Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 235 (2024) 588–597

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering (ICMLDE 2023)

Differentiating Parkinson's Disease from other Neuro Diseases and Diagnosis using Deep Learning with Nature Inspired Algorithms and Ensemble Learning

Anitha Rani Palakayala[a], Kuppusamy P[b,*]

[a,b] *VIT-AP University, Mandadam, Near AP Secretariat, Amaravati, Guntur (Dt), AP-522237, INDIA*

Abstract

A highly accurate automated approach using a deep structured neural network was proposed to detect Parkinson's disease throughvoice samples. This cost-effective and non-invasive method aims to enhance diagnostic precision and assess the disease's stage ofprogression. The study addresses two classification problems: binary and multi-category classification. The binary classification deep structured neural network achieved a diagnostic accuracy of 97.6%. For multiclass classification, a Deep Convolution Neural Network (DCNN) and a K- nearest neighbor technique for benchmarking, were utilized using a shared database. Nature Inspired Algorithm (NIA)-Genetic Algorithm was employed to optimize the initial feature set. The suggested deep structure neural network demonstrated a 93.7% accuracy in estimating the disease's stage, showing promising results. Further investigations are encouraged to explore the model's adaptability and pursue improved outcomes.

Keywords: Deep neural networks; Ensemble learning; Nature inspired algorithms; Neurodegeneration; Parkinson's disease; Voice analysis;

1. Introduction

Parkinson's disease (PD) is a neurological disorder that can be observed by a moderate decline in motor function andcognitive capacities [1, 2]. Globally, Alzheimer's disease occupied the first position to have high occurrence of neurological disorder, while PD stood in the second position [3]. PD has aglobal impact, affecting a substantial

* Kuppusamy P :+91-9843016837;
E-mail address: drpkscse@gmail.com

10.1016/j.procs.2024.04.058

population of 7 to 10 million individuals. Its prevalence is estimated to be between 1% to 2% among adults aged 60 years and above [4, 5]. The condition is attributed to the degeneration of certain clusters of neurons that are important for synthesizing neurotransmitters, including dopamine, acetylcholine, serotonin, and norepinephrine. Chemical molecules, known as neurotransmitters, are synthesized and secreted by neurons to facilitate communication within the nervous system. These neurotransmitters play a crucial role in regulating both movement and non-movement functions in the human anatomy. The consequences of their loss are characterized by significant impacts, including compromised motor control, leading to disruptions in movement coordination, instability in maintaining posture, diminished visual capabilities, and anomalies in voice production.

Fig. 1(a). Symptoms of PD [19]

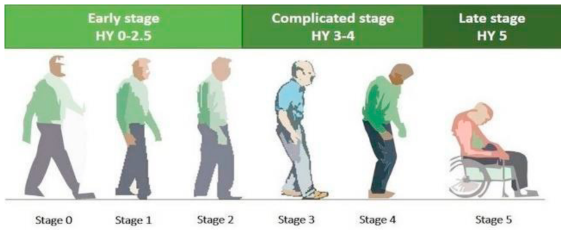

Fig. 1(b). Different stages of PD [20]

The movement and non-movement related symptoms as well as different stages of PD can be seen from Fig. 1(a) and Fig. 1(b) respectively. Speech abnormalities frequently manifest as a prevalent indication of movement disorders in persons diagnosed with PD. The disorders encompass dysarthria, characterized by challengesin articulating sounds, mumble and monotonic tone. Dysarthria, which is the most prevalent disorder, can be detected in around 90% of cases [6, 7]. At now, PD lacks a definitive treatment; nonetheless, the timely identification of the condition and the implementation of suitable therapeutic interventions may play an important role in mitigating the rate of symptom advancement. According to recent reports, voice issues have been observed in around 90% of individuals who have received a diagnosis, even at the initial phase of the condition [1,8]. Furthermore, it has been suggested that the efficacy of therapy can be assessed by analyzing voice cores [9]. Therefore, the study of vocal characteristics provides an uncomplicated and economical method for assessing PD [10, 11]. It also facilitates monitor the disease's progression. Therefore, the examination of vocal characteristics has the prospect not only to provide a straightforward and cost-effective means of diagnosing PD [10, 11], but also to track the progression of the condition. The development of neural networks has resulted in the emergence of novel topologies and associated algorithms, rendering them highly appealing for intricate classification tasks [12]. This research introduces an innovative method for automatically classifying audio samples to detect auditory patterns linked to PD. The approach utilizes aDCNN for the classification process. It also includes differentiating PD from other neuro disorders. The study has several compelling advantages, as follows:

1. The preliminary spotting of PD can be accomplished through audio techniques, eliminating the need for extensive clinical tests. This approach offers the advantages of reducing the associated costs and saving the valuable time in the diagnostic process.
2. A meticulous analogy of the proposed study conducted with the cutting-edge methodologies positions this study at the forefront in the field of research.

