IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 December 2024, accepted 14 December 2024, date of publication 19 December 2024, date of current version 6 January 2025.

Digital Object Identifier 10.1109/ACCESS.2024.3520482

RESEARCH ARTICLE

An Optimized Hybrid PSO-ELM for Parkinson's Disease Diagnosis

GUNAKALA ARCHANA AND AFZAL HUSSAIN SHAHID
School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh 522237, India
Corresponding author: Afzal Hussain Shahid (syedahshahid@gmail.com)

This work was supported by VIT-AP University, Amaravati, Andhra Pradesh.

ABSTRACT Parkinson's disease (PD) is a neurodegenerative disease that gradually causes movement impairment and various symptoms. It is difficult to precisely diagnose PD, especially in its early stages, because the signs and symptoms can resemble those of other medical conditions or normal age-related changes. This paper proposes a hybrid model combining Particle Swarm Optimization with Extreme Learning Machine (PSO-ELM) for PD diagnosis. This paper employs three feature ranking algorithms, namely ReliefF, minimum Redundancy Maximum relevance (mRMR), and Fisher, on six publicly available PD datasets. Various top-ranked feature subsets are created to identify the most discriminative features and enhance the performance of the proposed hybrid PSO-ELM model for all datasets. Furthermore, the efficiency of the proposed model is compared with basic models, namely ELM, SVM, RF, and the previous works. The results show that the proposed PSO-ELM model achieved the highest average accuracy, recall, precision, and F1-score of 100% each over the 3-fold cross-validation with a minimum number of features for all the six datasets. Therefore, the PSO-ELM model may be used as a highly accurate and efficient tool for PD diagnosis.

INDEX TERMS Classification, extreme learning machine, feature ranking, Parkinson's disease, particle swarm optimization.

I. INTRODUCTION

PD is a progressive neurological disorder that results in the loss of brain cells, leading to motor abnormalities [1]. PD is the most common neurological disease after Alzheimer's disease. PD predominantly affects older individuals, primarily striking those aged 60 years and above [2]. PD is characterized by two neuropathological elements: depletion of dopaminergic neurons in the substantia nigra pars compacta (an area in the midbrain, primarily in the basal ganglia) and the presence of cytoplasmic particles called Lewy bodies [3]. As the disease progresses, the nerve cells gradually lose their communication ability, leading to various nervous system disorders. The symptoms typically tend to develop gradually, and common early-stage signs include bradykinesia (slowed movement), tremors, muscle rigidity, impaired balance and posture, speech alterations, and changes in handwriting [4] The disease typically receives a diagnosis during its advanced stages, which are characterized by the degeneration of approximately 60% of neurons.

Consequently, initiating therapy at this point yields limited success in halting disease progression. Nevertheless, the current diagnostic rate still needs improvement due to overlapping symptoms with other pathologies [5]. Detecting the disease early, combined with appropriate medication, can effectively alleviate tremors and balance issues in patients, allowing them to lead a relatively better life.

Traditional diagnostic methods for PD typically include clinical assessments, neurological examinations, and neuroimaging. Although these methods are essential for diagnosis, they often lack sensitivity in detecting early-stage PD and may not effectively distinguish PD from similar disorders. Researchers have, therefore, explored various machine learning (ML) techniques to improve PD diagnosis. According to research [6], vocal cord impairment is an early-stage symptom in 90% of individuals with PD. In the initial phases

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

of PD, subtle voice abnormalities may exist, which might not be observable by the listeners. However, these can be assessed by analyzing voice signal acoustics [7]. In addition to speech patterns, how individuals write and draw can also be used to discern the disease. Examining the figures drawn by patients and healthy individuals can identify potential signs of the condition. In the early stages of PD, handwriting movements are affected by their kinematic aspects, such as size, speed, acceleration, and stroke length [8]. Therefore, this paper uses handwritten and speech data to diagnose PD. ML algorithms and their optimizations are currently being used to diagnose PD. Thus, employing different ML algorithms to identify an effective approach for analyzing voice and handwritten data related to PD becomes essential.

A. MOTIVATION

Several reasons have motivated this work. First, PD is the second largest neurological disease that affects humans with an increase in age. An early diagnosis can enhance the quality of life by halting the disease's progression or causes. Second, an efficient ML model with faster learning capability should be employed to diagnose PD with good performance. Third, a comprehensive analysis is required to determine the significance of various features in diagnosing PD.

B. CONTRIBUTIONS

The primary contributions of this paper are: (i) A hybrid PSO-ELM model is proposed to diagnose PD that learns extremely fast. (ii)The parameters of the PSO algorithm are optimized using the grid search method. (iii) To check for further improvement in the performance of the proposed hybrid PSO-ELM model, the number of neurons in the hidden layer is varied. (iv) Three well-known feature ranking algorithms, ReliefF, mRMR, and Fisher, are used to find the most discriminative features from six PD datasets. (v) A comprehensive analysis of PD diagnosis is done by creating 126 different feature subsets to determine the significance of various features in diagnosing PD. (vi) The proposed PSO-ELM model is compared to various basic models (ELM, SVM, and RF) and previous works.

An outline of this paper is given below: Section II reviews earlier research on diagnosing PD with handwritten and speech data using ML algorithms. Section III covers the methods and details of the dataset used in this work. Section IV discusses the results. Section V presents the conclusion.

II. LITERATURE REVIEW

PD is a severe global health concern that requires early and accurate diagnostic techniques. Researchers have diagnosed PD by using various types of datasets.

A. SPEECH BASED DIAGNOSIS

In [9], the authors investigated a vowel-centric artificial neural network (ANN) model for individual vowel phonation, showcasing its superiority with 91% accuracy, 99% sensitivity, 82% specificity, and an AUC of 91% compared to other models. Even though high sensitivity indicates a precise diagnosis of most PD patients, the focus on vowel phonation limits the model's ability to capture more complex speech patterns, such as prosody or articulation, which are better assessed using sentences [10]. In [11], the authors used various fuzzy classification methods—an Inductive Fuzzy Classifier and a Fuzzy Rough Classifier—to diagnose PD. Although fuzzy systems are flexible and helpful in handling uncertainty, they require the knowledge of an expert to define appropriate membership functions and rules for classifying the data. The authors of [12] introduced a multi-modal framework using various ML models, with LR and Linear SVM achieving the highest accuracy of 70% on speech data, emphasizing the value of vowel samples for PD diagnosis. Despite using multiple models, the focus on conventional methods limits the ability to capture more complex patterns in the data. Advanced models like deep learning or hybrid approaches may improve performance by identifying more discriminative features. In [13], the authors proposed a weighted local discriminant preservation projection embedded ensemble with ELM, using 5000 neurons, within their multi-modal framework, achieving 73.75% accuracy with linear SVM and 75.63% with ELM. ELM has a faster learning capacity. The large number of neurons improves the model's capacity to capture patterns and increases the risk of overfitting on small datasets.

The authors of [14] have focused on sample optimization by adopting a direct linear transformation and clustering using a multitype ensemble transformation algorithm (META). A joint structure consistency mechanism is used to combine the results of multiple classifiers. The experiment's results achieved an accuracy of 91.25% using the META With RF model. The proposed method is applied only to the speech data and, therefore, needs to be verified with different PD datasets. Different Feature selection techniques are used on the SpeechPD dataset in [15]. The features obtained are then sent to Classification and Regression Trees, ANN, and SVM to classify PD, and the result showed that SVM + RFE with seven features has achieved an accuracy of 93.84%. In [16] The authors proposed a Binary Improved Grey Wolf Optimizer (BIGWO) to diagnose PD and achieved an accuracy of 95.66% on the SpeechPD dataset. The model's performance heavily depends on selecting the optimal number of nearest neighbors (k) in the fitness function, so its robustness may be verified with different datasets.

The authors in [17] proposed an improvement to the smallest normalized difference associative memory algorithm to diagnose PD using SpeechPD and PMSR datasets and achieved an accuracy of 99.48% and 99.66%, respectively. In [18] PMSR dataset. Although the model outperforms conventional classifiers, more datasets are needed to validate its performance. Further testing is required to assess generalizability across varied PD datasets.

B. HANDWRITING BASED DIAGNOSIS

In most PD cases, the primary indicators typically involve movement-related challenges, specifically affecting tasks like writing and drawing. The research [19] uses three different handwritten data and proposes a fuzzy classifier method using metaheuristic algorithms in three stages, which involves generating the structure, selecting important features, and optimizing parameters. The authors extracted five features from the New Meander and New Spiral data by performing various statistical tests and achieved an accuracy of 79.14% and 78.14%, respectively. The authors used only handwritten data, which captures a subset of PD symptoms.

The authors in [20] optimized the convolutional neural network (CNN) with Harris Hawk's optimization (HHO) algorithm and achieved an accuracy of 94.2% and F1-Score of 94.11%. The authors focused only on optimizing hyperparameters of CNN and did not consider different CNN architectures, which may improve the results. Also, the traditional HHO algorithm has limited exploration ability as it gets completely exhausted when the escape energy is zero. It may suffer from local optima problems that may cause immature convergence during the exploration and exploitation phases [21]. In [22] the authors have utilized VGG16+ BGWO for feature extraction using the Meander, spiral, and circle drawings of the New HandPD dataset. The authors classified PD using the linear SVM and obtained 87.42%, 95.1%, and 96.2% accuracy with circle, meander, and spiral datasets.

In [23], the authors proposed a cascade ensemble of RF and two extremely random trees to diagnose PD using handwritten data. The dimensionality reduction was done using Principal component analysis (PCA). 77.38% and 80.99% accuracy is achieved with the Meander and Spiral datasets, respectively. The model may be optimized to achieve better performance. Also, the efficiency may be verified with other PD datasets. In the paper [24], the authors used different CNN models to extract features from handwritten spiral images and applied fusion techniques to extract the final features. The combined vectors of the extracted features are then sent to a linear SVM classifier, and an accuracy of 99.35% is obtained. In [25], the authors ensembled RF and PCA to classify healthy patients from PD patients, applied stratified k-fold cross-validation, and achieved 89.4% accuracy. The authors of [26] used Deep learning methods (AlexNet, VGG Net, GoogleNet, and Resnet) to diagnose PD using Meander, Spiral, New Meander, and New Spiral drawings of datasets. They achieved an accuracy of 89.19%, 86.49%, 88.46%, and 90.41% respectively. The authors of [27] used CNN architecture to diagnose PD. They achieved 88.6% accuracy on Meander data and 90.5% on spiral data. The authors of [4] used modified grey wolf optimization for feature selection with KNN, RF, and DT classifiers on four different datasets Meander, Spiral, SpeechPD, and PMSR. They achieved an accuracy of 93.04%, 92.41%, 93.87%, and 100%, respectively. In [28] the authors have used Quantum Mayfly Optimization-based feature subset selection with hybrid CNN with attention-based long short-term memory to diagnose PD using the four datasets Meander, Spiral, SpeechPD, and PMSR and achieved an accuracy of 96.7%,96.35%, 98.5%, 100% respectively.

Most of the existing works [9], [13], [23], [25], [26] have focused on either the handwritten data or the voice data for diagnosing PD. This paper uses six datasets for PD classification, including handwritten data (Meander, Spiral, New Meander, New Spiral) and voice data (SpeechPD and PMSR). To overcome the disadvantages of the existing works, this study proposed an optimized hybrid PSO-ELM model. The integration of PSO helps optimize the parameters of ELM (hidden weights and biases), resulting in faster learning and better generalization. The proposed PSO-ELM model was rigorously tested on all six datasets, demonstrating its robustness and versatility.

III. MATERIALS AND METHODS

Section III-A presents the various datasets used in this paper. Section III-B describes the feature selection and feature ranking algorithms. Section III-C describes the theoretical background and model development. Section III-D discusses the proposed hybrid PSO-ELM model.

A. DATASET DESCRIPTION

The datasets utilized in this paper have been sourced from the publicly available UCI and Kaggle repositories. The following sections provide detailed descriptions of the characteristics of various datasets used.

1) HANDPD DATASET

The HandPD dataset consists of handwritten examinations collected from two distinct groups of individuals: (i) a Healthy Control Group and (ii) a Patient Group. The Patient Group consists of individuals who have been diagnosed with PD. The dataset comprises 92 individuals, categorized into two groups: the Healthy Group, which consists of 18 individuals, and the Patient Group, which includes 74 individuals.

The Healthy Group has six males and 12 females aged 19 to 79. The average age in this group is 44.22 years, with a standard deviation of 16.5 years. Among these individuals, 16 are right-handed, and two are left-handed. In the Patient Group, males and females aged 38 to 78 are 59 and 15, respectively. The average age for this group is 58.75 years, with a standard deviation of 7.51 years. Within this group, 69 are right-handed, and five are left-handed. Each individual is asked to draw a Spiral and a Meander.

a: MEANDER DATASET

Each individual has provided four meander drawings. Therefore, the Healthy Group comprises 72 meander samples (4 samples from each of the 18 individuals), while the Patient Group contains 296 meander samples (4 samples from each of the 74 individuals). The Healthy Group includes 24 samples from males and 48 samples from females. On the

other hand, the Patient Group includes 236 samples from males and 60 from females. Among these meander samples, ten individuals are left-handed in the Patient Group, and four are left-handed in the Healthy Group. Right-handed individuals contribute the remaining meander drawings in both groups. The dataset comprises 12 features and one target variable indicating a healthy or PD class. Detailed descriptions of extracted features from meander drawings are described in Table 1.

TABLE 1. Description of meander dataset.

S.No	Attributes	Description
1	CLASS_TYPE	1 = Healthy Control group, 2 = Patient Group
2	GENDER	Male or Female
3	RIGHT/LEFT-HANDED	Right or left-hand writing
4	AGE	Age
5	RMS	Root mean square
6	MAX_BETWEEN_ET_HT	Maximum and minimum difference between exam template and handwritten trace radius
7	MIN_BETWEEN_ET_HT	
8	STD_DEVIATION_ET_HT	The standard deviation of the difference between exam template and handwritten trace radius
9	MRT	Mean relative tremor
10	MAX_HT	Maximum and minimum handwritten trace radius
11	MIN_HT	
12	STD_HT	Standard deviation of the handwritten radius
13	CNTP	No. of times the difference between exam template and handwritten trace radius changes from -ve to +ve, or vice-versa.

+ve= Positive, -ve = negative, no. of = number of, CNTP = CHANGES_FROM_NEGATIVE_TO_POSITIVE_BETWEEN_ET_HT.

b: SPIRAL DATASET

Like the meander drawings, each individual in the HandPD dataset also provided four spiral drawings. Consequently, the Healthy Group comprises 72 spiral samples (4 samples from each of the 18 individuals), while the Patient Group contains 296 spiral samples (4 samples from each of the 74 individuals). The Healthy Group consists of 24 spiral images from males and 48 from females. In contrast, the Patient Group includes 236 spiral samples from males and 60 from females. The features of this dataset are similar to those of the meander dataset, as described in Table 1.

2) NEW HANDPD DATASET

Prior studies [23] predominantly concentrate on the HandPD dataset for diagnosing PD. The New HandPD dataset [29] is an enhanced and more balanced collection of handwritten dynamics data assembled by Botucatu Medical School at São Paulo State University, Brazil. This dataset encompasses 66 participants categorized into the patient and healthy control groups. The patient group comprises 31 individuals, while the healthy control group comprises 35. Within the patient group, ten are female, and 21 are male, while the healthy group consists of 18 males and 17 females. Each participant was asked for 12 handwriting assessments with a smart pen. These assessments encompass drawing Meanders, spirals and circles in the air, circles on paper, and left and right-handed diadochokinesis (involves the person holding a pen with their arms straight and executing hand-wrist movements. Drawings are not part of this test; instead, the pen records the signals generated by these movements [29]). This paper primarily centers on the Meander and Spiral assessments to discern PD patients.

a: NEW SPIRAL DATASET

Each participant was asked to draw four spirals based on the given template, and the features were extracted from the drawings. The extracted features are similar to the HandPD dataset.

b: NEW MEANDER DATASET

Each subject is asked to draw 4 Meanders based on the given template, and the features are extracted based on the drawings. The extracted features are similar to those in the HandPD dataset.

TABLE 2. Description of SpeechPD dataset.

S. No	Attributes	Description
1	MDVP: Fo (Hz)	Average, Maximum, and Minimum values of vocal fundamental frequency
2	MDVP: Fhi (Hz)	
3	MDVP: Flo (Hz)	
4	MDVP: Jitter (%)	Fundamental frequency variation measures
5	MDVP: Jitter (Abs)	
6	MDVP: RAP	
7	MDVP: PPQ	
8	Jitter: DDP	
9	MDVP: Shimmer	Amplitude variation measures
10	MDVP: Shimmer (dB)	
11	Shimmer: APQ3	
12	Shimmer: APQ5	
13	MDVP: APQ	
14	Shimmer: DDA	
15	NHR	ratio of noise to tonal components in the voice
16	HNR	
17	RPDE	Non-linear dynamical complexity measures
18	D2	
19	DFA	Signal fractal scaling exponent
20	spread1	Three nonlinear measures of fundamental frequency variation
21	spread2	
22	PPE	
23	status	(1) PD (0) healthy

3) SpeechPD DATASET

The collection of recorded speech signals was taken from Max Little [30], [31]. Feature description of the speechPD

dataset is given in Table 2. This specific dataset comprises a diverse range of acoustic speech features sourced from 195 individuals, with 147 of them having PD. Each feature within the dataset characterizes an individual's distinct vocal measures, and each data entry corresponds to the total count of voice recordings conducted on these individuals. The main focus of the dataset is to differentiate between people who are healthy and those who are impacted by the disease, which is indicated by the ''status'' column. For healthy individuals, the value is set to 0, whereas it's 1 for those diagnosed with the disease.

