



Quantum mayfly optimization based feature subset selection with hybrid CNN for biomedical Parkinson's disease diagnosis

Romany F. Mansour¹

Received: 16 January 2022 / Accepted: 14 January 2024 / Published online: 22 February 2024
© The Author(s) 2024

Abstract

Parkinson's disease (PD) arises from brain cell damage and necessitates early detection for effective treatment and symptom management. While various methods such as voice, speech, and written exams have been explored, utilizing automated tools is crucial to enhance accuracy. Recent advancements in artificial intelligence (AI) and deep learning (DL) provide an opportunity for precise early-stage PD identification. This study introduces a novel approach known as Quantum Mayfly Optimization-based feature subset selection with hybrid convolutional neural network (QMFOFS-HCNN) to improve PD detection and classification. QMFOFS-HCNN is designed to identify optimal feature subsets and overcome the dimensionality challenge. It combines a quantum mayfly optimization approach for feature selection with a convolutional neural network with attention-based long short-term memory for PD detection and classification. Additionally, hyperparameter selection is optimized using the Nadam optimizer. Experimental validation using benchmark datasets yielded compelling results. The QMFOFS-HCNN technique achieved accuracy rates: 96.35% for HandPD Spiral, 96.7% for HandPD Meander, 98.5% for Speech PD, and a perfect 100% for Voice PD datasets. These quantitative findings underscore the potential of AI and DL to enhance early PD detection accuracy significantly. These results offer promising prospects for improving healthcare outcomes in managing PD and related neurological disorders.

Keywords Deep learning · Parkinson disease · Mayfly optimization algorithm · Feature selection

Abbreviations

AI	Artificial intelligence	QMFOFS-HCNN	Quantum mayfly optimization-based feature subset selection with hybrid convolutional neural network
DL	Deep learning		
ECSO	Expanded cat swarm optimization	RF	Random forest
EDF-EMD	Energy direction feature based empirical mode decomposition	SVM	Support vector machine
MGOA	Modified grasshopper optimization algorithm	K-NN	K-nearest neighbour
MGWO	Modified grey wolf optimizer	MLP	Multilayer perceptron
OCFA	Optimized cuttlefish algorithm	PD	Parkinson's disease
IFSO-DL	Improved sailfish optimization algorithm with deep learning	MF	Male and female flies
CNN-ALSTM	Convolutional neural network with attention-based long short-term memory	DNN	Deep neural network
		FAR	False alarm rate
		FS	Feature selection
		ANN	Artificial neural network
		IMF	Intrinsic mode function
		FF	Fitness function
		QMFO	Quantum mayfly optimization

✉ Romany F. Mansour
romanyf@sci.nvu.edu.eg

¹ Department of Mathematics, Faculty of Science, New Valley University, El-Kharga 72511, Egypt

1 Introduction

Parkinson's disease (PD), initially developed by James Parkinson, affects an individual's movement, leading to muscle stiffness, tremors, and changes in speech and writing skills [1]. This condition occurs when nerve cells produce a chemical called dopamine that breaks down, making nerve cells unable to transmit messages accurately, this condition occur may be due to genetic factors. It can result in depression, nervous disorders, and memory impairment [2]. Several authors have been conducted to diagnose the disease at its earlier stage, though not with great accomplishment. Identifying the disease in its earlier stages is significant so patients can live quality lives [3]. The disease at its advanced stage affects the day-to-day tasks, and the person might need help from others. The later stages of PD are sufficiently severe as the patient gets stiffness in the legs, making it impossible to stand or walk and might cause freezing on standing. Several methods have been utilized for identifying the disease correctly such as writing, speech, and voice exams [4]. The handwritten exam is widely employed for diagnosing PD, because it is easier to get data and inexpensive.

In recent times, data have been enhanced by the amount of features and instances that make data noisier [5]. The noisier data sets construct the algorithm to increase the computational cost, decrease the predicted accuracy, train the data slower, and increase the complexity. Thus, the feature selection (FS) method designed a significant process for the machine learning (ML) approach before training the model [6]. The task of processing and pre-processing data is a complex, as the increase in feature and instance count results in an increase in the quantity of data. Growth in data makes it more vulnerable to noise, which might result in degraded results and a drop in performance. Therefore, it becomes indispensable for treating the data [7]. Complexity tends and computational cost to increase when a large amount of data is used. Hence, the FS method plays an important role in building architecture in ML. In the FS method, also called parameter selection, a feature subset is selected from existing features.

The primary objective is to improve the algorithm's accuracy before and after FS. The FS method assists in resolving the problems by reducing the computation complexity and cost of datasets [8]. Information has been enhanced by using many instances and features, which makes data noisier. The noisier dataset causes the algorithm to reduce the accuracy predicted by models, increases computational costs, increases complexity, and trains the data slower [9]. Consequently, the FS method has become an essential process for ML before training the model. The FS approach focuses on finding a subset from

the entire set of features and less downgrade performance of the network; so, the subset of features forecasts the target with performance similar to the accuracy of the novel set of features and reduced of computation cost. Feature selection helps to understand the causes of disease, reduces the computational requirements, and prevents degradation in performance that contributes to better/faster convergence of the deep training method.

The FS model is classified into wrapper—and filter-based models [10]. The FS algorithms use MLs in wrapper-based approaches to check the accuracy of the selected subset of features with high accuracy. However, these approaches could be more effective with high dimensional datasets due to high training time [11]. Subsequently, filter-based approaches use statistical data dependency methods to reach the best subset faster. Filter-based approaches are less accurate, more scalable, faster, and less computationally expensive than wrapper-based approaches [12].

The QMFOFS-HCNN technique aims to improve Parkinson's disease (PD) detection and classification by utilizing Quantum Mayfly Optimization (QMFO) for feature selection, a Convolutional Neural Network with Attention Long Short Term Memory (CNN-ALSTM) for classification, and hyperparameter tuning with the Nadam optimizer. The contributions of the given study are: (i) It uses QMFO to select relevant features, enhancing classification accuracy and reducing computational complexity. (ii) It employs CNN-ALSTM for PD classification, which is well-suited for biomedical time series data with an attention mechanism to capture important information. (iii) It fine-tunes model parameters with Nadam optimizer, improving overall performance. (iv) It demonstrates superior accuracy and detection rates compared to existing methods on benchmark PD datasets. (v) It efficiently selects minimal features while maintaining high accuracy, which is crucial for real world applications. (vi) It is effective across various PD datasets, suggesting broader applicability.

Thus, this study develops a quantum mayfly optimization-based feature subset selection with a hybrid convolutional neural network (QMFOFS-HCNN) technique for PD detection and classification. The principal intention of the QMFOFS-HCNN technique is to identify the optimal feature subsets and enhance the classification accuracy of the PD diagnosis. The QMFOFS-HCNN technique initially designs a novel QMFO approach for the optimum feature choice and resolves the curse of dimensionality problem. In addition, an optimal CNN with attention long short-term memory (CNN-ALSTM) model is employed to detect and classify PD. In order to effectively boost the PD classification outcomes, the Nadam optimizer can be utilized to select the hyperparameters. The experimental validation

takes place using the benchmark datasets, and the results are assessed under several aspects.

2 Literature survey

The authors in [13] developed a cloud-based PD predictive model for making medical decisions that assist physicians in identifying the Parkinson-affected person from a remote place. An efficient expanded cat swarm optimization (ECSO)-based FS method has been examined to resolve the problems of data dimensionality. The classification method can considerably enhance the disease predictive performance by utilizing the FS method in the K-nearest neighbour (K-NN). Solana-Lavalle et al. [14] focused on increasing the accuracy and reducing the amount of selected vocal features in PD diagnoses while utilizing the most extensive and newest open-source dataset. While the number of features in this public dataset is 754, the number of selected features for classification ranges from 8 to 20 after utilizing the Wrapper feature subset selection. The KNN, multilayer perceptron (MLP), support vector machine (SVM), and random forest (RF) classifiers are employed for detecting vocal-based PD.