1.1 Motivation

The quality of life of a PD person is negatively impacted by all the movement and non-movement symptoms correlated with the senescent process [26]. In the initial stages of PD, it is essential to take the appropriate preventative measures, which can halt, slow down, or even reverse the progression of the disease. At present, there are no established hemoglobin test or examination procedure that can detect PD and its advancement. For the initial diagnosis of PD, doctors look at a patient’s case study and perform a cognitive assessment employing Unified Parkinson’s Disease Rating Scale (UPDRS) or Hoehn and Yahr (H&Y) scale [27]. However, these therapeutic assessment measures are dependent on the experience of skillful therapists, which results in subjective judgement being produced. Gathering all the pertinent information requires a huge amount of time, as well as the participation of multiple people.

1.2 Contributions

The major contributions made by this research for improving prediction accuracy are as follows:

- SMOTE (Synthetic Minority Oversampling TEchnique) helps to overcome the challenge of uneven class distribution in the dataset, by randomly generating synthetic samples of minority class.
- A remarkable increase in performance can be observed when features obtained from ensemble feature selection fusion process is applied on DCNN, rather than those obtained from individual feature selection methods.
- Implementing DCNN model in combination with NIA and conducting a thorough investigation and comparative analysis to evaluate the efficiency of the suggested model, in contrast to the most current research.
- Providing a comprehensive perspective on classification and analysis applications within the domain of PD diagnosis, intended to benefit PD diagnosis researchers and developers

The subsequent segments of this paper are meticulously structured as follows: The next section provides a concise summary of the contemporary cutting-edge developments, encompassing significant aspects and findings documented by other researchers who have pursued similar objectives. Later, a comprehensive explanation of the methodology for the proposed solution is followed and afterwards an evaluation process using commonly employed measures will be discussed in results section. Finally, in conclusion, the primary findings are discussed and potential avenues for future advancements have been provided.

2. Related Works

The identification of PD indicators in [3] depends on the binary categorization of voice recordings. Simply listening to patients' rough audio recordings was sufficient to make a diagnosis of PD. In addition to this, non-linear metrics of the variations in fundamental frequency were investigated. On the biomedical voice measurements of 147 patients and 48 non-patients collected from UCI, machine learning algorithms were implemented. The accuracy score can be enhanced to 99 percent with more enhancing on the model by using more statistical analysis and differently using values for 'k' in K-Fold cross validation. The model could be improved by rising the participant count, to more than 31 [14]. Deep neural networks were utilised in this method, which resulted in an improvement of the classification accuracy to 81.66 percent, as compared to the previously achieved accuracy of 44.3 percent. Implementing the model on a large data set that contains more cases (>5875) of each severity class is one way to further increase accuracy [15]. Using the ResNet algorithm and classifying the data based on the frequency-based features extracted from spectrograms led to much improved results. PC-GITA provides the source for the collection of voice recordings. ‘A’ was the only vowel that was taken into consideration; In addition to this, it was subjected only to a 10-fold validation and was applied on a limited data set [16]. Using the feature extension method, Principal Component Analysis (PCA) and Information Gain (IG) feature selection and reduction strategies were investigated to try and enhance classification accuracy. The realization of the feature expansion was made possible by the correlation analysis of several separate core feature sets. It is possible to accomplish this in two different ways: By expanding the set of features determined by PCA, and the other by expanding the set of features using iterative gradient method. In order to inflate the accuracy of classifiers, it is also possibleto investigate a variety of feature Selection methods, such as the chi-square test, random forest, and regularization, as well as reduction methods [17]. Solana et.al. identifies gender-specific factors in PD detection, emphasizing high-frequency voice content for women, while low, for men. They introduce feature variability for contextual interpretation, aiding classification understanding. Voice recordings are promoted as a non-intrusive, cost-effective PD diagnostic tool, aligned with neurobiology findings on gender-dependent features and guiding gender-specific diagnostics. The study highlights the merits of pattern recognition and machine learning for preliminary PD detection and improved patient care. However, potential limitations and challenges in clinical use are not discussed, and factors beyond gender influencing detection accuracy are unexplored [23]. Imran et.al employed a dataset with voice measurements collected from 31 individuals, where 23 are with PD and 8 belong to HC. It uses different machine learning classifiers. Random Forest achieved the highest accuracy at 97%. The study contributes to ParkDet 2.0, aiding PD diagnosis without complex systems. However, it lacks details on extracted features and patient demographics like age, gender, or disease stage [24]. Narendra et.al utilized the data set obtained from the PC-GITA speech database and made significant contributions by focusing on the detection of PD through the classification of speech signals. It highlights the use of traditional and advanced Glottal Inverse Filtering methods, to extract voice source information and 192 glottal features per utterance. The study explores detection system architectures using deep learning models and the study proved the enhancement of detection accuracy. Critically, the findings accentuate the pivotal role of voice source information in diagnosing PD and illuminate promising pathways for predicting the neurological status of PD patients. However, the paper has limitations, including a focus solely on voice-based features, the use of a specific Spanish database, reliance on deep learning models that require substantial labeled data, lack of performance comparison with other methods, and limited discussion on GIF method limitations [25]. With the usage of PCA, the number of false positives has been reduced drastically. The Speech Signals that were used were taken from previously conducted research, and the algorithm was tested only on 1040 voice recordings. In this work, binary as well as multi-class classification were studied. Table 1. provides a review of the existing research works, for PD diagnosis using speech signals.