4) PMSR DATASET

The Parkinson's Multiple Speech recordings (PMSR) dataset was gathered by Sarkar et al. [32]. The training set comprises records from 40 subjects, equally divided between 20 PD patients (6 female, 14 male) and 20 healthy individuals (ten male, ten female). Trust MC-1500 microphones were positioned 15cm away from subjects to capture voice recordings. Each subject contributed 26 samples, starting with three vowels (a, o, and u), followed by numbers 1 to 10 (samples 4 to 13), short sentences (samples 14 to 17), and concluding with words (samples 18 to 26). Hence, the training dataset has 40*26=1040 samples. Praat software [33] facilitated the extraction of 26 dysphonic features from each sample, and a description of all the features is provided in Table 3. The test dataset, gathered under identical conditions by the same medical professionals, involves 28 PD patients. Each patient articulated the sustained vowels ('a' and 'o') thrice, leading to 168 (28*6) samples. The same set of 26 dysphonic features was extracted from each of these test samples. This paper merged 1040 training and 168 test data samples, resulting in 1208 samples.

TABLE 3. Description of PMSR dataset.

S. No	Attributes	Description
1	Jitter (local)	Frequency Parameters
2	Jitter (local, absolute)	
3	Jitter (rap)	
4	Jitter (ppq5)	
5	Jitter (ddp)	
6	Shimmer (local)	Amplitude Parameters
7	Shimmer (local, dB)	
8	Shimmer (apq3)	
9	Shimmer (apq5)	
10	Shimmer (apq11)	
11	Shimmer (dda)	
12	AC –autocorrelation	Harmonicity Parameters
13	NTH- Noise –to-harmonic	
14	HTN – Harmonic –to-noise	
15	Median pitch	Pitch Parameters
16	Mean pitch	
17	Standard deviation	
18	Minimum pitch	
19	Maximum pitch	
20	Number of pulses	Pulse Parameters
21	Number of periods	
22	Mean period	
23	Standard deviation of period	
24	Fraction of locally unvoiced frames	Voicing parameters
25	Number of voice breaks	
26	Degree of voice breaks	

B. FEATURE SELECTION

One standard method for reducing the large number of features is feature selection. The objective is to choose a subset of the original dataset's important features based on predetermined relevance evaluation criteria. Better learning outcomes are typically the result of feature selection. Three key types of feature selection methods are wrappers, filters, and embedded algorithms [34]. Wrappers operate in conjunction with a classification algorithm, where each iteration's classification error contributes to a score for the wrapper's evaluation function.

In contrast, filters demand less computational resources than wrappers [35]. Filter-based feature selection algorithms are independent of the classifier's performance. Embedded methods, positioned between filters and wrappers, collaborate with a classifier while considering intermediate outcomes like weights, avoiding excessive computation [36].

According to the authors of [37], mRMR and ReliefF are considered the most reliable feature selection methods. On the other hand, Fisher is a quick and efficient approach that evaluates features individually. It determines the interclass separation and the variation within each class, and features are then ranked based on these calculations to select the top ones [38]. This paper uses three filter-based methods (ReliefF, mRMR, and Fisher) to select relevant features. Subsequent sections describe these methods. Tables 5 to 10 present the top-ranked features ranked by the following feature ranking techniques for each dataset.

1) ReliefF

ReliefF, an enhanced iteration of the original Relief algorithm, is tailored to accommodate multi-class scenarios. ReliefF introduces a method wherein a training sample is randomly selected. Subsequently, the algorithm identifies the samples resembling it and those differing from it within the same sample set. This process facilitates the computation of sample weights for each feature, indicating their ability to differentiate among proximate samples [39]. The steps below present the detailed procedure of the ReliefF algorithm.

1: Initialize weights of all the features(f) to 0

$$Wt\,(f) = 0$$

2: Select a random sample (s_r) from the data iteratively.
3: Find two kinds of k-nearest neighbors of the selected sample:
   a. neighbor belonging to the same class (Hit) - h_x
   b. neighbor belonging to a different class (Miss) – m_x

4: For each feature, update the weights of the feature w.r.t h, m, s_r as in (1)
   a. if h_x and s_r have different values, which implies that this feature separates two instances (h_x and s_r) from the same class, and therefore, $Wt\,(f_i)$ decreases
   b. if m_x and s_r have different values, that implies this feature separates two instances (h_x and s_r) from different classes, and therefore, $Wt\,(f_i)$ increases (1), as shown at the bottom of the next page.

where, D (f_i, s_a, s_b) calculates the difference between the values of fi for $s_a and s_b$, and can be given as in (2):

$$D\,(f_i, s_a, s_b) = \begin{cases} \dfrac{s_a\,[f_i] - s_b\,[f_i]}{\max\,(f_i) - \min\,(f_i)} & if\ f_i\ is\ continuous \\ 0 \quad if\ s_a\,[f_i] = s_b\,[f_i]\ and\ f_i\ is\ discrete \\ one \quad if\ s_a\,[f_i] \neq s_b\,[f_i]\ and\ f_i\ is\ discrete \end{cases} \tag{2}$$

5: Iterate from 2 to 4 until all the samples are considered
6: Return the Updated weights $Wt\,(f)$

The strengths of the ReliefF algorithm include its independence from heuristics, efficient low-order polynomial time complexity, robustness in handling noise and feature interactions, and applicability to both binary and continuous data. A limitation is that it assigns higher weights to features that exhibit strong correlations with the classification task, making it less effective at identifying and eliminating redundant features. However, it may overlook features with low weights that could improve classification results when combined with others.

2) FISHER

Fisher is a supervised feature ranking approach that uses filters and feature weights [40]. Finding a feature subset that minimizes the distances inside a class and maximizes the distances between data points in different classes within a data space spanned by the features that were selected is its primary objective [41]. With a training dataset denoted as X_{mxn}, where m represents the number of samples, n denotes the number of features, and there are c different classes in the dataset, the Fisher score of the i^{th} feature is determined as shown in Equation 3.

$$fs_i = \frac{\sum_{j=1}^{c} S_j(\mu_i^j - \mu_i)^2}{\sum_{j=1}^{c} S_j\,(\rho_{ij})^2} \tag{3}$$

where S_j is the size of samples in class j, μ_i^j is the j^{th} class's i^{th} feature's mean, μ_i is the i^{th} feature's mean in X (i.e., in all the classes), and ρ_{ij} is the ith feature's variance in the jth class. In (3), the numerator denotes inter-class scatter, and the denominator indicates intra-class scatter. Hence, the score denoted as fs_i serves as an indicator of the discriminatory capability of the ith feature. A higher fs_i value signifies a strong discriminatory ability of the ith feature.

Fisher feature ranking algorithm offers several advantages, such as identifying highly discriminative features, improving classification accuracy, and simplifying model interpretation while reducing overfitting and dimensionality. However, it assumes linearity and can be sensitive to imbalanced class distributions, potentially favoring majority classes. Its computational complexity can also be a challenge with large datasets, and it is primarily geared toward classification tasks. Furthermore, FS's independent evaluation of features may overlook potential feature interactions, and it may not accurately capture the data's underlying structure when data exhibits a manifold structure, impacting the accuracy of computed scatters [42].

3) MINIMUM REDUNDANCY MAXIMUM RELEVANCE

The mRMR technique is a filter-based approach aiming to find the most discriminative feature subsets while minimizing redundancy. Initially introduced by Peng et al. [43] for pattern classification tasks. The fundamental principle behind the mRMR feature ranking approach is to start by selecting features that strongly correlate with the target class while maintaining minimal correlation with the remaining features.

This algorithm handles the class label (C) vector and each attribute (x) as discrete events. Mutual information (MI), represented as MI (x, C), is used to measure the degree of similarity between two attributes or an attribute and the class label vector. MI is formally defined, as shown in (4).

$$MI\,(x, C) = \sum_{x \in X} \sum_{c \in C} Prob\,(x, c) \log \frac{Prob\,(x, c)}{Prob\,(x)\,Prob(c)} \tag{4}$$

$Prob\,(x)$ and $Prob\,(c)$ represent the marginal probability functions of specific features, while $Prob\,(x, c)$ denotes the joint probability distribution. The value of MI, as discussed in reference [44], equals 0 when two random variables are entirely independent. It's essential to note that this value is symmetric and cannot take on negative values. MI ensures that the selected features have the highest possible dependency on the classification variable through their joint distribution [45].

In finding the most discriminative set of features (X), the mRMR algorithm aims to satisfy two critical conditions when selecting attributes. The first condition emphasizes achieving maximum relevance, denoted as *MaxRel* as shown in (5), while the second condition stresses minimizing redundancy, denoted as *MinRed*, as shown in (6).

$$MaxRel = \frac{\sum_{x \in X} MI(x, C)}{|X|} \tag{5}$$

$$MinRed = \frac{\sum_{x,z \in X} MI(x, z)}{|X|^2} \tag{6}$$

Two approaches are commonly employed to integrate the abovementioned conditions when selecting features from a feature subset that comprehensively describes all features

except the ones selected. These approaches are Mutual Information Difference (MID), given as $max(MaxRel - MinRed)$, and Mutual Information Quotient (MIQ), defined as $max(MaxRel/MinRed)$ [46]. In this paper, MIQ is used to rank features.

The mRMR algorithm provides indices of the feature vector in the order of their scores. These feature scores are determined through a heuristic algorithm and are represented as scores. Higher score values indicate that the corresponding features are more critical predictors, while a decrease in the feature importance score suggests reduced confidence in selecting those features.

mRMR takes advantage of the benefits of the filter-based feature selection approach while actively discarding redundant features. During each iteration of the mRMR algorithm, it calculates a metric that gauges both redundancy and relevance among features, ultimately selecting the feature that optimally maximizes this metric.

C. THEORETICAL BACKGROUND AND MODEL DEVELOPMENT

This section describes the various classification and optimization techniques. Section III-C1 describes various basic ML models (ELM, SVM, and RF). Section III-C2 describes the optimization technique PSO.

1) CLASSIFICATION TECHNIQUES

Classification techniques provide a systematic approach to classifying PD, helping healthcare professionals to make diagnosis and treatment decisions. Patient records are used to diagnose the disease.

a: RANDOM FOREST

Breiman [47] introduced Random Forests (RF) that employ the concept of 'bagging' to create an ensemble of decision trees aimed at delivering precise predictions [48]. All the trees in RF depend on a randomly chosen subset of the training data. This method incorporates bootstrap sampling, allowing for an additional, untouched subset known as the out-of-bag data, which aids evaluation. Unlike individual decision trees, RF produces averaged estimations from various aggregations. For further insights into RF, refer to [47] and [49].

b: SUPPORT VECTOR MACHINE

SVM is based on statistical learning theory. This versatile method operates on classification and regression tasks and can handle linear or nonlinear datasets using kernel functions. Its surge in popularity stems from its robust theoretical foundations, efficiency with sizable datasets, flexibility enabled by kernel functions, and ability to yield highly accurate results. SVM identifies the widest margin among numerous potential linear functions to delineate data linearly. SVM uses kernel functions to map data to higher-dimensional spaces for nonlinear classification, facilitating the discovery of multiple planes with maximum margins.

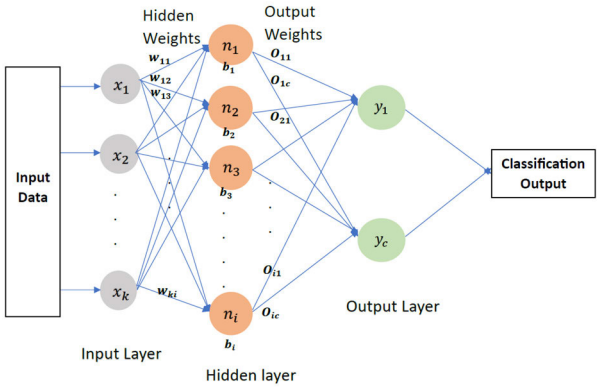

FIGURE 1. Structure of ELM.

c: EXTREME LEARNING MACHINE

The ELM is initially introduced by Huang et al. [50]. ELM is specifically designed to reduce the limitations associated with traditional learning algorithms when constructing single-layer feed-forward neural networks.

In contrast to conventional ML methods, ELM takes a distinct approach. The parameters of the hidden layer (hidden weights - wki and biases bi) are randomly initialized, and the output weights (O_{ic}) can be analytically determined using the least-square method. Theoretically, ELM is established to possess universal approximation capabilities when employing any non-constant, piecewise continuous activation function [51]. Fig. 1 depicts the ELM structure.

ELM provides many advantages [52]. In contrast to conventional approaches like the backpropagation algorithm, ELM exhibits very high learning speed. Moreover, ELM has excellent classification capability, making it suitable for many learning tasks. Also, it circumvents common issues such as falling into local minima or suffering from overfitting. These features make ELM a powerful and efficient tool for various machine-learning tasks.

2) OPTIMIZATION TECHNIQUES

Evolutionary computation techniques have gained widespread attention recently due to their remarkable optimization capabilities and extensive applicability to practical problem-solving [53]. Among these techniques, genetic algorithms aim to mimic the evolutionary processes seen in living

$$Wt\,(f_i) = Wt\,(f_{i-1}) + \frac{\left(\sum C \notin class\,(s_r)\left[\frac{Prob(C)}{1-Prob(class(S))}\sum_{x=1}^{k} D\,(f_i, s_r, m_x)\right] - \sum_{x=1}^{k} D\,(f_i, s_r, h_x)\right)}{n.k} \quad (1)$$

Algorithm 1 PSO-ELM

Inputs: **PSO Parameters**: the number of particles in the range (10,20,30,40,50) and the number of iterations in the range (20,40,60,80,100).
ELM Parameters: number of hidden neurons, number of features in the data subset

Output: Optimal hyperparameters of PSO-ELM (number of particles, number of iterations, hidden weights, hidden biases, and output weights), Best Accuracy.

```
Initialize the particle range (10, 20, 30, 40, 50) and iteration range (20, 40, 60, 80, 100) for Grid search
    for each value of particles in (10, 20, 30, 40, 50)
        Initialize the swarm with particles, number of hidden neurons, and number of features.
        for each value of iterations in (20, 40, 60, 80, 100)
            for fold = 1 to 3
                for i = 1 to iterations
                    for p = 1 to particles
                        Calculate the fitness value with the accuracy obtained by training ELM with training data
                        subsets, current hidden weights, and current hidden biases.
                        Update Particles Local best and Global best positions if necessary.
                    end for
                    Update Particle positions and velocities according to (7) and (8).
                end for
                Test the model using the optimized weights and biases obtained in the current solution.
            end for
            If current accuracy > previous accuracy
                Update the current average accuracy and store the optimized parameters
            end if
        end for
    end for
Display the best hyperparameter values obtained by PSO-ELM (Particles and Iterations).
Display the result achieved.
```

organisms. In contrast, particle swarm optimization (PSO) [54], [55] algorithms and ant colony algorithms draw inspiration from animal predation behaviors. Also, the fireworks algorithm [56] and simulated annealing algorithm has been introduced to replicate natural phenomena in optimization processes.

a: PARTICLE SWARM OPTIMIZATION

PSO is a stochastic optimization method introduced and developed by Eberhart and Kennedy [57]. The algorithm's fundamental concept draws inspiration from the behavior of social organisms that exist in swarms, like birds and fish. In the PSO technique, each entity is referred to as a ''particle'', representing a possible solution for the current problem. Collecting all the particles together constitutes what is termed ''population''. Each particle possesses two attributes: velocity and position. The velocity of each particle (7) is adapted according to its motion's impact as each particle strives to reach its best possible position (8).

$$V_{ij+1} = W.V_{ij} + Const_1.rand_1.\left(P_b - X_{ij}\right) + Const_2.rand_2.\left(G_b - X_{ij}\right) \quad (7)$$

$$X_{ij+1} = X_{ij} + V_{ij+1} \quad (8)$$

where W denotes the inertia weight, whose value ranges from [0, 1], P_b denotes the local best, and G_b denotes the particle's global best positions. Const1 and Const2 are typically two constants, while rand1 and rand2 are random numbers between 0 and 1. In this paper, the values considered for the above equations are W=0.5, Const1=1, and Const2=2.