Mathur et al. [15] use different ML methods, which could enhance the efficiency of data sets and play a significant part in making the earlier disease prediction. Afterward, the comparison of this algorithm selects the efficient one in terms of accuracy. The experiment outcomes show that the performance attained from the integrated effects of artificial neural network (ANN) and KNN algorithm is more effective than other approaches. The authors in [11] introduced two NN-based methods, voice impairment classifier and spectrogram detector that focus on helping people and doctors identify disease at earlier stages. A wide-ranging assessment of CNN has been conducted on a large image classifier of gait signal transformed to spectrogram image and deep dense ANN on the voice recording to forecast the disease. El Maachi et al. [12] developed a smart PD method-based deep learning (DL) method for analysing gait data. Then, 1D-Convnet is used to construct the deep neural network (DNN) classification. The presented method processes eighteen 1D signals from the foot sensor, evaluating the classifier. Haq et al. [16] introduced an ML and DNN-based non-invasive predictive model for timely and accurate diagnoses of PD. The ML prediction methods, namely SVM, linear regression (LR), and DNN, have been utilized for classifying healthier people and PD. Zhang et al. [17] proposed an energy direction feature-based empirical mode decomposition (EDF-EMD) feature to display the distinct features of voice signals among healthy and PD patients. At first, the intrinsic mode function (IMF) was attained by using the

decomposition of voice signal with empirical mode decomposition.

In Parkinson's disease (PD) research, several previous studies have aimed to diagnose the condition in its early stages but have had limited success [18, 19]. Detecting PD early on is crucial for improving the quality of life for patients. Existing approaches have explored methods such as handwriting analysis, speech assessment, and voice examinations, with handwriting being a preferred choice due to its ease of data collection and affordability [20–22]. However, contemporary data sets have grown in size and complexity, introducing noise that can hinder algorithm accuracy, increase computational costs, and slow down data processing [23]. Researchers have turned to feature selection (FS) techniques to address these challenges as a critical step in machine learning (ML) model development. FS helps optimize algorithm performance by selecting a subset of relevant features, reducing computational complexity, and mitigating data noise. The QMFOFS-HCNN technique presented in this study represents a significant advancement in PD detection and classification. It leverages Quantum Mayfly Optimization (QMFO) for feature selection, employs a Convolutional Neural Network with Attention Long Short Term Memory (CNN-ALSTM) for classification, and fine-tunes model parameters using the Nadam optimizer. The key contributions of this research include improved accuracy, feature subset optimization, and enhanced classification performance. Importantly, this technique efficiently selects minimal features while maintaining high accuracy, making it suitable for real-world applications across various PD datasets. Compared to prior work, this study introduces a comprehensive and innovative approach to PD detection, offering the potential for more accurate and efficient diagnoses. While previous research has explored various machine learning methods and feature selection techniques [24], the QMFOFS-HCNN method stands out for its superior accuracy, computational efficiency, and adaptability across diverse PD datasets.

3 Material and methods

3.1 Dataset

The proposed method has been employed with datasets related to Parkinson's disease, encompassing diverse types of sound recordings, as well as data from Parkinson's HandPD, which are as follows:

In the Speech PD dataset, a set of biomedical voice measurements has been gathered from 23 individuals. Each dataset column corresponds to a particular voice measurement, and each dataset row links to one of the 195 recordings of voice taken from these individuals. The main

aim of this dataset is to classify between healthy individuals (coded as 0) and those with PD (coded as 1). The dataset was curated by Max Little from the University of Oxford, in partnership with the National Centre for Voice and Speech in Denver, Colorado, where the speech signals were acquired.

In the Voice PD dataset, the training data comprises of records from 20 individuals with PD (14 male and 6 female) and 20 healthy (10 male and 10 female) individuals, who were seen at the department of neurology in Cerrahpasa faculty of medicine, Istanbul University. In data acquisition process, 28 PD patients were suggested to repeat the vowels 'a' and 'o' three times each, ensuing total 168 voice recordings. This dataset acts as a valuable independent test set for results validation obtained from the training dataset.

In the HandPD meander dataset, data has been collected from a total of 158 individuals, including 74 in patient group and 18 in healthy group. Dataset comprises 632 data instances encompassing 13 distinct features. Furthermore, the dataset involves 632 images of meanders drawn by the patients. These individuals represented the age ranges, from 14 to 79 years old. The handwritten examinations were comprised at Botucatu Medical School, São Paulo State University, Brazil.

In the HandPD spiral dataset, participants were recommended to sketch spirals instead of meanders. This dataset comprises of data from 158 individuals. It involves 632 data instances and contains 13 distinct features. The handwritten examinations were collected at Botucatu Medical School, São Paulo State University, Brazil.

3.2 Methods

3.2.1 Design of QMFOFS-HCNN model

This study has developed a novel QMFOFS-HCNN technique for detecting and classifying PD. It aimed to identify the optimal feature subsets and optimize the classification performance of the PD diagnosis. The suggested QMFOFS-HCNN technique encompasses several processes such as QMFO-based feature subset selection, CNN-ALSTM-based classifier, and Nadam-based hyper-parameter tuning. Using QMFOFS and Nadam techniques helps boost the PD classification outcomes effectively. Figure 1 depicts the entire working process of the proposed QMFOFS-HCNN technique. Figure 1 depicts the entire working process of the proposed QMFOFS-HCNN technique. First, the Parkinson dataset has been given as input for its pre-processing in order to remove artefacts. Subsequently, the dataset is divided into the training and testing datasets for providing training and testing. Afterward, a novel QMFO-based feature selection (FS) method has been

used to resolve the curse of dimensionality problem by reduction of computational complexity. Information has been enhanced by using many instances and features, which makes data noisier. The noisier dataset causes the algorithm to reduce the accuracy predicted by models, the computational costs, increases the complexity and slows the training process.

Consequently, the FS approach focuses on finding an appropriate subset from the entire set of features with high performance of the network and less computational cost. In order to optimize the efficiency of the MFO algorithm, the QMFO technique is derived; for details, refer to [25]. Subsequently, the Nadam optimizer has been used to boost the classification outcome for hyperparameter tuning. At this stage, the CNN-ALSTM model is employed for PD classification. The CNN-ALSTM is a hybrid DL approach for extracting features in the raw information and implementing predicting utilizing the LSTM-NN [26]. The CNN uses LSTM for optimum extracting the features of experimental data. The attention method is a procedure for allocating weight. Thus, the proposed work develops a QMFOFS-HCNN technique for PD detection and classification. The primary intention of the proposed technique is to identify the optimal feature subsets and enhance the classification accuracy of the PD diagnosis.

3.2.2 Algorithmic design of QMFOFS technique

The MFO algorithm derives from the social activity of MFs [27]. MFs were generated by adults, and afterward, the fittest lived. Two sets of populations were primarily created. It can signify both males as well as female populations. The candidate is signified by d dimension vector $x = (x_1, \dots, x_d)$. The fitness of candidates is estimated by computing the fitness function (FF) $f(x)$. The velocity $v = (v_1, \dots, v_d)$ has been modified from the candidate place. All the candidates alter their trajectory based on their optimum place (pbest) and an optimum place for every MF (gbest).

Collecting male MFs reflects all males' knowledge from defining their place in terms of $neighbors^l$ places determining x_i^t as present place of candidate solutions i at time t , the place was changed by adding a velocity v_i^{t+1} as [28]:

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (1)$$

With $x_i^0 U(x_{\min}, x_{\max})$. Considering the minimum velocity of the male population, the velocity is computed as follows:

$$v_{ij}^{t+1} = v_{ij}^t + a_1 e^{-\beta r_p^2} (pbest_{ij} - x_{ij}^t) + a_2 e^{-\beta r_g^2} (gbest_i - x_{ij}^t), \quad (2)$$

