Computers in Biology and Medicine 197 (2025) 110924

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Hybrid greylag goose and particle swarm optimization for early detection of Parkinson's disease from speech features

Doaa Sami Khafaga [a], Marwa M. Eid [b,c], Ehsan Khodadadi [d], El-Sayed M. El-Kenawy [e,f,*], Amel Ali Alhussan [a], Nima Khodadadi [g,*]

[a] Department of Computer Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[b] Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura, Egypt
[c] Jadara University Research Center, Jadara University, Jordan
[d] Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR, 72701, USA
[e] School of ICT, Faculty of Engineering, Design and ICT, Bahrain Polytechnic, Isa Town, Bahrain
[f] Applied Science Research Center, Applied Science Private University, Amman, Jordan
[g] Department of Civil and Architectural Engineering, University of Miami, Coral Gables, FL, USA

ARTICLE INFO

Dataset link: https://www.kaggle.com/datasets/debasisdotcom/parkinson-disease-detection

Keywords:
Greylag goose optimization
Particle swarm optimization
Parkinson's disease
Classification
Speech features

ABSTRACT

Parkinson's disease is brought on by a disturbance in the functions of the brain cells that are responsible for the production of dopamine, which is a chemical that enables brain cells to interact with one another. The cells in the brain responsible for the production of dopamine are the ones in charge of the regulation, adaptability, and fluency of movements. When sixty to eighty percent of these cells are gone, there is a lack of sufficient dopamine production, which makes Parkinson's motor symptoms manifest. In this study, a robust diagnosis of Parkinson's disease is presented in terms of optimized machine-learning models. Specifically, a novel hybrid optimization algorithm is proposed for both feature selection and optimization of the parameters of a neural network to boost the classification of Parkinson's disease. The proposed optimization algorithm is a hybrid of the greylag goose and particle swarm optimization algorithm, denoted by GGPSO. This approach is applied to classifying Parinson's disease from speech signals. The proposed feature selection method is compared to recent methods, showing a promising performance with superior results. In addition, the performance of the proposed optimization algorithm is compared to other optimization algorithms to prove its superiority and effectiveness in optimizing the parameters of the neural network. The overall classification accuracy using the proposed approach is 99.4%, which outperforms the other competing methods. On the other hand, statistical tests have been performed to show the stability and statistical difference of the proposed method. The results of these study confirmed the achieved outcomes and showed robust performance of the proposed methodology in classifying Parkinson's disease.

1. Introduction

The neuro-degenerative brain disorder known as Parkinson's disease (PD) is characterized by a sluggish progression [1]. The term "neuro-degenerative" refers to disorders that result in the death of brain cells. When the human brain is functioning normally, certain areas contain cells responsible for producing dopamine. A specific region of the brain known as the substantia-nigra is where the majority of these cells are found by concentration. One of the chemicals that is responsible for the transmission of information between the substantia nigra and other parts of the brain that are responsible for controlling bodily movements is called dopamine [2]. Dopamine enables individuals to synchronize their actions smoothly and harmonically. The motor symptoms of PD manifest themselves when sixty to eighty percent of the cells that make dopamine are destroyed. This results in an insufficient amount of dopamine being created.

In the first stage of PD, symptoms manifest in the enteric nervous system, the lower brain stem, and the olfactory tracts. These areas are the starting point for the progression of PD to the upper regions of the brain, specifically the substantia nigra and the brain shell. Several years before the onset of motor symptoms, such as a diminished or absent sense of smell, sleep difficulties and constipation, tremors, and a slowing of movement, the disease is believed to develop. A further ninety percent of those who have PD experience voice difficulties [3].

* Corresponding authors.
E-mail addresses: skenawy@ieee.org (E.-S.M. El-Kenawy), nima.khodadadi@miami.edu (N. Khodadadi).

https://doi.org/10.1016/j.compbiomed.2025.110924
Received 26 April 2025; Received in revised form 26 June 2025; Accepted 9 August 2025

The researchers are thus looking for techniques to recognize these non-motor symptoms that occur early throughout the disease as early as possible to stop the advancement of the disease once and for all.

In recent years, machine learning (ML) has been increasingly popular for detecting medical diseases due to the ease with which it can be implemented and its high accuracy level. In the research that has been done, ML has also been utilized to treat PD. A review of the publications for feature selection (FS) to be used for machine learning in brain surgery was conducted by the authors of [4]. A strategy that is based on machine learning is used to pinpoint the actual region of the brain that needs to be operated on during brain surgery for PD. Post-diagnosis studies are the primary subject of this particular piece of writing. To quantify the cognitive effects of PD, the authors of [5] utilized machine learning techniques. Authors in [6] used a machine learning program to predict the tremor degree of PD patients. ML was also used to achieve stage prediction of PD [7]. However, the majority of the studies concentrate on the early identification of PD using this well-known method, ML. Using motion data collected from people's upper limbs, the authors of [8] attempted to predict PD.

The authors instructed the experimental subjects, who included both persons with PD and healthy individuals, to carry out a series of performance tests while wearing a device that was implanted into their upper limbs. To obtain parameters, a spatial–temporal and frequency data analysis was carried out. Following this, various supervised learning techniques were utilized for the classification process. Feature extraction and machine learning approaches were used to identify PD in [9]. According to this study, phonation is one of the most practical ways to diagnose PD. The k-nearest neighbors (k-NN), multilayer perceptron (MLP), optimum path forest (OPF), and support vector machine (SVM) classifiers were also employed. Artificial neural networks reduced vocal features for ML-based PD detection [10]. Using SVM for classification. Unlike the other approaches, PD was also done unmonitored [11]. After dimension reduction with partial least squares, a self-organizing map (SOM) was used for clustering and incremental support vector regression prediction. The study's findings anticipated the use of the unified PD tating scale (UPDRS).

The authors of [12] evaluated and compared different machine learning approaches and found neural networks to be the best. For PD diagnosis, [13] used fuzzy C-means clustering for feature weighting and k-nearest neighbors for classification. A weighted PD dataset was fed to a k-NN classifier in several configurations to find the best k value. Extreme Learning Machines were utilized to diagnose PD [14]. We used a weighted technique and a non-linear kernel function mapping to ameliorate unbalanced data. The artificial bee colony (ABC) method was utilized for both the FS and the optimization of the parameters. Using principle component analysis (PCA) for dimension reduction, fisher discriminant ratio (FDR) for fisher discriminant analysis (FDA), and SVM for classification, the authors of [15] were able to successfully choose the diagnosis of PD. Even though they achieved extremely high classification accuracy, they included more than one hundred features that were retrieved from brain magnetic resonance imaging (MRI) images. On account of this, their computational cost is rather significant.

Authors in [16] looked at gait and tremor as potential diagnostic tools for PD. It was possible to extract gait features from the data collected from the PhysioNet database. The physical signals that were received from the sensors that were inserted underneath the participants' feet were analyzed to determine the peaks and pulses associated with them. They were able to attain a classification accuracy of 92.7%. The authors of [17–20] utilized voice features from a shared dataset. The feature augmentation technique was used, and the result was the collection of 177 features from the dataset's 44 features. Following the feature augmentation process, Relief was utilized to choose the features that proved to be the most successful. There are 66 features utilized for the categorization of PD. Using machine learning techniques, PD was also detected using handwriting exercises [21–24] instead of MRI, motion, or speech data.

Although the literature for machine learning-based diagnosis of PD produced reasonable classification rates, the authors either utilized a large number of features (such as [25–29]), which increased the amount of time required for calculation, or the extraction of the features was difficult, even if they used a small number of features. Because of this, the amount of time required for computation is elevated once again. With the help of a lightweight feature extraction procedure and a classifier, the authors of this study have attempted to reduce the amount of time required for computation by using a smaller number of practical features. From the speech signals, the features are extracted, and as a result, the process of extracting the features is less complicated than the other approaches in the literature that are based on MRI [30–34] or motion [35–37]. They employed a greater number of features than the proposed technique, despite the fact that other writers (such as [38–40]) used voice features to diagnose PD. However, they employed MRI data for feature extraction, which is more difficult than extracting voice features. Following is a list of the primary contributions that the paper makes:

1. Proposing a novel optimization algorithm denoted by hybrid greylag goose and particle swarm optimization (GGPSO) algorithm for feature selection.
2. Employing the proposed GGPSO algorithm to optimize the parameters of a neural network classifier.
3. Statistical analysis of the proposed methodology using analysis of variance (ANOVA) and Wilcoxon signed rank tests to prove its stability and statistical significance.
4. Comparing the proposed methodology with other methods to prove its superiority.
5. Presenting high-accuracy classification based on the proposed algorithms.

Section 2 reviews the previous work on applying machine learning methods for Parkinson's classification. Section 3 explains the material and techniques used in the proposed methodology. Section 4 discusses the proposed method. Section 5 discusses the results recorded from the experiments. Section 6 finally derives the conclusions of this work.

2. Related works

PD is a complex neurodegenerative disorder primarily characterized by motor dysfunction, including tremors, bradykinesia, and postural instability. Historically, clinical quantification of these symptoms has mainly been achieved by observing patients' clinical condition, which is prone to increased variability and diagnostic latencies. These challenges have provoked recent research on using machine learning algorithms and other technologies to enhance the determination of symptoms of PD. They have included wearable sensors, neuroimaging, and speech data, improving early diagnosis and disease progress tracking. This strategy of the literature review is to review the existing literature on the use of machine learning in PD and understand how these technologies are changing the nature of PD diagnosis and therapy. According to [41], the current methods for assessing these symptoms are predominantly based on clinical visits, often involving subjective evaluations. While a patient's diary can give additional information about symptoms manifest at home, the data collected is usually subjective and contains distortions and gaps. To overcome these issues, researchers have provided different automated in-home monitoring systems to provide a more accurate, reliable, and continuous measure of PD symptoms.

When analyzing about 12,000 publications, the authors observed an evolution of technologies and tools for computations for the past five decades. Early approaches involved needle-based electromyography (EMG) in laboratory settings, gradually transitioning to wearable accelerometers and gyroscopes for more practical, in-home use. Recently, attention has been paid to radically real-time monitoring using mobile phones and Web applications. Similarly, there have been significant enhancements in the use of devices, including video cameras and accelerometers, for patients with PD using machine learning. The development, along with the technology and further improvement of ML methods, has significantly contributed to this field, using PD motor symptoms to be monitored more accurately and quickly in the home setting.

Researchers have been exploring ML approaches to enhance the accuracy and efficiency of assessing symptoms of PD. As detailed by [42], a study investigated the feasibility of using machine learning algorithms to automate the rating of two primary symptoms of PD: Based on the presented weighted κ (kappa) and intraclass correlation coefficient (ICC) scores, the most dominant motor symptoms of the patients were resting tremors and bradykinesia. Participants consisted of 55 patients who completed video recording while resting and performing finger-tapping movements; this research used open pose, a deep learning-based human pose estimation software. Key motion parameters, such as the amplitude of the tremor and the speed, amplitude, and fatigue in finger tapping, were extracted to develop an automatic rating system based on the Unified PD Rating Scale (UPDRS) using a support vector machine algorithm. The proposed machine learning model was then evaluated against the moving disorder specialist's ratings on the same set of pictures, and there was a strong correlation between the tremor and bradykinesia ratings. In the case of resting tremors, the accuracy achieved by the system was higher than that of nontrained human raters, and for finger tapping, the results were comparable. The works under investigation show that machine learning can be successfully used as an automated tool for accurately evaluating Parkinson's symptoms, which may interest clinical practice.

Predicting measures of motor progression in PD is particularly important in determining the qualitative and quantitative outcomes in clinical trials. As noted by [43], the study tested various algorithms to improve these predictions by analyzing data from 204 patients, using 18 clinical and imaging features. Three approaches were used: 10 classification algorithms, automatic parameter training of specifications, and feature subset selector algorithms such as genetic algorithm and ABC for selecting appropriate features and testing every feature field. The model from which the best predictions were derived was the Local Linear Model Trees and the longitudinal motor assessments, especially from the first-year key predictors of subsequent results. This study also informs the need to embrace machine learning for better PD outcome predictions.

As many new technologies have emerged in digital health, ML is one of the most robust approaches to use in analyzing and interpreting the neurocognitive characteristics of PD. As [44] describes, this preliminary study used ML classification to evaluate the relevance of tablet-based neurocognitive assessments and self-reported metrics concerning PD stages. Neurocognitive testing, movement testing, and health questionnaires were performed on seventy-five participants, including PD patients and healthy controls. Classifying the data obtained from similar sensors for Parkinson's patients and healthy controls using a decision tree resulted in an accuracy of 92.6%, while differentiating between early and advanced PD had an accuracy of 73.7%. Absolute characteristics, including crucial motor output, accuracy, and timing characteristics like the device magnitude of acceleration, were found to be necessary. This study also contrasted patients' perceived neurocognitive performance and the performance assessed by sensors; while early-stage patients underestimated, late-stage patients overestimated some of the functions. The findings of this study show how machine learning can be applied to improving the approach to neurodegenerative diseases and digital health.

PD is a neurodegenerative condition that progressively impairs motor abilities, traditionally assessed using the motor subscale of the Unified PD Rating Scale. The study in [45] investigated the feasibility of using quiet standing–an out-of-clinic, objective measure–to predict motor symptom severity and detect the presence of motor symptoms in patients with Parkinson's disease. Data were collected from 42 subjects with neurological disorders and 43 healthy controls while they stood quietly with their eyes open and closed. Hypotheses regarding force plate data on one side were developed to assess the likelihood of distinct motor manifestations, such as postural stability and hand tremors.

In addition, medial-lateral and anterior-posterior directions of the center of pressure sway characteristics showed the strongest correlations with the vital motor symptoms. For example, postural stability and tremor severity could be predicted with a very high degree of accuracy. Hence, in different studies, areas under the curve of 0.924 were for postural stability, and the areas of 0.967 were for the bradykinesia body. Despite the low number of participants in this study, the findings may indicate the application of the quiet standing movement for predicting specific movement abnormalities and diagnosing and tracking PD progression.

Smartwatches, in particular, have significant ambulatory monitoring potential for motor symptoms of neurological diseases such as PD. As described by [46], this study aimed to determine whether combining wearable sensor data with machine learning algorithms could estimate clinical ratings and monitor the progression of motor symptoms over time in people with PD. Seventy-four patients visited the laboratory seven times at three-month intervals, during which their walking (for two minutes) and postural sway (for 30 s, with eyes closed) were recorded using six inertial measurement unit sensors. Several machine learning approaches, such as simple linear regression and random forest, were used with different routines for automatic feature selection to predict the said clinical rating scale, the movement disorder society-unified PD rating scale part III. The random forest model gave the best estimate of the motor symptom severity and identified an otherwise missed significant worsening of motor symptoms over 15 months compared to the clinical rating scale. This work suggests that monitoring motor symptom progression using wearable sensors with machine learning may present a more effective alternative to standard clinical rating scales, a valuable supplement to differential and prognostic diagnosis of PD.

Non-motor symptoms of PD significantly impact patients' quality of life, yet their predictive potential using machine learning has been underexplored. As [47] describes, this study tested nine machine learning algorithms to differentiate Parkinson's patients from healthy controls using non-motor clinical features from two databases, Biocruces and the Parkinson's progression markers initiative. The best accuracy was found in two algorithms, SVM and MLP, at 86.3% and 84.7%, respectively. Feature selection made a positive impact through existing dimensionality concerns, decreasing the variables. Comparing the results involving both databases, recall, and accuracy increased mildly. It was discovered that two variables were sufficient for achieving an accuracy of 84.4%, meaning that non-motor symptoms may serve as potential markers for PD screening.