D. PROPOSED METHODOLOGY

Integrating evolutionary algorithms (EAs) with conventional ML models leverages the performance of ML [58]. This paper proposes a hybrid PSO-ELM model to improve the classification performance. PSO is the commonly used optimization technique to optimize the parameters of ML models [59] The PSO optimizes the ELM model's hidden weights and biases. The parameters of the PSO algorithm are optimized by varying the particles and iterations used to get the best hidden weights and biases of the ELM model.

Feature ranking techniques such as Fisher, mRMR, and ReliefF are applied separately to all six datasets: Meander, Spiral, New Meander, New Spiral, SpeechPD, and PMSR. The selected features are divided into 22 different feature subsets for the Meander, Spiral, New Meander, New Spiral, and SpeechPD datasets and 16 feature subsets for the PMSR dataset.

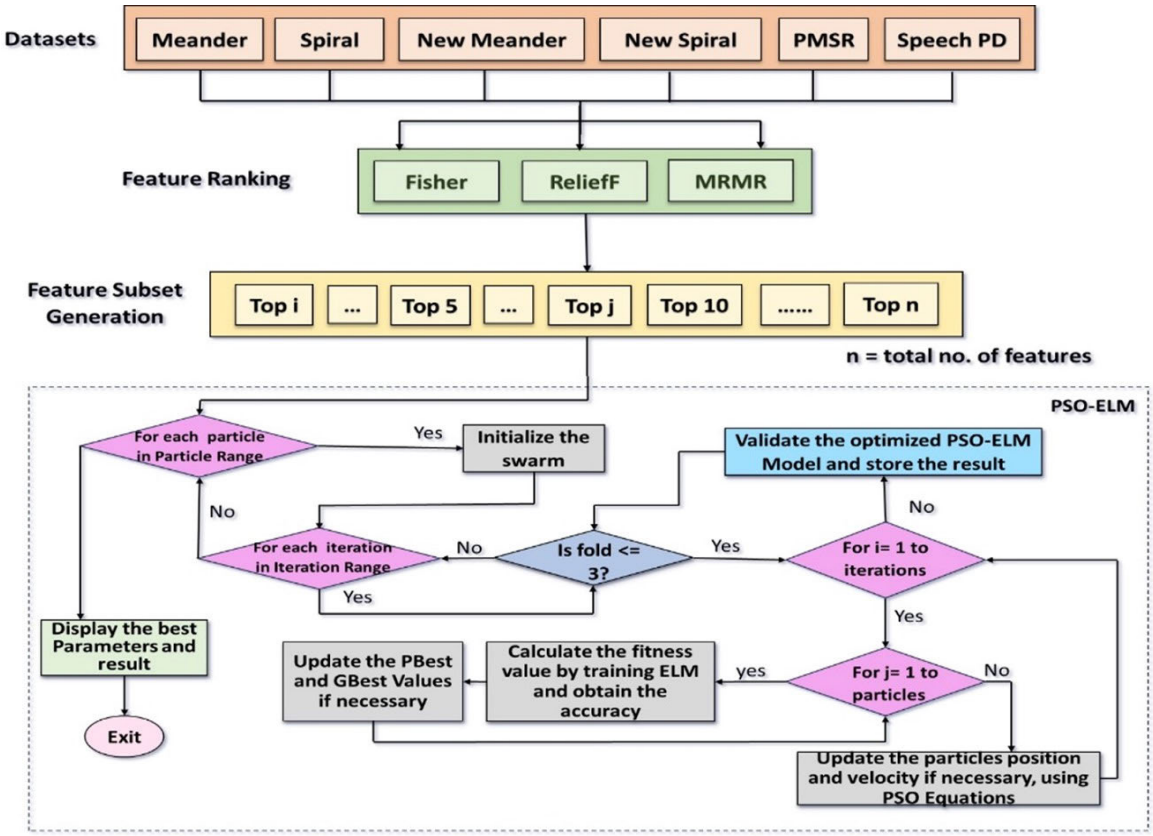

FIGURE 2. Workflow of the proposed model.

Each dataset's features and selected feature subsets are sent separately to the proposed hybrid PSO-ELM model. A grid search technique is applied to determine the best parameters of PSO (i.e., number of particles and number of iterations) suitable for improving ELM performance. The range of the number of particles considered for grid search is [10, 20, 30, 40, 50], and the corresponding iterations considered for these particles lie in the range [20, 40, 60, 80, 100]. Algorithm 1 provides a detailed process of the proposed hybrid PSO-ELM model. For each data subset, the swarm is initialized with 10, 20, 30, 40, and 50 number of particles. Each particle in the swarm is initialized with random numbers ranging from 0 to 1, representing the weights and biases of the hidden layer in the ELM network. If x is the number of features, then each particle's total number of values (weights + biases) is given as $(x \times n)+b$. Where n and b are the number of neurons and bias values. Number of bias values is equal to number of neurons (i.e., $b = n$).

The proposed PSO-ELM model is evaluated using stratified 3-fold cross-validation (3-fold CV). For each fold, the particles in the swarm move through multiple iterations (up to the number specified in the range [20, 40, 60, 80, 100]). For each iteration, the fitness value of each particle is determined with the accuracy obtained by training the ELM network with that particle. Each particle's local and global best values are updated according to the fitness value. In each iteration, based on the accuracy obtained, the particle's velocity and position are updated using (7) and (8). If the fitness value (the accuracy of ELM on training data) is maximum. In that case, the current solution (the particle values that contain weights and biases of the ELM hidden layer) is stored as the best solution. After completing all the iterations, the model is tested using the optimized weights and biases found during the training process. After all combinations of particles and iterations, the ELM is validated with the test data using the best parameters obtained during the training process. The average accuracy of all the 3-folds is calculated and displayed along with the optimized parameters. Accuracy, Precision, Recall, F1-Score, and AUC are the metrics used to evaluate the proposed model's performance. The complete workflow of this paper can be seen in Fig. 2.

IV. RESULTS AND DISCUSSION

This section describes the experimental results of various classifiers: ELM, RF, SVM, and the proposed hybrid PSO-ELM.

A. EXPERIMENTAL SETUP

This experiment used Python 3.10.9 in the Windows 11 environment for feature selection. All the classifiers, namely ELM, RF, SVM, and the proposed model, are implemented

in MATLAB R2022b. The computer hardware environment is configured as follows: 64-bit system with a Windows 11 operating system, 12th Gen Intel(R) Core (TM) i5-1240P 1.70 GHz processor, and 16 GB RAM.

B. PERFORMANCE METRICS

Accuracy, Precision, Recall, F1-Score, and Area Under the Curve (AUC) are frequently used as performance metrics to evaluate the effectiveness of classification models, particularly in the context of disease diagnosis. The confusion matrix is utilized to derive these metrics, as illustrated in Table 4. Here, TP denotes the count of true positives, indicating cases accurately identified as PD. FN represents the false negatives, indicating individuals with PD are incorrectly classified as healthy. TN is the count of true negatives, signifying correctly classified healthy controls, and FP stands for the count of false positives, representing healthy individuals incorrectly classified as patients with PD.

Accuracy (ACC), Precision, Recall, and F1-Score can be defined as:

$$ACCURACY = \frac{TP + TN}{TP + TN + FP + FN}$$
$$PRECISION = \frac{TP}{TP + FP}$$
$$RECALL = \frac{TP}{TP + FN}$$
$$F - SCORE = 2\frac{TP}{2TP + FP + FN}$$
$$AUC = \frac{RECALL}{2} * \frac{TN}{TN + FP}$$

The AUC signifies the area under the Receiver Operating Characteristic (ROC) curve. The ROC curve illustrates the trade-off between true and false positive rates. A classifier with a higher AUC is considered better than one with a lower AUC. An ideal classifier achieves an AUC equal to 1. AUC is a robust method for comparing classifiers in two-class problems [60].

TABLE 4. Confusion matrix.

Actual	Predicted	
	PD	**HC**
PD	TP	TN
HC	FP	FN

C. STATISTICAL SIGNIFICANCE TEST

Tests for statistical significance assess the likelihood that an observed relationship in the data is due to chance or if the variables are genuinely unrelated in the overall population. These tests can help dismiss unlikely hypotheses. This paper compares the statistical significance of different models by using two non-parametric tests: Kruskal Wallis and Dunn's posthoc.

1) KRUSKAL-WALLIS TEST

The Kruskal-Wallis test is a non-parametric method used to check if there are significant differences between the medians of three or more independent groups [61]. It helps to determine if the distributions of these groups are different from each other. To perform the test, values from all groups are ranked from lowest to highest, and the test statistic is found by adding the ranks within each group. The null hypothesis(H_0) states that there are no significant differences between the groups, while the alternative hypothesis suggests that at least one group differs significantly. The test hypothesis(H) can be calculated as shown in equation (9).

$$\boldsymbol{H = \frac{12}{n\,(n+1)} \sum_{i=1}^{n_c} \frac{T_i^2}{n_i^2} - 3\,(n+1)} \tag{9}$$

where $\boldsymbol{n_i}$ is the number of observations of each group, $\boldsymbol{n}$ is the total number of observations, $\boldsymbol{T_i}$ is the sum of the ranks of i^{th} group. When there are n_c groups, the Kruskal-Wallis test statistic is compared to the $\chi 2$ distribution with n_c-1 degrees of freedom, provided each group has a sufficiently large sample size. This statistic is calculated for all features to enable ranking based on their $\chi 2$ values. If the Kruskal-Wallis test shows a significant result, indicating that at least one group is different, a post-hoc test can be done to find out which specific groups are different.

2) DUNN'S POSTHOC TEST

Dunn's post hoc test is used to compare multiple pairs of mean accuracies after the statistical tests [62]. It adjusts the significance level (alpha) to account for the number of comparisons made, reducing the risk of finding differences by chance. This adjustment helps control the ''multiple comparisons problem'' and ensures more reliable results when identifying significant differences between pairs of accuracies of two models. This paper uses the dunn-sidak correction type in Dunn's post hoc test, which is considered to be optimal [63] when the variances are either known or unknown but equal.

D. FEATURE RANKING AND SUBSET SELECTION RESULTS

Three feature ranking algorithms, ReliefF, Fisher, and mRMR, are employed on all six datasets: Meander, Spiral, New Meander, New Spiral, Speech PD, and PMSR. The ranking of the features by different ranking algorithms is listed in Tables 5 to 10. This section discusses the appropriateness of various features in diagnosing PD with different PD datasets and their subsets.

Table 5 shows that AGE is identified as the top-ranked feature by ReliefF and Fisher feature ranking algorithms for the meander dataset. mRMR has ranked AGE as the top-5th feature for the meander dataset. Also, it can be observed that the top features selected by Fisher are the same as those of the ReliefF algorithm. The top-5th ranked feature by ReliefF and Fisher is 'MAX_BETWEEN_ET_HT' and 'STD_DEVIATION_ST_HT', respectively. 'MAX_HT' is

ranked as the topmost feature by mRMR, but it is ranked 6th by ReliefF and 11th by Fisher feature ranking algorithms. That indicates that 'MAX_HT' is particularly important for minimizing redundancy and maximizing relevance, which mRMR aims to achieve. The sum of ranks of all three feature ranking algorithms shows that the top-5 contributing features are 'AGE', 'GENDER', 'MRT', 'RMS', 'CNTP', which can be observed from Table 5.

TABLE 5. Feature ranking of meander dataset.

S.No	Features	Feature Ranking Algorithms		
		ReliefF	mRMR	Fisher
1	AGE	1	5	1
2	GENDER	2	7	3
3	MRT	3	9	2
4	RMS	4	8	4
5	CNTP	8	3	6
6	MAX_HT	6	1	11
7	MIN_HT	10	4	7
8	STD_HT	12	2	10
9	MAX_BETWEEN_ET_HT	5	11	9
10	STD_DEVIATION_ST_HT	11	10	5
11	RIGH/LEFT-HANDED	7	12	8
12	MIN_BETWEEN_ET_HT	9	6	12

Table 6 shows the ranking of features for the spiral dataset using different feature ranking algorithms. It can be seen that the 'AGE' is ranked 1, 6, and 4 by ReliefF, mRMR, and Fisher, respectively. 'AGE' and 'GENDER' are generally important features across different algorithms, making them the most discriminative features for PD diagnosis. 'MAX_HT' and 'STD_HT' are highly ranked by mRMR, indicating they might be crucial for reducing redundancy in the data. 'MAX_BETWEEN_ET_HT' is consistently ranked 5, 11, and 12 by ReliefF, mRMR, and Fisher, respectively. Therefore, 'MAX_BETWEEN_ET_HT' is not a crucial feature for diagnosing PD. The top-5 contributing features obtained by the sum of ranks of each algorithm can be given as 'AGE', 'GENDER', 'MAX_HT', 'MIN_BETWEEN_ET_HT', 'RIGH/LEFT-HANDED'.

TABLE 6. Feature ranking of spiral dataset.

S.No	Features	Feature Ranking Algorithms		
		ReliefF	mRMR	Fisher
1	AGE	1	6	4
2	GENDER	2	7	3
3	MAX_HT	4	1	8
4	MIN_BETWEEN_ET_HT	8	5	5
5	RIGH/LEFT-HANDED	6	12	1
6	RMS	3	8	9
7	CNTP	7	3	10
8	MIN_HT	11	4	6
9	STD_HT	10	2	11
10	STD_DEVIATION_ET_HT	12	10	2
11	MRT	9	9	7
12	MAX_BETWEEN_ET_HT	5	11	12

Table 7 indicates the ranking of features of New Meander data by different feature ranking algorithms, namely ReliefF, mRMR, and Fisher. Table 7 shows that 'STD_HT' and 'MAX_HT' are highly relevant features, especially for mRMR and Fisher, indicating their significance in reducing redundancy and improving relevance. 'AGE' and 'MRT' are ranked as 1 and 2, respectively, by ReliefF, suggesting they capture essential aspects of the dataset relevant to PD diagnosis. 'MIN_HT' has moderate to lower rankings, such as 9, 4, and 9 by ReliefF, mRMR, and Fisher. Therefore, 'MIN_HT' is not a crucial feature for diagnosing PD. Similarly, 'MIN_BETWEEN_ET_HT', 'RIGH/LEFT-HANDED', and 'STD_DEVIATION_ET_HT' have consistently lower ranks across all the algorithms, indicating the less relevant features. The top-5 contributing features obtained by the sum of ranks of each algorithm can be given as 'STD_HT', 'MAX_HT', 'GENDER', 'MRT', and 'AGE' for the New Meander dataset.

TABLE 7. Feature ranking of new meander dataset.

S.No	Features	Feature Ranking Algorithms		
		ReliefF	mRMR	Fisher
1	STD_HT	3	2	2
2	MAX_HT	12	1	1
3	GENDER	6	7	3
4	MRT	2	8	6
5	AGE	1	5	11
6	MAX_BETWEEN_ET_HT	4	10	4
7	CNTP	8	3	7
8	RMS	5	11	5
9	MIN_HT	9	4	9
10	MIN_BETWEEN_ET_HT	7	6	10
11	RIGH/LEFT-HANDED	11	12	8
12	STD_DEVIATION_ET_HT	10	9	12

TABLE 8. Feature ranking of new spiral dataset.

S.No	Features	Feature Ranking Algorithms		
		ReliefF	mRMR	Fisher
1	STD_HT	2	2	2
2	AGE	1	5	4
3	MAX_HT	5	1	5
4	CNTP	6	3	6
5	GENDER	8	7	3
6	MIN_HT	9	4	7
7	STD_DEVIATION_ET_HT	12	9	1
8	MRT	4	8	10
9	RMS	3	12	8
10	MIN_BETWEEN_ET_HT	10	6	9
11	MAX_BETWEEN_ET_HT	7	10	12
12	RIGH/LEFT-HANDED	11	11	11

Table 8 represents the ranking of features for the New Spiral dataset with different feature ranking algorithms. 'STD_HT' is ranked 2nd by all three algorithms (ReliefF, mRMR, and Fisher), indicating it is a highly significant feature across different selection methods. 'AGE' also shows strong relevance, ranking 1st by ReliefF, 5th by mRMR, and 4th by Fisher. Also, Table 8 shows that 'CNTP' is ranked 6,3, and 6, and 'GENDER' is ranked 8, 7, and 3 by ReliefF, mRMR, and Fisher, respectively, showing moderate importance across different methods. 'RIGH/LEFT-HANDED' and

'MAX_BETWEEN_ET_HT' are consistently low-ranked, indicating they are less critical for diagnosing PD using the New Spiral dataset. The top-5 contributing features obtained by the sum of ranks of each algorithm can be given as 'STD_HT', 'AGE', 'MAX_HT', 'CNTP', and 'GENDER'.

TABLE 9. Feature ranking of new SpeechPD dataset.