, where v_{ij}^t refers to the velocity of MFs i , x_{ij}^t signifies the

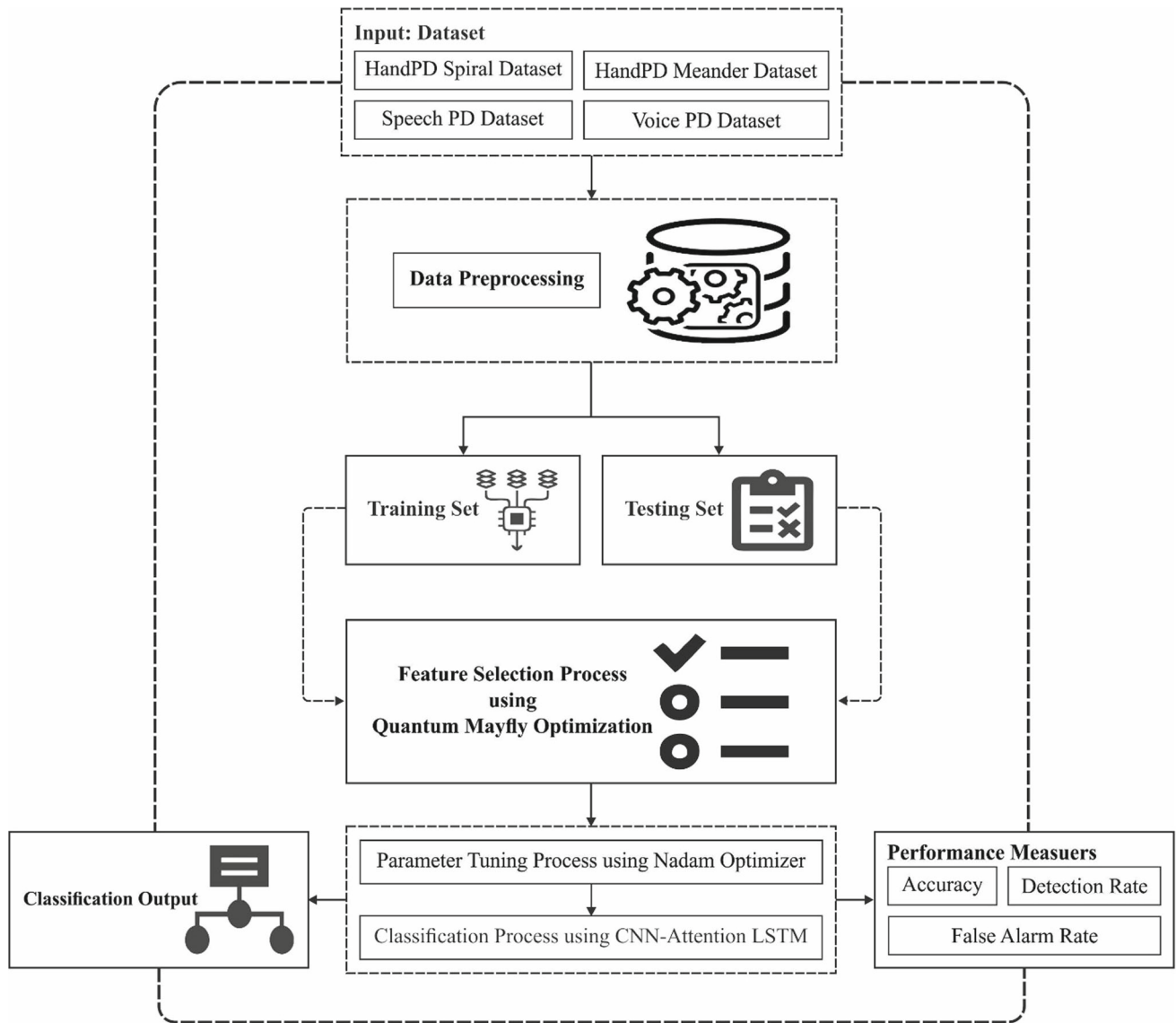


Fig. 1 Overall working process of QMFOFS-HCNN technique

place of MFs i , a_1 , and a_2 are determined as positive constants signifying the attractive. $pbest_i$ stands for the optimum place that candidate solution i had always obtained, and $pbest_{ij}$ at the subsequent step $t + 1$ was defined in Eq. (3).

$$pbest_i = \begin{cases} x_i^{t+1}, & \text{iff } (x_i^{t+1}) < f(pbest_i) \\ \text{same as before,} & \text{otherwise} \end{cases} \quad (3)$$

, where $f : \mathbb{R}^n \Rightarrow \mathbb{R}$ refers to the function minimizing, $gbest$ signifies the global optimum attained from the issue ever at time t . The co-efficient in Eq. (2) limits the population's visibility. r_p implies the distance among x_i and $pbest_i$. In the meantime, r_g determines the distance in x_i to $gbest$. r_p and r_g are defined in Eq. (4).

$$\|x_i - X_i\| = \sqrt{\sum_{j=1}^n (x_{ij} - X_{ij})^2} \quad (4)$$

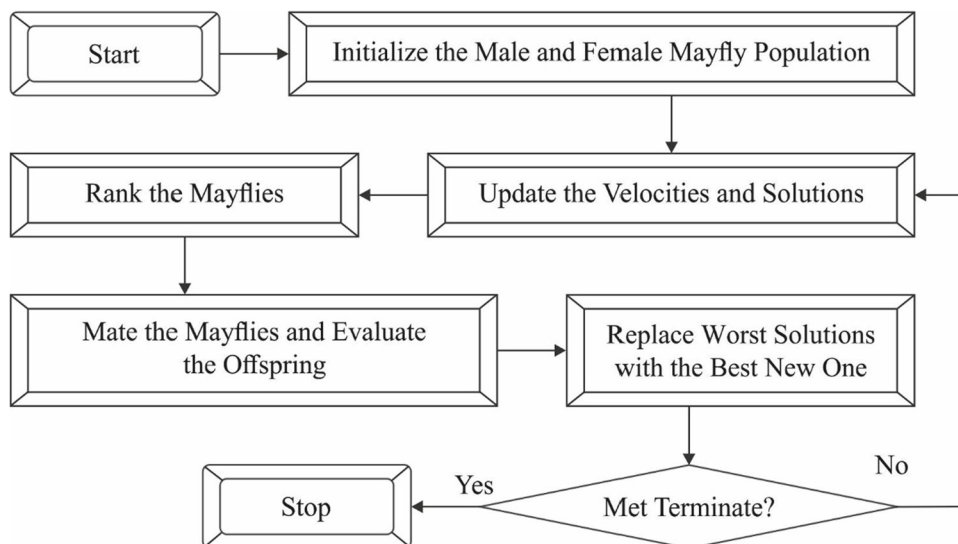
, where x_{ij} refers the j^{th} component of i^{th} candidate. X_i is connected to $pbest$.

An optimum fit candidate keeps implementing up and down motions by different velocities. The velocity is defined as in Eq. (5).

$$v_{ij}^{t+1} = v_{ij}^t + d \times r \quad (5)$$

, where d denotes the co-efficient compared with up and down motions, and r represents the arbitrary value between -1 and 1 . Figure 2 demonstrates the flowchart of the MFO technique.

Fig. 2 Flowchart of MFO algorithm



The female MFs do not gather, but they move near males. Assume that y_i^t is the present place of female MF i at time t . The alteration from the place was computed as:

$$y_i^{t+1} = y_i^t + v_i^{t+1} \tag{6}$$

With $y_i^0 U(x_{\min}, x_{\max})$. The female MFs' velocity is defined as in Eq. (7).

$$v_{ij}^{t+1} = \begin{cases} v_{ij}^t + a_2 e^{-\beta r_{mf}^2} (x_{ij}^t - y_{ij}^t), & \text{if } f(y_i) > f(x_i) \\ v_{ij}^t + fl \times r, & \text{if } f(y_i) \leq f(x_i) \end{cases} \tag{7}$$

, where v_{ij}^t refers to the velocity of i^{th} female at time t , y_{ij}^t signifies the place of i^{th} female candidate solution at time t , a_2 signifies the positive constants, β stands for the set coefficient, r_{mf} indicates the distance between the male candidate solution and female ones that are calculated utilizing in Eq. (4), fl signifies the co-efficient that relates the female which is not attractive. r implies the arbitrary number between -1 and 1 . The mating was demonstrated by an operator that is a crossover operator. The pair of male, as well as female parents are selected.

$$\begin{aligned} \text{off spring 1} &= L \times \text{male} + (1 - L) \times \text{female} \\ \text{off spring 2} &= L \times \text{female} + (1 - L) \times \text{male} \end{aligned} \tag{8}$$

, where L refers to the arbitrary number. Primarily, the velocity of offspring is equivalent to 0. In order to optimize the efficiency of the MFO algorithm, the QMFO technique is derived [25]. With the quantization of grasshopper individuals, the feature search space has improved to balance exploitation and exploration. A vital unit of QC is qubit. The two important forms $|0\rangle$ and $|1\rangle$ way a qubit which has been formulated as a linear grouping of these two essential forms as:

$$|Q\rangle = \alpha|0\rangle + \beta|1\rangle. \tag{9}$$

$|\alpha|^2$ refers the probability of identifying form $|0\rangle$, $|\beta|^2$ signifies the probability of detecting state $|1\rangle$, where $|\alpha|^2 + |\beta|^2 = 1$. The quantum is composed of n qubits. Because of the form of quantum superposition, all quantum has 2^n probable values.

$$\Psi = \sum_{x=0}^{2^n-1} C_x |x\rangle, \sum_{x=0}^{2^n-1} |C_x|^2 = 1. \tag{10}$$

Quantum gates have modified the state of qubits as Hadamard, rotation, and NOT gates, among others. The rotation gate was explained as a mutation function to make quanta model optimal solutions and finally determined the global optimal solutions.