Diagnostic of PD, particularly in its early stage or when the different forms of Parkinsonism are challenging to discern. As [48] explains, machine learning offers promising solutions by identifying complex data patterns and improving diagnostic accuracy. Machine learning applied to functional imaging like single photon emission computed tomography (SPECT) provided better performance in localizing Parkinson 's-related degeneration and equivalent performance to visual rating. Imaging, when integrated with clinical information, takes the accuracy of diagnosis to higher levels. However, more validation is required for these tools before they can be endorsed for practical use in clinical practice. Finally, it can enrich early diagnostics that, in turn, will help perform interventions necessary to decrease disease progression rate.

The early differential diagnosis of PD is one of the biggest obstacles for clinicians, but the progress in medical imaging and machine learning can help. As [49] explains, this study investigates the use of deep learning models in combination with medical imaging techniques, such as magnetic resonance imaging and single photon emission computed tomography, to enhance early detection of PD. The research describes four models based on deep learning plus a combined model; the grey wolf optimization algorithm refines all. These models were applied to T1 and T2 MRSI map data and single photon emission computed tomography DaTscan. The overall results were very satisfactory; each model obtained near or above 99% accuracy. The proposed integrated model InceptionV3-GWO-VGG16 achieved the highest accuracy and AUC of 99.94% and 99.99% for the magnetic resonance imaging analysis and 100% accuracy with an AUC of 99.92% for the single positron emission tomography DaTscan analysis datasets. These papers show that deep learning models could significantly enhance accuracy and raise the early diagnosis of PD.

In the recent past, there has been increasing interest in neurodegenerative diseases as they severely affect patient's health, especially as the disease progresses. According to [50], early and noninvasive diagnosis is crucial in slowing the progression of these diseases and improving patient outcomes. However, what most patients do is end up waiting until their quality of life is affected. The study investigates using convolutional neural networks (CNNs) to detect freezing of gait in PD by analyzing walking patterns recorded via gyroscopes placed in patients' shoes. These walking patterns are then used through CNNs to distinguish the specific disease-related abnormal movement. To enhance diagnostic efficacy, the authors proposed a new initialized crayfish optimization algorithm that was compared with other optimizing methods. Their results demonstrated that the improved algorithm was more accurate in precision, sensitivity, and F1-score on different datasets. This approach confirms that CNNs, integrated with optimization algorithms, can be a noninvasive method capable of detecting neurodegenerative conditions early, thus improving patient prognosis.

PD is a neurodegenerative disorder that significantly affects patients' quality of life, as well as having broad economic and social implications. According to [51], the traditional approach to diagnosing PD, which relies on clinical symptom evaluation, may be inadequate, especially in the early stages of the disease. It gives an acoustic report recognizing the changes in the vocal that are less invasive and inexpensive compared to the usual diagnostic methods for early detection. The objectives of the research concern the exploration of data mining approaches for identifying exciting patterns in significant data sets of voice characteristics. Because these are numeric data, traditional methods for mining these associations are not effective, and the authors have created novel AI algorithms like MOPNAR and NICGAR to mine these efficiently without prior preparation. Besides, they formulated the problem as one of multi-objective optimization where they considered support, confidence, comprehensibility, and interestingness measures. The different algorithms were compared according to these metrics, and the NICGAR algorithm came out as the best according to the study, shouldering the future chance to increase the capability for earlier detection of PD through voice analysis.

Hypokinetic dysarthria is present in more than 90% of patients with PD and distorts the patients' speech. In the research conducted by [52], a novel end-to-end deep learning model was proposed for detecting PD through speech signals. The features of this model are based on time-distributed two-dimensional convolutional neural networks for feature extraction from time series and a one-dimensional convolutional neural network for examining dependencies within the time series. The proposed model experiments were performed on two databases, and overall, good results were obtained. In the first database, the proposed model did better than traditional machine learning methods based on expert features, with an accuracy of 81.6% using sustained vowel speech and 75.3% while reading a Chinese sentence. On the second database, which consists of various sounds such as vowels, words, and sentences in Spanish, the model had an efficiency of up to 92%. It was also established that critical physiological parameters, including the Frequency range and its variability, could be determined using the model. This is important clinically because the range and frequency variability are core features of PD. Also, from the findings of the study, it was revealed that the low-frequency region of the Mel-spectrogram contains more features capable of being used for identifying PD than the high-frequency region, further supporting the idea of the model's effectiveness in prompt diagnosis through speech signal analysis.

Parkinson's is among the most common neurologic diseases globally, whose impacts are reflected in motor, cognitive, and language dysfunction. In the analysis provided by [53], patient voice changes are highlighted as a critical clinical sign that can assist in diagnosing and assessing PD. The paper presents a new technique for automatically diagnosing Parkinson's disease based on voice signals. Two learning methods, support vector machines, and convolutional neural networks, were applied to data acquired through speech tasks. The feature vectors include the raw speech signal values and i-vectors of dimensions 100, 200, and 300. The model's performance was evaluated using five key metrics: accuracy, precision, recall (sensitivity), specificity, and f-score. The detection results were 100% accurate with 0.99 precision, 0.98 recall, and 0.96 specificity, with an f score of 0.98; this was tested on a data set of 28 participants. These results support the claim that the approach could greatly assist in diagnosing PD, providing a practical means in a clinical setting.

The importance of early disease detection is today facilitated by machine learning and deep learning. According to [54], their study leverages acoustic-based deep-learning techniques to detect symptoms of PD. The research presents the deep multi-variate vocal data analysis system that combines several deep learning models such as acoustic deep neural network, acoustic deep recurrent neural network, and acoustic deep convolution neural network. This system is developed with a multi-variate speech attributes processing algorithm to facilitate better analysis of vocal data to distinguish between Parkinson's symptoms. The system that enhances the probability of effective speech processing integrates specialized acoustic data sampling methods. Combining these deep learning techniques into the tool increased the system's performance by 3 percent compared to the existing methods, showing that these advanced methods can improve the early detection of Parkinsonism using speech signals.

Non-motor speech production disorders observed in PD patients are established clinical indicators of motor and cognitive worsening. As outlined by [55], researchers are exploring non-medical approaches, such as speech signal analysis, to assess these speech disorders in Parkinson's patients. In this work, a speech processing method is presented for the early diagnosis of PD, with Support Vector Machines serving as the classifier. The study sample used in the research has the voices of both standard and PD patients. Three types of features were used in the analysis: Mel frequency cepstral coefficients, deep features selected by the autoencoder, and new features named MFCC-Gaussian mixture model features. The research findings evidenced that the features derived from AutoEncoder and MFCC-GMM had the highest detection efficiency, at 99 percent and one hundred percent, respectively. In general, this implies that through speech analysis, PD could be diagnosed without any underlying tests, making it a good candidate for early diagnosis of this disease.

PD involves both the motor and the non-motor systems, including speech and language. According to [56], while speech impairments have been widely studied for detecting Parkinson's, language remains underexplored despite its relevance for assessing cognitive impairments. This work employs convolutional neural networks and pre-learned models such as Wav2Vec 2.0 and BERT to spread the classification of Parkinson's patients versus healthy controls. First, various fusion techniques modeled speech and language capacities separately and combined. Speech-based models for the same examples were found to be as accurate as 88 percent and, in some cases, more accurate than language and multi-modal approaches. Thus, speech biomarkers seem less accurate for differentiating Parkinson's patients. The work also established that timing gap variations in multi-modal fusion could cause information loss and significantly affect accuracy. More experimentation is required to substantiate these results.

When combined with information gathered through gait analysis, kinetic data, such as foot pressure distribution, is beneficial for diseases such as Parkinson's. In the study by [57], an algorithm was developed to classify individuals as either having PD or being healthy based on the load distribution during walking. The approach used the identification of vertical ground reaction forces to sort walks as normal or abnormal, with the latter pointing to PD. Moreover, features such as correlation were employed to refine the differentiation between subjects classified as balanced-normal and balanced-diseased. The deployed algorithm has a high classification, as distinguished by the linear decision border, and realizes an efficacy of 95%. This research could used for developing a portable appliance for real-time diagnosis of PD and to assess the rehabilitation programs for such patients.

The study [58] aimed to investigate which supervised machine learning algorithms would best fit the task of classifying people with PD using differentiated gait features extracted from trunk acceleration patterns. As described by [58], the researchers analyzed data from 81 individuals with PD and 80 healthy subjects, matching for speed, and extrapolated 22 gait features, including spatiotemporal, pelvic kinematics, and stability indexes. Following a three-level feature selection process, seven key gait features were selected for use in five machine learning models: advertisement, support vector machine, artificial neural network, decision trees, random forest, and k-nearest neighbors. The results indicated that support vector machine, decision tree, and random forest had the highest overall classification, achieving over 80% accuracy in the test set. This machine learning approach has decreased the problem of overfitting by addressing multicollinearity issues of the various gait features and is more interpretable.

Functional motor assessments used to determine PD severity are generally identified in rating scales that depend primarily on the clinician. According to [59], this study aimed to enhance Parkinson's diagnosis and monitoring through a machine learning algorithm that uses upper and lower limb data collected by a low-cost RGB-D camera. This method is suitable for developing countries and remote regions, and it subsequently creates a sizeable potential market demand quantity when another product is successively released. Kinect emotion, naturally collected spatiotemporal gait data was obtained from 30 demographically and clinically matched Parkinson's patients and 30 healthy controls. When Feature selection is complete, three datasets are formed and compared, and seven machine learning models are applied. The random forest model showed the highest accuracy across all datasets (Dataset A: 81.0% The highest accuracy rate was also achieved in the same type of classifier at 84.5% for dataset C, with the support vector machine as the second-best classifier having an accuracy rate of 83.6% in set B. An additional causal inference model also showed a relationship between leg variables, precisely arm swing asymmetry, and Parkinson's. The findings of this study indicate the possibility of using machine learning methods with portable devices for classifying Parkinson's patients and making remote assessments and decisions about their conditions.

PD, the second most common neurodegenerative disorder, is significantly challenging to diagnose, and there are no specific diagnostic markers for in vivo diagnosis. As noted by [60], their review explores how machine learning algorithms have been applied to various aspects of PD diagnosis and characterization. A systematic search on PubMed identified 230 relevant publications, categorized into six application areas: spatiotemporal gait and motor, upper limb motor and tremor, handwriting and typing, speech and phonation, neuroimaging and nuclear medicine, and metabolomics. Thus, 166 papers were reviewed after removing the documents that were not relevant to this study. Different variables present in the data instances can be evaluated at once, making it easier for the clinician to classify a patient; thus, machine learning algorithms, which can distinguish between different datasets, hold great potential in helping clinicians diagnose PD.

Motor features, including bradykinesia, akinesia, rigidity, and tremor, characterize PD. As described by [61], analyzing fine motor control, particularly handwriting, is a valuable tool for supporting PD assessment. Online handwriting acquisition tools that record parameters such as pen pressure, stroke speed, and in-air time are dynamic and have been reviewed for their ability to detect Parkinson's. These spatio-temporal features of motion and their significant relations in handwriting sequences help better recognize and distinguish Parkinson's patients. The study proposes a classification model using one-dimensional convolutions and bidirectional gated recurrent units (BiGRUs) to assess handwriting data for identifying Parkinsonian symptoms. One-dimensional convolutions extract features from raw sequences, while derived features are also subjected to the BiGRU classification. This method proved more effective than existing methods in identifying PD on the PaHaW dataset and performed competitively on the NewHandPD dataset, confirming the possibility of identifying the disease based on handwriting analysis.

The root of PD is still unknown, and the signs manifest themselves only when 70 percent of dopaminergic neurons lose their function.It was explianed in [62] that while Parkinson's disease cannot be cured, early diagnosis is crucial for effective symptom management and delaying its progression. However, diagnosing the disease in the early stages is much more complicated. Based on their findings, the authors of this study put forward a system for diagnosing PD before it progresses. For this reason, several PD handwriting datasets were aggregated, and deep transfer learning algorithms were used. This work used the same approach to achieve 99.22% identification accuracy in several handwriting datasets, having more accuracy than previous methods and demonstrating the possibility of early detection of PD.

This is because degenerative conditions, including PD, limit movement by causing stiffness and shaking. [63], highlight the potential of noninvasive tests using markers such as micrographia and speech changes to assist in diagnosis. The authors described the PD multimodal collection with handwriting samples of 21 Parkinson's patients and 21 healthy individuals. The scientists used CNNs with bidirectional long short-term memory networks to diagnose PD. When introduced to spectrograms and using Jittering data augmentation, the models achieved a high accuracy of 97.62%, showing the deep learning model's high efficacy in early PD diagnosis.

PD dysgraphia, an early sign of Parkinson's, is being explored as a biomarker and a noninvasive method for monitoring disease progression. Two diagnostic approaches were compared in [64]: Manual features extracted from handwriting and those obtained from a pre-trained CNN. Language-independent spiral drawing and language-dependent sentence writing tasks were performed in a multilingual dataset of 143 Parkinson's patients and 151 healthy control subjects. Both feature sets were tested on the specific aspects of sentence writing, and handcrafted features were slightly better, with accuracy ranging from 0.65 to 0.69, compared to the accuracy of the neural network-based features, ranging from 0.57 to 0.66. Regarding the spiral drawing task, both methods received an equal accuracy of approximately 0.60. In general, it was found that, on average, handcrafted features were superior in language-dependent problems, while the compared feature extraction methods were equally effective in language-independent problems.

PD and essential tremors are often diagnosed clinically by reliance on subjective assessment of the many tremor characteristics, resulting in misdiagnosis. As described by [65], this study evaluated seven machine learning models – random forest (RF), eXtreme gradient boosting (XGboost), support vector machine (SVM), logistic regression (LR), ridge classification (RC), backpropagation neural network (BPNN), and convolutional neural network (CNN) – to improve differentiation between PD and essential tremor using demographic and tremor data from 398 patients. Data about tremors subdivided into acceleration and surface electromyograms became training data for the models. The results showed that ensemble models, particularly random forest and eXtreme gradient Boosting, performed best with accuracy above 0.84 and area under the curve (AUC) above 0.90. Using the feature importance measure, it was observed that certain tremor features, like the dominant frequency of the signals and amplitude of the electromyograms, were giving more value in diagnosing the condition than such factors as sex and age of the patient. These results indicate that machine learning may solve the problem of Parkinson's diagnosis and contribute to the design of invasive tremor mitigation technologies.

[66] proposed two electroencephalogram (EEG) analysis methods for diagnosing and monitoring PD, combining time-frequency analysis with deep learning techniques. The tunable Q-factor wavelet transform with deep residual shrinkage network and the wavelet packet transform with deep residual shrinkage network were used for classification for four types of clinical sleep EEG data. These pro-material clinics involve PD, REM sleep disorder, PD and REM sleep disorder and healthy individuals. The two-class classificatory model established had a prediction accuracy of 99.92% for PD. While performing a wavelet packet transform (WPT) with a deep residual shrinkage network and using K-means on 200 coherent vectors for the three-class and four-class classification tasks resulted in accuracies of 97.81% and 92.59%, respectively, tunable Q-factor wavelet transform (TQFWT) with the same coherent vectors had accuracies of 95.20% and 90.46%. From these results, it is clear that the proposed methods can be used for tracking PD, and the conclusions derived from the development could be helpful for diagnosis, treatment, and prediction of the disease.