Table 1. Exploration of the available methods for PD detection using voice signals

Ref/Author	Methodology	Database	Recordings	Accuracy (%)
[14] Dinesh et. al.	Machine Learning	UCIML	295	91.21
[15] Srishti et. al.	Deep neural networks	UCIML	5875	81.66
[16] Marek et. al.	Deep Learning	PC-GITA	HC-50; PD-50;	91.7
[17] Mei Jie et. al.	Machine Learning	UCIML	PD-20; HC-20;	Logistic regression :76.03 Linear Kernel SVM :75.49
[23] Solana et. al	Machine Learning	UCIML	PD-188; HC-64;	95.9
[24] Imran et. al	Deep Learning	UCIML	PD-23; HC-8;	RF-97
[25] Narendra et. al	Machine Learning	PC-GITA	PD-50; HC-50;	Pipeline systems-67.93 End-to-end systems-68.56
[28] Sousa et. al	Deep learning	UCIML	PD-20; HC-20;	Binary Classification: without PCA: 91.80; with PCA:93.44 Multi class Classification: without PCA:69.93; with PCA:84.67
Proposed Approach	**Deep Learning + Nature Inspired Algorithm**	**UCIML** **PPMI**	**PD-20; HC-20;** **PD-53; HC-8; Others-20;**	**Binary classification:97.6** **Multi-class classification: 93.7**

3. Methodology

3.1. *Proposed System*

The objective of the proposed method is to detect PD by analyzing voice samples using acousticanalysis. In the initial step, the existence of the disease will be identified. In the subsequent phase, PD will be differentiated from other diseases, which exhibits a greater level of difficulty in comparison to the previous phase. Therefore, a more intricate network architecture is employed to improve the discrimination of data, by making use of NIA for obtaining optimal parameters in training the DCNN model. As a novel aspect, Average Weighted Rank Feature Selection (AWRFS) ensemble method using feature selection fusion mechanism was implemented, for the very first time and can be seen from Fig. 2(a). The workflow of the current study can be observed from Fig. 2(b) and 2(c).

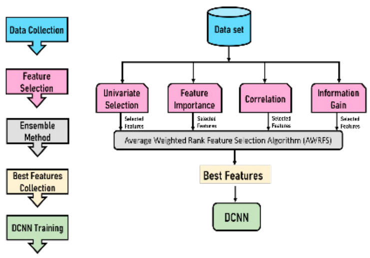

Fig. 2(a). Ensemble mechanism for feature selection

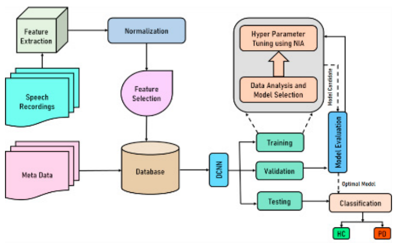

Fig. 2(b). Proposed methodology for binary classification

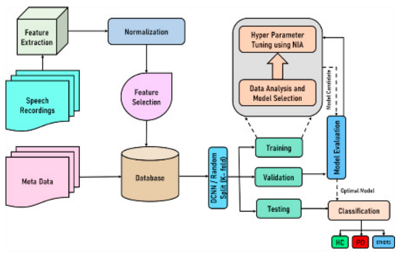

Fig. 2(c). Proposed methodology for multi-class classification

3.2. Materials and Tools

The data for this work for binary classification, was collected from University of California Irvine Machine Learning (UCIML) database (*https://archive.ics.uci.edu/*), a public available database [13]. It contains voice samples collected from 40 participants: 20 with PD and 20 HC. 96 kHz frequency was employed to sample the recordings. Note that UPDRS rating scales were collected by laboratory personnel, who had completed online training and not a board-certified neurologist. Each participant provided 26 voice samples (standard *vowels a,o,u; numbers 1 to 10; short sentences and words*), which were extracted using linear and time-frequency techniques. The total number of recordings obtained for training were 1040. For multi-class classification, the data was collected from a public database, PPMI. This database is composed of 81 speakers: 53 with PD contributing 177 voice samples, 8 HC contributing 48 voice samples and others suffering from other neurological diseases are 20, contributing to a sum of 20 speech samples. Table 2 describes the data used in this study while Table. 3 Presents an elaborated description of the data.