S.No	Features	Feature Ranking Algorithms		
		ReliefF	mRMR	Fisher
1	PPE	4	1	2
2	RPDE	6	4	4
3	spread1	1	6	8
4	HNR	7	5	7
5	DFA	17	2	6
6	Shimmer: DDA	9	12	5
7	MDVP: Shimmer	10	9	12
8	MDVP: Shimmer(dB)	12	10	10
9	Shimmer: APQ3	8	13	11
10	Shimmer: APQ5	13	7	14
11	MDVP: Flo (Hz)	2	14	20
12	MDVP: Jitter (Abs)	16	19	3
13	MDVP: APQ	14	3	21
14	MDVP: Jitter (%)	18	20	1
15	spread2	5	16	18
16	MDVP: Fo (Hz)	3	22	19
17	D2	15	15	15
18	MDVP: RAP	20	18	9
19	NHR	22	8	17
20	Jitter: DDP	21	11	16
21	MDVP: PPQ	19	17	13
22	MDVP: Fhi(Hz)	11	21	22

From Tables 5, 6, 7, and 8, it can be observed that the top contributing features for handwritten datasets can be 'AGE' and GENDER, as they are lower-ranked (between 1 to 5) features in all the handwritten datasets with all three feature ranking algorithms. 'MAX_HT' is the next important feature in three (New Meander, Spiral, New Spiral) of the four handwritten datasets. 'MRT', 'STD_HT', and 'CNTP' are common in the top five features for two (Meander and New Spiral) out of four handwritten datasets.

Table 9 shows the features of the SpeechPD dataset ranked with ReliefF, mRMR, and Fisher feature ranking algorithms. Table 9 shows that the 'spread1', 'PPE', and 'MDVP: Jitter (%)' are the top-ranked features by ReliefF, mRMR, and Fisher feature ranking algorithms, respectively. 'PPE' is ranked consistently high (4th by ReliefF, 1st by mRMR, and 2nd by Fisher), indicating it is a crucial feature for PD diagnosis across all feature ranking methods. Similarly, 'RPDE' also shows strong relevance, with rankings of 6th by ReliefF, 4th by mRMR, and 4th by Fisher, indicating it is an important feature. 'DFA' is ranked as the top 17th feature by ReliefF but much higher by mRMR (2nd) and Fisher (6th), indicating its strong relevance in minimizing redundancy and maximizing relevance. From Table 9, it can also be observed that the 'D2' feature is given an equal rank of 15 across all three feature ranking methods. 'NHR' and 'MDVP(Hz)' are consistently ranked low, suggesting they are less critical for the SpeechPD dataset.

TABLE 10. Feature ranking of new PMSR dataset.

S.No	Features	Feature Ranking Algorithms		
		ReliefF	mRMR	Fisher
1	Number of pulses	11	7	1
2	Jitter (local, absolute)	5	2	14
3	Fraction of locally unvoiced frames	1	8	18
4	Mean period	9	13	5
5	Shimmer (local, dB)	7	5	17
6	Standard deviation of period	18	10	3
7	Shimmer (apq3)	23	1	7
8	Maximum pitch	3	14	15
9	NTH	6	15	11
10	HTN	4	21	8
11	Shimmer (apq11)	8	19	6
12	Jitter (rap)	16	6	12
13	Degree of voice breaks	21	11	2
14	Number of periods	10	16	10
15	Shimmer (local)	14	4	19
16	Jitter (local)	12	3	24
17	Number of voice breaks,	25	17	4
18	Standard deviation,	20	9	20
19	AC,	2	23	25
20	Minimum pitch,	19	12	22
21	Median pitch,	13	25	16
22	Mean pitch,	15	18	21
23	Jitter (ddp),	17	26	13
24	Shimmer (dda),	24	24	9
25	Jitter (ppq5),	22	20	26
26	Shimmer (apq5)	26	22	23

Table 10 shows the feature ranking of the PMSR dataset with the three feature ranking algorithms: ReliefF, mRMR, and Fisher. From Table 10, it can be observed that the top-ranked features include 'Number of pulses,' which is ranked one by Fisher and seven by mRMR, 'Jitter (local, absolute),' which is ranked five by ReliefF and two by mRMR, and 'Fraction of locally unvoiced frames' ranked one by ReliefF and eight by mRMR. The least ranked features include 'Shimmer (dda),' ranked 24 by both ReliefF and mRMR feature ranking algorithms; 'Jitter (ppq5)', ranked 22 by ReliefF and 20 by mRMR; and 'Shimmer (apq5)', ranked 26 by ReliefF, 22 by mRMR and 23 by Fisher, indicating it as less contributing feature for PD diagnosis. Table 10 shows that the ReliefF algorithm prioritizes the features 'Number of pulses' and 'Jitter,' which indicate speech patterns and vocal quality. The mRMR algorithm, on the other hand, considers the features 'Standard deviation of period' and 'Maximum pitch' as crucial features. These features belong to the tone and rhythm of speech. The Fisher algorithm shows a mix of these trends shown by ReliefF and mRMR, ranking 'Maximum pitch' and 'Jitter (local, absolute)' high while ranking the 'Number of pulses' and 'Maximum pitch' low. Overall, some consistently highly ranked features across the algorithms include 'Fraction of locally unvoiced frames', 'AC', 'Maximum pitch', 'Jitter

TABLE 11. Performance of the proposed model on the meander dataset with 100 neurons.

FS Algorithm	Features	Average Accuracy (%)				Performance Metrics of PSO-ELM (%)				Best Parameters	
		ELM	SVM	RF	PSO-ELM	AUC	Precision	Recall	F1-Score	Particles	Iterations
All	12	85.33	83.15	93.47	99.46	0.986	100	97.22	98.58	20	20
ReliefF	4	85.05	92.66	94.56	99.73	0.993	100	98.61	99.29	10	20
	5	85.33	88.86	94.29	99.46	0.991	98.67	98.61	98.61	50	20
	6	85.06	86.68	93.48	**100**	1	100	100	100	**10**	**20**
	7	82.07	85.32	92.66	98.92	0.977	98.48	95.83	97.10	10	20
	8	83.96	83.97	93.75	**100**	1	100	100	100	40	20
	9	84.79	83.97	93.20	**100**	1	100	100	100	10	20
	10	83.96	83.42	93.75	99.73	0.993	100	98.61	99.29	10	20
mRMR	4	73.90	80.71	79.35	98.37	0.964	98.61	93.06	95.68	10	20
	5	78.25	85.32	90.49	99.46	0.986	100	97.22	98.58	10	40
	6	79.34	84.79	89.40	99.45	0.991	98.61	98.61	98.61	10	40
	7	86.68	89.40	95.10	99.73	0.993	100	98.61	99.29	10	20
	8	85.60	87.50	93.75	99.73	0.993	100	98.61	99.29	20	20
	9	86.42	86.68	92.39	**100**	1	100	100	100	50	20
	10	82.88	86.41	93.75	**100**	1	100	100	100	10	20
Fisher	5	88.04	90.77	95.12	99.73	0.993	100	98.61	99.29	10	20
	6	85.32	88.05	94.57	**100**	1	100	100	100	50	20
	7	84.23	87.78	95.38	**100**	1	100	100	100	10	40
	8	85.05	87.23	95.38	**100**	1	100	100	100	10	20
	9	84.50	84.79	94.30	**100**	1	100	100	100	10	20
	10	86.40	84.79	94.03	**100**	1	100	100	100	10	20

(local, absolute)', and 'Number of pulses'. However, the ranking differs for many other features across the algorithms.

Based on the feature ranks of ReliefF, mRMR, and Fisher, different data subsets are generated for the Meander, Spiral, New Meander, and New Spiral datasets having top-4 to top-10 features and the original dataset, resulting in 22 feature subsets for each of the four datasets. The speech PD dataset is divided into top-5, top-8, top-10, top-13, top-14, top-15, and top-20 subsets for each applied algorithm, and the original dataset resulting in 22 different feature subsets. Similarly, top-5, top-10, top-15, top-20, and top-23 subsets were created for the PMSR dataset, resulting in 16 feature subsets.

E. CLASSIFICATION RESULTS

In this paper, all the results are evaluated using the 3-fold CV technique by dividing the samples of each dataset into three equal parts. One of the challenges in PD diagnosis is to find the most significant features that are crucial for the diagnosis of PD. Three different feature ranking algorithms, ReliefF, mRMR, and Fisher, are employed on six different PD datasets: Meander, Spiral, New Meander, New Spiral, SpeechPD, and PMSR. Each subset of features is sent to three basic models: ELM, SVM, RF, and the proposed hybrid PSO-ELM model. In addition to optimizing the hidden weights and biases, the number of neurons in the hidden layer of the ELM network in the proposed PSO-ELM method is also varied to check the neurons' impact on the proposed model. The proposed hybrid PSO-ELM model is trained and tested with 126 datasets and their subsets (6 original datasets+120 data subsets created from 6 original datasets). 22 different feature subsets are created from the Meander dataset by selecting a different number of top-ranked features ranked by ReliefF, mRMR, and Fisher Feature ranking algorithms, as listed in Table 11.

Similarly, for each dataset, namely, New Meander, Spiral, and New Spiral datasets, 22 different feature subsets are created, as listed in Tables 12, 13, and 14, respectively. Also, for the SpeechPD dataset, 22 different feature subsets are created, as listed in Table 15. For the PMSR dataset, 16 feature subsets are created, as listed in Table 16.

Various numbers of neurons are selected for the hidden layer of the ELM network in the Meander dataset to further improve the performance of the proposed PSO-ELM model. These are 70, 80, 90, and 100. From Fig. 3(d), it can be observed that the performance of the proposed PSO-ELM model is gradually increasing with an increase in the number of neurons.

Finally, when the number of neurons is 100, the proposed PSO-ELM model achieved 100% accuracy with the top-6, top-8, and top-9 feature subsets selected by the ReliefF feature ranking algorithm. The corresponding particles & iterations are 10 and 20, respectively. With top-9 and top-10 feature subsets selected by the mRMR feature ranking algorithm, the proposed PSO-ELM achieved 100% accuracy. Also, with the top-6, 7, 8, 9, and 10 features selected by the Fisher feature ranking algorithm, the same 100% accuracy is achieved, which can also be observed from Table 11. So, the 7th, 8th, 9th and 10th ranked features by the Fisher feature ranking algorithm are not affecting the accuracy. Therefore, these features are not significant. It can be seen from Table 11 that the highest accuracy is 100% for both the top-6 features subsets ranked by ReliefF ('AGE', 'GENDER', 'MRT', 'RMS', 'MAX_BETWEEN_ET_HT', 'MAX_HT') and Fisher ('AGE', 'MRT', 'GENDER', 'RMS', 'STD_DEVIATION_ST_HT', 'CNTP').

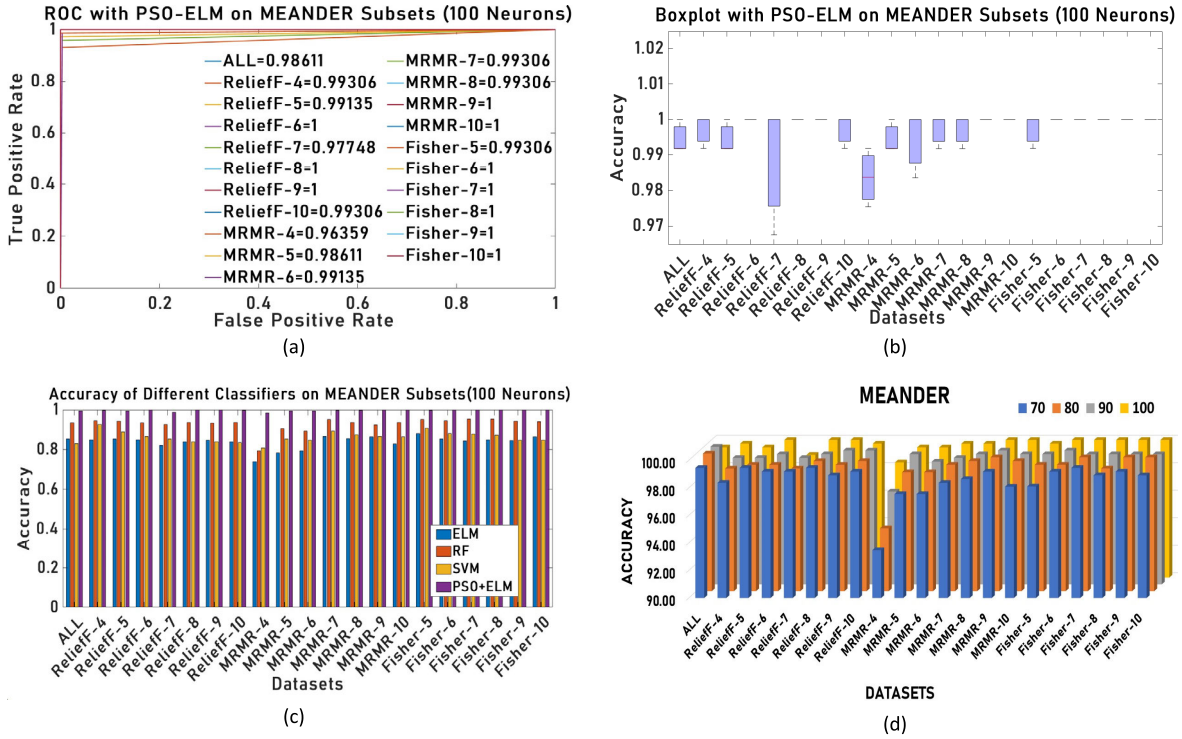

FIGURE 3. MEANDER dataset: (a)ROC-AUC Curve of the proposed model (b) Box Plot indicating the accuracies obtained in 3-fold CV (c) Bar Graph comparing traditional ML models and proposed model (d) Impact of neurons on the performance of the proposed model.

TABLE 12. Performance of the proposed model on the new meander dataset with 80 neurons.

FS Algorithm	Features	Average Accuracy (%)				Performance Metrics of PSO-ELM (%)				Best Parameters	
		ELM	SVM	RF	PSO-ELM	AUC	Precision	Recall	F1-Score	Particles	Iterations
All	12	76.14	73.86	84.85	99.62	1	100	99.28	99.63	10	20
ReliefF	4	67.80	81.06	82.20	99.62	0.996	100	99.28	99.63	10	20
	5	72.73	82.20	82.95	98.86	0.989	99.28	98.57	98.92	10	40
	6	69.70	85.23	83.71	99.24	0.993	99.29	99.29	99.28	10	40
	7	73.48	83.33	85.23	**100**	1	100	100	100	10	40
	8	75.00	78.41	84.85	99.62	0.996	100	99.28	99.63	10	60
	9	74.24	79.55	85.61	**100**	1	100	100	100	10	40
	10	77.65	79.17	83.33	99.62	0.997	100.00	99.29	99.64	10	40
mRMR	4	68.56	73.48	76.52	98.86	0.989	99.28	98.58	98.92	10	60
	5	74.62	76.52	79.92	**100**	1	100	100	100	**20**	**20**
	6	75.38	79.17	79.92	99.62	0.996	99.31	100	99.65	30	20
	7	76.89	78.03	82.20	98.86	0.989	99.29	98.58	98.93	30	20
	8	80.68	78.41	80.68	**100**	1	100.00	100	100	10	40
	9	78.03	78.03	81.82	99.24	0.993	100	98.58	99.28	10	40
	10	76.14	77.27	81.82	**100**	1	100	100	100	10	40
Fisher	4	60.23	64.39	69.70	98.86	0.988	98.60	99.29	98.94	10	20
	5	57.20	68.18	69.70	99.24	0.993	100	98.58	99.28	10	40
	6	66.29	68.18	73.86	98.86	0.989	99.29	98.57	98.92	20	20
	7	65.91	68.56	76.89	99.62	0.997	100	99.29	99.64	10	20
	8	65.53	67.80	76.14	99.24	0.992	99.29	99.29	99.29	10	20
	9	65.53	68.18	76.52	99.24	0.992	99.29	99.29	99.29	10	20
	10	63.64	66.67	78.03	99.62	0.996	99.31	100	99.65	20	20

The top 4 features of ReliefF and Fisher are the same. The proposed model's performance also outperforms all the feature subsets compared to the basic ELM, SVM, and RF, as can also be observed in Fig. 3(c).