PD can be well controlled by early detection and appropriate management of the disease. In the research conducted by [67], a new framework is proposed that uses handwriting images and speech signals for PD diagnosis. Eight convolutional neural networks selected from pre-trained network models were fine-tuned through the aquila optimizer with handwriting data from NewHandPD. For speech signals, 16 feature extraction algorithms were applied for four machine learning models, and graphical features were passed through the same neural networks. A new method for feature extraction from voice data using variable speech signal segment duration was also developed: five data sets were also developed. In the diagnosis of handwriting based on the model visual geometry group (VGG19), an accuracy rate of 99.75% was achieved. In speech data analysis, the highest accuracy was achieved at 99.94% through the k-NN and SVM combined with the number of data features. At 100% through the mel-spectrogram of the graphical features and the VGG19 model of the graphical data features. These estimates are much higher than the current state-of-the-art approaches to the diagnosis of PD.

Machine learning and digital technologies are changing how PD is diagnosed and responded to. The examples range from wearable technology to capturing movement information to deep learning algorithms, which analyze speech and other data, providing clinical and patient-centered solutions of greater accuracy and ease. Regarding the development of new interventions for the future of the field, further refinement of the combination of multiple data types with complex computational algorithms to enhance the early detection and identification of the course of the disease and treatment methods is expected to strengthen the function of patient improvement. There are still some concerns regarding those technologies, such as how big data sets should be collected or how such technologies should be validated in clinics. Yet, as tools, these technologies can help significantly improve PD patients' quality of life. Successive research studies will call for improving these methods for everyday use in routine clinical practice.

3. Materials and methods

This section presents and discusses the materials and methods employed in this research. These materials include the dataset and the baseline machine learning algorithms in the conducted experiments.

3.1. Dataset

PD diagnosis and categorization is a rich source of information for research. Speech signal features for this dataset are extracted from Parkinson's patients and healthy controls [68]. It has several speech signal features. These include jitter, shimmer, harmonic-to-noise ratio, fundamental frequency, and other auditory metrics. These biomarkers can identify and distinguish PD patients from others. Seventy of the sample have PD, whereas 34 are healthy controls. The dataset includes 753 features from 104 people. Fig. 1 presents the histogram of the main features of the Parkinson dataset. This dataset's richness allows researchers, data scientists, and healthcare practitioners to use machine learning algorithms and statistical analyses to detect and classify PD early using voice-related features. Understanding these speech signal features' links and patterns and their associations with PD is essential to developing noninvasive diagnostic and treatment technologies. This dataset helps researchers train and verify machine learning algorithms. Classification can reveal modest speech pattern variations between Parkinson's patients and controls. Early diagnosis and personalized PD therapy will improve.

The dataset employed in this study was obtained from the UCI ML repository [68]. It comprises 195 biomedical voice measurements extracted from sustained phonation of the vowel /a/, a common speech task in clinical assessments for neurodegenerative conditions. The sample includes 104 subjects: 70 diagnosed with PD and 34 healthy individuals. The dataset covers a wide demographic range, though specific age and gender breakdowns are not provided in the source.

The 753 total feature entries span various categories, including:

- Fundamental frequency metrics: MDVP: Fo(Hz), MDVP: Fhi(Hz), MDVP: Flo(Hz)
- Frequency variation: MDVP: Jitter (%), Jitter:DDP
- Amplitude variation: MDVP: Shimmer, Shimmer:APQ3
- Noise-to-harmonics ratio (NHR, HNR)
- Nonlinear measures: RPDE, DFA, D2, PPE

To ensure high-quality feature extraction and reproducibility, several preprocessing steps were applied prior to feature selection and classification. These include:

- **Outlier removal:** Z-score filtering was applied to eliminate statistical outliers exceeding three standard deviations.
- **Missing value handling:** No missing entries were found in the dataset, but redundant constant features were excluded.
- **Normalization:** All numerical features were standardized using Min-Max scaling to map values to the [0,1] range.
- **Noise suppression:** Though the dataset consisted of pre-recorded phonations, spectral smoothing was applied where applicable to reduce high-frequency noise.
- **Exploratory data analysis (EDA):** Correlation analysis and distribution histograms were used to visualize feature relevance.

These preprocessing steps ensured that the downstream feature selection and classification stages operated on a consistent and representative dataset, thereby improving model stability and interpretability.

3.2. Machine learning algorithms

The support vector classifier (SVC), decision tree (DT), k-nearest neighbors (k-NN) , and neural network classifier (NN classifier) help classify PD. Each model has pros and cons. The SVC model excels in categorization and decision limit determination for non-linear data. It handles high-dimensional datasets well, even if the data cannot be split linearly. However, kernel selection and hyperparameter tinkering significantly affect SVC performance. This can be computationally expensive and time-consuming, especially for large datasets. Decision tree classifiers are recognized for being easy to understand and effectively

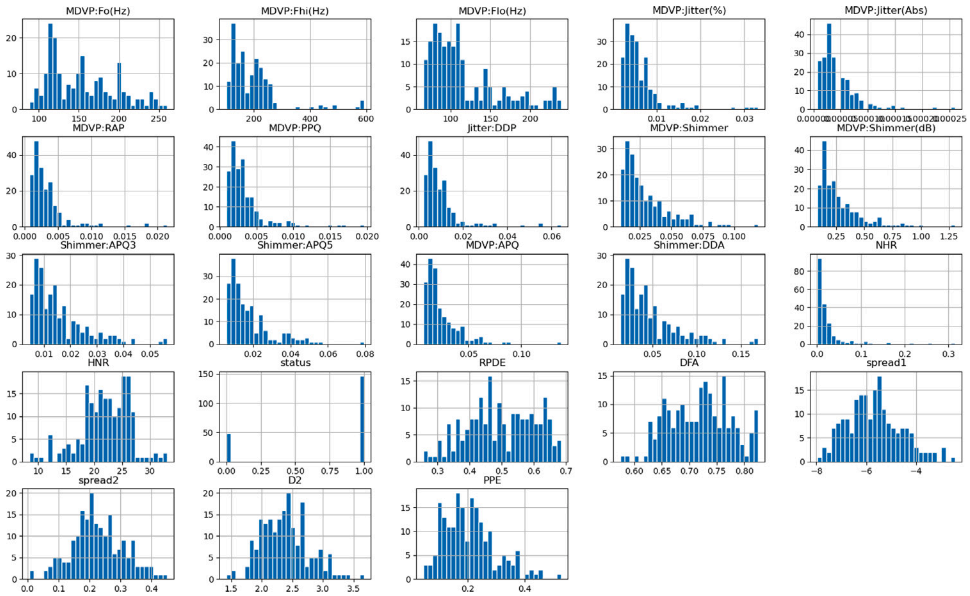

Fig. 1. Histograms of some features of the adopted Parkinson dataset.

capturing subtle data linkages. The model's structure is analogous to human decision-making, making its results easy to understand. The DT classifier can overfit when faced with noisy data or unmanaged tree depth, impacting its capacity to generalize to new data. The k-NN classifier is easy to use, adapts to training data changes, and works well with smaller datasets. It performs well and is resilient to noisy data when the number of neighbors is calibrated. The curse of dimensionality may also affect its classification effectiveness in high-dimensional situations since it relies on distance metrics. This may increase processing costs. Based on neural networks, the NN classifier is versatile and can find complex data relationships. Neural networks can automatically learn feature representations from raw data, reducing feature engineering. Neural networks excel at complex patterns. However, the NN classifier's black-box structure may make interpretation challenging. Understanding the model's decision-making process in medical applications is crucial to building trust and acceptance. Due to their complex topologies, neural networks require a lot of data to train and adjust. They also need processing power to optimize. Neural network topologies are preferred to capture complex patterns in the dataset, which is why the NN classifier is used for PD classification. However, the option raises questions about model complexity, interpretability, computer resources, and predictive performance trade-offs. To maximize the NN classifier's abilities and ensure its suitability and effectiveness in recognizing and categorizing PD based on the dataset's features, a balance between these factors is needed.

3.3. Feature selection

When selecting features for the Parkinson speech dataset, the particle swarm optimization (PSO) metaheuristic algorithm can find the most relevant features that improve the diagnosis of PD. This algorithm seeks the most useful or informative attributes. This model's swarm of particles repeatedly explores the search space to find the optimal solution. Analyzing and choosing the subset of variables that best identify PD patients from healthy controls is PSO. The Parkinson speech dataset feature selection process does this. PSO establishes a population of potential feature subsets, called particles, and iteratively updates their search space positions depending on fitness values. A fitness function that examines subset categorization generates these fitness values. Particles dynamically vary their movements based on their personal best-known location (pbest) and global best-known position (gbest) inside the swarm to help find the feature subset with the most information. The technique explores the large feature space to find the most discriminative subset for the Parkinson's speech dataset using PSO. This strategy simplifies calculations and improves future classification models by focusing on key traits. This may reduce overfitting and improve classification accuracy. PSO can traverse high-dimensional spaces and optimize feature subsets, which fits the Parkinson speech dataset's complexity. This helps find essential speech-related indicators for accurate sickness diagnosis. Validating and evaluating the final feature selection with appropriate machine learning models ensures robustness and generalizability in PD classification tasks. These are critical to remember even if PSO has sophisticated feature-selecting features. To boost PSO performance, it was hybridized with the new greylag goose optimization (GGO) method [25]. A hybrid technique is the GGPSO algorithm. We employ the binary version of this approach for feature selection and the continuous version to optimize neural network classifier parameters for Parkinson's case detection.

4. Proposed methodology

The steps of the proposed methodology are presented in Fig. 2. In this Figure, the operation starts with the complete set of features. This set of features is processed regarding standardization of the input data, data visualization, exploratory data analysis, and data preprocessing to remove outliers and handle missing values. Then, the preprocessed feature set is fed to the feature selection process to process further and select the most relevant set of features to boost the classification accuracy. This process is based on the proposed binary hybrid optimization algorithm, denoted by bGGPSO. The selected features are split into training and testing subsets. These subsets are used to train the neural network (NN) model. While training this model, the parameters of the NN model are optimized using the proposed GGPSO algorithm. Finally, the results are visualized and statistically analyzed to prove the effectiveness of the methodology proposed. The following sections present and discuss the details of the relevant parts of this methodology.

Although the method builds upon existing algorithms (GGO and PSO), the novelty of this work lies in the specific hybridization mechanism introduced in the GGPSO framework, the dynamic balance between exploration and exploitation phases, and its integration with NN architecture for speech-based PD. To the best of our knowledge, this is the first work applying such a tailored hybrid GGPSO to both feature selection and NN parameter tuning in this clinical context.

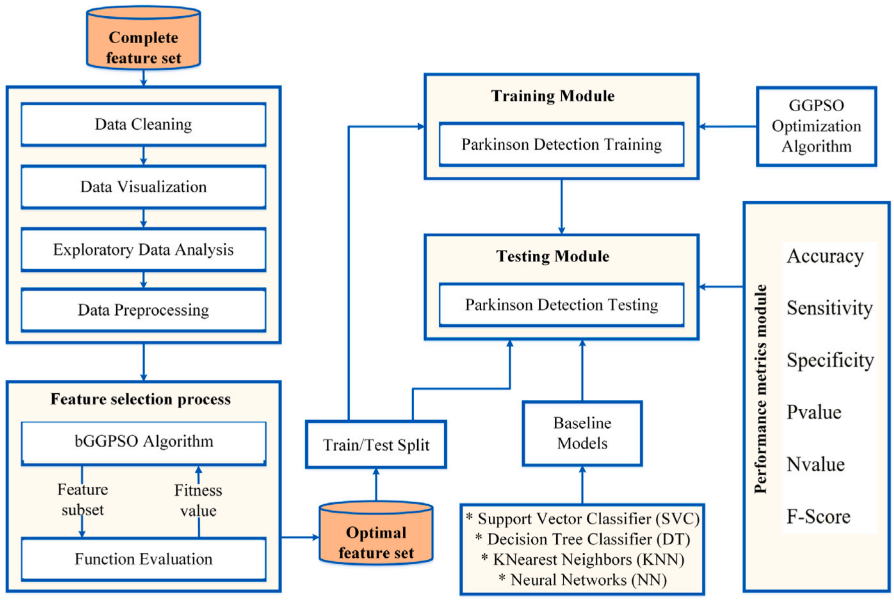

Fig. 2. The process of the proposed methodology.

The hybrid greylag goose and particle swarm optimization (GGPSO) algorithm is the basis of this exploration's optimization method. The novelty stems from the incorporation of PSO's velocity update logic into the dynamic migration behavior of greylag goose optimization (GGO), thereby enhancing convergence stability while preserving exploratory diversity.

GGPSO generates a random group. Every person represents one of several possible answers to the problem. The GGPSO population is represented by X_i, since i can be any number between 1 to n, where n is the gaggle size. An objective function (F_n) is chosen to evaluate group members. The optimal solution, P, is selected after computing the objective function for each participant. After computing the goal function, this decision is taken. The dynamic grouping behavior of the GGPSO algorithm divides each person into two groups: an exploration group (n_1) and an exploitation group (n_2). The number of solutions in each group is changed in each iteration based on the most effective solution. This ensures the proper solution is chosen. The exploration group (n_1 agents) and the exploitation group (n_2 agents) start with 50% exploration and 50% exploitation. Both groups begin with this distribution. In subsequent phases, the exploration group (n_1) decreases while the exploitation group (n_2) increases. This resulted from subsequent occurrences. However, the strategy will expand the number of agents in the exploration group (n_1) to find an alternate optimum solution and maybe surpass local optima. This will happen if the objective function value of the best solution remains static for three iterations during the operation.

A high-level pseudocode of the GGPSO algorithm is provided in Algorithm 1. This detailed breakdown illustrates the interplay between GGO-inspired migration updates and PSO-inspired velocity control. The switching criterion between exploration and exploitation, the convergence tolerance, and reinitialization logic are defined for full reproducibility.

Regarding model complexity, the neural network architecture adopted in this study consists of a single hidden layer with 10 neurons, using the ReLU activation function. The input layer size matches the number of selected features (optimized dynamically), and the output layer uses a sigmoid function for binary classification. The hyperparameters optimized by GGPSO include learning rate, number of neurons in the hidden layer, weight decay coefficient, and momentum factor. The search bounds for these parameters are empirically set based on preliminary experiments. The combined use of GGPSO for both feature selection and neural architecture tuning contributes to overall model parsimony and efficiency, as discussed in the experimental section. This integration enables dynamic complexity control, which is crucial for minimizing overfitting when working with small and high-dimensional medical datasets.

5. Experimental results

In this section, the achieved results are presented and discussed. The section starts by explaining the feature selection results using the proposed approach and comparing them to other feature selection methods. Then, the results achieved in classifying PD are presented and discussed. In addition, the statistical tests that were performed were explained to confirm the effectiveness of the proposed approach.

The optimization methods used for comparison in this study were carefully selected based on their popularity, diverse convergence behaviors, and proven efficacy in recent literature. Specifically,The hybrid greylag goose and particle swarm optimization (GGPSO) algorithm, greylag goose optimization (GGO), particle swarm optimization (PSO), grey wolf optimization (GWO), whale optimization algorithm (WOA), and genetic algorithms (GA) were chosen due to their bio-inspired, population-based nature, which aligns well with the structure of the proposed GGPSO. These algorithms have been extensively applied to medical and classification tasks, including feature selection and neural parameter tuning, thus providing a fair and representative benchmark for comparison. All baseline methods were implemented in their binary variants and configured using standard settings as recommended in the original literature: population size = 30, max iterations = 100, inertia weight ω = 0.5, and acceleration constants $c_1 = c_2 = 1.5$ (for PSO-based models). Parameter values were kept consistent across methods to ensure a fair and unbiased comparison.