Table 2. Data Split

Type	Binary		Multi-Class	
Dataset	UCIML		PPMI	
PD	20	520	53	177
HC	20	520	8	48
OTHERS	--	--	20	20
TOTAL	**40**	**1040**	**81**	**245**

Table 3. Data set Description

Class	Types of Parameters used
Binary Classification / UCIML	Frequency Parameters Amplitude Parameters Pitch Parameters Harmonicity Parameters Pulse Parameters
Multi-Class Classification / PPMI	Average, Maximum, Minimum vocal frequency Variation measures in fundamental frequency Variation measures in amplitude Measures of the ratio of noise to harmonicity in voice

3.2 Feature Selection

To optimize the efficiency of the proposed algorithm, different feature selection mechanisms like univariate selection, feature importance, correlation and information gain [22] are employed to obtain the scores of importance for all features in the data sets corresponding to binary as well as multi-class classification. From these, the top most 'T' important features were selected as a next step, in the process. Now, an ensemble mechanism is applied on the outputs obtained from the four methods, contributing to AWRFS algorithm. All the output features of each feature selection method are arranged in falling order of their score of importance and each sample is assigned a rank as well as weight. The feature with the maximum score of importance is assigned the highest rank 'n' ('n' is the total number of features obtained using the method) and the maximum weight of '1'. The sample with next highest score is assigned the next rank 'n-1' and weight of '(n-1)/n'. The process is continued until the last feature is assigned the lowest rank '1' and the least weight of '1/n'. for all the other features, ranks and weights will be assigned as 'zero' to indicate their absence for consideration. Now, the average weighted rank operation is performed on each feature and the corresponding scores are obtained. These are compared against a specified threshold value (here, threshold=0.5), and those features with scores equal to or beyond the threshold are selected as the best features, while neglecting the remaining (values < 0.5). The AWRFS algorithm can be observed from algorithm-1.

3.3 DCNN

The initial stage involved in binary classification process is extracting the characteristics from the audio recordings, followed by data normalization using a z-score algorithm. The best features are selected using ensemble mechanism during feature selection process. In the pursuit of the highest standards for model evaluation, we adhere to a meticulous methodology. This process involves the random generation of three distinct subsets during each session: a training subset for the precise refinement of model parameters, a validation subset focused on the identification of the optimal model, and a stringent test subset for the comprehensive and rigorous evaluation of the selected model's performance. A meticulously designed Deep Convolution Neural Network (DCNN) was

Algorithm-1: The AWRFS Algorithm

1. Input: The Feature Selection methods ($FS_1, FS_2, \dots \dots FS_n$) were selected to be implemented on the voice dataset. The methods Produced the outcomes $[O_1, O_2, O_3, \dots \dots O_n]$; total number of methods 'm; Best Features set, 'BF={ }';
2. Output: Set of the Best Features 'BF={ best features }' using AWRFS algorithm.
3. Begin
   - 3.1 For each ($m_i \in M$), where $\mathbf{m_i}$ is the feature selection method and 'M' is total number of methods employed
     - 3.1.1. Arrange the features ($F_1, F_2, \dots \dots F_n$) obtained as outputs, in descending order of their scores of importance. Assign Ranks ($R_1, R_2, \dots \dots R_n$) to these features and Weights ($W_1, W_2, \dots \dots W_n$) in such a way that the feature with the highest score of importance (FS_i, say), will be assigned the highest rank 'n' ($R_i = n$) and highest weight ($W_i = 1$), the model with the next highest performance (FS_x, say),will be assigned the next rank '(n-1)' ($R_x = n - 1$) and weight of '(n-1)/n' ($W_x = [(n-1)/n]$) and so on.
     - 3.1.2 Repeat the process until the feature with the least score of importance will be assigned the lowest rank '1' and lowest weight '1/n'.
   - 3.2 Repeat step 3.1 until all methods are done.
   - 3.3 As a subsequent step, calculate the Average Weighted Rank of each feature from the corresponding 'M' methods as the ratio of sum of the product of feature Rank ($R_1, R_2, \dots \dots R_M$) and their associated weights ($W_1, W_2, \dots \dots W_M$) to the total number of feature selection methods 'M', as follows:

$$\textbf{Average Weighted Rank} = \frac{\textbf{sum of the product of feature rank and feature weight}}{\textbf{total number of feature selection methods}}$$

$$\textbf{Average Weighted Rank} = \frac{\sum_{i=1}^{M} \boldsymbol{R_i * W_i}}{\mathbf{M}}$$

   - 3.4 If the value obtained in the previous step is greater than threshold value, 'th', add the feature to the set of Best Features; BF={ }+F_i ; otherwise, Neglect it.
   - 3.5 Repeat steps 3.3 and 3.4, until all the features are processed.
4. End.

incorporated for classification of PD from HC. The input layer is designed with sufficient neurons to support the number of selected features obtained through ensemble mechanism. Leveraging the Rectified Linear Unit (ReLU) activation function in both the input and hidden layers captures intricate non-linear patterns, activating output exclusively for positive input values. This also addresses the vanishing gradient problem, inherent in multi-layer models further enhancing training optimization, outperforming previous neural network architectures. The output layer, housing a single neuron activated by the sigmoid function, yields a binary classification outcome, complemented by a rounding operation for precision. Conscientiously implementing dropout after the input layer plays a pivotal role in mitigating overfitting, randomly deactivating half of the input layer neurons during each training epoch for better generalization. Five concealed layers are utilized, each with the same number of units as the input layer, and similarly activated using ReLU. The model's parameters are carefully chosen using the Genetic Algorithm (GA), a nature inspired algorithm, which is a powerful optimization and search technique inspired by the process of natural selection and evolution. The basic idea behind a Genetic Algorithm is to simulate the process of natural evolution to evolve a population of potential solutions to a problem over successive generations. The DCNN architecture employed for binary classification can be depiced from Fig. 3(a)