The rotation gate is shown as follows:

$$\begin{aligned} \begin{bmatrix} \alpha^d(t+1) \\ \beta^d(t+1) \end{bmatrix} &= \begin{bmatrix} \cos(\Delta\theta^d) & -\sin(\Delta\theta^d) \\ \sin(\Delta\theta^d) & \cos(\Delta\theta^d) \end{bmatrix} \begin{bmatrix} \alpha^d(t) \\ \beta^d(t) \end{bmatrix} \text{ for } d \\ &= 1, 2, \dots, n. \end{aligned} \tag{11}$$

$\Delta\theta^d = \Delta \times S(\alpha^d, \beta^d)$, $\Delta\theta^d$ stands for the rotation angle of qubit, whereas Δ and $S(\alpha^d, \beta^d)$ are size and way of rotation correspondingly.

The mathematical model of the QMFOFS approach was established. Generally, some data sets' classification (i.e. supervised learning) is size $N_S \times N_F$, whereas N_S refers to the number of instances, and N_F implies the number of features. An important objective of the FS issue is to select a subset of features S in the entire amount of features (N_F) in which the size of S is lesser than N_F . It is obtained by minimizing the subsequent primary function:

$$\text{Fit} = \lambda \times \gamma_S + (1 - \lambda) \times \left(\frac{|S|}{N_F} \right) \tag{12}$$

, where γ_S denotes the classification error utilizing S and $|S|$ is the amount of chosen features. λ is utilized for balancing among $\left(\frac{|S|}{N_F} \right)$ and γ_S .

3.2.3 The process involved in CNN-ALSTM-based classification

At this stage, the CNN-ALSTM model is employed for PD classification. The CNN-ALSTM is a hybrid DL approach for extracting features in the raw information and implementing predicting utilizing the LSTM-NN [26]. The CNN layer has been utilized for extracting the suitable features in the time series data, demonstrating extra hidden data has the potential for improving the forecast accuracy. The experimental outcomes illustrate that the CNN layer comprises one $16 \ 3 \times 1$ convolutional kernel layer and one $32 \ 3 \times 1$ convolutional kernel layer, which optimizes the forecast efficiency. The feature vector attained in the secondary layer of CNN is input to the LSTM layer to forecast. All the elements of feature vectors are similar to most 32 units from the LSTM layer. The attention process sets the superior weight to feature quantity, undoubtedly associated with the present output. Eventually, the FC layer managed the resultant vector of the attention process utilizing the unfolding function. The forecasted value of AC2 at the following moment was the outcome. The LSTM is well suited to forecast experimental time series data. The recent mechanism depicts the maximum predicting efficiency relating CNN and LSTM to distinct applications. The CNN uses LSTM for optimum extracting the features of experimental data. The attention method is a procedure for allocating weight. Inverse normalized prediction power was attained based on Eq. (13).

$$\text{Pr}_P = \text{Pr}_{Iac2} * (P_{\max} - P_{\min}) + P_{\min}, \tag{13}$$

where Pr_P refers to the forecasted value of powers and Pr_{Iac2} signifies the forecasted value of AC2.

The presence of LSTM cell infrastructure efficiently solves the gradient explosion or vanishing issues. There are four essential components from the flowchart of the LSTM technique: cell status, output, input, and forget gates. Those gates were utilized to control the upgrading, maintaining, and deleting of data from cell status. The forward computation procedure is referred to as:

$$\begin{aligned} f_z &= \sigma(W_f \cdot [h_{z-1}, x_z] + b_f), \\ i_z &= \sigma(W_j \cdot [h_{z-1}, x_z] + b_i), \\ O_z &= \sigma(W_o \cdot [h_{z-1}, x_z] + b_o), \\ \tilde{C}_z &= \tanh(W_C \cdot [h_{z-1}, x_z] + b_c), \\ C_z &= f_z \cdot C_{z-1} + i_z \cdot \tilde{C}_z \\ h_z &= O_z \cdot \tanh(C_z), \end{aligned} \tag{14}$$

, where W_f , W_j , and W_o refer to the weight matrix of forgetting, input, and output gates correspondingly; b_f , b_j , and b_o signifies the offset item of forget, input, and output gates correspondingly; σ signifies the sigmoid activation functions; \tanh denotes the hyperbolic tangent activation functions.

The attention process is a brain signal-processing method peculiar to human vision. It rapidly scans the global image to obtain the destination region, which requires attention and ignores other regions of unnecessary data. The attention process technique was effectively executed and implemented to train the model and other connected areas. The proposed model utilizes the LSTM hidden neuron resultant vector $H = \{h_1, h_2, \dots, h_t\}$ as input of the attention process, and the attention process will determine the attention weight α_i of h_i that is computed as shown in Eq. (15).

$$\begin{aligned} e_i &= \tanh(W_h h_i + b_h), \\ \alpha_i &= \frac{\exp(e_i)}{\sum_{i=1} \exp(e_i)}, \end{aligned} \tag{15}$$

whereas α_i signifies attention to weight, W_h refers to the weight matrix of h_j , and b_h represents the bias.

3.3 Hyperparameter tuning

For optimally tuning the hyperparameters of the CNN-ALSTM model, the Nadam optimizer is used. Nadam is an extended version of Adam optimizer [29], which can be applied to optimize the efficiency of the DL approaches. The upgrading rules of the Adam optimizer can be attained using the following equations:

$$g_t = \nabla_{\theta} J(\theta_t) \tag{16}$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{17}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{18}$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{19}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{20}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}} + \varepsilon} \hat{m}_t \quad (21)$$

, where g_t indicates the gradient vector of the CNN-ALSTM model at the time of training; η denotes the learning rate of the CNN-ALSTM model training; $J(\theta_t)$ is the divider function of the CNN in the CNN-ALSTM model; ∇_{θ_t} is the partial derivative of $J(\theta_t)$ and θ , m_t and v_t denotes 1st and 2nd order moment of the gradient at the time of training the CNN-ALSTM model; m_t and \hat{v} represents the deviation corrections of m_t and v_t , that can be utilized for offsetting the variation; β_1 and β_2 indicate exponential decay rate of m_t and v_t , ε is the correction variable used for ensuring that the denominator is not zero; t represents the number of iterations involved in the training process of the CNN-ALSTM model. Utilizing Eq. (17) into Eqs. (19) and (21) provides,

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}} + \varepsilon} \left(\frac{\beta_1 m_{t-1}}{1 - \beta_1^t} + \frac{(1 - \beta_1) g_t}{1 - \beta_1^t} \right) \quad (22)$$

The $m_{t-1}/1 - \beta_1^t$ presents the deviation correction estimation of the momentum vector at an earlier moment of the CNN-ALSTM model that can be attained by substituting m_{t-1} with:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}} + \varepsilon} \left(\beta_1 \hat{m}_{t-1} + \frac{(1 - \beta_1) g_t}{1 - \beta_1^t} \right) \quad (23)$$

With the addition of the Nesterov momentum, the deviation correction estimation m_t of the present momentum vector of the CNN-ALSTM model is straightaway utilized for replacing the deviation corrected estimates m_{t-1} of the earlier momentum that results in the updating rule of the Nadam, as provided below.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}} + \varepsilon} \left(\beta_1 \hat{m} + \frac{(1 - \beta_1) g_t}{1 - \beta_1^t} \right) \quad (24)$$

The conventional momentum approach has the demerit that the learning rate remains the same in the training procedure and utilizes an individual learning rate for updating weights.