The RF classifier has performed better with almost all the selected feature subsets than the other two models (ELM and SVM). Based on the observation, the top contributing features of the Meander dataset are AGE, GENDER, MRT, RMS, MAX_BETWEEN_ET_HT and MAX_HT. Performance metrics like precision and recall are also high, with the proposed model indicating its efficacy in identifying PD patients. The F1 score is 100%, which shows a good

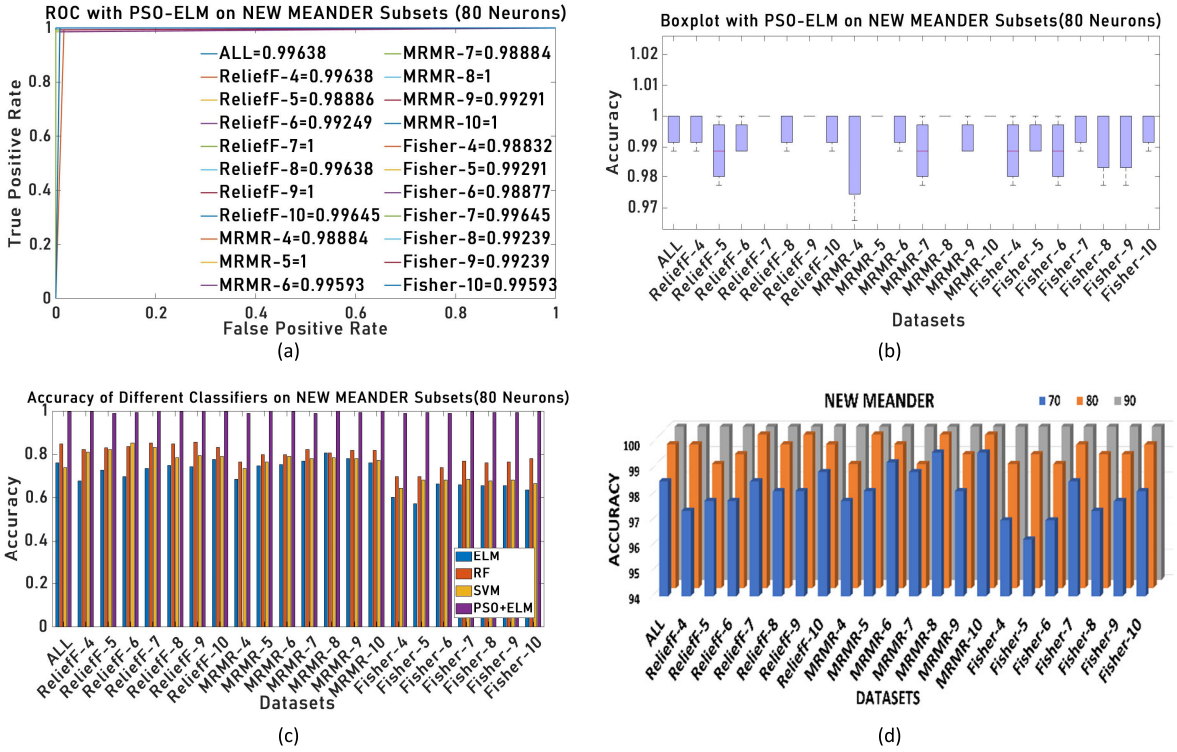

FIGURE 4. NEW MEANDER dataset: (a)ROC-AUC Curve of the proposed model (b) Box Plot indicating the accuracies obtained in 3-fold CV (c) Bar Graph comparing traditional ML models and proposed model (d) Impact of neurons on the performance of the proposed model.

balance between precision and recall. The ROC-AUC Curve is shown in Fig. 3(a), indicating the highest AUC of 1 with the ReliefF's top-6 feature subset. Fig. 3(b) shows the boxplot on different subsets of Meander data, indicating the range of accuracies across the 3-fold CV. The boxplot shows that the proposed model achieved 100% accuracy with the top-6 features of both the ReliefF and Fisher in all three folds. In contrast, adding the seventh feature (RIGHT/LEFT-HANDED) resulted in a more extensive Inter Quartile Range (IQR).

The New Meander dataset was divided into 22 feature subsets based on the ranks given by the three feature ranking algorithms, ReliefF, mRMR, and Fisher, as shown in Table 12. A different number of neurons (70, 80, and 90) are chosen to check any enhancement in the performance of the proposed model. The impact of neurons on the proposed model with all the 22 subsets is shown in Fig. 4(d). It can be observed from Fig. 4(d) that the performance of the proposed PSO-ELM model is increasing gradually with an increase in the number of neurons, and the best performance is obtained with 80 neurons. Table 12 compares the performance of the proposed PSO-ELM model, basic ELM, and two other machine learning models (SVM and RF) on the New Meander dataset using 80 neurons. It can be observed that the proposed model achieved the highest accuracy of 100% with the top-5 (MAX_HT, STD_HT, CNTP, MIN_HT, AGE), top-8, and top-10 feature subsets selected by the mRMR feature ranking algorithm and the corresponding best particles and iterations are 20 and 20 respectively. Also, with top-7 and top-9 feature subsets of the ReliefF feature ranking algorithm, the proposed PSO-ELM model achieved 100% accuracy.

It can also be observed that the performance metrics are also high for the top-5 feature subset selected by the mRMR feature ranking algorithm with an AUC of 1 and 100% precision, recall, and F1-score. The proposed PSO-ELM model outperformed the basic ELM, SVM, and RF models in terms of accuracy, as seen in Fig. 4(c). The ROC-AUC Curve shown in Fig. 4(a) indicates the highest AUC of 1 with the top-5 features ranked by the mRMR feature ranking algorithm, and the second highest AUC of 0.996 is obtained by ReliefF's top-10 and Fisher's top-7 feature subsets. Fig. 4(b) shows the boxplot on different subsets of New Meander data, indicating the range of accuracies across the 3-fold CV. The boxplot shows that mRMR with top-4 features has a larger IQR when compared with other subsets. It can also be observed that the median line of each feature subset of the New Meander dataset is above 0.98, indicating the good performance of the proposed model. 22 feature subsets are created for the Spiral dataset based on the ranks given by three feature ranking algorithms: ReliefF, mRMR, and Fisher. Different numbers of neurons (70, 80, and 90) are chosen for the proposed PSO-ELM model to achieve better performance. The impact of neurons is shown in Fig. 5(d). From Fig. 5(d), it can be observed that the

TABLE 13. Performance of the proposed model on the spiral dataset with 80 neurons.

FS Algorithm	Features	Average Accuracy (%)				Performance Metrics of PSO-ELM (%)				Best Parameters	
		ELM	SVM	RF	PSO-ELM	AUC	Precision	Recall	F1-Score	Particles	Iterations
All	**12**	73.48	70.83	81.06	**100**	1	100	100	100	10	20
ReliefF	4	67.80	81.06	82.20	99.62	0.996	100	99.27	99.63	10	20
	5	72.73	82.20	82.95	98.86	0.989	99.27	98.57	98.92	10	40
	6	69.70	85.23	83.71	99.24	0.993	99.29	99.29	99.28	10	40
	7	73.48	83.33	85.23	**100**	1	100	100	100	10	40
	8	75.00	78.41	84.85	99.62	0.996	100	99.27	99.63	10	60
	9	74.24	79.55	85.61	**100**	1	100	100	100	10	40
	10	77.65	79.17	83.33	99.62	0.996	100	99.291	99.64	10	40
mRMR	4	68.56	73.48	76.52	98.86	0.989	99.27	98.58	98.92	10	60
	5	74.62	76.52	79.92	**100**	1	100	100	100	**20**	**20**
	6	75.38	79.17	79.92	99.62	0.996	99.30	100	99.65	30	20
	7	76.89	78.03	82.20	98.86	0.989	99.29	98.58	98.93	30	20
	8	80.68	78.41	80.68	**100**	1	100	100	100	10	40
	9	78.03	78.03	81.82	99.24	0.993	100	98.58	99.28	10	40
	10	76.14	77.27	81.82	**100**	1	100	100	100	10	40
Fisher	4	60.23	64.39	69.70	98.86	0.988	98.6	99.29	98.94	10	20
	5	57.20	68.18	69.70	99.24	0.993	100	98.58	99.28	10	40
	6	66.29	68.18	73.86	98.86	0.989	99.29	98.57	98.92	20	20
	7	65.91	68.56	76.89	99.62	0.996	100	99.29	99.64	10	20
	8	65.53	67.80	76.14	99.24	0.992	99.29	99.29	99.29	10	20
	9	65.53	68.18	76.52	99.24	0.992	99.29	99.29	99.29	10	20
	10	63.64	66.67	78.03	99.62	0.996	99.30	100	99.64	20	20

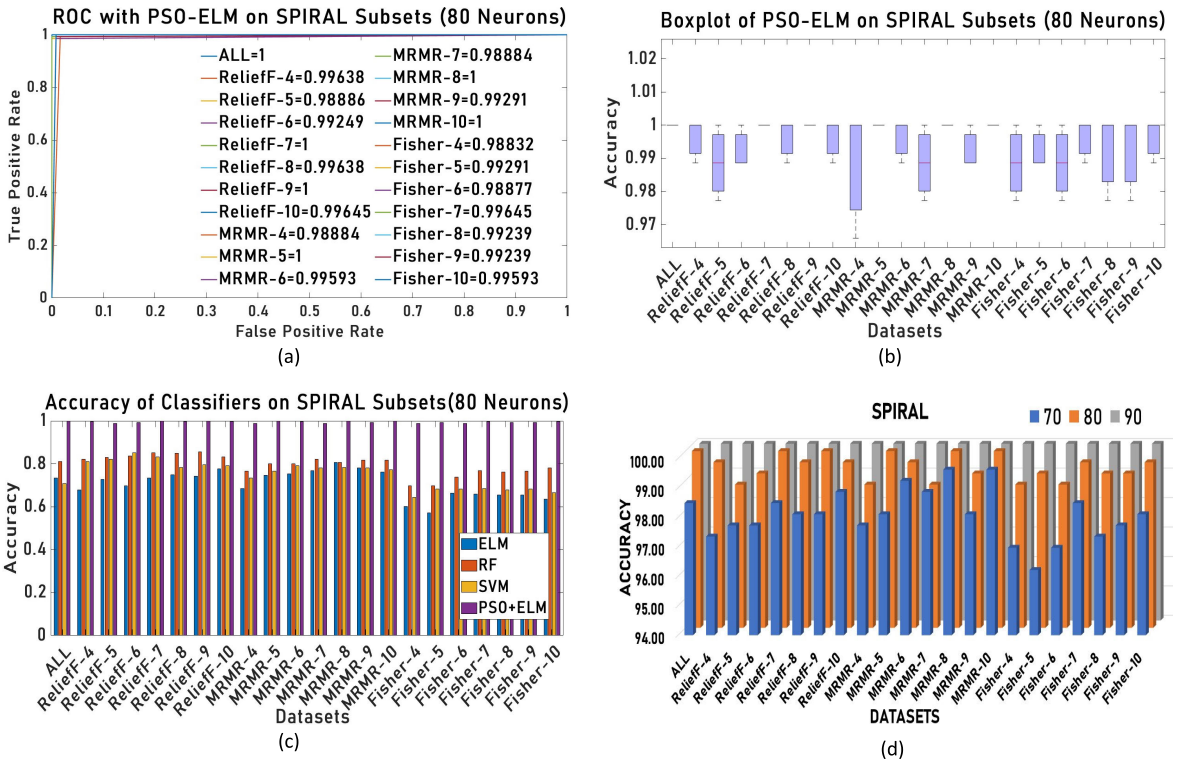

FIGURE 5. SPIRAL dataset: (a)ROC-AUC Curve of the proposed model (b) Box Plot indicating the accuracies obtained in 3-fold CV (c) Bar Graph comparing traditional ML models and proposed model (d) Impact of neurons on the performance of the proposed model.

highest accuracy of 100% is achieved with 80 neurons. The performance of the spiral dataset with 80 neurons is shown in Table 13. The highest accuracy of 100 % with the proposed hybrid PSO-ELM model is achieved with the top-5, top-8, and top-10 feature subsets of mRMR, and the corresponding particles and iterations are 20 and 20, respectively. Also, the proposed model achieved the highest accuracy of 100% when no feature selection was applied, along with the top-7, top-9 features of ReliefF and the top-8, top-10 features of mRMR. Fig. 5(a) gives the ROC-AUC curve of the feature subsets of

TABLE 14. Performance of the proposed model on the new spiral dataset with 80 neurons.

FS Algorithm	Features	Average Accuracy (%)				Performance Metrics of PSO-ELM (%)				Best Parameters	
		ELM	SVM	RF	PSO-ELM	AUC	Precision	Recall	F1-Score	Particles	Iterations
All	**12**	68.18	75.76	85.23	**100**	1	100	100	100	10	40
	4	66.67	78.79	80.68	98.48	0.984	98.61	98.57	98.58	10	20
	5	73.48	82.95	83.71	97.73	0.977	98.58	97.15	97.86	10	60
	6	74.24	81.82	81.82	99.24	0.992	100.00	98.58	99.28	20	20
ReliefF	**7**	73.11	81.82	84.85	**100**	1	100	100	100	10	80
	8	73.11	80.30	84.85	**100**	1	100	100	100	10	40
	9	77.65	79.55	82.58	99.24	0.992	99.28	99.28	99.28	20	40
	10	73.48	76.89	82.95	99.24	0.992	99.28	99.28	99.28	50	20
	4	57.95	71.97	65.15	99.24	0.992	98.60	100	99.29	10	40
	5	78.41	80.30	79.55	98.11	0.981	98.61	97.86	98.22	10	20
	6	74.62	79.92	84.09	**100**	1	100	100	100	**30**	**20**
mRMR	7	72.73	79.17	82.95	99.62	0.996	100.00	99.28	99.63	10	40
	8	76.14	79.17	82.20	98.86	0.988	98.61	99.29	98.93	50	20
	9	73.86	79.17	82.58	99.62	0.996	100	99.29	99.64	10	20
	10	74.24	77.27	83.33	99.62	0.996	100	99.28	99.63	50	20
	4	71.59	80.68	79.17	99.24	0.993	100	98.57	99.28	20	20
	5	77.65	81.06	80.68	99.62	0.996	100	99.28	99.63	40	20
	6	75.76	82.20	83.33	98.86	0.989	99.28	98.57	98.92	10	20
Fisher	7	75.76	82.20	81.44	99.24	0.992	99.29	99.29	99.29	10	20
	8	74.62	78.79	82.95	99.62	0.996	99.31	100	99.65	30	20
	9	73.11	77.27	84.47	**100**	1	100	100	100	20	20
	10	75.76	79.17	85.23	**100**	1	100	100	100	10	20

the spiral dataset with the proposed hybrid PSO-ELM model. An AUC of 1 is obtained with all the features even though no feature selection algorithm is applied. The box plot is shown in Fig. 5(b), indicating the distribution of accuracies obtained in a 3-fold CV for each subset of data. It can be observed that the lowest accuracy is obtained with mRMR's top-4 feature subset. Also, the IQR is highest with mRMR's top-4 feature subset. It can be seen from Fig. 5(b) that the IQR of almost all the feature subsets of mRMR and Fisher are between 98% and 100%, indicating good performance of the proposed model. Fig. 5(c) compares the average accuracy of the proposed PSO-ELM model with the basic ELM, SVM, and RF models. Fig. 5(c) indicates that the proposed model performed better than the basic ELM, SVM, and RF models.

Table 14 shows 22 feature subsets created based on the rankings given by the three feature ranking algorithms on the New Spiral dataset. The impact of the neurons (70, 80, 90) on the proposed PSO-ELM model is shown in Fig. 6(d). It shows that the better accuracy of the proposed PSO-ELM model is obtained with 80 neurons on the New Spiral dataset. Table 14 compares the performance of the proposed PSO-ELM model with the basic ELM, SVM, and RF models. The highest accuracy obtained by the proposed PSO-ELM model is 100%, with the top 6 features ranked by the mRMR feature ranking algorithm, and the corresponding particles and iterations are 30 and 20, respectively.

The proposed model also achieved 100% accuracy with the top-7 and top-8 feature subsets ranked by ReliefF, the top-9 and ten feature subsets of Fisher, and all the features. Four features are common among these feature subsets. These are AGE, STD_HT, MAX_HT, and CNTP. Therefore, these features are the most discriminant features of New Spiral data for diagnosing PD. Table 14 shows that the precision, recall, and F1-score are 100% each, indicating the good performance of the proposed model. Fig. 6(a) shows an AUC of 1 for feature subsets of ReliefF with top-7 and top-8 features, mRMR with top-6 features, and with all the features. Also, the corresponding boxplots have zero IQR, as shown in Fig. 6(b). The top-4 and top-5 feature subsets selected by ReliefF have larger IQRs, indicating a lower accuracy than other feature subsets. Fig. 6(C) compares the accuracies of the proposed PSO-ELM model with the basic ELM, SVM, and RF models. It can be observed that the proposed PSO-ELM model achieved the highest accuracy than the basic ELM, SVM, and RF models. Also, in most subsets, RF performed better than ELM and SVM.

22 feature subsets are created based on the top-ranked features obtained by ReliefF, mRMR, and Fisher feature ranking algorithms on the Speech PD dataset. The performance of the proposed PSO-ELM and basic ELM models is checked by varying the number of neurons (40, 50, and 60). The proposed PSO-ELM model achieved the highest accuracy of 100% with 40 neurons, as seen in Fig. 7(d). However, with all the feature subsets, the proposed model achieved 100% accuracy when the number of neurons was 60. Table 15 compares the performance of the proposed PSO-ELM model with basic ELM, SVM, and RF on the SpeechPD dataset with 40 neurons. The highest accuracy of 100% with ReliefF's top-8 ('spread1', 'MDVP:Flo(Hz)', 'MDVP:Fo(Hz)', 'PPE', 'spread2', 'RPDE', 'HNR', 'Shimmer:APQ3') and top-10 feature subsets is achieved by the proposed model. The corresponding particles and iterations are 50 and 20, respectively.