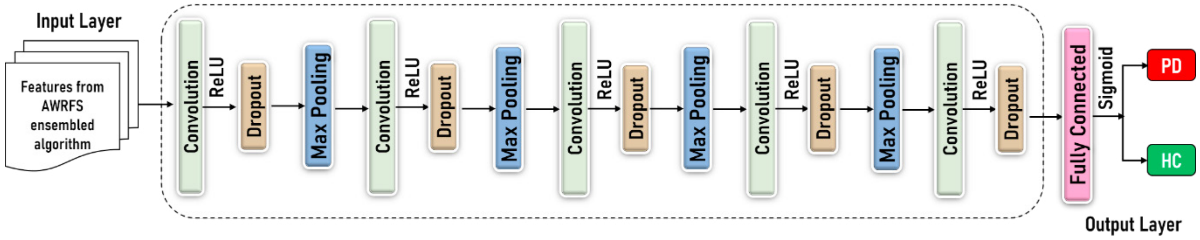

Fig. 3(a) DCNN architecture for binary classification

In addressing the multiclass problem, characteristics are extracted from the audio recordings, followed by data normalization using a z-score algorithm. The best features are selected using ensemble mechanism during feature selection process. A similar network to the one used in binary classification process, deviating slightly from the conventional approach, has been employed, here. The input layer remains unchanged, adjusting the neurons employed, to match the features and ReLU activation function was employed. A dropout is observed after the initial hidden layer. Five concealed layers are utilized, each with units equal to those used in the input layer, and similarly activated using ReLU. The output layer comprises 3 neurons, representing 3 distinct classes of illness, with the softmax function as the chosen activation. The likelihood of each feature set belonging to the classes is determined, and the class with the maximum score is assigned. The DCNN architectures employed for multi-class classification can be observed from Fig. 3(b).

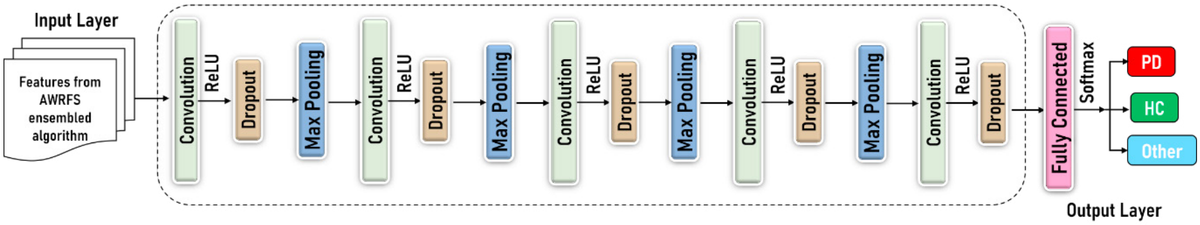

Fig. 3(b) DCNN architecture for multi-class classification

Fig. 4 and Fig. 5 represent the convolution layered architectures employed for binary and multi-class classification schemes respectively.

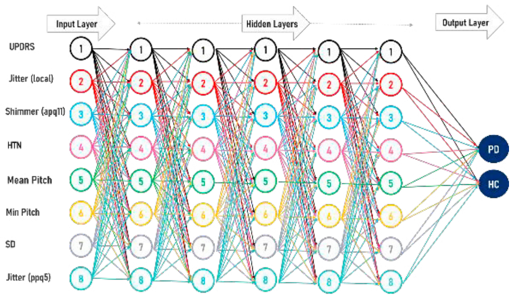

Fig. 4. binary classification layered architecture

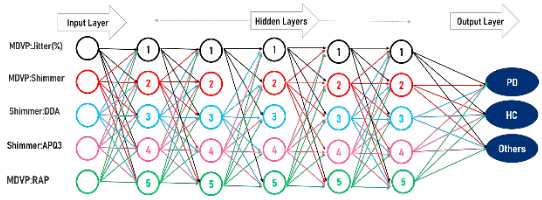

Fig. 5. multi-class classification layered architecture

A K-Nearest Neighbors (KNN) architecture is designed for similarity, utilizing the complete training dataset to identify ‘k’ similar instances for prediction. In scenarios involving categorical data or binary nature, the similarity between instances is ascertained by computing the Euclidean distance. Hence, when there arises a need to forecast outcomes for unfamiliar data, the model diligently scrutinizes the training dataset, identifying the k samples that bear the closest resemblance. Subsequently, the model predicts the title that appears most frequently among these k samples, facilitating accurate and informed predictions for new, unexplored data. The KNN algorithm is recognized as a non-linear classifier, offering simplicity and favorable performance on distinctly clustered datasets. To thoroughly assess the effectiveness of the proposed methodologies, k-fold cross-validation was employed during the training, validation, and testing phases. The method entailed partitioning the training dataset into 'k' subsets of equal size. One subset was dedicated for validation purposes, while the remaining 'k-1' subsets were utilized for training the model. This iterative process was repeated 'k' times, with each subset acting as the validation set once, ensuring a comprehensive evaluation of the methodology performance. The training and testing process is iterated for every subgroup, resulting in a cumulative count of ‘k’ unique runs. There are many primary benefits associated with the use of k-fold method are: Firstly, during the training phase, regularization can effectively reduce the likelihood of overfitting. Secondly, during the testing phase, regularization allows for the evaluation of the model's robustness in the face of parameter alterations and weight initialization, as well as the detection of potential local minima [21].