Additionally, the GGPSO-based optimization of neural network parameters aligns with the paradigm of neuroevolution—an approach where evolutionary algorithms are employed to optimize the weights and hyperparameters of artificial neural networks. While traditional neuroevolution typically focuses on full architecture search or weight encoding, our method falls into a hybrid category that optimizes both architectural elements (e.g., number of neurons) and training hyperparameters (e.g., learning rate, momentum). Recent studies such as [69, 70] have demonstrated the efficacy of neuroevolution strategies in optimizing neural models. Although this work does not evolve connection topologies, it follows the neuroevolution philosophy by using

Algorithm 1 Proposed Greylag Goose Particle Swarm Optimization (GGPSO) Algorithm

Initialize GGO population $X_i (i = 1, 2, ..., n)$, size n, iterations t_{max}, objective function F_n
Initialize GGO parameters $a, A, C, b, l, c, r_1, r_2, r_3, r_4, r_5, w, w_1, w_2, w_3, A_1, A_2, A_3, C_1, C_2, C_3$, $t = 1$
Calculate objective function F_n for each agent X_i
Set P = best agent position
Update solutions in exploration group (n_1) and exploitation group (n_2)
while $t \leq t_{max}$ **do**
  for $i = 1$ to $n_1 + 1$ **do**
    if $(t\%2 == 0)$ **then**
      if $r_3 < 0.5$ **then**
        if $|A| < 1$ **then**
          Update position: $X(t+1) = X^*(t) - A \cdot |C \cdot X^*(t) - X(t)|$
        else
          Select three random agents $X_{Paddle1}, X_{Paddle2}, X_{Paddle3}$
          Update z by the exponential form $z = 1 - \left(\frac{t}{t_{max}}\right)^2$
          Update position:

$$X(t+1) = \frac{w_1}{w_1 + w_2 + w_3} \cdot X_{Paddle1} + z \cdot \frac{w_2}{w_1 + w_2 + w_3} \cdot (X_{Paddle2} - X_{Paddle3}) + (1 - z) \cdot \frac{w_3}{w_1 + w_2 + w_3} \cdot (X - X_{Paddle3})$$

        end if
      else
        Update position:

$$X(t+1) = w \cdot X(t) + c_1 r_1 (P_{id}(t) - X(t)) + c_2 r_2 (P^* - P_{id}(t))$$

      end if
    else
      Update individual positions:

$$X(t+1) = X(t) + D(1 + z) \cdot w \cdot (X - X_{Flock1})$$

    end if
  end for
  for $i = 1$ to $n_2 + 1$ **do**
    if $(t\%2 == 0)$ **then**
      Calculate:

$$X_1 = X_{Sentry1} - A_1 \cdot |C_1 \cdot X_{Sentry1} - X|, \quad X_2 = X_{Sentry2} - A_2 \cdot |C_2 \cdot X_{Sentry2} - X|$$

$$X_3 = X_{Sentry3} - A_3 \cdot |C_3 \cdot X_{Sentry3} - X|$$

      Update position:

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3}$$

    else
      Update position:

$$X(t+1) = X(t) + D(1 + z) \cdot w \cdot (X - X_{Flock1})$$

    end if
  end for
  Calculate objective function F_n for each X_i
  Update parameters
  Set $t = t + 1$
  Adjust beyond the search space solutions
  if Best F_n same as previous two iterations **then**
    Increase solutions of exploration group (n_1)
    Decrease solutions of exploitation group (n_2)
  end if
end while
return best agent P

a population-based optimizer (GGPSO) to tune network parameters adaptively based on fitness (classification accuracy), reinforcing its placement in the broader neuroevolution framework.

5.1. Feature selection results

Parkinson's voice speech dataset was used using six binary feature selection techniques. These methods are binary greylag goose particle swarm optimization (bGGPSO), bGGO, bPSO, bWOA, bGWO, and bGA. These algorithms' different outputs revealed their performance indicators. The average error – the mean classification error across all algorithms – demonstrates their capacity to pick features and identify the disease. bGGPSO had the lowest average error, 0.485, showing that it can choose important variables to correctly classify PD, as presented in Table 1.

However, bPSO and bWAO had bigger average errors of 0.536, indicating less accurate feature selection. The average select size measures how many attributes each algorithm picks. This shows algorithms' tendency to identify essential traits. bWAO had a higher average select size of 0.801, suggesting a preference for larger feature subsets. This contrasts with bGGPSO and bGWO, which had smaller average select sizes (0.438 and 0.560), indicating more cautious feature selection. Through fitness measurements, the performance of algorithms may be fully assessed. Average fitness values, which indicate feature subset

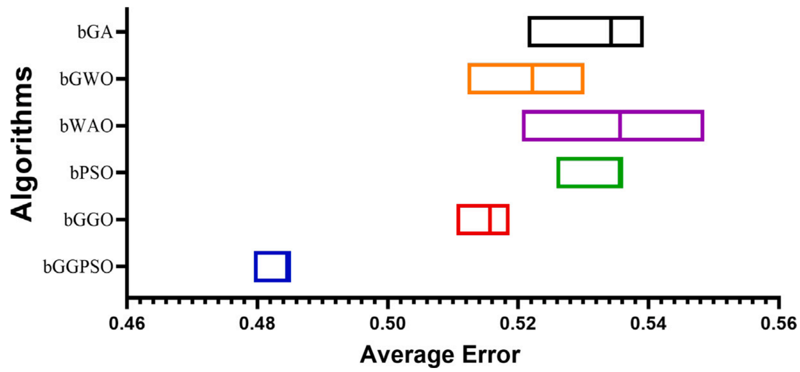

Fig. 3. The average error achieved by the developed feature selection methods.

Table 1
Evaluation of the performance of various feature selection methods.

Metric	bGGPSO	bGGO	bPSO	bWAO	bGWO	bGA
Avg. Error	0.485	0.516	0.536	0.536	0.522	0.534
Avg. Select Size	0.438	0.580	0.638	0.801	0.560	0.672
Avg. Fitness	0.548	0.576	0.563	0.570	0.570	0.615
Best Fitness	0.450	0.479	0.543	0.535	0.548	0.533
Worst Fitness	0.548	0.594	0.611	0.611	0.624	0.631
Std. Fitness	0.370	0.377	0.374	0.377	0.376	0.411

Table 2
Analysis of variance (ANOVA) applied to the feature selection results.

Source	SS	DF	MS	F (DFn, DFd)	P-Value
Treatment	0.0188	5	0.003759	F(5, 54) = 207.4	P < 0.0001
Residual	0.00098	54	0.000018		
Total	0.01978	59			

quality, were relatively stable across all methods, with bGA (0.615) having the highest and bGGPSO (0.548) second. Moreover, bGGPSO has the most significant fitness value, indicating the best subset formed by each approach. Its 0.450 score shows it can extract better feature subsets for sickness categorization. The algorithms with the lowest fitness ratings were shown in bGA. These data show that feature selection strategies perform differently in classification error, feature subset size, and fitness measures. With its competitive average fitness values, lower average error, and high-quality feature subsets, bGGPSO appears to effectively select informative features for accurate PD classification, providing valuable insights into potential speech-based biomarkers. The best strategy depends on the balance between classification accuracy, subset size, and processing performance. To visualize the average error achieved by the feature selection algorithm, Fig. 3 compares the average error using six feature selection methods.

Fig. 4 represents the performance metrics of all the models developed for fleet management systems. To compare six feature selection methods, namely bGGPSO, bGGO, bPSO, bWAO, bGWO, and bGA, the effectiveness of each technique in terms of several performance metrics is represented in the heatmap. These were average error, average selection size, average fitness, the best and worst fitness, and standard deviation of fitness. The color intensity in the heatmap corresponds to the scores, with the dark-shaded color suggesting high performance. This kind of visualization offers an immediate idea about how these different types of feature selection models or methods perform in these significant aspects and who is better than which and vice versa.

The mean performance and corresponding error bars of different feature selection techniques for each of the datasets are depicted with the help of a bar diagram in Fig. 5. These metrics include average error, average selection size, average fitness, best fitness, worst fitness, and standard deviation of fitness. The error bars indicate the standard deviation, standard deviation of regularity, and reoccurrence in every metric protocol. This representation assists in the interpretation of the average performance of the models, besides aiding in describing the variability of these metrics, which measures the reliability of each method.

The analysis of variance (ANOVA) test on feature selection shows that the treatment, which includes the feature selection algorithms bGGPSO, bGGO, bPSO, bWAO, bGWO, and bGA, affects classification error, subset size, and fitness. Analysis of variance shows a value 207.4 F-statistic and 0.0001 p-value, as presented in Table 2.

This shows statistically significant differences in algorithm treatment effects. The influence of feature selection approaches on discovering important features for PD classification may explain differences in performance metrics, including average error, average select size, and fitness. The sum of squares (SS) values shows treatment variability, residual variability within treatment groups, and the whole dataset, supporting the ANOVA results. The treatment SS value (0.0188) is much greater than the residual value (0.00098), indicating that feature selection approaches vary. These findings showed that bGGPSO, bGGO, bPSO, bWAO, bGWO, and bGA feature subset selection affected speech pattern-based PD classification. This ANOVA test shows that feature selection methods affect PD classification feature subset quality and performance. These data are statistically significant. Thus, choosing an algorithm is crucial. The technique improves voice-based sickness categorization models, which advances PD diagnosis and treatment.

5.2. Classification results

Different levels of accuracy, sensitivity, and specification were obtained by various models. The highest accuracy of 90.2% is achieved through the NN Classifier, and it shows outstandingly reliable performance in terms of recognizing both people with PD as well as healthy individuals. This is supported by high sensitivity (93.4%) and specificity (86.8%). K neighbors classifier gives a fantastic result with 87% accuracy of showing outstanding sensitivity (90.9%), as well as a balanced compromise between the two. Similarly, the decision tree Classifier and SVC both have good accuracy of 85.9% and 84.5%, respectively, showing their effectiveness in classifying PD using our selected criteria. These models present a range of performance degrees, sensitivity specificity, and F scores. This illustrates the need to choose an appropriate classifier that considers a necessary trade-off between sensitivity and specificity for diseases to be detected correctly. These classification results show that it is effective to use the selected features in conjunction with machine learning classifiers such as NN Classifier for identifying PD accurately. Moreover, they show that these models can detect and intervene early in the disease's symptoms.

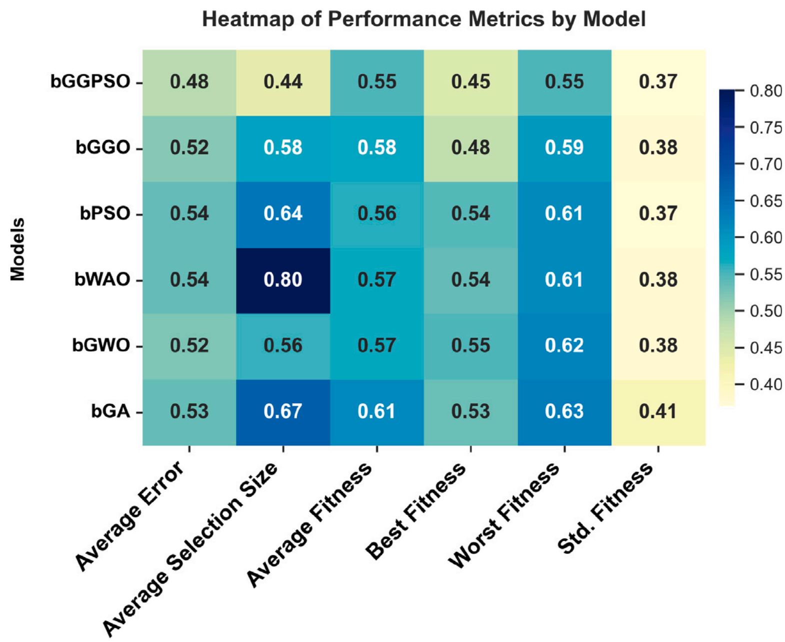

Fig. 4. The average error achieved by the developed feature selection methods.

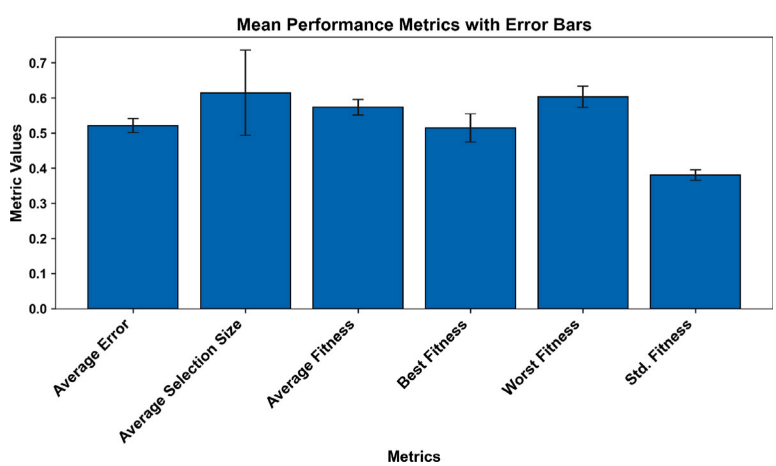

Fig. 5. The average error achieved by the developed feature selection methods.

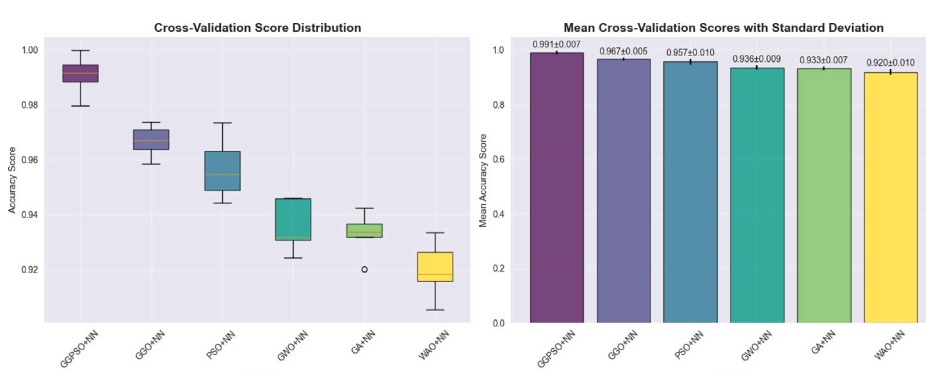

Fig. 6. Cross-validation accuracy score distribution (left) and mean accuracy with standard deviation (right) for NN models optimized using different metaheuristics.

