Interspeech 2020, Shanghai, China, 25–29 October 2020(2020).