4. Results and Discussion

The proposed methods were developed using Python version 3.5, executed on a Windows i5 PC with 12 GB of memory. The GA works by maintaining a population of individuals (also known as chromosomes or solutions), each representing a potential solution to the problem at hand. A total of 100 epochs, a learning rate of 0.001, a decay rate of 0.01, andamomentum of 0.5 shown promising results for binary classification. The Adam optimizer combined with the binary cross-entropy loss function, further elevated the model's overall performance. This thoughtfully crafted architecture seamlessly blends the cutting-edge deep learning advancements with established techniques, delivering remarkable results for the binary classification problem. The study of multi-class classification uses specific parameter values as tuned by GA, including 80 epochs, a learning rate of 0.01, decay rate of 0.1, momentum value of 0.3, and Adam optimizer. The categorical cross-entropy loss function is employed for analysis. While employing KNN mechanism, a value of k = 10 was selected as a trade-off between the level of statistical analysis accuracy and the duration required for training and testing. Moreover, to detect potential instances of model overfitting, the k-fold methodology was adopted for evaluating both, the test data and the validation data. By comparing the resulting outcomes and their corresponding statistical measures, such as the mean and standard deviation, more insights were gained into the model's generalization performance and signs of overfitting can be identified, if any. To rigorously evaluate and compare the results obtained, accuracy was chosen as the primary metric. The NIA (GA) results over different hyper parameter values can be observed from Table. 4. The finalized hyper parameters can be seen in Table 5.

Table 4. NIA (GA) results over different hyper parameter values

Optimizer	Learning rate	Binary Classification				Multi-Class Classification			
		Training		Validation		Training		Validation	
		loss	acc	loss	acc	loss	acc	loss	acc
SGD	0.1	0.5	84.9	0.58	81.4	0.41	81.76	0.47	78.64
Adam		0.41	90.11	0.5	87.71	0.25	89.85	0.36	83.48
SGD	**0.01**	0.45	88.4	0.51	85.6	0.26	86.82	0.31	84.92
Adam		0.29	93.4	0.33	90.8	**0.18**	**93.7**	**0.25**	**91.43**
SGD	**0.001**	0.31	92.5	0.38	89.85	0.38	81.29	0.42	79.19
Adam		**0.21**	**97.6**	**0.30**	**94.75**	0.23	88.08	0.31	84.62
SGD	0.0001	0.41	87.2	0.49	85.32	0.5	77.07	0.56	75.67
Adam		0.31	94.7	0.36	90.45	0.37	82.16	0.45	77.19

Table 5. NIA(GA) Hyper Parameters

Parameter	Binary Classification	Multi-Class Classification
Epochs	100	80
Learning Rate	0.001	0.01
Decay rate	0.01	0.1
Momentum	0.5	0.3
Optimizer	Adam	Adam