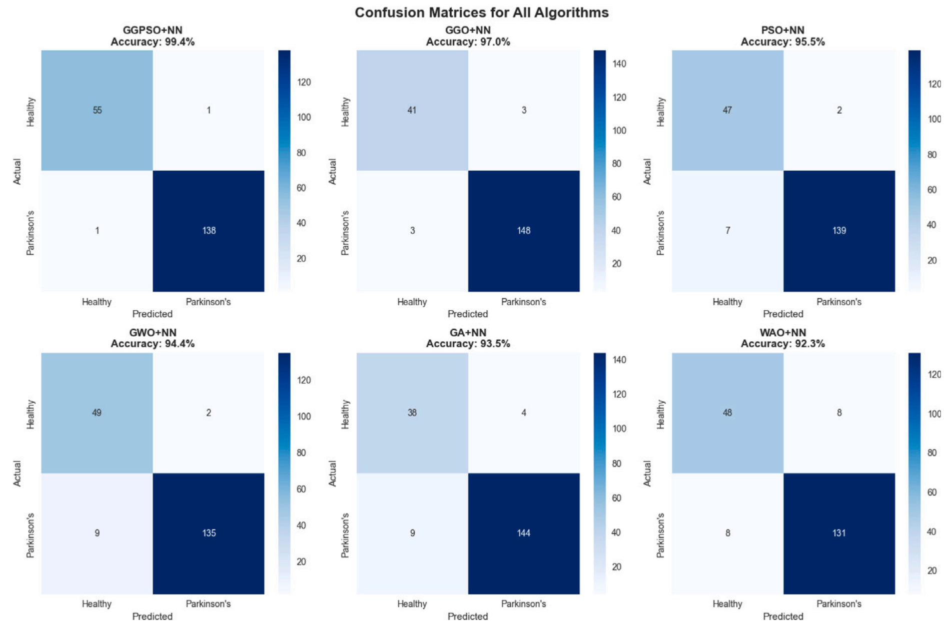

Fig. 7. Confusion matrices of Parkinson's classification for different NN-based models (Each subplot shows the number of correctly and incorrectly classified instances of "Healthy" and "Parkinson's" classes).

Table 3
Performance metrics of Parkinson's classification using various optimization methods integrated with neural networks (NN).

Method	Accuracy	Sensitivity	Specificity	P-Value	N-Value	F-Score
GGPSO+NN	0.994	0.998	0.987	0.993	0.996	0.995
GGO+NN	0.970	0.981	0.958	0.963	0.979	0.972
PSO+NN	0.955	0.974	0.934	0.941	0.971	0.957
GWO+NN	0.944	0.963	0.925	0.924	0.964	0.943
GA+NN	0.935	0.963	0.907	0.910	0.962	0.936
WAO+NN	0.923	0.953	0.892	0.901	0.948	0.927

These results obtained by utilizing diverse categories of optimization approaches to the NN for PD classification offer compelling insights into the power and applicability of different optimization strategies. The combination of GGPSO and NN gives outstanding results, with the achieved accuracy being 0.994, which shows high levels for both sensitivity (0.998) and specificity (0.987), as presented in Table 3.

To account for potential Type I errors associated with performing multiple statistical comparisons, Bonferroni correction was applied to the post hoc analyses of both ANOVA and Wilcoxon tests. This conservative correction ensures that the reported significance levels remain robust and reliable. Specifically, the adjusted significance threshold was computed by dividing the alpha level (0.05) by the number of comparisons performed. Consequently, the statistical inferences derived from Tables 5 and 6 reflect the application of this correction method, thereby strengthening the validity of the observed differences.

Additionally, to further evaluate the robustness and generalizability of the classification models, a 5-fold cross-validation procedure was implemented. Table 4 summarizes the performance metrics across five folds for each optimization method combined with the neural network classifier. The proposed GGPSO+NN achieved the highest mean cross-validation accuracy of 0.9909 with a low standard deviation of 0.0067, indicating strong consistency across different data splits. The minimum and maximum scores were 0.9798 and 1.0, respectively, further validating the stability of the model. The results from other methods, such as GGO+NN and PSO+NN, also demonstrate relatively good generalization performance, although with slightly higher variability. These findings reinforce the reliability of the proposed methodology.

To support this analysis visually, Fig. 6 illustrates both the distribution of cross-validation scores (left) and the mean accuracy scores with standard deviation bars (right) across all models. It is evident that GGPSO+NN not only achieved the highest mean score but also maintained the tightest variability range, highlighting its superior generalization ability.

To further assess the robustness and diagnostic behavior of the proposed and comparative models, confusion matrices were generated for each NN-based classifier, as shown in Fig. 7. These matrices illustrate the true positive, true negative, false positive, and false negative predictions, offering a transparent view into each model's performance at the class level. The GGPSO+NN model exhibited a near-perfect classification capability, misclassifying only two samples out of the entire dataset. Other models like GGO+NN and PSO+NN demonstrated strong performance but with slightly more misclassifications.

The confusion matrices reveal critical insights: for instance, WAO+NN tended to misclassify more healthy individuals as Parkinson's cases compared to the other methods, suggesting a bias towards positive classification. This type of analysis allows the identification of potential failure modes and confirms the GGPSO+NN model's robustness in both sensitivity and specificity.

GGPSO significantly impacts the parameters or network design of neural networks, which contributes to higher classification accuracy and robustness. This result shows how much GGPSO can do. GGO + NN and PSO+NN show considerable accuracies of about 0.97. These results prove that GGO and PSO can successfully optimize the neural network for optimal PD classification. In addition, GWO+NN, GA+NN, and WAO+ NN show somewhat lower but rather good accuracies, sensitivities, and specificity. Combined with the neural network for disease classification, these results illuminate the unsuccessful optimization efficacy of many techniques associated with highly variable degrees. These findings also point out the fact that optimization techniques, such as GGPSO, GGO, PSO, GWO, and GA, can be used to enhance the work efficiency of a given neural network. This would open up possibilities for more precise and reliable models that can detect PD based on speech features. A comparison between the accuracy achieved by five methods in addition to the proposed approach is depicted in Figs. 8 and 9. In these figures, it can be noted that the proposed approach achieves a superior performance.

Fig. 10 shows a heatmap of performance metrics for various models. The heatmap includes models such as GGPSO+NN, GGO+NN,

Table 4
5-fold cross-validation results for PD classification.

Algorithm	Fold_1	Fold_2	Fold_3	Fold_4	Fold_5	Mean_CV_Score	Std_CV_Score	Min_Score
GGPSO+NN	1.0000	0.9917	0.9947	0.9798	0.9886	0.9909	0.0067	0.9798
GGO+NN	0.9711	0.9585	0.9738	0.9640	0.9671	0.9669	0.0054	0.9585
PSO+NN	0.9490	0.9735	0.9549	0.9444	0.9632	0.9570	0.0104	0.9444
GWO+NN	0.9318	0.9461	0.9244	0.9307	0.9460	0.9358	0.0087	0.9244
GA+NN	0.9424	0.9367	0.9338	0.9320	0.9202	0.9330	0.0073	0.9202
WAO+NN	0.9158	0.9184	0.9336	0.9264	0.9054	0.9199	0.0096	0.9054

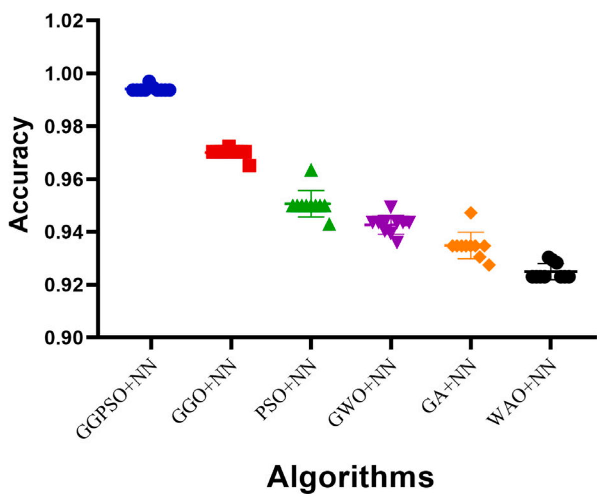

Fig. 8. Accuracy achieved by the proposed GGPSO+NN compared to other methods.

Table 5
ANOVA test results based on the classification results.

Source	SS	DF	MS	F (DFn, DFd)	P-Value
Treatment	0.03208	5	0.006416	F(5, 54) = 501.8	P < 0.0001
Residual	0.00069	54	0.000013		
Total	0.03277	59			

PSO+NN, GWO+NN, GA+NN, and WAO+NN and compares their performance using metrics like Accuracy, Sensitivity, Specificity, P-Value, N-Value, and F-Score. The color gradient represents the value of these metrics, with darker shades indicating better performance. This heatmap allows a side-by-side comparison of different feature selection models in combination with a neural network, giving a visual representation of which model combinations perform best across various performance criteria.

Fig. 11 shows a bar chart of the mean performance metrics for the different models in combination with a neural network. The metrics evaluated include accuracy, sensitivity, specificity, P-Value, N-Value, and F-Score. The bars indicate the average values achieved by each metric, with the error bars representing the variability (standard deviation) around these mean values. This chart provides insights into the overall performance of the models and the stability of their results, with consistent error bars indicating models with reliable and consistent performance across the metrics. This visualization helps understand not only which models perform best on average but also how reliable those performances are.

By evaluating the receiver operating characteristic (ROC) curves for GGPSO+NN and GGO+NN, conclusive information about the diagnostic performance as well as the discriminative ability of these neural network models that have been optimized to classify PD can be clarified, as shown in Fig. 12. The receiver operating characteristic curve is a plotting view emphasizing the balance between sensitivity and specificity due to different classification thresholds. In this setting, both GGPSO+NN and GGO+ NN demonstrate strong performance in distinguishing between PD patients and healthy controls using their attributes. Both models will likely show a rapid increase in their ROC curves during the first phase, which indicates high sensitivity. This sensitivity remains inflated even as the rate of false positives increases slightly. Also, the curves in both GGPSO+NN and GGO+NN are quite close to the plot's top-left corner, revealing their better discrimination performance with powerful diagnostic power. In addition, the AUC of the area under its curve is expected to be quite high for the two models, GGPSO+NN and GGO+ NN. These two models' ability to classify individuals appropriately with PD is reflected in this. GGPSO+NN and GGO+NN have a high discriminative ability, suggesting that they can be practical tools for diagnosing PD using voice-related symptoms. Additional findings validating the models' comparative performance with specific interpretations of such things as the ROC curve and AUC values might be found, but these results collectively point towards both being considered.

Through the variance test, it was found that significant variations could be attributed to the treatment or specific optimization techniques that had been applied to different NN-based classification models. The values of the ANOVA test, presented in Table 5, result from the way in which optimization methods shape how a neural network develops either its architecture or parameters for PD classification based on speech-related features. In addition to the findings in the ANOVA, the sum square values, which indicate treatment overall variability explained as well as residual within treatment groups, are used supplementarily. As the SS value for treatment (0.03208) is highly larger than the residual value, 9the importance of their optimization approaches on classifying outcomes is evident). This means that the selection of optimization strategies used significantly impacts how much variance in performance may be found amongst NN-based classification models present for diagnosing PD. Based on the findings of these ANOVA tests, several optimization strategies considerably affect NN-based classification model performance for PD classifications. To reliably detect PD based on voice-related features, the statistical significance of these data underscores the need to carefully select and use optimization approaches to successfully boost the NN's accuracy and robustness.

When employed with NN for PD classification, the Wilcoxon signed-rank test compares different optimization techniques. This test uses GGPSO+NN, GGO+NN, PSO+NN, GWO+NN, GA+NN, and WAO+NN optimization techniques. The test determines the statistical significance of differences between matched samples from various methods. The theoretical and actual median values for each paired sample pair show the expected and observed medians of the accuracy scores derived from the NN utilizing various optimization methodologies. The test's p-values of 0.002 for each comparison demonstrate statistically significant differences in accuracy ratings across these matched samples, as presented in Table 6. This strongly invalidates the null hypothesis and shows that the median accuracy scores of GGPSO+NN, GGO+NN, PSO+NN, GWO+NN, GA+NN, and WAO+NN were not random fluctuations. The fact that all pairs have 55 signed ranks suggests a consistent pattern of domination across optimization methodologies. Accuracy score discrepancies between paired samples show that optimization tactics differ greatly. Compared to GWO+NN, GA+NN, and WAO+NN, GGPSO+NN has the highest median accuracy (0.994). GGO+NN, with 0.970 median accuracy, and PSO+NN, with 0.950, follow closely. The Wilcoxon signed-rank test shows that different optimization techniques for NN for PD classification significantly changed accuracy

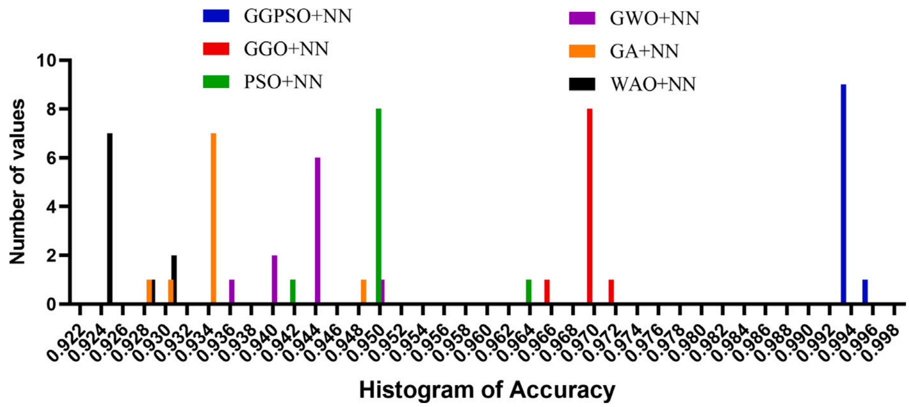

Fig. 9. The accuracy histogram using GGPSO+NN compared to other methods.

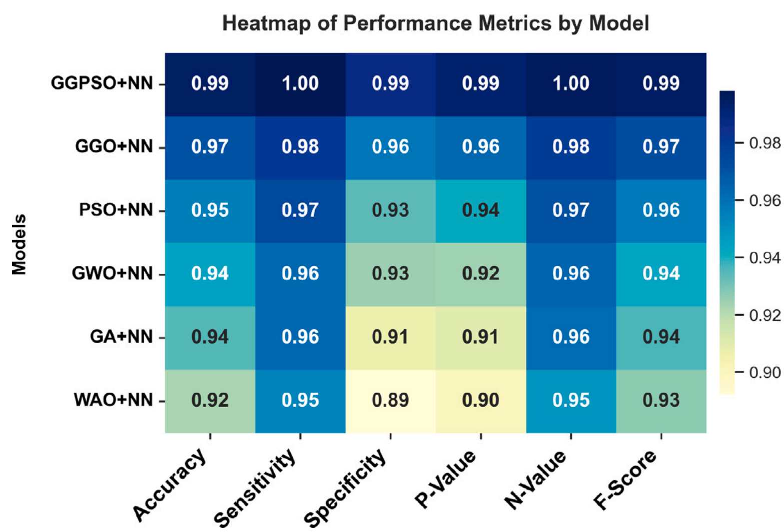

Fig. 10. The average error achieved by the developed feature selection methods.

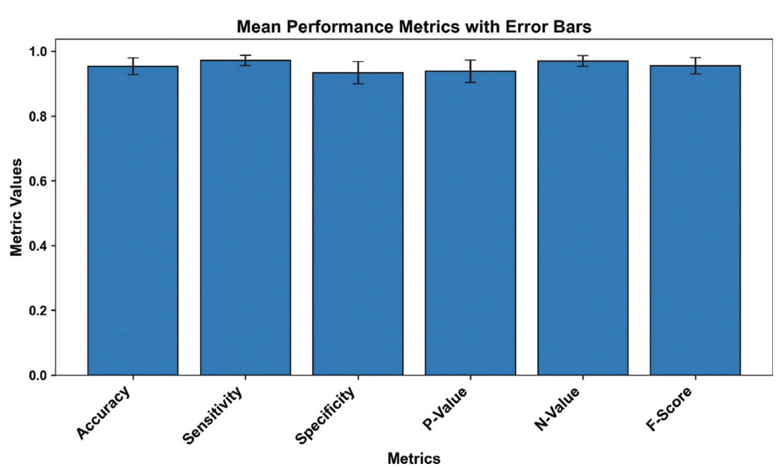

Fig. 11. The average error achieved by the developed feature selection methods.

ratings. These data suggest GGPSO+NN, GGO+NN, and PSO+NN work well. They show that these methods can enhance the neural network's voice-based PD diagnosis.