4 Experimental validation

The performance validation of the QMFOFS-HCNN technique uses four benchmark datasets: HandPD meander, HandPD spiral, voice PD, and speech PD [30] using various evaluation metrics. The metrics used for performance evaluation are accuracy, detection, and false alarm rate. The accuracy rate is defined as the proportion of observations that have been correctly classified. A detection rate is defined as an outcome where the model correctly predicts the positive class. It measures the percentage of actual

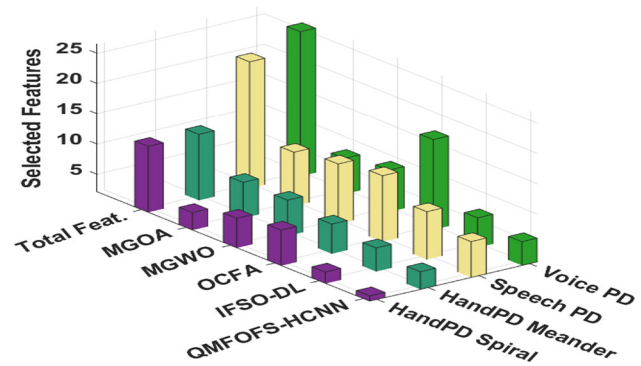


Fig. 3 FS analysis of QMFOFS-HCNN technique

positives that are correctly identified. A false alarm rate (FAR) is defined as an outcome where the model incorrectly predicts the positive class.

Figure 3 shows the FS results of the QMFOFS-HCNN system with existing methods on four data sets. The results showed that the QMFOFS-HCNN technique has shown an effectual outcome by selecting the least number of features. For instance, under HandPD Spiral dataset, the QMFOFS-HCNN technique has elected three features while the modified grasshopper optimization algorithm (MGOA) [31], modified grey wolf optimizer (MGWO) [32], optimized cuttlefish algorithm (OCFA) [30], and improved sailfish optimization algorithm with deep learning (IFSO-DL) [33] systems have selected 5, 7, 8, and 4 features correspondingly. Similarly, under the voice PD dataset, the QMFOFS-HCNN technique has picked six features, while the MGOA, MGWO, OCFA, and IFSO-DL systems have elected 8, 9, 17, and 7 features, respectively.

Table 1 demonstrates the comparative PD detection analysis of the QMFOFS-HCNN system with existing approaches on the HandPD spiral and HandPD Meander dataset [31, 33].

Figure 4 exhibits the comparative acc_y analysis of the QMFOFS-HCNN system with existing techniques on HandPD spiral and HandPD Meander datasets. The results show that the QMFOFS-HCNN technique has accomplished enhanced classification outcomes with higher accuracy than the other techniques on both datasets. For instance, on HandPD spiral datasets, the QMFOFS-HCNN technique has reached to maximum acc_y of 96.35%, whereas the MGOA-KNeN, MGOA-RANDF, MGOA-DT (C4.5), MGWO-KNeN, MGWO-RANDF, MGWO-DT (C4.5), and IFSO-DL techniques have obtained minimum acc_y values of 75.54%, 92.62%, 89.88%, 74.13%, 92.62%, 92.03%, and 93.61%, respectively.

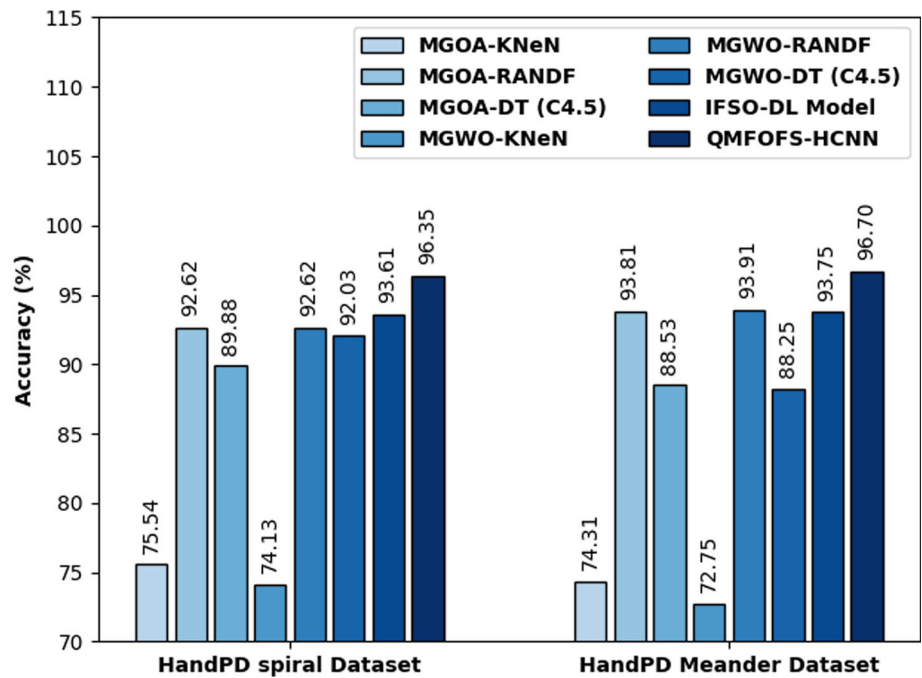
Figure 5 demonstrates the comparison study of the QMFOFS-HCNN technique with recent models in terms of detection rate d_{rate} on HandPD spiral and HandPD Meander datasets. The experimental values indicated that the

Table 1 Results analysis of existing with proposed model on HandPD spiral dataset and HandPD meander dataset

Methods	Accuracy	Detection rate	False alarm rate (FAR)
HandPD spiral dataset			
MGOA-KNeN	75.54	84.89	24.46
MGOA-RANDF	92.62	97.99	08.72
MGOA-DT (C4.5)	89.88	95.58	10.22
MGWO-KNeN	74.13	82.54	25.87
MGWO-RANDF	92.62	94.99	08.07
MGWO-DT (C4.5)	92.03	93.65	08.52
IFSO-DL Model	93.61	98.04	07.96
QMFOFS-HCNN	96.35	99.22	04.12
HandPD meander dataset			
MGOA-KNeN	74.31	86.34	25.69
MGOA-RANDF	93.81	99.64	06.19
MGOA-DT (C4.5)	88.53	91.01	11.47
MGWO-KNeN	72.75	85.04	27.25
MGWO-RANDF	93.91	98.35	06.19
MGWO-DT (C4.5)	88.25	91.39	11.71
IFSO-DL Model	93.75	99.71	07.25
QMFOFS-HCNN	96.70	99.89	09.11

Bold indicates the output of our method

Fig. 4 Accy analysis of QMFOFS-HCNN technique under HandPD spiral and HandPD meander datasets



QMFOFS-HCNN system has demonstrated improved classifier results with the maximum d_{rate} values over the other techniques on both datasets. For instance, on HandPD spiral dataset, the QMFOFS-HCNN technique has offered increased d_{rate} of 99.22%, whereas the MGOA-KNeN, MGOA-RANDF, MGOA-DT (C4.5), MGWO-KNeN, MGWO-RANDF, MGWO-DT (C4.5), and IFSO-DL techniques have resulted in reduced d_{rate} values of 84.89%,

97.99%, 95.58%, 82.54%, 94.99%, 93.65%, and 98.04%, respectively.