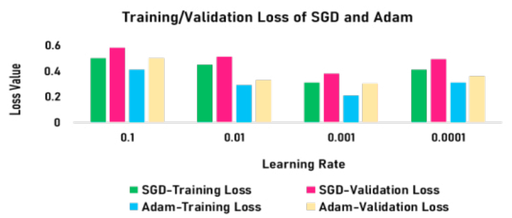

Fig. 6(a) Loss values of binary classification

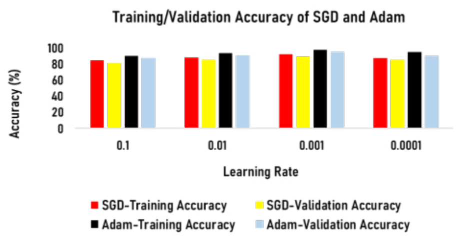

Fig. 6(b) Accuracy values of binary classification

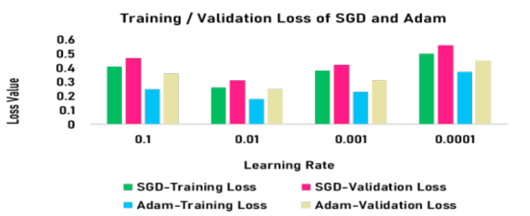

Fig. 6(c) Loss values of multi-class classification

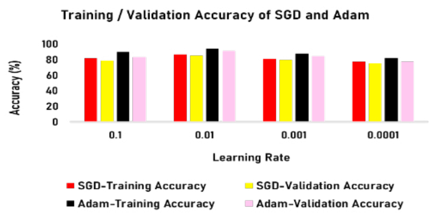

Fig. 6(d) Accuracy values of multi-class classification

The loss and accuracy obtained during binary as well as multi-class classification can be depicted from Fig. 6(a), 6(b), 6(c) and 6(d) respectively. In order to assess the learning capacity of the model and the efficacy of the associated parameters, the process was initiated by graphically depicting the accuracy and loss curves across different learning rates. This is done to verify the appropriate progression of the learning process. Based on the examination of Fig. 6(a) and 6(c), it is evident that, throughout the training phase, the network weights undergo repetitive adjustments with the objective of minimizing the loss function. The aforementioned progression of the weights has also been demonstrated to result in increased levels of accuracy (Fig. 6(b) and 6(d)). The network's performance demonstrates a tendency to stabilize after around 75 training epochs. Furthermore, the analysis of the accuracy and loss metrics calculated for both the training and validation datasets can be observed from Table 6. It provides evidence that the trained model did not experience any issues related to overfitting. The study performance is compared with state-of-the-art works and can be observed from Fig. 7

Table 6. Performance of binary and multi-class classification

Classification	Binary		Multi-Class			
Method	DCNN		KNN		DCNN	
Operation	Without GA	With GA	Without GA	With GA	Without GA	With GA
Train	96.18	99.78	81.25	88.57	92.66	97.11
Validation	93.12	95.62	78.61	84.12	85.02	95.05
Test	92.05	**97.67**	74.56	81.55	81.64	**93.74**

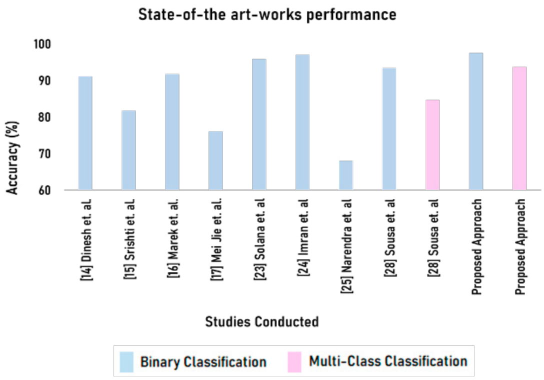

Fig. 7 State-of-the-art-works performance comparison

5. Conclusion and Future Work

This work presents a novel approach for the characterization of PD through the analysis of speech samples, including deep structured learning techniques and nature inspired algorithms. The application of the ensemble approach using AWRFS algorithm gave the best features as output with which DCNN and NIA together, has yielded remarkable results. An exceptional accuracy of 97.67% for binary classification and 93.7% for multi-class classification were achieved. This remarkable progress underscores the potential of this approach for accurate PD diagnosis, primarily based on speech signals. Nonetheless, it is imperative to acknowledge certain limitations within this study. Notably, patient characteristics such as age, geographical location, dietary habits, and gender were not factored into the classification process, which could potentially influence the overall performance. Another constraint is the inability to determine the disease stage of PD patients using the H&Y scale. It is recommended that a combination of metrics or comprehensive statistical tests be employed, as a single clinical metric may not be sufficient to accurately track PD progression [29]. This study, while currently centered on vocal signals, underscores the necessity of future research to encompass various other motor and non-motor signals and to develop more efficient architectures for PD detection. Future research endeavors should prioritize the inclusion of these factors to enhance the accuracy and practicality of the proposed method, alongside the development of imaging biomarkers for PD and related conditions. As a conclusion, this study marks a significant stride forward, offering valuable insights to both researchers and clinicians in the domain of voice signals evaluation through deep learning. The objective identification of PD is poised to become increasingly accessible and reliable for clinicians, owing to the ongoing advancements in deep learning techniques, promising a bright future for the field.

References

[1] Ho, A.K., Iansek, R., Marigliani, C., Bradshaw, J.L., Gates, S. (1998) "Speech impairment in a large sample of patients with Parkinson's disease." *Behav. Neurol.* **11,** 131–137,

[2] Ramig, L.O., Fox, C., Sapir, S.: Speech treatment for Parkinson's disease. Expert Rev. Neurother. 8, 297–309 (2008)

[3] Khemphila, A., Boonjing, V. (2012) "Parkinsons disease classification using neural network and feature selection." *Int. J. Math. Phys. Electr. Comput. Eng*. **6,** 377–380,

[4] Hirsch, L., Jette, N., Frolkis, A., Steeves, T., Pringsheim, T. (2016) "The incidence of Parkinson's disease: a systematic review and meta- analysis." *Neuroepidemiology* **46**, 292–300,

[5] "Neurological disorders: public health challenges." World Health Organization, (2006).

[6] Müller, J., Wenning, G.K., Verny, M., McKee, A., Chaudhuri, K.R., Jellinger, K., Poewe, W., Litvan, I. (2001) "Progression of dysarthria anddysphagia in postmortem-confirmed parkinsonian disorders." *Arch. Neurol*. **58**, 259–264,

[7] Liu, L., Luo, X.-G., Dy, C.-L., Ren, Y., Feng, Y., Yu, H.-M., Shang, H., He, Z.-Y. (2015) "Characteristics of language impairment in Parkinson's disease and its influencing factors." *Transl. Neurodegener*. **4, 2**.