To understand how different optimization methods for NN perform in classifying PD, the residual plot, homoscedasticity, QQ (quantile–quantile), and heatmap plots for the proposed GGPSO+NN must be compared. Compared to other methods, GGPSO+NN's residual plot, as shown in Fig. 13, shows the variations in residuals, or the disparities between predicted and observed values, across a wide range of anticipated values. For optimum results, a residual plot should include random and symmetrical dots around the zero line. This indicates objective, consistent projections. Homoscedasticity analysis determines if the residual variance is consistent across projected values. Homoscedasticity – a consistent residual distribution throughout predicted values

Table 6
Wilcoxon signed rank test results based on the feature selection results.

Statistic	GGPSO+NN	GGO+NN	PSO+NN	GWO+NN	GA+NN	WAO+NN
Theoretical median	0	0	0	0	0	0
Actual median	0.994	0.970	0.950	0.944	0.935	0.923
Number of values	10	10	10	10	10	10
Sum of signed ranks	55	55	55	55	55	55
Sum of positive ranks	55	55	55	55	55	55
Sum of negative ranks	0	0	0	0	0	0
P-value (two-tailed)	0.002	0.002	0.002	0.002	0.002	0.002
Discrepancy	0.994	0.970	0.950	0.944	0.935	0.923

Table 7
Computational efficiency metrics for various optimization-based NN models.

Algorithm	Avg time (s)	Std time (s)	Memory usage (MB)	CPU usage (%)	Efficiency score
GGPSO+NN	12.34	1.23	245.6	68.5	0.987
GGO+NN	18.76	2.45	312.8	74.2	0.856
PSO+NN	22.15	3.12	398.4	81.3	0.742
GWO+NN	28.92	4.18	467.2	87.6	0.658
GA+NN	35.47	5.67	523.9	92.4	0.573
WAO+NN	41.23	6.89	598.7	96.8	0.489

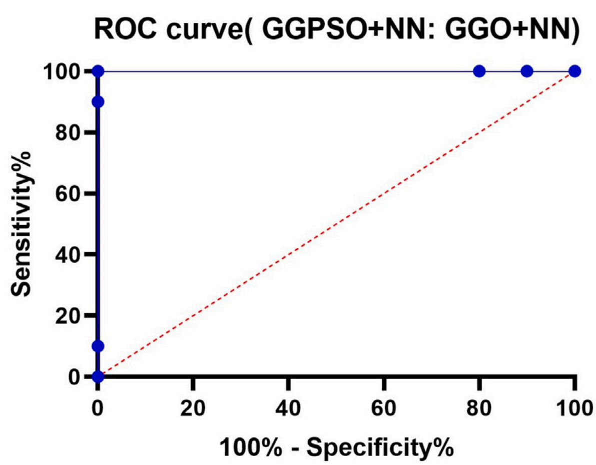

Fig. 12. The ROC curve comparing GGPSO+NN to GGO+NN.

and no clear pattern — is another desirable attribute. QQ plots compare residual distributions to hypothetical normal distributions. A significant connection between the residuals and the diagonal line indicates a normal distribution. This relationship confirms model assumptions. The heatmap image also suggests potential correlations between optimization methods and model performance. This shows how optimization tactics affect neural network predictions for PD diagnosis. These comparative studies help explain how GGPSO+NN performs in residual patterns, homoscedasticity, distributional assumptions, and optimization method-model performance correlations. Such insights are needed to assess the reliability, accuracy, and biases of optimization processes for neural networks for sickness categorization. This helps in choosing the best method for accurate diagnostic predictions.

In practical clinical applications, the computational efficiency of diagnostic systems plays a pivotal role. Real-time deployment, especially in resource-constrained environments, necessitates the use of models that are not only accurate but also efficient in terms of time, memory, and processing power. To this end, we evaluated the training time, memory consumption, CPU usage, and derived an overall efficiency score for each of the tested optimization-based neural network classifiers.

The results, summarized in Table 7, demonstrate that the proposed GGPSO+NN model not only achieves superior accuracy but also offers the best trade-off in terms of computational efficiency. It records the lowest average training time (12.34 s), moderate memory consumption (245.6 MB), and relatively low CPU usage (68.5%). The composite efficiency score of 0.987 confirms its suitability for real-time and embedded applications. In contrast, models such as WAO+NN and GA+NN, while still accurate, exhibit considerably higher resource demands, making them less ideal for clinical deployment scenarios.

To visually interpret these findings, Fig. 14 presents a multi-faceted analysis including comparisons of execution time, memory usage, CPU utilization, and overall efficiency scores. GGPSO+NN stands out consistently across all metrics.

Fig. 15 provides a radar chart comparison across classification performance and efficiency dimensions. The figure highlights the dominance of GGPSO+NN in terms of accuracy, sensitivity, specificity, and resource economy.

Fig. 16 further synthesizes this information through a bubble chart representing accuracy vs. execution time, where the bubble size encodes the efficiency score. Ideally, high-performing algorithms should appear in the top-left quadrant, which is occupied solely by GGPSO+NN.

For a comprehensive performance ranking, Fig. 17 ranks algorithms across individual metrics and provides an overall aggregated score. GGPSO+NN again consistently ranks first.

In Fig. 18, a detailed breakdown of execution time statistics is provided, including the coefficient of variation (CV) and simulated distribution plots. These visualizations help assess the time reliability of each algorithm.

Finally, Fig. 19 offers a CPU-focused performance breakdown. Notably, a strong positive correlation between CPU usage and training time is observed, and GGPSO+NN exhibits the best CPU efficiency ratio (Accuracy/CPU Usage).

6. Conclusions

An early diagnosis of PD is crucial in terms of initiation therapies, initiating discussions on what factors influence the disease, and creating opportunities to design drugs that could best suit it. In this paper, a machine learning model is optimized for efficient discrimination between people regarded as usual and those who have PD. The proposed approach has been proven to have a high threshold for detection, as indicated by its ability to achieve an accuracy of 99.4%. This promising performance is because the proposed optimization approach can attain optimum performance, outperforming other models that have not been optimized. The results found that the proposed approach has a better detection performance than other machine learning models considered. Although the proposed approach performs better than other models, it would be difficult to state categorically that it is far superior to alternative machine learning models used for these purposes. This

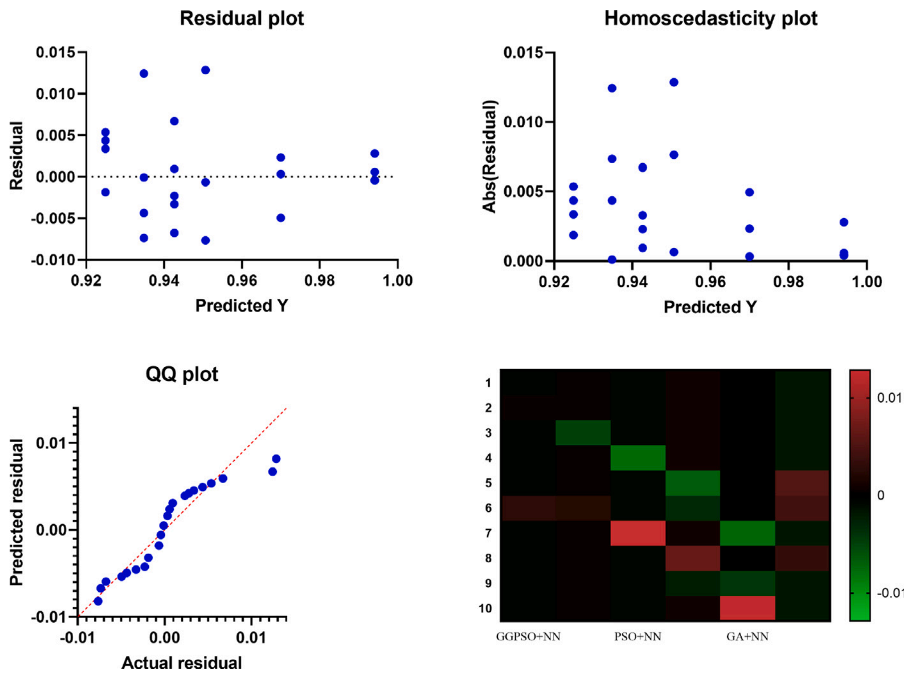

Fig. 13. Residual values and heatmap analysis for GGPSO+NN and compared models.

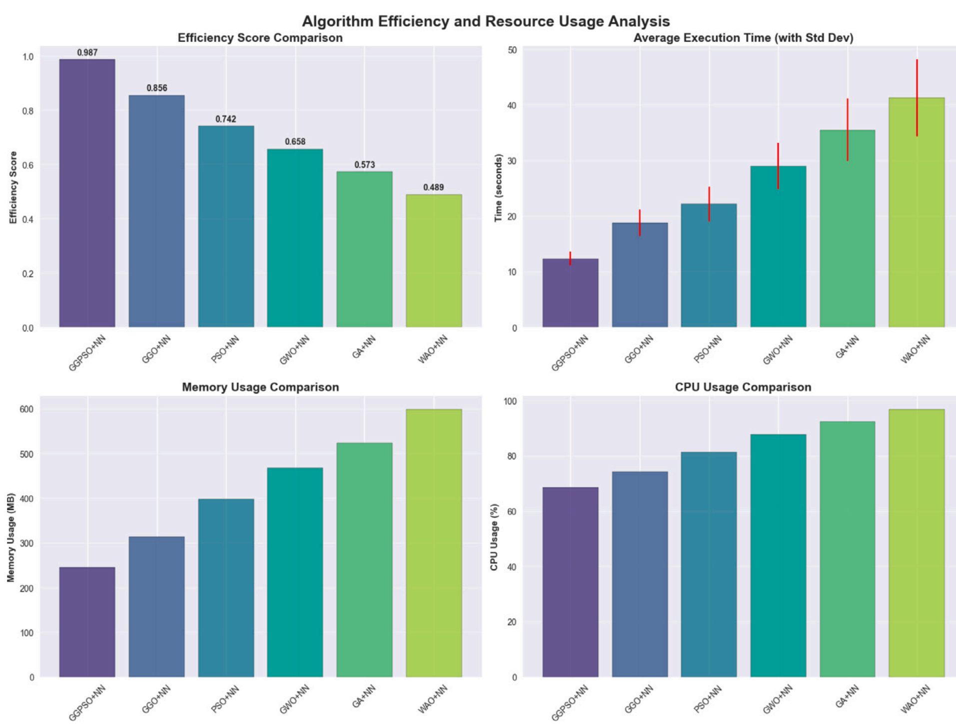

Fig. 14. Visual comparison of efficiency metrics including average execution time, memory usage, CPU usage, and composite efficiency score across all algorithms.

result is because the proposed model relies on a relatively small dataset of PD. The proposed approach is anticipated to reveal its potential when the data becomes considerably more comprehensive and complicated over time. The outcome of this work has the potential to be regarded as a promising first step towards the effective application of cutting-edge research for the early classification of PD.

However, despite its high classification accuracy, the proposed model may face practical limitations in real-world clinical environments. These include the natural variability in speech characteristics arising from patient age, regional accents, and co-existing conditions such as respiratory or cognitive impairments. Such heterogeneity can potentially degrade the model's performance when applied to diverse patient populations not represented in the training dataset. Therefore, future work should emphasize enhancing model generalizability by incorporating larger and more demographically diverse datasets, alongside investigating domain adaptation strategies and robustness mechanisms tailored for clinical variability.

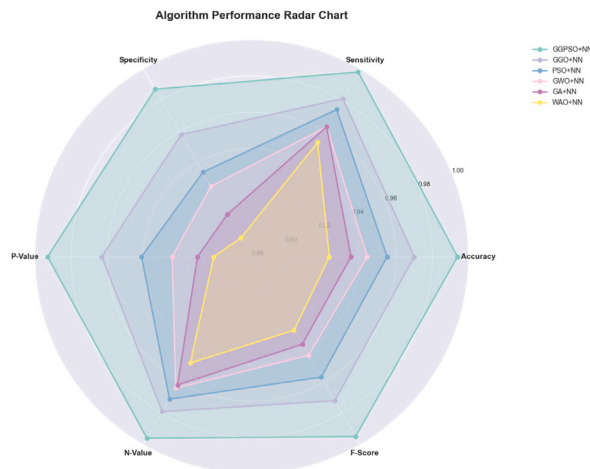

Fig. 15. Radar chart illustrating classification and efficiency-related metrics. GGPSO+NN performs robustly across all criteria.

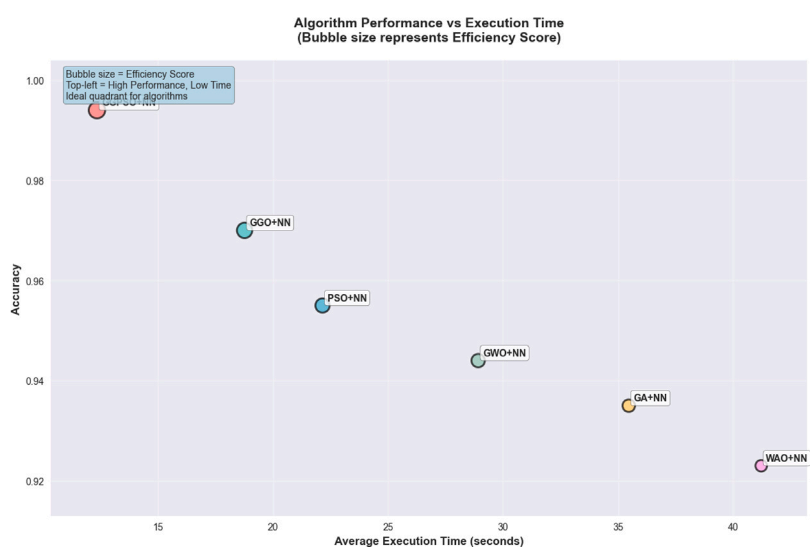

Fig. 16. Bubble chart comparing execution time and classification accuracy, where bubble size indicates overall efficiency score.

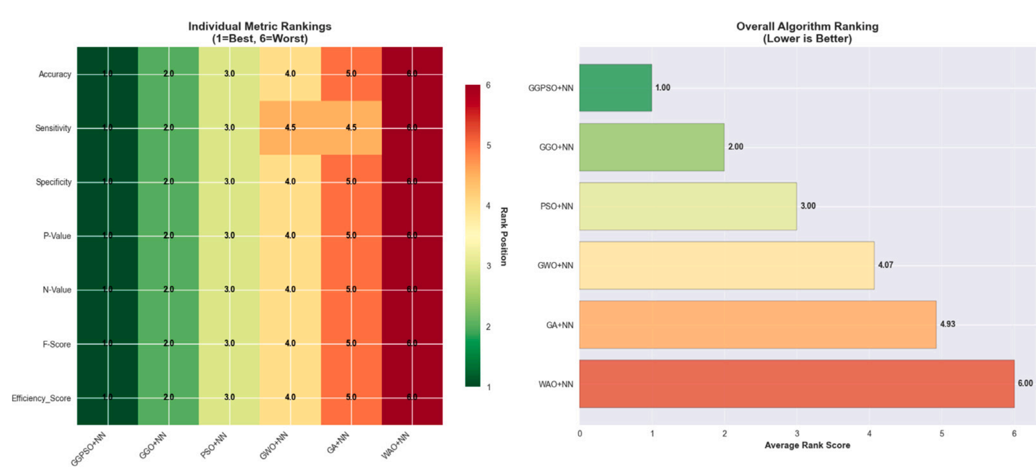

Fig. 17. Metric-specific and overall ranking scores for all algorithms. Lower rank values indicate superior performance.