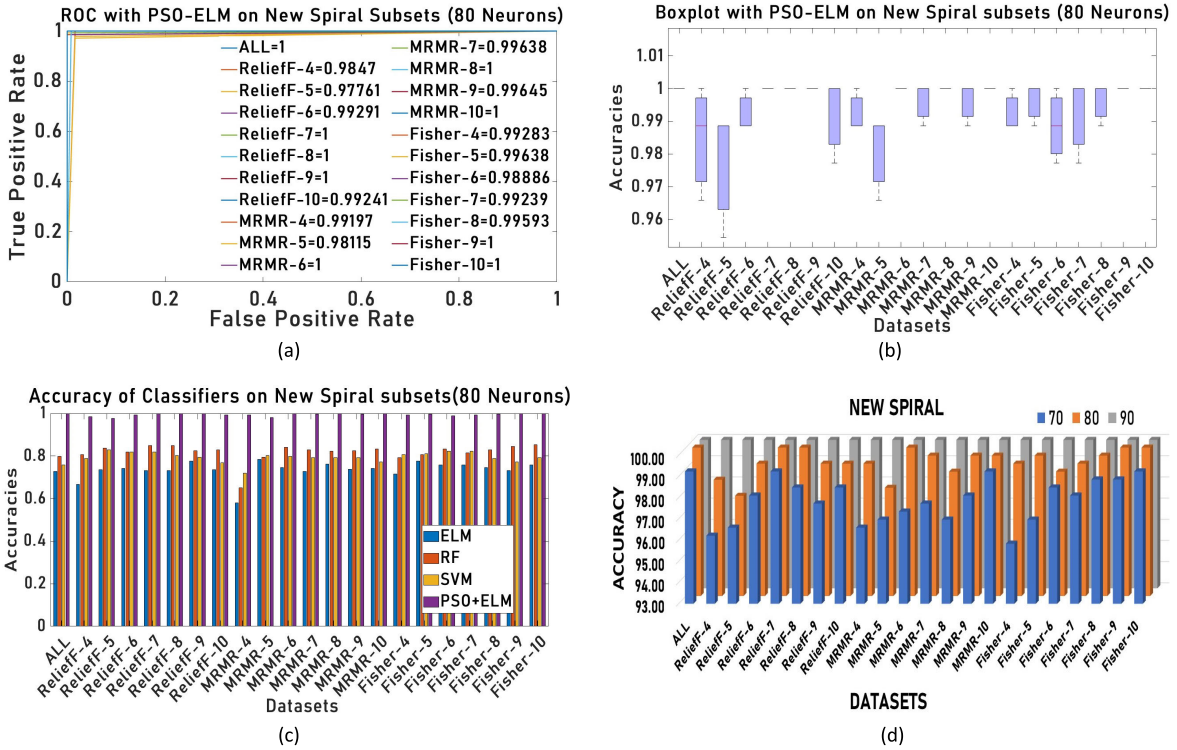

FIGURE 6. NEW SPIRAL dataset: (a)ROC-AUC Curve of the proposed model (b) Box Plot indicating the accuracies obtained in 3-fold CV (c) Bar Graph comparing traditional ML models and proposed model (d) Impact of neurons on the performance of the proposed model.

TABLE 15. Performance of the proposed model on the SpeechPD dataset with 40 neurons.

FS Algorithm	Features	Average Accuracy (%)				Performance Metrics of PSO-ELM (%)				Best Parameters	
		ELM	SVM	RF	PSO-ELM	AUC	Precision	Recall	F1-Score	Particles	Iterations
ALL	**22**	86.67	80.51	90.26	**100**	1	100	100	100	30	20
	5	87.69	89.23	88.72	97.44	0.955	97.33	99.32	98.32	20	20
	8	88.72	91.28	92.31	**100**	1	100	100	100	**50**	**20**
	10	87.69	91.28	92.31	**100**	1	100	100	100	10	40
ReliefF	13	85.13	87.69	93.33	99.49	0.99	99.33	100	99.66	10	20
	14	84.62	87.69	90.77	98.46	0.969	98.08	100	99.01	10	20
	15	90.77	81.54	91.79	98.46	0.969	98.08	100	99.01	10	20
	20	82.56	81.54	90.26	98.46	0.976	98.65	99.32	98.98	10	20
	5	84.62	86.15	88.21	97.44	0.948	96.77	100	98.34	10	20
	8	83.59	85.13	90.26	97.95	0.965	98	99.32	98.65	10	20
	10	83.08	85.64	90.26	96.92	0.945	96.69	99.32	97.99	10	40
mRMR	13	83.59	86.67	88.21	96.41	0.941	96.74	98.64	97.66	20	60
	14	82.05	90.26	89.23	97.95	0.965	98.01	99.32	98.65	10	40
	15	87.18	84.10	89.74	99.49	0.997	100	99.32	99.66	30	20
	20	83.08	81.54	90.26	99.49	0.99	99.33	100	99.66	10	40
	5	83.59	86.67	88.72	96.92	0.945	96.69	99.32	97.99	30	20
	8	85.13	89.74	87.69	98.46	0.969	98.03	100	99	10	20
	10	82.05	89.23	88.21	98.46	0.976	98.65	99.32	98.98	20	20
Fisher	13	85.13	88.72	87.69	96.92	0.945	96.73	99.32	97.99	10	20
	14	78.97	88.72	86.15	96.92	0.952	97.33	98.64	97.97	50	20
	15	85.13	85.13	88.72	97.44	0.969	98.63	97.96	98.29	20	40
	20	86.67	82.56	91.79	98.97	0.993	100	98.64	99.31	10	20

From Table 15, it can be observed that the proposed PSO-ELM model achieved 100% accuracy with all the features. The second highest accuracy of 99.49% is obtained with the top-13 features of ReliefF and the top-15 and top-20 features of mRMR feature ranking algorithms. All the top 8 features of ReliefF are common in the top 20 features of mRMR. Ten features are common among the top 13 features ranked by ReliefF and the top 15 by the mRMR feature ranking algorithm. The performance metrics like AUC, Precision, Recall, and F1-score are also high, with 100% each

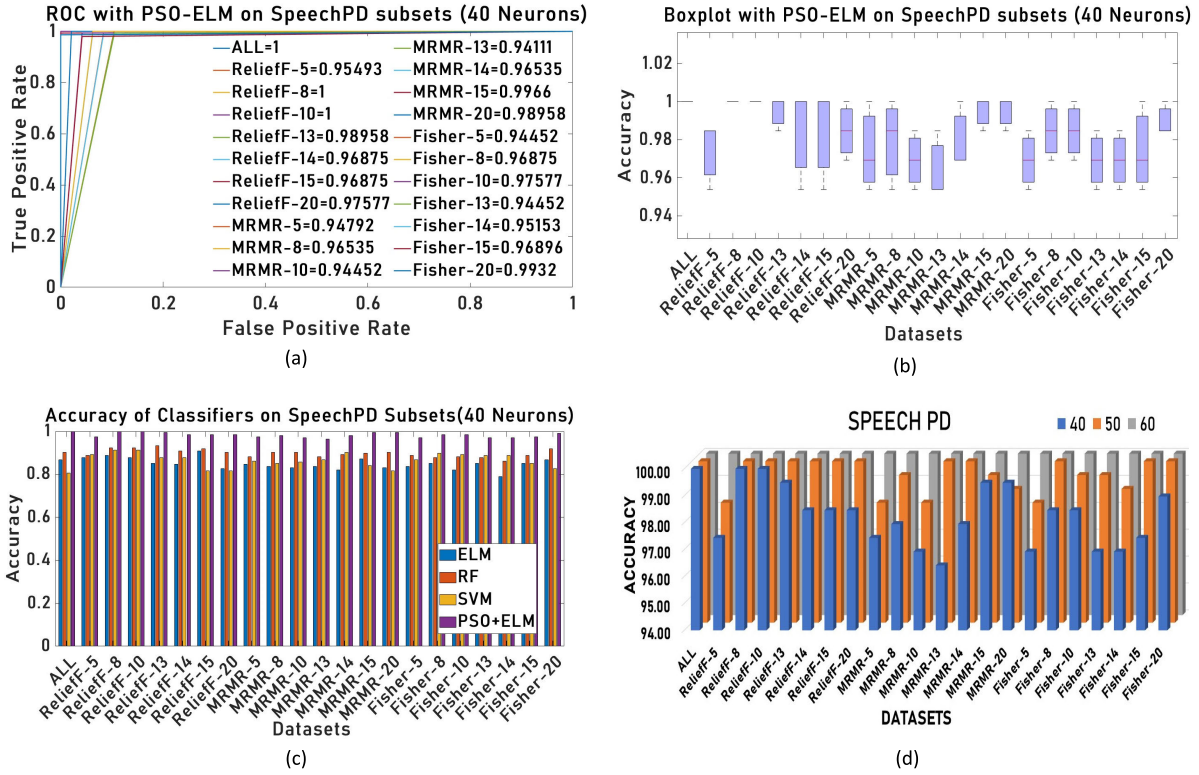

FIGURE 7. SPEECHPD dataset: (a)ROC-AUC Curve of the proposed model (b) Box Plot indicating the accuracies obtained in 3-fold CV (c) Bar Graph comparing traditional ML models and proposed model (d) Impact of neurons on the performance of the proposed model.

for the top-8 and top-10 feature subsets ranked by ReliefF. The F1-score for all the feature subsets is between 97.66% and 100%, indicating a strong balance between precision and recall, as seen in Table 15. The ROC curve for each subset of features with the proposed PSO-ELM model is shown in Fig. 7(a). Fig. 7(a) shows that the top-8 and top-10 feature subsets ranked by the ReliefF feature ranking algorithm have an AUC of 1. The accuracies obtained in each fold of a 3-fold CV technique are plotted with the boxplot in Fig. 7(b). Fig. 7(b) reveals that the IQRs for top-15 and top-20 mRMR feature subsets and ReliefF's top-13 feature subsets are very narrow. Also, the median value of the IQR for most feature subsets shows more than 96% accuracy. Therefore, all these feature subsets are relevant. Fig. 7(c) compares the proposed PSO-ELM model and the basic ELM, SVM, and RF models. It can be seen from Fig. 7(c) that the proposed model outperformed other models.

The PMSR dataset is divided into 16 feature subsets based on the ranks given by feature ranking algorithms: ReliefF, mRMR, and Fisher. Different numbers of neurons (90, 100, 120, 150, 200, 300, 350, and 370) are chosen for the performance enhancement of the proposed PSO-ELM model. The impact of neurons on the proposed PSO-ELM model can be seen in Fig. 8(d). From Fig. 8(d), it can be observed that the performance of the proposed PSO-ELM model is increasing gradually with the increase in the number of neurons, and the best performance is achieved at 370 neurons. Table 16 compares the performance of the proposed PSO-ELM model (370 neurons) with the basic ELM, SVM, and RF models. The highest accuracy of 100% is obtained with top-10 (Fraction of locally unvoiced frames, AC, Maximum pitch, HTN, Jitter (local, absolute), NTH, Shimmer (local, dB), Shimmer (apq11), Mean period, Number of periods), top-15, top-20 and top-23 feature subsets of ReliefF feature ranking algorithm. The corresponding particles and iterations are 10 and 20, respectively. From Table 16 it can be observed that the highest accuracy of 100% is also achieved with the top-10(Shimmer (apq3), Jitter (local, absolute), Jitter (local), Shimmer (local), Shimmer (local, dB), Jitter (rap), Number of pulses, Fraction of locally unvoiced frames, Standard deviation, Standard deviation of period), top-20 and top-23 feature subsets selected by mRMR and top-20 and top-23 feature subsets selected by Fisher feature ranking algorithms. Also, the proposed PSO-ELM model achieved the highest accuracy of 100% with all the 26-features. The top-10 features ranked by ReliefF and mRMR have three common features ('Fraction of locally unvoiced frames', 'Jitter (local, absolute)', 'Shimmer (local, dB)'), which indicates that these features have a significant impact on the accuracy of the proposed PSO-ELM model. The remaining seven different features from the top 10 features of ReliefF and mRMR are also very significant as these features help achieve the best accuracy with the proposed PSO-ELM model. Fig. 8(a) shows that the AUC is 1 for all the feature subsets that

TABLE 16. Performance of the proposed model on the PMSR dataset with 370 neurons.

FS Algorithm	Features	Average Accuracy (%)				Performance Metrics of PSO-ELM (%)				Best Parameters	
		ELM	SVM	RF	PSO-ELM	AUC	Precision	Recall	F1-Score	Particles	Iterations
All	**26**	65.81	62.50	72.93	**100**	1	100	100	100	10	40
ReliefF	5	61.92	66.06	67.05	99.83	0.998	99.86	99.86	99.85	10	60
	10	64.07	69.21	70.03	**100**	1	100	100	100	**10**	**20**
	15	66.97	66.97	71.11	**100**	1	100	100	100	20	20
	20	65.40	65.81	72.10	**100**	1	100	100	100	20	20
	23	63.41	65.07	71.36	**100**	1	100	100	100	10	20
mRMR	5	58.69	62.59	61.18	99.67	0.996	99.57	99.86	99.71	10	20
	10	65.56	67.55	68.71	**100**	1	100	100	100	**10**	**20**
	15	63.83	65.73	70.70	99.83	0.998	99.85	99.85	99.85	30	20
	20	65.98	65.90	70.86	**100**	1	100	100	100	10	40
	23	67.72	63.50	72.10	**100**	1	100	100	100	40	20
Fisher	5	60.84	63.91	63.16	99.50	0.995	99.57	99.56	99.56	20	60
	10	62.83	67.39	68.38	99.75	0.997	99.71	99.85	99.78	10	20
	15	64.57	67.97	69.54	99.83	0.998	99.86	99.85	99.85	10	40
	20	67.47	66.14	71.52	**100**	1	100	100	100	10	20
	23	65.56	65.73	73.10	**100**	1	100	100	100	10	20

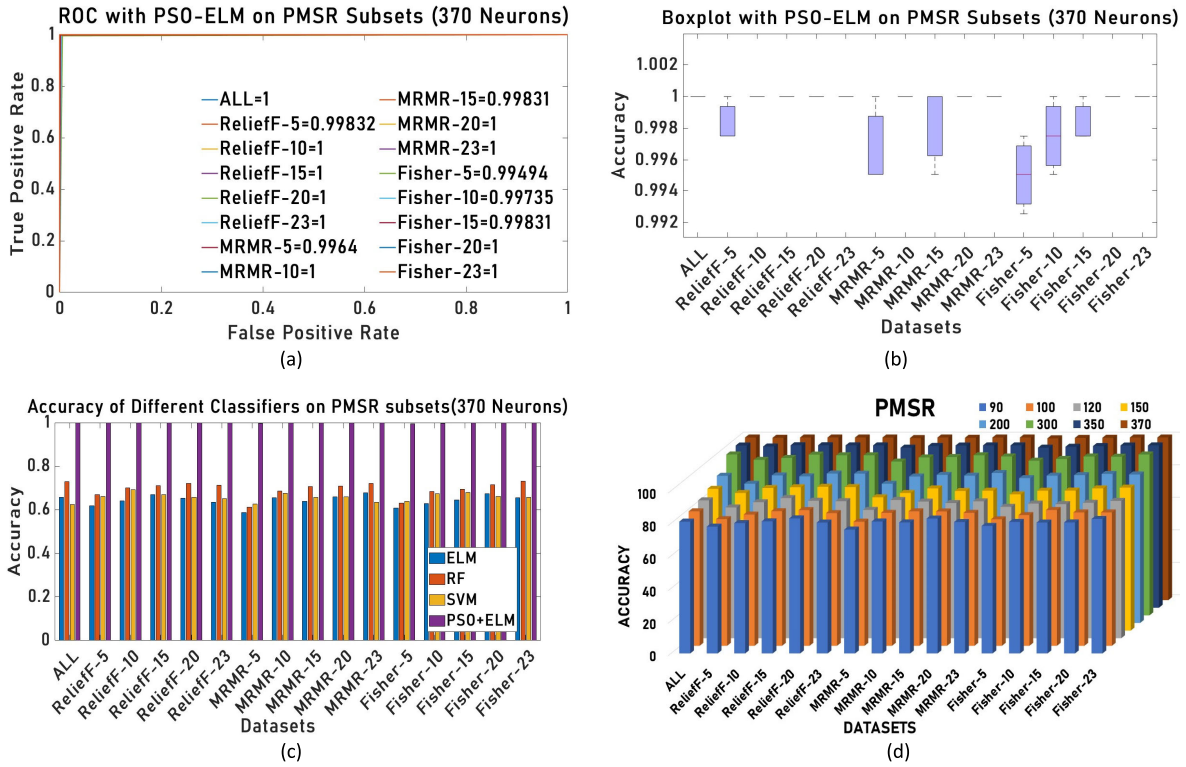

FIGURE 8. PMSR dataset: (a)ROC-AUC Curve of the proposed model (b) Box Plot indicating the accuracies obtained in 3-fold CV (c) Bar Graph comparing traditional ML models and proposed model (d) Impact of neurons on the performance of the proposed model.

achieved 100% accuracy. The second highest AUC obtained is 0.998, with the top-5 features of ReliefF. From Table 16, it can be observed that the precision, recall, and F1-score are also achieved as 100% each with top-10, top-15, top-20, top-23 feature subsets of ReliefF, top-10, top-20, top-23 feature subsets of mRMR and top-20, top-23 feature subsets of Fisher algorithm. From Fig. 8(b), the mRMR's top-5, top-15 feature subsets and Fisher's top-5, top-10, and top-15 feature subsets IQR is higher when compared to all other feature subsets, indicating higher variability in classification performance among folds. ReliefF's top-10, top-15, top-20, and top-23 feature subsets have zero IQR, indicating more consistent performance across the 3-fold CV. Similarly, Fisher's top-20 and top-23 feature subsets have zero IQR. Fig. 8(c) illustrates a performance comparison between the proposed model and the basic ELM, SVM, and RF. The results show that the proposed model performed better than all the three ML models across all feature subsets.