[8] Teixeira, J., Soares, L., Martins, P., Coelho, L., Lopes, C. (2017) "Towards an objective criterion for the diagnosis of Parkinson disease based on speech assessment", *XXXV Congresso Anual de la Sociedad Espanola de Ingenierıa Biomedica, Bilbao*.

[9] Norel, R., Agurto, C., Rice, J.J., Ho, B.K., Cecchi, G.A. (2018) "Speech-based identification of L- DOPA ON/OFF state in Parkinson's Disease subjects." *BioRxiv Prepr*. **420422.**

[10] Zhang, Y.N. (2017) "Can a smartphone diagnose Parkinson disease? A deep neural network method and telediagnosis system implementation." *Park. Dis.*

[11] Braga, D., Madureira, A.M., Coelho, L., Abraham, A. (2018) "Neurodegenerative diseases detection through voice analysis." *Hybrid Intelligent Systems*. Springer, pp. 213–223.

[12] LeCun, Y., Bengio, Y., Hinton, G. (2015) "Deep learning". *Nature* **521**, 436–444.

[13] Erdogdu Sakar, B., Isenkul, M., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., Kursun, O. (2013) "Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings." *IEEE Journal of Biomedical and Health Informatics*, **vol. 17(4),** pp. 828-834.

[14] Dinesh, Akshaya, and Jennifer. (2017) "Using machine learning to diagnose Parkinson's disease from voice recordings." *2017 IEEE MIT Undergraduate Research Technology Conference*, IEEE, pp. 1-4.

[15] Srishti Grover, Saloni Bhartia, Abhilasha Yadav, KR Seeja. (2018) "Predicting Severity of Parkinson's Disease Using Deep Learning", *Procedia Computer Science,* **132**, pp. 1788-1794.

[16] Marek Wodzinski1, Andrzej Skalski1, Daria Hemmerling1, Juan Rafael Orozco-Arroyave2, Elmar Noth. (2019) "Deep Learning Approach toParkinson's Disease Detection Using Voice Recordings and Convolutional Neural Network Dedicated to Image Classification." *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 717–720.

[17] Mei Jie, Desrosiers Christian, Frasnelli Johannes. (2021) "Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature." *Frontiers in Aging Neuroscience*, **Vol. 13**, pp. 1663-4365.

[18] Filippini, Massimo, and Lester C. Hunt. (2011) "Energy demand and energy efficiency in the OECD countries: a stochastic demand frontier approach." *Energy Journal* **32** (**2**): 59–80.

[19] https://stanfordmedicine25.stanford.edu/the25/parkinsondisease.html

[20] Claesson. (2018). "Better Balance with Somatosensory Exercises-a Parkinson Perspective." *Ingrid.*

[21] Bhatt, Chandradeep, Indrajeet Kumar, V. Vijayakumar, Kamred Udham Singh, and Abhishek Kumar. (2021) "The state of the art of deep learning models in medical science and their challenges." *Multimedia Systems* **27, no. 4**, 599-613.

[22] Chaurasia, Vikas, and Aparna Chaurasia. (2023). "Detection of Parkinson's Disease by Using Machine Learning Stacking and Ensemble Method." *Biomedical Materials & Devices*, 1-13.

[23] Solana-Lavalle, Gabriel, and Roberto Rosas-Romero. (2021). "Analysis of voice as an assisting tool for detection of Parkinson's disease and its subsequent clinical interpretation." *Biomedical Signal Processing and Control* **66** :102415**.**

[24] Ahmed, Imran, Sultan Aljahdali, Muhammad Shakeel Khan, and Sanaa Kaddoura. (2022). "Classification of Parkinson disease based on patient's voice signal using machine learning." *Intelligent Automation and Soft Computing*, **32, no. 2**: 705.

[25] N. P. Narendra, B. Schuller and P. Alku. (2021). "The Detection of Parkinson's Disease from Speech Using Voice Source Information." *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **vol. 29,** pp. 1925-1936.

[26] Chaudhuri KR, Schapira AH. (2009). "non-motor symptoms of Parkinsons disease: dopaminergic pathophysiology and treatment." *Lancet Neurol*. https://doi.org/10.1016/S1474-4422(09)70068-7

[27] Xu J, Zhang M. (2019). "Use of magnetic resonance imaging and artificial intelligence in studies of diagnosisof PD." *ACS Chem Neurosci*. https://doi.org/10.1021/acschemneuro.9b00207

[28] Sousa, Lígia, Diogo Braga, Ana Madureira, Luis Coelho, and Francesco Renna. (2018). "Characterizing Parkinson's Disease from Speech Samples Using Deep Structured Learning", *International Conf. on Soft Computing and Pattern Recognition,* Springer, pp. 137-146.

[29] Huang, D., Sherman, B. & Lempicki, R. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." *Nat Protoc* **4**, 44–57. https://doi.org/10.1038/nprot.2008.211