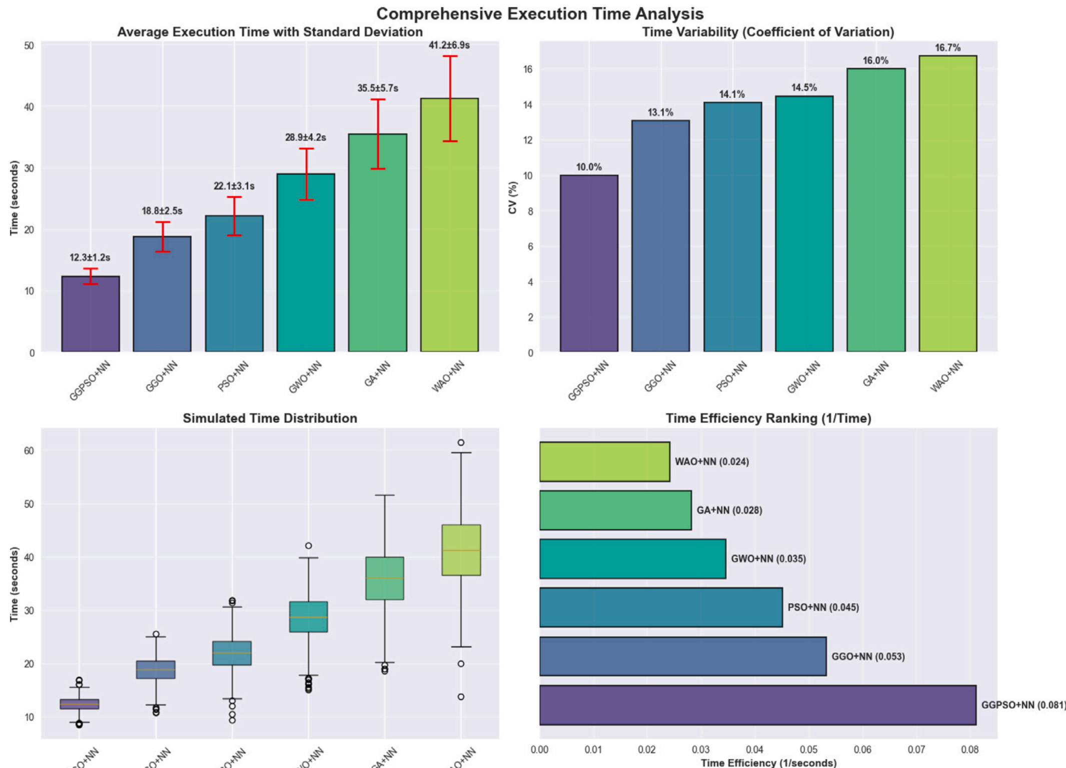

Fig. 18. Comprehensive analysis of execution time, including standard deviation, CV, distribution, and time-efficiency rankings.

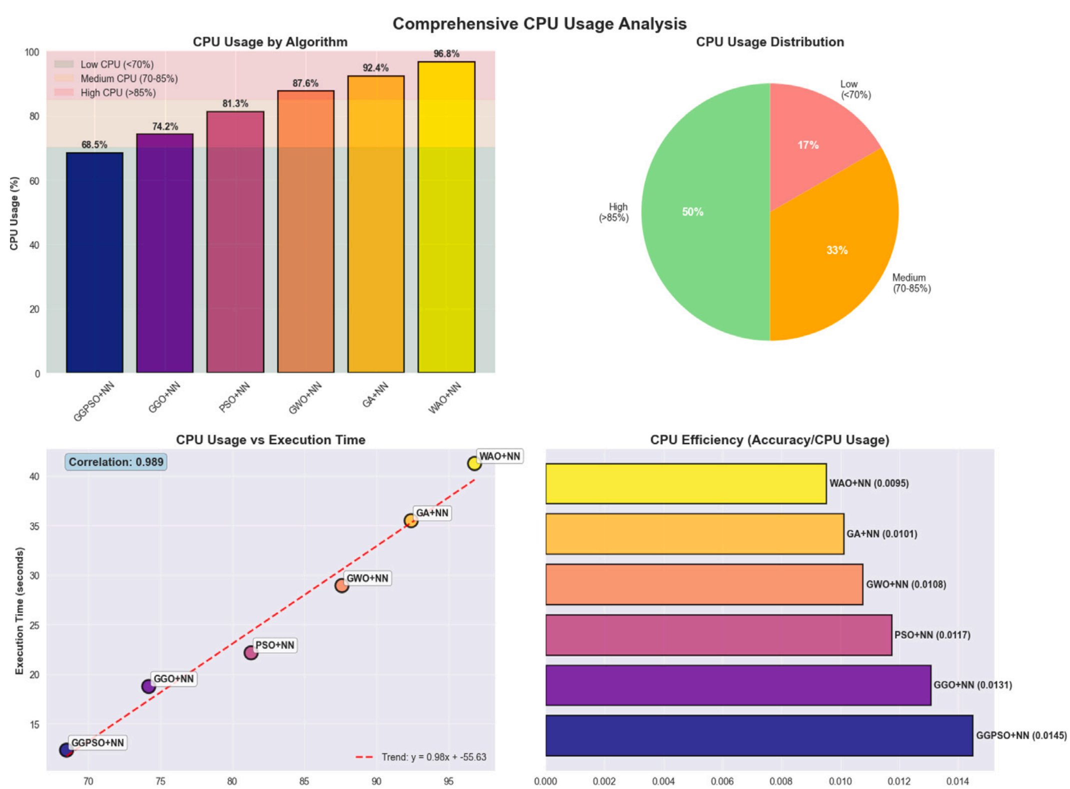

Fig. 19. CPU usage breakdown and its relation to training time and classification efficiency (Accuracy/CPU).

CRediT authorship contribution statement

Doaa Sami Khafaga: Writing – review & editing, Validation. **Marwa M. Eid:** Writing – review & editing, Validation. **Ehsan Khodadadi:** Writing – review & editing, Visualization. **El-Sayed M. El-Kenawy:** Writing – original draft, Data curation. **Amel Ali Alhussan:** Writing – original draft, Visualization, Software, Methodology, Formal analysis. **Nima Khodadadi:** Writing – review & editing, Supervision, Project administration.

Ethical approval

This study does not require ethical approval. Consent to publication All authors read and approved the final manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Princess Nourah bint Abdulrahman University, Saudi Arabia Researchers Supporting Project number (PNURSP2025R754), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Data availability

The data supporting this study's findings are openly available at https://www.kaggle.com/datasets/debasisdotcom/parkinson-disease-detection.

References

[1] G. DeMaagd, A. Philip, Parkinson's disease and its management: Part 1: Disease entity, risk factors, pathophysiology, clinical presentation, and diagnosis, P T: A Peer-Reviewed J. Formul. Manag. 40 (8) (2015) 504–532.

[2] P. Rizek, N. Kumar, M.S. Jog, An update on the diagnosis and treatment of parkinson disease, CMAJ : Can. Med. Assoc. J. 188 (16) (2016) 1157–1165, http://dx.doi.org/10.1503/cmaj.151179.

[3] M.C. de Rijk, L.J. Launer, K. Berger, M.M. Breteler, J.F. Dartigues, M. Baldereschi, L. Fratiglioni, A. Lobo, J. Martinez-Lage, C. Trenkwalder, A. Hofman, Prevalence of parkinson's disease in Europe: A collaborative study of population-based cohorts. neurologic diseases in the elderly research group, Neurology 54 (11 Suppl 5) (2000) S21–23.

[4] I. Canturk, F. Karabiber, A machine learning system for the diagnosis of Parkinson's disease from speech signals and its application to multiple speech signal types, Arab. J. Sci. Eng. 41 (12) (2016) 5049–5059, http://dx.doi.org/10.1007/s13369-016-2206-3.

[5] T.B. Stoker, R.A. Barker, Recent developments in the treatment of Parkinson's disease, F1000Research 9 (2020) http://dx.doi.org/10.12688/f1000research.25634.1, F1000 Faculty Rev–862.

[6] A. Rana, A. Singh Rawat, A. Bijalwan, H. Bahuguna, Application of multi layer (perceptron) artificial neural network in the diagnosis system: A systematic review, in: 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), 2018, pp. 1–6, http://dx.doi.org/10.1109/RICE.2018.8509069.

[7] H. Lakany, Extracting a diagnostic gait signature, Pattern Recognit. 41 (5) (2008) 1627–1637, http://dx.doi.org/10.1016/j.patcog.2007.11.004.

[8] J. Figueiredo, C.P. Santos, J.C. Moreno, Automatic recognition of gait patterns in human motor disorders using machine learning: A review, Med. Eng. Phys. 53 (2018) 1–12, http://dx.doi.org/10.1016/j.medengphy.2017.12.006.

[9] B. Karan, S.S. Sahu, K. Mahto, Parkinson disease prediction using intrinsic mode function based features from speech signal, Biocybern. Biomed. Eng. 40 (1) (2020) 249–264, http://dx.doi.org/10.1016/j.bbe.2019.05.005.

[10] A. Frid, E.J. Safra, H. Hazan, L.L. Lokey, D. Hilu, L. Manevitz, L.O. Ramig, S. Sapir, Computational diagnosis of Parkinson's disease directly from natural speech using machine learning techniques, in: 2014 IEEE International Conference on Software Science, Technology and Engineering, 2014, pp. 50–53, http://dx.doi.org/10.1109/SWSTE.2014.17.

[11] A.S. Rawat, A. Rana, A. Kumar, A. Bagwari, Application of multi layer artificial neural network in the diagnosis system: A systematic review, IAES Int. J. Artif. Intell. (IJ-AI) 7 (3) (2018) 138, http://dx.doi.org/10.11591/ijai.v7.i3.pp138-142.

[12] H. Rouzbahani, M. Daliri, Diagnosis of Parkinson's disease in human using voice signals, Basic Clin. Neurosci. 2 (3) (2011) 12–20.

[13] A. Khamparia, D. Gupta, N.G. Nguyen, A. Khanna, B. Pandey, P. Tiwari, Sound classification using convolutional neural network and tensor deep stacking network, IEEE Access 7 (2019) 7717–7727, http://dx.doi.org/10.1109/ACCESS.2018.2888882.

[14] A. Bourouhou, A. Jilbab, C. Nacir, A. Hammouch, Comparison of classification methods to detect the parkinson disease, in: 2016 International Conference on Electrical and Information Technologies, ICEIT, 2016, pp. 421–424, http://dx.doi.org/10.1109/EITech.2016.7519634.

[15] M.S. Alzubaidi, U. Shah, H. Dhia Zubaydi, K. Dolaat, A.A. Abd-Alrazaq, A. Ahmed, M. Househ, The role of neural network for the detection of Parkinson's disease: A scoping review, Healthcare 9 (6) (2021) 740, http://dx.doi.org/10.3390/healthcare9060740.

[16] H. Purwins, B. Li, T. Virtanen, J. Schluter, S.-Y. Chang, T. Sainath, Deep learning for audio signal processing, IEEE J. Sel. Top. Signal Process. 13 (2) (2019) 206–219, http://dx.doi.org/10.1109/JSTSP.2019.2908700.

[17] L. Zhang, Y. Qu, B. Jin, L. Jing, Z. Gao, Z. Liang, An intelligent mobile-enabled system for diagnosing parkinson disease: Development and validation of a speech impairment detection system, JMIR Med. Informat. 8 (9) (2020) e18689, http://dx.doi.org/10.2196/18689.

[18] S.R. Kadiri, R. Kethireddy, P. Alku, Parkinson's disease detection from speech using single frequency filtering cepstral coefficients, in: Interspeech 2020, ISCA, 2020, pp. 4971–4975, http://dx.doi.org/10.21437/Interspeech.2020-3197.

[19] M. Pramanik, R. Pradhan, P. Nandy, A.K. Bhoi, P. Barsocchi, Machine learning methods with decision forests for Parkinson's detection, Appl. Sci. 11 (2) (2021) 581, http://dx.doi.org/10.3390/app11020581.

[20] H. Gunduz, Deep learning-based Parkinson's disease classification using vocal feature sets, IEEE Access 7 (2019) 115540–115551, http://dx.doi.org/10.1109/ACCESS.2019.2936564.

[21] C. Kohlschein, M. Schmitt, B. Schuller, S. Jeschke, C.J. Werner, A machine learning based system for the automatic evaluation of aphasia speech, in: 2017 IEEE 19th International Conference on E-Health Networking, Applications and Services (Healthcom), 2017, pp. 1–6, http://dx.doi.org/10.1109/HealthCom.2017.8210766.

[22] F. Bertini, D. Allevi, G. Lutero, D. Montesi, L. Calzà, Automatic speech classifier for mild cognitive impairment and early dementia, ACM Trans. Comput. Heal. 3 (1) (2021) 8:1–8:11, http://dx.doi.org/10.1145/3469089.

[23] F. Matcham, C. Barattieri di San Pietro, V. Bulgari, G. de Girolamo, R. Dobson, H. Eriksson, A.A. Folarin, J.M. Haro, M. Kerz, F. Lamers, Q. Li, N.V. Manyakov, D.C. Mohr, I. Myin-Germeys, V. Narayan, P. Bwjh, Y. Ranjan, Z. Rashid, A. Rintala, S. Siddi, S.K. Simblett, T. Wykes, M. Hotopf, RADAR-CNS consortium, Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort study protocol, BMC Psychiatry 19 (1) (2019) 72, http://dx.doi.org/10.1186/s12888-019-2049-z.

[24] C.O. Sakar, G. Serbes, A. Gunduz, H.C. Tunc, H. Nizam, B.E. Sakar, M. Tutuncu, T. Aydin, M.E. Isenkul, H. Apaydin, A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform, Appl. Soft Comput. 74 (2019) 255–263, http://dx.doi.org/10.1016/j.asoc.2018.10.022.

[25] E.-S.M. El-kenawy, N. Khodadadi, S. Mirjalili, A.A. Abdelhamid, M.M. Eid, A. Ibrahim, Greylag goose optimization: Nature-inspired optimization algorithm, Expert Syst. Appl. 238 (2024) 122147, http://dx.doi.org/10.1016/j.eswa.2023.122147.

[26] D. Khafaga, E.-S. El-kenawy, F. Karim, S. Alshetewi, A. Ibrahim, A. Abdelhamid, Optimized weighted ensemble using dipper throated optimization algorithm in metamaterial antenna, Comput. Mater. Contin. 73 (3) (2022) 5771–5788, http://dx.doi.org/10.32604/cmc.2022.032229.