Figure 6 provides the accuracy and loss graph analysis of the QMFOFS-HCNN system under HandPD spiral and HandPD meander datasets. The outcomes shown that the accuracy value tends to be higher, and the loss value tends to decrease with an increase in epoch count. It is also observed that the training loss is low, and validation

Fig. 5 Detection rate analysis of QMFOFS-HCNN technique under HandPD spiral and HandPD meander datasets

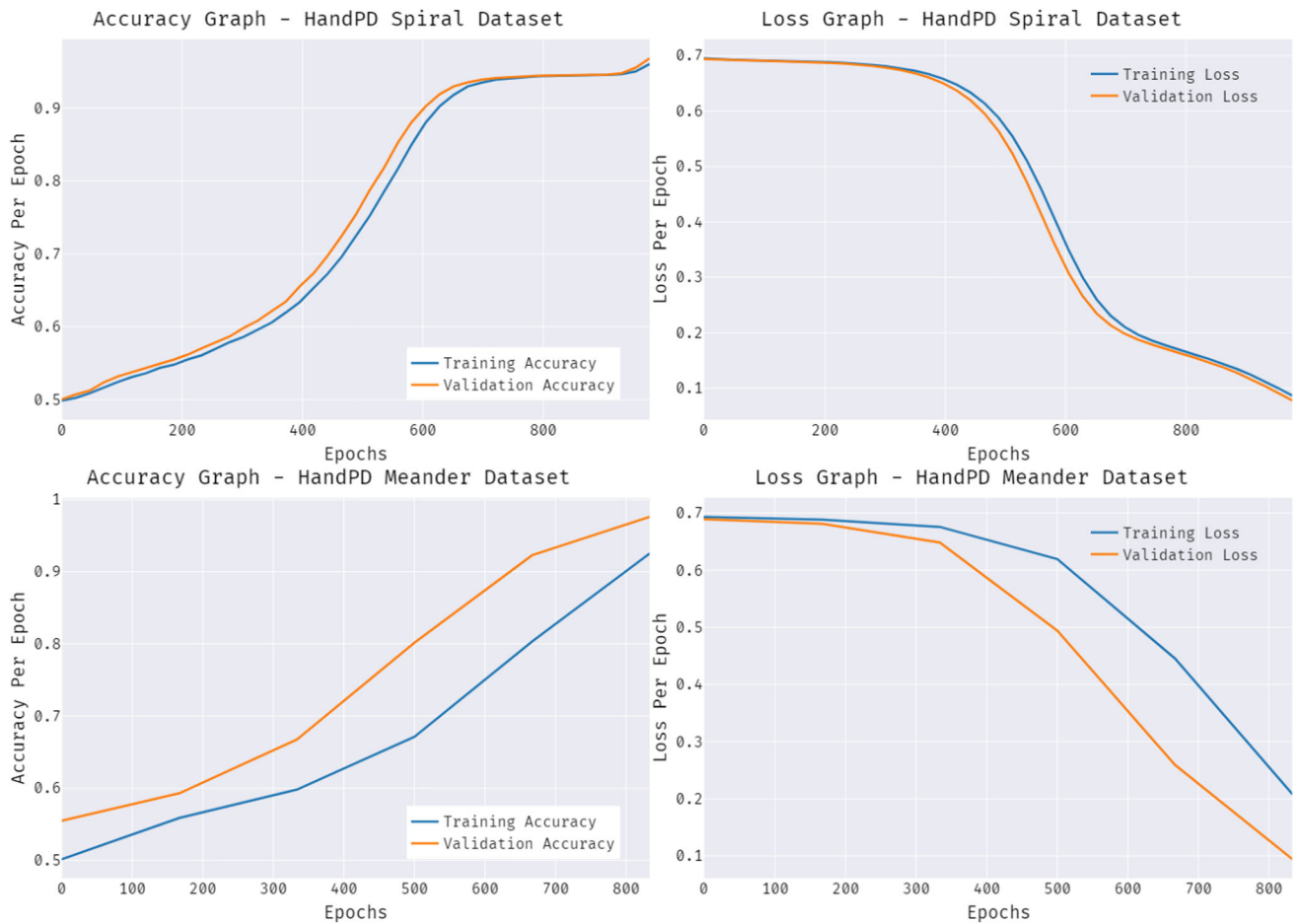
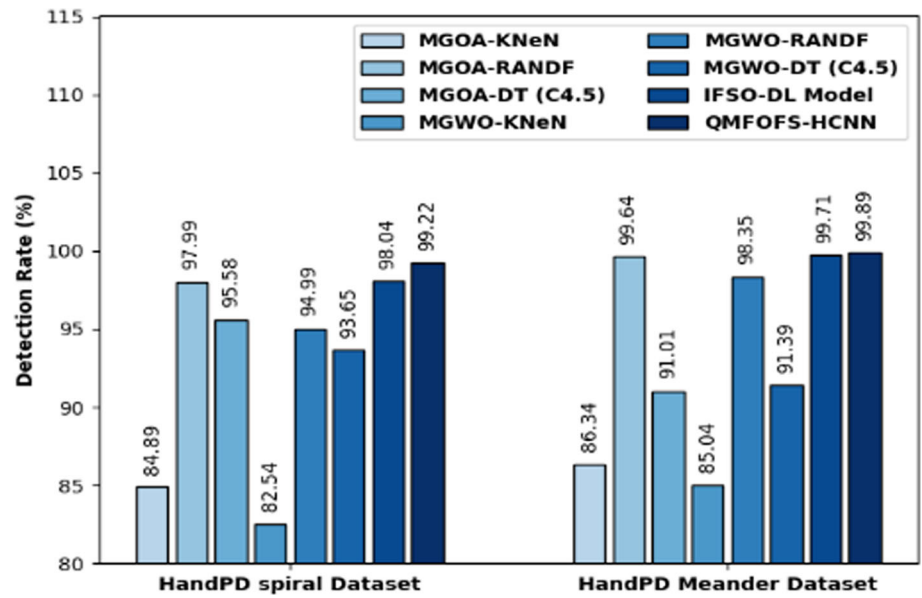


Fig. 6 Accuracy and Loss analysis of QMFOFS-HCNN method under HandPD spiral and HandPD meander datasets

Table 2 Results analysis of existing with proposed model on speech PD dataset and voice PD dataset

Methods	Accuracy	Detection rate	False alarm rate
Speech PD dataset			
MGOA-KNeN	89.69	96.56	31.00
MGOA-RANDF	95.56	99.16	21.52
MGOA-DT (C4.5)	85.53	90.17	29.76
MGWO-KNeN	92.35	97.21	30.52
MGWO-RANDF	93.64	99.95	30.57
MGWO-DT (C4.5)	90.18	94.28	30.25
IFSO-DL Model	96.19	99.16	18.32
QMFOFS-HCNN	98.50	99.98	12.45
Voice PD dataset			
MGOA-KNeN	91.70	83.50	00.89
MGOA-RANDF	100	100	–
MGOA-DT (C4.5)	100	100	–
MGWO-KNeN	85.70	80.20	08.20
MGWO-RANDF	100	100	–
MGWO-DT (C4.5)	100	100	–
IFSO-DL Model	100	100	–
QMFOFS-HCNN	100	100	–

Bold indicates the output of our method

accuracy is maximum on HandPD spiral and HandPD meander datasets. Table 2 demonstrates the comparative PD detection result analysis of the QMFOFS-HCNN technique with existing approaches on the speech PD and

voice datasets. Figure 7 depicts the comparative accuracy analysis of the QMFOFS-HCNN technique with existing methods on speech PD and voice PD datasets. The results showed that the QMFOFS-HCNN system has accomplished enhanced classification outcomes with higher accuracy than the other techniques on both datasets. For instance, on the speech PD dataset, the QMFOFS-HCNN technique has reached to maximum accuracy of 98.50%, whereas the MGOA-KNeN, MGOA-RANDF, MGOA-DT (C4.5), MGWO-KNeN, MGWO-RANDF, MGWO-DT (C4.5), and IFSO-DL approaches have obtained lesser accuracy values of 89.69%, 95.56%, 85.53%, 92.35%, 93.64%, 90.18%, and 96.19%, correspondingly.

Figure 8 examines the comparison study of the QMFOFS-HCNN approach with recent models in terms of detection rate d_{rate} on speech PD and voice PD datasets. The experimental values indicated that the QMFOFS-HCNN system had outperformed higher classifier results with higher d_{rate} values over the other techniques on both datasets. For instance, on speech PD dataset, the QMFOFS-HCNN approach has offered increased d_{rate} of 99.98%, whereas the MGOA-KNeN, MGOA-RANDF, MGOA-DT (C4.5), MGWO-KNeN, MGWO-RANDF, MGWO-DT (C4.5), and IFSO-DL systems have resulted in lower d_{rate} values of 96.56%, 90.17%, 97.21%, 99.95%, 94.28%, 99.16%, and 99.98%, correspondingly.