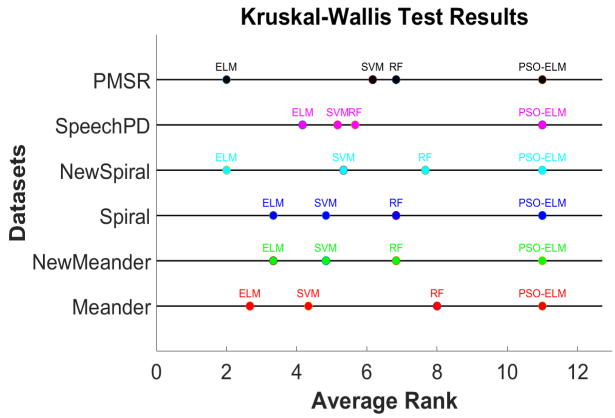

FIGURE 9. Average ranks of each dataset using the Kruskal-Wallis test.

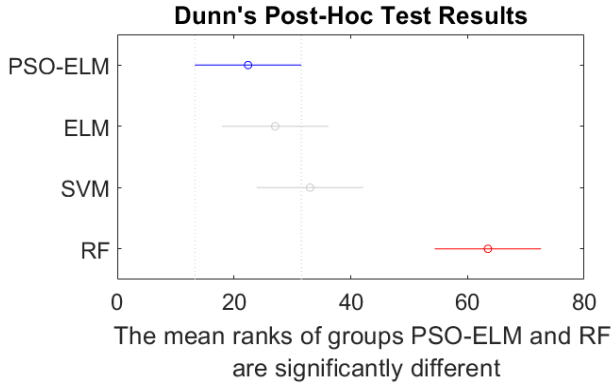

FIGURE 10. Dunn's post hoc test results with multiple comparisons on six datasets.

F. SIGNIFICANCE TEST RESULT

The Kruskal-Wallis test is commonly used to check whether the models applied (PSO-ELM, ELM, SVM, and RF) are significantly different or not.The significance level of $\boldsymbol{\alpha = 0.05}$ is used. The input data consists of testing accuracy values from the 3-fold cross-validation of the subset that achieved the highest accuracy for each dataset. Table 17 shows the p-value and Chi-square statistics for each data subset. The test results showed varying significance levels across datasets, with p-values ranging from 0.0162 to 0.0834 on different datasets. Since all the p-values are nearer to zero, it can be concluded that the models are significantly different. The average ranks obtained in the Kruskal-Wallis test shown in Fig. 9 reveal that the proposed PSO-ELM model consistently achieved higher ranks than other models across the six datasets, indicating superior performance.

To investigate these differences further, Dunn's post-hoc test is applied. Dunn's post hoc test results are shown in Fig. 10. Each horizontal line indicates the rank range of the ML models, and the circle represents the mean rank of each model. The blue line indicates the proposed PSO-ELM model and the red line indicates the models that are significantly different from the proposed model. Fig. 10 shows that the mean ranks of PSO-ELM and RF were significantly different, confirming that the PSO-ELM model outperformed RF in terms of classification accuracy. The ranks of the ELM and SVM models slightly overlap with those of the PSO-ELM model. This analysis supports the conclusion that PSO-ELM is more effective for the datasets used, with statistically significant improvements in performance compared to other machine learning models.).

Several significant aspects contribute to the PSO-ELM model's performance, as seen by the findings presented and statistical tests performed. First, PSO effectively optimizes the ELM's hidden weights and biases, guaranteeing that the model converges on an optimal solution by iteratively refining these parameters to enhance classification accuracy. In contrast, traditional ELM chooses parameters at random, resulting in less consistent performance. Second, the model incorporates filter-based feature selection approaches (Fisher, mRMR, and ReliefF) ensures that only the most relevant and informative feature subsets are employed, decreasing redundancy and improving predictive capability. Additionally, using grid search to customize PSO's hyperparameters, such as the number of particles and iterations, enables the model to adapt successfully to different datasets with varying properties. These variables combine to allow the PSO-ELM model to consistently achieve statistically significant improvement in classification accuracy over baseline models such as ELM, SVM, and RF across six PD datasets.

TABLE 17. ANOVA tables for the Kruskal-Wallis test with each data subset.

Data Subset	p-value	Chi-square value
Meander_ReliefF_6	0.0203	9.8038
New Meander_mRMR_5	0.0506	7.7900
Spiral_mRMR_5	0.0506	7.7900
NewSpiral_mRMR_6	0.0162	10.2984
SpeechPD_ReliefF_8	0.0834	6.6631
PMSR_ReliefF_10	0.0227	9.5647

G. COMPARISION OF THE PROPOSED MODEL WITH THE PREVIOUS WORKS

Table 18 compares the proposed PSO-ELM model and previous works across various datasets and their respective number of features. The proposed PSO-ELM model outperforms existing works on all six datasets using a 3-fold CV. In [27], the authors used the CNN model on the Meander dataset and achieved accuracy and F1-score of 88.6% and 82.3%, respectively. The authors of [26] used AlexNet, VGG16, and VGG19 models on the same dataset and achieved an accuracy of 89.19% using a 90%-10% train-test split. In [4], the authors used the Modified Grey Wolf optimization algorithm for feature selection and achieved an accuracy of 93.04%. However, the proposed PSO-ELM model (with 100 neurons) achieved the highest accuracy and F1-Score of 100%, each with top-6 feature subsets ranked by ReliefF and Fisher feature ranking algorithms.

TABLE 18. Comparison of the proposed PSO-ELM model with the previous works.

Ref	Year	Technique	Dataset Used	# of features	Acc (%)	F1-Score	Validation Method
[27]	2024	CNN	Meander	-	88.6	82.3	-
[26]	2021	AlexNet, VGG16, VGG19		-	89.19	-	90%-10% split
[4]	2018	MGWO using RF and MGWO using decision trees		<10	92.41	-	70%-30% split
Proposed	**-**	**PSO_ELM (100 Neurons) with ReliefF, Fisher**		**6**	**100**	**100**	**Stratified 3-fold CV**
[27]	2024	CNN	Spiral	-	90.5	91.8	-
[26]	2021	AlexNet		-	86.49	-	90%-10% split
[4]	2018	MGWO using RF		<10	93.04		70%-30% split
Proposed	**-**	**PSO-ELM (80 Neurons) with mRMR**		**5**	**100**	**100**	**Stratified 3-fold CV**
[19]	2023	K-means with BGOA	New Meander	5	82.5	-	10-fold CV
[26]	2021	AlexNet		-	92.31	-	90%-10% split
[23]	2020	PCA with an Ensemble of RF and extremely random trees		200 c	78.18	78.85	Stratified 5-fold CV
[25]	2020	RF with PCA		200 c	82.5	70.6	Stratified 5-fold CV
Proposed	**-**	**PSO-ELM (80 Neurons) with mRMR**		**5**	**100**	**100**	**Stratified 3-fold CV**
[19]	2023	AEVFC with BGW	New Spiral	5	82.77	-	10-fold CV
[26]	2021	VGG16, VGG19, ResNet50		-	88.46	-	90%-10% split
[23]	2020	PCA with an Ensemble of RF and extremely random trees		200 c	81.17	80.51	Stratified 5-fold CV
[25]	2020	RF with PCA		200 c	80.2	80.3	Stratified 5-fold CV
Proposed	**-**	**PSO-ELM (80 Neurons) with mRMR**		**6**	**100**	**100**	**Stratified 3-fold CV**
[64]	2023	local dynamic feature selection fusion method	Speech PD	5	98.2	-	10-fold CV
[16]	2022	Binary Improved Grey Wolf Optimization with adaptive k-NN		4	95.66	-	K-fold CV (k value Varied from 3 to 20)
[13]	2021	Weighted Local Discriminant Preservation Projection Embedded Ensemble with ELM (5000 neurons)		-	89.17	-	Hold-out CV with 33% of train, test and validation split.
[15]	2020	SVM with RFE		13	93.84	-	-
[4]	2018	MGWO using RF		>10	93.87	-	70%-30% split
Proposed	**-**	**PSO-ELM (40 Neurons) with ReliefF**		**8**	**100**	**100**	**Stratified 3-fold CV**
[9]	2022	Vowel-centric ANN model	PMSR	-	91	0.992	70%-30% split
[13]	2021	Weighted Local Discriminant Preservation Projection Embedded Ensemble with ELM (5000 neurons)		-	76.25	0.992	33% of train, test and validation split.
[4]	2018	MGWO using RF		>10	100	0.996	70%-30% split
Proposed	**-**	**PSO-ELM (370 neurons) with ReliefF, mRMR**		**10**	**100**	**100**	**Stratified 3-fold CV**

= number; Acc = accuracy; CNN = Convolutional neural network; MGWO = modified grey wolf optimization; BGOA = Binary grasshopper optimization algorithm; AEVFC =Algorithm of extreme values of features in classes; ANN =artificial neural network; PCA = Principal component analysis; RFE = Recursive feature elimination, c=Components of PCA.

Similarly, for the Spiral dataset, the authors of [26] and [27] achieved an accuracy of 90.5% and 86.49 %, respectively. In [4], the authors used the Modified Grey Wolf optimization algorithm for feature selection and achieved an accuracy of 92.41%. However, the proposed PSO-ELM model (with 80 neurons) achieved the highest accuracy and F1-score of 100% each for the spiral dataset.

Accuracy and an F1-score of 100% each on the New Meander dataset are achieved by the proposed model with 80 neurons and a top-5 feature subset ranked by the mRMR feature ranking algorithm. This outperforms most previous works, including [19], which employed K-means and the Binary grasshopper optimization algorithm (BGOA) for selecting the features and achieved an accuracy of 82.5%. The authors of [26] used AlexNet and achieved an accuracy of 92.31%. In [23], the authors employed PCA with an ensemble of RF and extremely random trees and obtained an accuracy of 78.18% and an F1-score of 78.85%. Furthermore, in the paper [25] Using RF and PCA with 200 components, the authors obtained an accuracy of 82.5% and an F1-score of 70.6%. However, the proposed PSO-ELM model used only five features and achieved the highest accuracy of 100%, which is much better than the previous works. Therefore, these features may be the most relevant for PD diagnosis.

Similarly, for the New Spiral dataset, in [19] the authors used the extreme values of features in the classification algorithm and achieved an accuracy of 82.77%. The authors of [26] achieved an accuracy of 88.46%. In [23], the authors used 200 PCA components with an RF ensemble and extremely random trees. The authors achieved accuracy and F1-score of 81.17% and 80.51%, respectively, with a stratified 5-fold CV. The authors of [25] used RF with PCA and achieved F1-score and accuracy of 80.3% and 80.2%, respectively, with a stratified 5-fold CV. However, the proposed PSO-ELM model (with 80 neurons) achieved the highest accuracy and F1-score of 100%, each with the top-6 feature subsets ranked by the mRMR feature ranking algorithm.

For the SpeechPD dataset, the authors of [64] employed the Local dynamic feature selection fusion method and

achieved 98.2% accuracy with five features. The authors of [16] used Binary improved grey wolf optimization to select four relevant features of the SpeechPD dataset and achieved an accuracy of 95.66% using k-fold CV by varying the k-value from 3 to 20. In [13] an accuracy of only 89.17% is achieved with the Weighted local discriminant preservation projection embedded ensemble with ELM. The authors employed 5000 neurons in the hidden layer of ELM. The authors of [15] employed Recursive feature elimination (RFE) with SVM and achieved an accuracy of 93.84% with 13 features. In [4], the authors employed modified grey wolf optimization with RF and achieved an accuracy of 93.87% with more than ten features. In contrast, the proposed model achieved the highest accuracy and F1-score of 100%, each with only 40 neurons in the hidden layer of the PSO-ELM model with a top-8 feature subset ranked by the ReliefF feature ranking algorithm for the same dataset.

Finally, for the PMSR dataset, in paper [13], the authors used a Weighted Local Discriminant Embedded Ensemble with ELM (5000 neurons) and achieved an accuracy of 76.25%. The author of [9] has divided the samples based on the multiple speech recordings, such as vowels, numbers, and words in the dataset, concluding that better performance is obtained only with vowel data. They achieved an accuracy of 91% with vowel data. The authors of [4] used a modified grey wolf optimization algorithm with RF and achieved an accuracy of 100%. The authors employed a 70%-30% train-test split for the model evaluation. However, the proposed PSO-ELM model with 370 neurons has achieved the highest accuracy of 100% with only the top-10 features ranked by the ReliefF feature ranking algorithm. The same highest accuracy of 100% is achieved by the top-10 features ranked by the mRMR feature ranking algorithm with multiple speech samples. As discussed in Section IV-D, the PMSR dataset contains various relevant features for diagnosing PD, indicating the proposed PSO-ELM model's ability to manage diverse data efficiently to achieve the highest accuracy.

Also, the proposed model achieved better results with other evaluation metrics (precision, recall, and AUC) than the existing works. Table 18 shows that the proposed PSO-ELM model outperformed the existing methods with almost all the datasets, indicating its robustness.

V. CONCLUSION AND FUTURE WORKS

This paper proposed a PSO-ELM model for diagnosing PD on six publicly available datasets: Meader, Spiral, New Meander, New Spiral, SpeechPD, and PMSR. Three feature ranking algorithms, ReliefF, mRMR, and Fisher, are employed to find the most discriminative features for all six PD datasets. The proposed model was evaluated on 126 different feature subsets created from the six datasets to better understand their impact on the performance of PD diagnosis. For all six datasets, the highest accuracy obtained is 100%. The proposed PSO-ELM model's performance is compared with that of basic ML models (ELM, SVM, and RF) and previous works. Among the basic models, RF performed better than other ML models.

The proposed hybrid PSO-ELM model outperformed the previous works regarding the accuracy of the PD diagnosis with all six datasets. Also, in most cases, the proposed PSO-ELM model requires a smaller number of features than previous works. The proposed model's advantage is that it learns quickly and works well in the presence of impressions that occur very commonly in medical datasets. However, as a variation of a neural network model, the explainability is limited.

In the future, the proposed PSO-ELM model's performance may be validated with other datasets. Adapting the model for real-time processing and combining it with other diagnostic methods could enhance its utility in clinical settings.

VI. DECLARATIONS

A. ETHICAL APPROVAL

This manuscript reports studies that do not involve human participants, data, tissue, or animals. It utilizes only publicly available data and patient details are not disclosed.

B. CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest related to the content of this article.

C. AUTHORS CONTRIBUTIONS

Gunakala Archana contributed to the conceptualization, methodology, software development, data curation, formal analysis, and drafting of the original manuscript. Afzal Hussain Shahid was involved in the conceptualization, methodology, supervision, result analysis, writing, review, and editing of the final draft and validation.

D. FUNDING

No specific funding was provided for this research by any agency.

E. AVAILABILITY OF DATA AND MATERIALS

Data used in this study is publicly available in the UCI repository and Kaggle database and can be downloaded from the website:

https://archive.ics.uci.edu/datasets?search=parkinson

https://www.kaggle.com/datasets/claytonteybauru/spiral-handpd

ACKNOWLEDGMENT

The authors acknowledge the financial assistance provided, which made this work possible.

REFERENCES

[1] J. Goyal, P. Khandnor, and T. C. Aseri, "A hybrid approach for Parkinson's disease diagnosis with resonance and time-frequency based features from speech signals," *Expert Syst. Appl.*, vol. 182, Nov. 2021, Art. no. 115283, doi: 10.1016/j.eswa.2021.115283.

[2] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Syst. Appl.*, vol. 137, pp. 22–28, Dec. 2019, doi: 10.1016/j.eswa.2019.06.052.

[3] I. Kaya, A. Nilsson, D. Luptáková, Y. He, T. Vallianatou, P. Bjärterot, P. Svenningsson, E. Bezard, and P. E. Andrén, "Spatial lipidomics reveals brain region-specific changes of sulfatides in an experimental MPTP Parkinson's disease primate model," *NPJ Parkinson's Disease*, vol. 9, no. 1, p. 118, Jul. 2023, doi: 10.1038/s41531-023-00558-1.