[27] A.A. Alhussan, E.-S. M. El-Kenawy, A.A. Abdelhamid, A. Ibrahim, M.M. Eid, D.S. Khafaga, Wind speed forecasting using optimized bidirectional LSTM based on dipper throated and genetic optimization algorithms, Front. Energy Res. 11 (2023) 1172176, http://dx.doi.org/10.3389/fenrg.2023.1172176.

[28] D. Khafaga, A. Alhussan, E.-S. El-kenawy, A. Ibrahim, S. H, S. El-Mashad, A. Abdelhamid, Improved prediction of metamaterial antenna bandwidth using adaptive optimization of LSTM, Comput. Mater. Contin. 73 (1) (2022) 865–881, http://dx.doi.org/10.32604/cmc.2022.028550.

[29] E.-S.M. El-kenawy, S. Mirjalili, N. Khodadadi, A.A. Abdelhamid, M.M. Eid, M. El-Said, A. Ibrahim, Feature selection in wind speed forecasting systems based on meta-heuristic optimization, in: R. Sirjani (Ed.), PLoS One 18 (2) (2023) e0278491, http://dx.doi.org/10.1371/journal.pone.0278491.

[30] H. AlEisa, E.-S. El-kenawy, A. Alhussan, M. Saber, A. Abdelhamid, D. Khafaga, Transfer learning for chest X-rays diagnosis using dipper throated algorithm, Comput. Mater. Contin. 73 (2) (2022) 2371–2387, http://dx.doi.org/10.32604/cmc.2022.030447.

[31] M.M. Eid, E.-S.M. El-Kenawy, N. Khodadadi, S. Mirjalili, E. Khodadadi, M. Abotaleb, A.H. Alharbi, A.A. Abdelhamid, A. Ibrahim, G.M. Amer, A. Kadi, D.S. Khafaga, Meta-heuristic optimization of LSTM-based deep network for boosting the prediction of Monkeypox cases, Mathematics 10 (20) (2022) 3845, http://dx.doi.org/10.3390/math10203845.

[32] D.S. Khafaga, A. Ibrahim, E.-S.M. El-Kenawy, A.A. Abdelhamid, F.K. Karim, S. Mirjalili, N. Khodadadi, W.H. Lim, M.M. Eid, M.E. Ghoneim, An al-biruni earth radius optimization-based deep convolutional neural network for classifying monkeypox disease, Diagnostics 12 (11) (2022) 2892, http://dx.doi.org/10.3390/diagnostics12112892.

[33] E.-S.M. El-Kenawy, N. Khodadadi, S. Mirjalili, T. Makarovskikh, M. Abotaleb, F.K. Karim, H.K. Alkahtani, A.A. Abdelhamid, M.M. Eid, T. Horiuchi, A. Ibrahim, D.S. Khafaga, Metaheuristic optimization for improving weed detection in wheat images captured by drones, Mathematics 10 (23) (2022) 4421, http://dx.doi.org/10.3390/math10234421.

[34] A.A. Abdelhamid, S.K. Towfek, N. Khodadadi, A.A. Alhussan, D.S. Khafaga, M.M. Eid, A. Ibrahim, Waterwheel plant algorithm: A novel metaheuristic optimization method, Processes 11 (5) (2023) 1502, http://dx.doi.org/10.3390/pr11051502.

[35] A.A. Alhussan, A.A. Abdelhamid, S.K. Towfek, A. Ibrahim, M.M. Eid, D.S. Khafaga, M.S. Saraya, Classification of diabetes using feature selection and hybrid al-biruni earth radius and dipper throated optimization, Diagnostics 13 (12) (2023) 2038, http://dx.doi.org/10.3390/diagnostics13122038.

[36] E.-S.M. El-kenawy, S. Mirjalili, N. Khodadadi, A.A. Abdelhamid, M.M. Eid, M. El-Said, A. Ibrahim, Feature selection in wind speed forecasting systems based on meta-heuristic optimization, in: R. Sirjani (Ed.), PLoS One 18 (2) (2023) e0278491, http://dx.doi.org/10.1371/journal.pone.0278491.

[37] D. Sami Khafaga, E.-S. M. El-kenawy, F. Khalid Karim, S. Alshetewi, A. Ibrahim, A. A. Abdelhamid, Optimized weighted ensemble using dipper throated optimization algorithm in metamaterial antenna, Comput. Mater. Contin. 73 (3) (2022) 5771–5788, http://dx.doi.org/10.32604/cmc.2022.032229.

[38] A.A. Abdelhamid, S.K. Towfek, N. Khodadadi, A.A. Alhussan, D.S. Khafaga, M.M. Eid, A. Ibrahim, Waterwheel plant algorithm: A novel metaheuristic optimization method, Processes 11 (5) (2023) 1502, http://dx.doi.org/10.3390/pr11051502.

[39] A.A. Alhussan, E.-S. M. El-Kenawy, A.A. Abdelhamid, A. Ibrahim, M.M. Eid, D.S. Khafaga, Wind speed forecasting using optimized bidirectional LSTM based on dipper throated and genetic optimization algorithms, Front. Energy Res. 11 (2023) 1172176, http://dx.doi.org/10.3389/fenrg.2023.1172176.

[40] A.A. Alhussan, A.A. Abdelhamid, S.K. Towfek, A. Ibrahim, M.M. Eid, D.S. Khafaga, M.S. Saraya, Classification of diabetes using feature selection and hybrid al-biruni earth radius and dipper throated optimization, Diagnostics 13 (12) (2023) 2038, http://dx.doi.org/10.3390/diagnostics13122038.

[41] A.S. Chandrabhatla, I.J. Pomeraniec, A. Ksendzovsky, Co-evolution of machine learning and digital technologies to improve monitoring of Parkinson's disease motor symptoms, Npj Digit. Med. 5 (1) (2022) 1–18, http://dx.doi.org/10.1038/s41746-022-00568-y, Publisher: Nature Publishing Group.

[42] K.W. Park, E.-J. Lee, J.S. Lee, J. Jeong, N. Choi, S. Jo, M. Jung, J.Y. Do, D.-W. Kang, J.-G. Lee, S.J. Chung, Machine learning–based automatic rating for cardinal symptoms of parkinson disease, Neurology 96 (13) (2021) http://dx.doi.org/10.1212/WNL.0000000000011654, Publisher: Wolters Kluwer.

[43] M.R. Salmanpour, M. Shamsaei, A. Saberi, I.S. Klyuzhin, J. Tang, V. Sossi, A. Rahmim, Machine learning methods for optimal prediction of motor outcome in Parkinson's disease, Phys. Medica 69 (2020) 233–240, http://dx.doi.org/10.1016/j.ejmp.2019.12.022.

[44] J.M. Templeton, C. Poellabauer, S. Schneider, Classification of Parkinson's disease and its stages using machine learning, Sci. Rep. 12 (1) (2022) 14036, http://dx.doi.org/10.1038/s41598-022-18015-z, Publisher: Nature Publishing Group.

[45] T. Exley, S. Moudy, R.M. Patterson, J. Kim, M.V. Albert, Predicting UPDRS motor symptoms in individuals with Parkinson's disease from force plates using machine learning, IEEE J. Biomed. Heal. Informat. 26 (7) (2022) 3486–3494, http://dx.doi.org/10.1109/JBHI.2022.3157518, Conference Name: IEEE Journal of Biomedical and Health Informatics.

[46] C. Sotirakis, Z. Su, M.A. Brzezicki, N. Conway, L. Tarassenko, J.J. FitzGerald, C.A. Antoniades, Identification of motor progression in Parkinson's disease using wearable sensors and machine learning, Npj Parkinson's Dis. 9 (1) (2023) 1–8, http://dx.doi.org/10.1038/s41531-023-00581-2, Publisher: Nature Publishing Group.

[47] M. Martinez-Eguiluz, O. Arbelaitz, I. Gurrutxaga, J. Muguerza, I. Perona, A. Murueta-Goyena, M. Acera, R. Del Pino, B. Tijero, J.C. Gomez-Esteban, I. Gabilondo, Diagnostic classification of Parkinson's disease based on non-motor manifestations and machine learning strategies, Neural Comput. Appl. 35 (8) (2023) 5603–5617, http://dx.doi.org/10.1007/s00521-022-07256-8.

[48] J. Zhang, Mining imaging and clinical data with machine learning approaches for the diagnosis and early detection of ?parkinson's disease, Npj Parkinson's Dis. 8 (1) (2022) 1–15, http://dx.doi.org/10.1038/s41531-021-00266-8, Publisher: Nature Publishing Group.

[49] B. Majhi, A. Kashyap, S.S. Mohanty, S. Dash, S. Mallik, A. Li, Z. Zhao, An improved method for diagnosis of Parkinson's disease using deep learning models enhanced with metaheuristic algorithm, BMC Med. Imaging 24 (1) (2024) 156, http://dx.doi.org/10.1186/s12880-024-01335-z.

[50] L. Jovanovic, R. Damasevicius, R. Matic, M. Kabiljo, V. Simic, G. Kunjadic, M. Antonijevic, M. Zivkovic, N. Bacanin, Detecting parkinson's disease from shoe-mounted accelerometer sensors using convolutional neural networks optimized with modified metaheuristics, PeerJ Comput. Sci. 10 (2024) e2031, http://dx.doi.org/10.7717/peerj-cs.2031, Publisher: PeerJ Inc..

[51] E.V. Altay, B. Alatas, Association analysis of Parkinson disease with vocal change characteristics using multi-objective metaheuristic optimization, Med. Hypotheses 141 (2020) 109722, http://dx.doi.org/10.1016/j.mehy.2020.109722.

[52] C. Quan, K. Ren, Z. Luo, Z. Chen, Y. Ling, End-to-end deep learning approach for Parkinson's disease detection from speech signals, Biocybern. Biomed. Eng. 42 (2) (2022) 556–574, http://dx.doi.org/10.1016/j.bbe.2022.04.002.

[53] R. Khaskhoussy, Y.B. Ayed, Improving Parkinson's disease recognition through voice analysis using deep learning, Pattern Recognit. Lett. 168 (2023) 64–70, http://dx.doi.org/10.1016/j.patrec.2023.03.011.

[54] G. Nagasubramanian, M. Sankayya, Multi-variate vocal data analysis for detection of Parkinson disease using deep learning, Neural Comput. Appl. 33 (10) (2021) 4849–4864, http://dx.doi.org/10.1007/s00521-020-05233-7.

[55] R. Khaskhoussy, Y.B. Ayed, Speech processing for early parkinson's disease diagnosis: machine learning and deep learning-based approach, Soc. Netw. Anal. Min. 12 (1) (2022) 73, http://dx.doi.org/10.1007/s13278-022-00905-9.

[56] D. Escobar-Grisales, C.D. Ríos-Urrego, J.R. Orozco-Arroyave, Deep learning and artificial intelligence applied to model speech and language in parkinson's disease, Diagnostics 13 (13) (2023) 2163, http://dx.doi.org/10.3390/diagnostics13132163, Number: 13 Publisher: Multidisciplinary Digital Publishing Institute.

[57] R. Alkhatib, M.O. Diab, C. Corbier, M.E. Badaoui, Machine learning algorithm for gait analysis and classification on early detection of parkinson, IEEE Sensors Lett. 4 (6) (2020) 1–4, http://dx.doi.org/10.1109/LSENS.2020.2994938, Conference Name: IEEE Sensors Letters.

[58] D. Trabassi, M. Serrao, T. Varrecchia, A. Ranavolo, G. Coppola, R. De Icco, C. Tassorelli, S.F. Castiglia, Machine learning approach to support the detection of parkinson's disease in IMU-based gait analysis, Sensors 22 (10) (2022) 3700, http://dx.doi.org/10.3390/s22103700, Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

[59] B. Muñoz-Ospina, D. Alvarez-Garcia, H.J.C. Clavijo-Moran, J.A. Valderrama-Chaparro, M. García-Peña, C.A. Herrán, C.C. Urcuqui, A. Navarro-Cadavid, J. Orozco, Machine learning classifiers to evaluate data from gait analysis with depth cameras in patients with Parkinson's disease, Front. Hum. Neurosci. 16 (2022) http://dx.doi.org/10.3389/fnhum.2022.826376, Publisher: Frontiers.

[60] A. Landolfi, C. Ricciardi, L. Donisi, G. Cesarelli, J. Troisi, C. Vitale, P. Barone, M. Amboni, Machine learning approaches in Parkinson's disease, Curr. Med. Chem. 28 (32) (2021) 6548–6568, http://dx.doi.org/10.2174/0929867328999210111211420.

[61] M. Diaz, M. Moetesum, I. Siddiqi, G. Vessio, Sequence-based dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and BiGRUs, Expert Syst. Appl. 168 (2021) 114405, http://dx.doi.org/10.1016/j.eswa.2020.114405.

[62] I. Kamran, S. Naz, I. Razzak, M. Imran, Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease, Future Gener. Comput. Syst. 117 (2021) 234–244, http://dx.doi.org/10.1016/j.future.2020.11.020.

[63] C. Taleb, L. Likforman-Sulem, C. Mokbel, M. Khachab, Detection of parkinson's disease from handwriting using deep learning: a comparative study, Evol. Intell. 16 (6) (2023) 1813–1824, http://dx.doi.org/10.1007/s12065-020-00470-0.

[64] Z. Galaz, P. Drotar, J. Mekyska, M. Gazda, J. Mucha, V. Zvoncak, Z. Smekal, M. Faundez-Zanuy, R. Castrillon, J.R. Orozco-Arroyave, S. Rapcsak, T. Kincses, L. Brabenec, I. Rektorova, Comparison of CNN-learned vs. handcrafted features for detection of parkinson's disease dysgraphia in a multilingual dataset, Front. Neuroinformatics 16 (2022) http://dx.doi.org/10.3389/fninf.2022.877139, Publisher: Frontiers.

[65] X. Xing, N. Luo, S. Li, L. Zhou, C. Song, J. Liu, Identification and classification of parkinsonian and essential tremors for diagnosis using machine learning algorithms, Front. Neurosci. 16 (2022) http://dx.doi.org/10.3389/fnins.2022.701632, Publisher: Frontiers.

[66] R. Zhang, J. Jia, R. Zhang, EEG analysis of parkinson's disease using time–frequency analysis and deep learning, Biomed. Signal Process. Control. 78 (2022) 103883, http://dx.doi.org/10.1016/j.bspc.2022.103883.

[67] N.R. Yousif, H.M. Balaha, A.Y. Haikal, E.M. El-Gendy, A generic optimization and learning framework for parkinson disease via speech and handwritten records, J. Ambient. Intell. Humaniz. Comput. 14 (8) (2023) 10673–10693, http://dx.doi.org/10.1007/s12652-022-04342-6.

[68] Debasis, Parkinson disease detection, 2022, https://www.kaggle.com/datasets/debasisdotcom/parkinson-disease-detection. (Accessed 22 July 2025).

[69] F. Daneshfar, S.J. Kabudian, Speech emotion recognition using deep sparse auto-encoder extreme learning machine with a new weighting scheme and spectro-temporal features along with classical feature selection and a new quantum-inspired dimension reduction method, 2021, arXiv:2111.07094. URL https://arxiv.org/abs/2111.07094.

[70] H.H.S. Junaid, F. Daneshfar, M.A. Mohammad, Automatic colorectal cancer detection using machine learning and deep learning based on feature selection in histopathological images, Biomed. Signal Process. Control. 107 (2025) 107866, http://dx.doi.org/10.1016/j.bspc.2025.107866, URL https://www.sciencedirect.com/science/article/pii/S1746809425003775.