Figure 9 offers the accuracy and loss graph analysis of the QMFOFS-HCNN methodology under speech PD and voice PD Datasets. The outcomes outperformed that the accuracy value tends to increase, and the loss value tends to

Fig. 7 Acc. analysis of QMFOFS-HCNN technique under speech PD and voice PD datasets

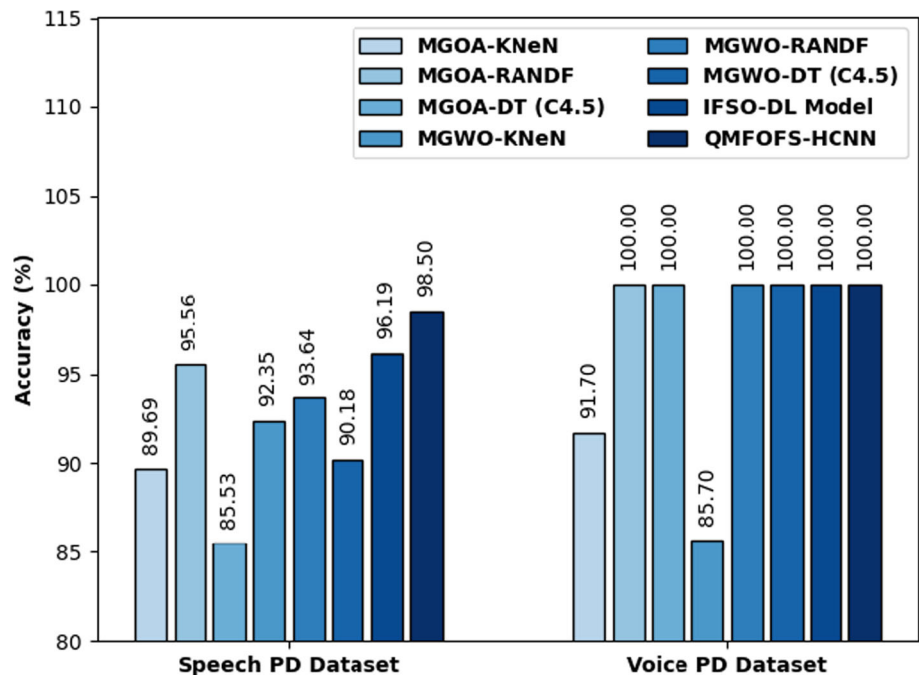


Fig. 8 Detection rate analysis of QMFOFS-HCNN technique under speech PD and voice PD datasets

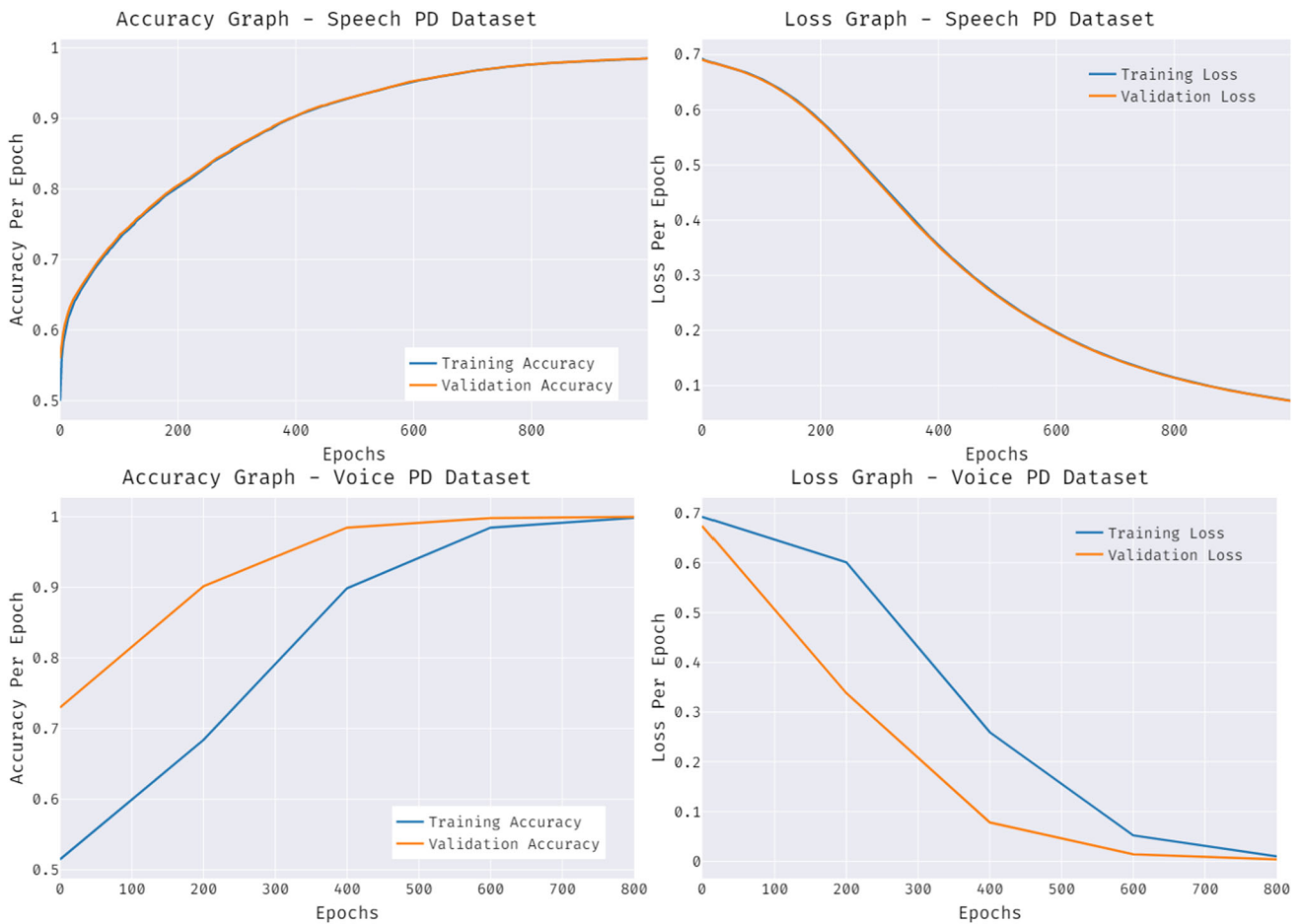
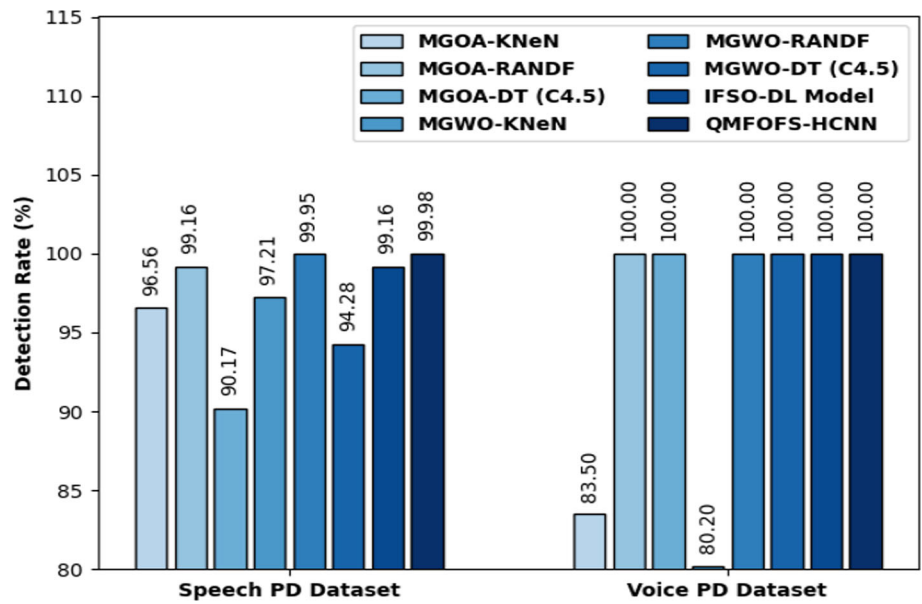


Fig. 9 Accuracy and Loss analysis of QMFOFS-HCNN technique under speech PD and voice PD datasets

reduce with a higher epoch count. It can also be observed that the training loss is lesser, and validation accuracy is high on speech PD and voice PD Datasets. From these results, it is ensured that the proposed model is superior to other methods of PD classification.

The study compared the QMFOFS-HCNN technique to existing methods for Parkinson's disease (PD) detection using four benchmark datasets. The QMFOFS-HCNN technique demonstrated several strengths: QMFOFS-HCNN selected fewer features while maintaining or improving classification performance, reducing data dimensionality, thus, providing Efficient Feature Selection. It consistently outperformed existing methods in accuracy, enhancing PD patient classification and, thus, high accuracy. It achieved higher detection rates, crucial for accurate PD diagnosis, and exhibited lower FAR, reducing the risk of misdiagnosis. It consistently outperformed existing methods across various datasets, demonstrating its versatility. The model showed increasing accuracy and decreasing loss during training, indicating effective learning.