[4] P. Sharma, S. Sundaram, M. Sharma, A. Sharma, and D. Gupta, "Diagnosis of Parkinson's disease using modified grey wolf optimization," *Cogn. Syst. Res.*, vol. 54, pp. 100–115, May 2019, doi: 10.1016/j.cogsys.2018.12.002.

[5] R. Guatelli, V. Aubin, M. Mora, J. Naranjo-Torres, and A. Mora-Olivari, "Detection of Parkinson's disease based on spectrograms of voice recordings and extreme learning machine random weight neural networks," *Eng. Appl. Artif. Intell.*, vol. 125, Oct. 2023, Art. no. 106700, doi: 10.1016/j.engappai.2023.106700.

[6] A. Govindu and S. Palwe, "Early detection of Parkinson's disease using machine learning," *Proc. Comput. Sci.*, vol. 218, pp. 249–261, Jun. 2023, doi: 10.1016/j.procs.2023.01.007.

[7] B. T. Harel, M. S. Cannizzaro, H. Cohen, N. Reilly, and P. J. Snyder, "Acoustic characteristics of parkinsonian speech: A potential biomarker of early disease progression and treatment," *J. Neurolinguistics*, vol. 17, no. 6, pp. 439–453, Nov. 2004, doi: 10.1016/j.jneuroling.2004.06.001.

[8] O. Tucha, L. Mecklinger, J. Thome, A. Reiter, G. L. Alders, H. Sartor, M. Naumann, and K. W. Lange, "Kinematic analysis of dopaminergic effects on skilled handwriting movements in Parkinson's disease," *J. Neural Transmiss.*, vol. 113, no. 5, pp. 609–623, May 2006, doi: 10.1007/s00702-005-0346-9.

[9] W. Liu, J. Liu, T. Peng, G. Wang, V. E. Balas, O. Geman, and H.-W. Chiu, "Prediction of Parkinson's disease based on artificial neural networks using speech datasets," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 10, pp. 13571–13584, Oct. 2023, doi: 10.1007/s12652-022-03825-w.

[10] L. Moro-Velazquez, J. Gomez-Garcia, N. Dehak, and J. I. Godino-Llorente, "New tools for the differential evaluation of Parkinson's disease using voice and speech processing," *Int. Speech Commun. Assoc.*, vol. 2021, pp. 165–169, Mar. 2021, doi: 10.21437/iberspeech.2021-36.

[11] N. K. Dastjerd, O. C. Sert, T. Ozyer, and R. Alhajj, "Fuzzy classification methods based diagnosis of Parkinson's disease from speech test cases," *Current Aging Sci.*, vol. 12, no. 2, pp. 100–120, Nov. 2019, doi: 10.2174/1874609812666190625140311.

[12] L. Ali, S. U. Khan, M. Arshad, S. Ali, and M. Anwar, "A multi-model framework for evaluating type of speech samples having complementary information about Parkinson's disease," in *Proc. Int. Conf. Electr., Commun., Comput. Eng. (ICECCE)*, Swat, Pakistan, Jul. 2019, pp. 1–5.

[13] Y. Liu, Y. Li, X. Tan, P. Wang, and Y. Zhang, "Local discriminant preservation projection embedded ensemble learning based dimensionality reduction of speech data of Parkinson's disease," *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102165, doi: 10.1016/j.bspc.2020.102165.

[14] Y. Li, C. Liu, P. Wang, H. Zhang, A. Wei, and Y. Zhang, "Envelope multi-type transformation ensemble algorithm of Parkinson speech samples," *Appl. Intell.*, vol. 53, no. 12, pp. 15957–15978, Jun. 2023, doi: 10.1007/s10489-022-04345-y.

[15] Z. K. Senturk, "Early diagnosis of Parkinson's disease using machine learning algorithms," *Med. Hypotheses*, vol. 138, May 2020, Art. no. 109603, doi: 10.1016/j.mehy.2020.109603.

[16] R. Ramasamy Rajammal, S. Mirjalili, G. Ekambaram, and N. Palanisamy, "Binary grey wolf optimizer with mutation and adaptive K-nearest neighbour for feature selection in Parkinson's disease diagnosis," *Knowl.-Based Syst.*, vol. 246, Jun. 2022, Art. no. 108701, doi: 10.1016/j.knosys.2022.108701.

[17] I. Luna-Ortiz, M. Aldape-Pérez, A. V. Uriarte-Arcia, A. Rodríguez-Molina, A. Alarcón-Paredes, and E. Ventura-Molina, "Parkinson's disease detection from voice recordings using associative memories," *Healthcare*, vol. 11, no. 11, p. 1601, May 2023, doi: 10.3390/healthcare11111601.

[18] R. Nijhawan, M. Kumar, S. Arya, N. Mendirtta, S. Kumar, S. K. Towfek, D. S. Khafaga, H. K. Alkahtani, and A. A. Abdelhamid, "A novel artificial-intelligence-based approach for classification of Parkinson's disease using complex and large vocal features," *Biomimetics*, vol. 8, no. 4, p. 351, Aug. 2023, doi: 10.3390/biomimetics8040351.

[19] K. Sarin, M. Bardamova, M. Svetlakov, N. Koryshev, R. Ostapenko, A. Hodashinskaya, and I. Hodashinsky, "A three-stage fuzzy classifier method for Parkinson's disease diagnosis using dynamic handwriting analysis," *Decis. Anal. J.*, vol. 8, Sep. 2023, Art. no. 100274, doi: 10.1016/j.dajour.2023.100274.

[20] S. Hadadi and S. P. Arabani, "A novel approach for Parkinson's disease diagnosis using deep learning and Harris hawks optimization algorithm with handwritten samples," *Multimedia Tools Appl.*, vol. 83, no. 34, pp. 81491–81510, Mar. 2024, doi: 10.1007/s11042-024-18584-3.

[21] B. K. Tripathy, P. K. R. Maddikunta, Q.-V. Pham, T. R. Gadekallu, K. Dev, S. Pandya, and B. M. ElHalawany, "Harris hawk optimization: A survey onVariants and applications," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–20, Jun. 2022, doi: 10.1155/2022/2218594.

[22] S. Agrawal and S. P. Sahu, "Image-based Parkinson disease detection using deep transfer learning and optimization algorithm," *Int. J. Inf. Technol.*, vol. 16, no. 2, pp. 871–879, Feb. 2024, doi: 10.1007/s41870-023-01601-3.

[23] S. Xu, Z. Zhu, and Z. Pan, "A cascade ensemble learning model for Parkinson's disease diagnosis using handwritten sensor signals," *J. Phys., Conf. Ser.*, vol. 1631, no. 1, Sep. 2020, Art. no. 012168, doi: 10.1088/1742-6596/1631/1/012168.

[24] S. Naz, I. Kamran, S. Gul, F. Hadi, and F. Khalifa, "Multi-model fusion of CNNs for identification of Parkinson's disease using handwritten samples," *IEEE Access*, vol. 11, pp. 135600–135608, 2023, doi: 10.1109/ACCESS.2023.3337804.

[25] S. Xu and Z. Pan, "A novel ensemble of random forest for assisting diagnosis of Parkinson's disease on small handwritten dynamics dataset," *Int. J. Med. Informat.*, vol. 144, Dec. 2020, Art. no. 104283, doi: 10.1016/j.ijmedinf.2020.104283.

[26] I. Kamran, S. Naz, I. Razzak, and M. Imran, "Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease," *Future Gener. Comput. Syst.*, vol. 117, pp. 234–244, Apr. 2021, doi: 10.1016/j.future.2020.11.020.

[27] P. Sridevi, S. Himaja, S. B. Nandini, and M. Bhargavi. (Apr. 2024). *Early Detection of Parkinson's Disease Using Convolutional Neural Networks on Handwritten Patterns*. [Online]. Available: https://www.researchgate.net/publication/380075243

[28] R. F. Mansour, "Quantum mayfly optimization based feature subset selection with hybrid CNN for biomedical Parkinson's disease diagnosis," *Neural Comput. Appl.*, vol. 36, no. 15, pp. 8383–8396, May 2024, doi: 10.1007/s00521-024-09516-1.

[29] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics," in *Proc. 29th SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Oct. 2016, pp. 340–346, doi: 10.1109/SIBGRAPI.2016.054.

[30] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009.

[31] M. Little. (2008). *Parkinsons*. UCI Mach. Learn. Repository. Accessed: Aug. 11, 2023. [Online]. Available: https://archive.ics.uci.edu/dataset/174/parkinsons

[32] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 4, pp. 828–834, Jul. 2013, doi: 10.1109/JBHI.2013.2245674.

[33] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomed. Eng. OnLine*, vol. 6, no. 1, p. 23, 2007, doi: 10.1186/1475-925x-6-23.

[34] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, *Feature Extraction: Foundations and Applications*. Springer, 2006. [Online]. Available: https://link.springer.com/book/10.1007/978-3-540-35488-8 and https://www.academia.edu/76998247/Feature_Extraction

[35] V. Kumar, "Feature selection: A literature review," *Smart Comput. Rev.*, vol. 4, no. 3, pp. 211–229, Jun. 2014. [Online]. Available: https://faculty.cc.gatech.edu/~hic/CS7616/Papers/Kumar-Minz-2014.pdfs, doi: 10.6029/smartcr.2014.03.007.

[36] B. Seijo-Pardo, I. Porto-Díaz, V. Bolón-Canedo, and A. Alonso-Betanzos, "Ensemble feature selection: Homogeneous and heterogeneous approaches," *Knowl.-Based Syst.*, vol. 118, pp. 124–139, Feb. 2017, doi: 10.1016/j.knosys.2016.11.017.

[37] V. Bolón-Canedo, D. Rego-Fernández, D. Peteiro-Barral, A. Alonso-Betanzos, B. Guijarro-Berdiñas, and N. Sánchez-Maroño, "On the scalability of feature selection methods on high-dimensional data," *Knowl. Inf. Syst.*, vol. 56, no. 2, pp. 395–442, Aug. 2018, doi: 10.1007/s10115-017-1140-3.

[38] A. H. Shahid and M. P. Singh, "A novel approach for coronary artery disease diagnosis using hybrid particle swarm optimization based emotional neural network," *Biocybern. Biomed. Eng.*, vol. 40, no. 4, pp. 1568–1585, Oct. 2020, doi: 10.1016/j.bbe.2020.09.005.
[39] L. Fan, X. Shi, Z. Wang, R. Zhang, and J. Zhang, "Disease identification method based on graph features between pulse cycles," *Biomed. Signal Process. Control*, vol. 83, May 2023, Art. no. 104670, doi: 10.1016/j.bspc.2023.104670.
[40] L. Sun, X.-Y. Zhang, Y.-H. Qian, J.-C. Xu, S.-G. Zhang, and Y. Tian, "Joint neighborhood entropy-based gene selection method with Fisher score for tumor classification," *Appl. Intell.*, vol. 49, no. 4, pp. 1245–1259, Apr. 2019, doi: 10.1007/s10489-018-1320-1.
[41] L. Sun, T. Wang, W. Ding, J. Xu, and Y. Lin, "Feature selection using Fisher score and multilabel neighborhood rough sets for multilabel classification," *Inf. Sci.*, vol. 578, pp. 887–912, Nov. 2021, doi: 10.1016/j.ins.2021.08.032.
[42] M. Gan and L. Zhang, "Iteratively local Fisher score for feature selection," *Appl. Intell.*, vol. 51, no. 8, pp. 6167–6181, Aug. 2021, doi: 10.1007/s10489-020-02141-0.
[43] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
[44] R. Smith, "A mutual information approach to calculating nonlinearity," *Stat*, vol. 4, no. 1, pp. 291–303, Feb. 2015.
[45] A. R. Subhani, W. Mumtaz, N. Kamil, N. M. Saad, N. Nandagopal, and A. S. Malik, "MRMR based feature selection for the classification of stress using EEG," in *Proc. 11th Int. Conf. Sens. Technol. (ICST)*, Dec. 2017, pp. 1–4.
[46] G. Gulgezen, Z. Cataltepe, and L. Yu, "Stable and accurate feature selection," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2009, pp. 455–468.
[47] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[48] J. Fan, W. Yue, L. Wu, F. Zhang, H. Cai, X. Wang, X. Lu, and Y. Xiang, "Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China," *Agricult. Forest Meteorol.*, vol. 263, pp. 225–241, Dec. 2018, doi: 10.1016/j.agrformet.2018.08.019.
[49] A. Gunakala and A. H. Shahid, "A comparative study on performance of basic and ensemble classifiers with various datasets," *Appl. Comput. Sci.*, vol. 19, no. 1, pp. 107–132, Mar. 2023, doi: 10.35784/acs-2023-08.
[50] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006, doi: 10.1016/j.neucom.2005.12.126.
[51] J. Zhang, W. Xiao, S. Zhang, and S. Huang, "Device-free localization via an extreme learning machine with parameterized geometrical feature extraction," *Sensors*, vol. 17, no. 4, p. 879, Apr. 2017, doi: 10.3390/s17040879.
[52] O. Kisi and M. Alizamir, "Modelling reference evapotranspiration using a new wavelet conjunction heuristic method: Wavelet extreme learning machine vs wavelet neural networks," *Agricult. Forest Meteorol.*, vol. 263, pp. 41–48, Dec. 2018, doi: 10.1016/j.agrformet.2018.08.007.
[53] Y. Xue, H. Zhu, and F. Neri, "A feature selection approach based on NSGA-II with ReliefF," *Appl. Soft Comput.*, vol. 134, Feb. 2023, Art. no. 109987, doi: 10.1016/j.asoc.2023.109987.
[54] M. I. K. Khalil, I. U. Rahman, M. Zakarya, and M. Khan, "A neighborhood-aware multi-Markovian switching particle swarm optimization technique for solving complex and expensive problems," *Soft Comput.*, vol. 28, nos. 9–10, pp. 6517–6536, May 2024, doi: 10.1007/s00500-023-09517-7.
[55] M. I. K. Khalil, I. U. Rahman, M. Zakarya, A. Zia, A. A. Khan, M. R. C. Qazani, M. Al-Bahri, and M. Haleem, "A multi-objective optimisation approach with improved Pareto-optimal solutions to enhance economic and environmental dispatch in power systems," *Sci. Rep.*, vol. 14, no. 1, p. 13418, Jun. 2024, doi: 10.1038/s41598-024-62904-4.
[56] Y. Tan and Y. Zhu, "Fireworks algorithm for optimization," in *Proc. 1st Int. Conf. Adv. Swarm Intell.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, Jan. 2010, pp. 355–364, doi: 10.1007/978-3-642-13495-1_44.
[57] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Int. Symp. Micro Mach. Hum. Sci.*, Oct. 1995, pp. 39–43.
[58] H. Li, C.-M. Pun, F. Xu, L. Pan, R. Zong, H. Gao, and H. Lu, "A hybrid feature selection algorithm based on a discrete artificial bee colony for Parkinson's diagnosis," *ACM Trans. Internet Technol.*, vol. 21, no. 3, pp. 1–22, Aug. 2021, doi: 10.1145/3397161.
[59] A. G. Gad, "Particle swarm optimization algorithm and its applications: A systematic review," *Arch. Comput. Methods Eng.*, vol. 29, no. 5, pp. 2531–2561, Aug. 2022, doi: 10.1007/s11831-021-09694-4.
[60] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.
[61] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *J. Amer. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, Dec. 1952, doi: 10.1080/01621459.1952.10483441.
[62] O. J. Dunn, "Multiple comparisons using rank sums," *Technometrics*, vol. 6, no. 3, p. 241, Aug. 1964.
[63] C. E. Agbangba, E. S. Aide, H. Honfo, and R. Glèlè Kakai, "On the use of post-hoc tests in environmental and biological sciences: A critical review," *Heliyon*, vol. 10, no. 3, Feb. 2024, Art. no. e25131, doi: 10.1016/j.heliyon.2024.e25131.
[64] Z. Xue, H. Lu, T. Zhang, J. Xu, and X. Guo, "A local dynamic feature selection fusion method for voice diagnosis of Parkinson's disease," *Comput. Speech Lang.*, vol. 82, Jul. 2023, Art. no. 101536, doi: 10.1016/j.csl.2023.101536.

GUNAKALA ARCHANA received the B.Tech. degree in computer science and engineering from the Nalanda Institute of Engineering and Technology, Andhra Pradesh, India, in 2014, and the M.Tech. degree in computer science engineering from the RVR and JC College of Engineering, Andhra Pradesh, in 2017. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh. She is working on machine learning and soft computing. She has a teaching experience of six years at reputed colleges in Andhra Pradesh.

AFZAL HUSSAIN SHAHID received the B.Tech. degree in computer science and engineering from Netaji Subhash Engineering College, Kolkata, India, in 2010, and the M.Tech. and Ph.D. degrees in computer science and engineering from the National Institute of Technology Patna, India, in 2013 and 2020, respectively. He is currently an Assistant Professor with the School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India. His research interests include artificial intelligence, machine learning, deep learning, soft computing, and evolutionary algorithms.