5 Conclusion

This study has developed a novel QMFOFS-HCNN method for detecting and classifying PD. It aimed to identify the optimal feature subsets and enhance the classification accuracy of the PD diagnosis. The proposed QMFOFS-HCNN technique encompasses several processes, such as QMFO-based feature selection, CNN-ALSTM based classification, and Nadam-based hyperparameter tuning. Using QMFOFS classify Nadam techniques helps to boost the PD classification outcomes effectively. The experimental validation takes place using the benchmark datasets, and the results are assessed under several aspects. The comparative results indicated the QMFOFS-HCNN technique's promising performance in several evaluation metrics. Therefore, the QMFOFS-HCNN technique can be utilized as a proficient tool for PD detection and classification. It offers a proficient PD detection and classification tool, contributing to medical diagnostics. However, it is important to acknowledge some limitations of this study. Firstly, the performance evaluation was conducted on benchmark datasets, and the real-world applicability of the technique may require further validation with diverse and more extensive datasets. Secondly, while the QMFOFS-HCNN method shows promise, it may benefit from additional optimization and fine-tuning to achieve even higher accuracy levels.

In the future, outlier detection techniques can be incorporated into the QMFOFS-HCNN technique to improve the classifier results and robustness. Additionally, exploring

the integration of real time data collection and analysis for PD diagnosis could improve the practicality and timeliness of the method. Overall, this study lays the foundation for more advanced and effective PD diagnostic tools, and further refinement and validation in clinical settings will be essential for its successful implementation.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Data availability The dataset used in this study is publicly available via the following link: <https://www.fc.unesp.br/~papa/pub/datasets/Handpd/>.

Declarations

Conflict of interest The author declares that I have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Ali L, Zhu C, Golilarz NA, Javeed A, Zhou M, Liu Y (2019) Reliable Parkinson's disease detection by analyzing handwritten drawings: construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model. *Ieee Access* 7:116480–116489
2. Pahuja G, Nagabhushan TN (2021) A comparative study of existing machine learning approaches for Parkinson's disease detection. *IETE J Res* 67(1):4–14
3. Zahid L, Maqsood M, Durrani MY, Bakhtyar M, Baber J, Jamal H, Mehmood I, Song OY (2020) A spectrogram-based deep feature assisted computer-aided diagnostic system for Parkinson's disease. *IEEE Access* 8:35482–35495
4. Senturk ZK (2020) Early diagnosis of Parkinson's disease using machine learning algorithms. *Med Hypotheses* 138:109603
5. Wang W, Lee J, Harrou F, Sun Y (2020) Early detection of Parkinson's disease using deep learning and machine learning. *IEEE Access* 8:147635–147646
6. Wroge TJ, Özkanca Y, Demiroglu C, Si D, Atkins DC, Ghomi RH (2018) Parkinson's disease diagnosis using machine learning and voice. In: 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), pp 1–7. IEEE.

7. Lahmiri S, Dawson DA, Shmuel A (2018) Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures. *Biomed Eng Lett* 8(1):29–39
8. Rovini E, Maremmanni C, Moschetti A, Esposito D, Cavallo F (2018) Comparative motor pre-clinical assessment in Parkinson's disease using supervised machine learning approaches. *Ann Biomed Eng* 46(12):2057–2068
9. Belić M, Bobić V, Badža M, Šolaja N, Đurić-Jovičić M, Kostić VS (2019) Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease—a review. *Clin Neurol Neurosurg* 184:105442
10. Ouhmida A, Raihani A, Cherradi B, Terrada O (2021) A novel approach for Parkinson's disease detection based on voice classification and features selection techniques. *Int J Online & Biomed Eng* 17(10):111
11. Johri A, Tripathi A (2019) Parkinson disease detection using deep neural networks. In: 2019 Twelfth International Conference on Contemporary Computing (IC3), pp 1–4. IEEE.
12. El Maachi I, Bilodeau GA, Bouachir W (2020) Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait. *Expert Syst Appl* 143:113075
13. Jayaram R, Senthil Kumar T (2022) Cloud-based Parkinson disease prediction system using expanded cat swarm optimization. *IoT and analytics for sensor networks*. Springer, Singapore, pp 299–309
14. Solana-Lavalle G, Galán-Hernández JC, Rosas-Romero R (2020) Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features. *Biocybern Biomed Eng* 40(1):505–516
15. Mathur R, Pathak V, Bandil D (2019) Parkinson disease prediction using machine learning algorithm. *Emerging trends in expert applications and security*. Springer, Singapore, pp 357–363
16. Haq AU, Li J, Memon MH, Khan J, Din SU, Ahad I, Sun R, Lai Z, (2018) Comparative analysis of the classification performance of machine learning classifiers and deep neural network classifier for prediction of Parkinson disease. In: 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), pp 101–106. IEEE.
17. Zhang T, Zhang Y, Sun H, Shan H (2021) Parkinson disease detection using energy direction features based on EMD from voice signal. *Biocybern Biomed Eng* 41(1):127–141
18. Vidya B, Sasikumar P (2022) Parkinson's disease diagnosis and stage prediction based on gait signal analysis using EMD and CNN–LSTM network. *Eng Appl Artif Intell* 114:105099
19. Klaar ACR, Stefenon SF, Seman LO, Mariani VC, Coelho LDS (2023) Optimized EWT-Seq2Seq-LSTM with attention mechanism to insulators fault prediction. *Sensors* 23(6):3202
20. Borré A, Seman LO, Camponogara E, Stefenon SF, Mariani VC, Coelho LDS (2023) Machine fault detection using a hybrid CNN–LSTM attention-based model. *Sensors* 23(9):4512
21. Balaji E, Brindha D, Elumalai VK, Vikrama R (2021) Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network. *Appl Soft Comput* 108:107463
22. Koundal D, Jain DK, Guo Y, Ashour AS, Zaguia A (Eds.) (2023) *Data analysis for neurodegenerative disorders*. Springer Nature.
23. Oktay AB, Kocer A (2020) Differential diagnosis of Parkinson and essential tremor with convolutional LSTM networks. *Biomed Signal Process Control* 56:101683
24. Khan YF, Kaushik B, Koundal D (2023) Machine learning models for alzheimer's disease detection using medical images. *Data analysis for neurodegenerative disorders*. Singapore Springer Nature, Singapore, pp 165–182
25. Wang D, Chen H, Li T, Wan J, Huang Y (2020) A novel quantum grasshopper optimization algorithm for feature selection. *Int J Approx Reason* 127:33–53
26. Sethi M, Ahuja S, Rani S, Koundal D, Zaguia A, Enbeyle W (2022) An Exploration: alzheimer's disease classification based on convolutional neural network. *BioMed Res Int*. <https://doi.org/10.1155/2022/8739960>
27. Zervoudakis K, Tsafarakis S (2020) A mayfly optimization algorithm. *Comput Ind Eng* 145:106559
28. Shaheen MA, Hasanien HM, El Moursi MS, El-Fergany AA (2021) Precise modeling of PEM fuel cell using improved chaotic MayFly optimization algorithm. *Int J Energy Res* 45(13):18754–18769
29. Bera S, Shrivastava VK (2020) Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification. *Int J Remote Sens* 41(7):2664–2683
30. Gupta D, Julka A, Jain S et al (2018) Optimized cuttlefish algorithm for diagnosis of Parkinson's disease. *Cogn Syst Res* 52:36–48
31. Sehgal S, Agarwal M, Gupta D, Sundaram S, Bashambu A (2020) Optimized grass hopper algorithm for diagnosis of Parkinson's disease. *SN Appl Sci* 2(6):1–18
32. Sharma P, Sundaram S, Sharma M, Sharma A, Gupta D (2019) Diagnosis of Parkinson's disease using modified grey wolf optimization. *Cognit Syst Res* 54:100–115
33. Zhang Y, Mo Y (2022) Chaotic adaptive sailfish optimizer with genetic characteristics for global optimization. *J Supercomput* 78:10950–10996. <https://doi.org/10.1007/s11227-021-04255-9>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.