



Particle Swarm Optimization-Based Extreme Learning Machine for COVID-19 Detection

Musatafa Abbas Abbood Albadr¹ · Sabrina Tiun¹ · Masri Ayob¹ · Fahad Taha AL-Dhief²

Received: 19 September 2021 / Accepted: 5 October 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

COVID-19 (coronavirus disease 2019) is an ongoing global pandemic caused by severe acute respiratory syndrome coronavirus 2. Recently, it has been demonstrated that the voice data of the respiratory system (i.e., speech, sneezing, coughing, and breathing) can be processed via machine learning (ML) algorithms to detect respiratory system diseases, including COVID-19. Consequently, many researchers have applied various ML algorithms to detect COVID-19 by using voice data from the respiratory system. However, most of the recent COVID-19 detection systems have worked on a limited dataset. In other words, the systems utilize cough and breath voices only and ignore the voices of the other respiratory system, such as speech and vowels. In addition, another issue that should be considered in COVID-19 detection systems is the classification accuracy of the algorithm. The particle swarm optimization-extreme learning machine (PSO-ELM) is an ML algorithm that can be considered an accurate and fast algorithm in the process of classification. Therefore, this study proposes a COVID-19 detection system by utilizing the PSO-ELM as a classifier and mel frequency cepstral coefficients (MFCCs) for feature extraction. In this study, respiratory system voice samples were taken from the Corona Hack Respiratory Sound Dataset (CHRSD). The proposed system involves thirteen different scenarios: breath deep, breath shallow, all breath, cough heavy, cough shallow, all cough, count fast, count normal, all count, vowel a, vowel e, vowel o, and all vowels. The experimental results demonstrated that the PSO-ELM was capable of attaining the highest accuracy, reaching 95.83%, 91.67%, 89.13%, 96.43%, 92.86%, 88.89%, 96.15%, 96.43%, 88.46%, 96.15%, 96.15%, 95.83%, and 82.89% for breath deep, breath shallow, all breath, cough heavy, cough shallow, all cough, count fast, count normal, all count, vowel a, vowel e, vowel o, and all vowel scenarios, respectively. The PSO-ELM is an efficient technique for the detection of COVID-19 utilizing voice data from the respiratory system.

Keywords Mel frequency cepstral coefficients · Particle swarm optimization-extreme learning machine

Introduction

COVID-19 (coronavirus disease 2019), which is caused by SARS-COV-2 (severe acute respiratory syndrome coronavirus 2), was declared a worldwide pandemic on 11 March 2020 by the World Health Organization (WHO). COVID-19 is considered a new infectious disease, but it is similar to other diseases that are caused by some coronaviruses such

as Middle East respiratory syndrome coronavirus (MERS-COV) and severe acute respiratory syndrome coronavirus (SARS-COV), which caused disease outbreaks in both years (2002 and 2012) [1]. Fever, dry coughs, and fatigue are the most common COVID-19 symptoms [2]. In addition, the other COVID-19 symptoms include muscle pain, joint pain, gastrointestinal symptoms, shortness of breath, and loss of taste or smell [3]. At the writing time of this manuscript, there were 209.9 million active COVID-19 cases worldwide, and there were 4.4 million deaths. The USA has recorded the greatest number of new COVID-19 cases (37.1 million) and 622,263 deaths [4]. The scale of the pandemic has caused several health systems to be overrun due to case management and the need for testing.

Several efforts have been made to detect early COVID-19 symptoms by applying artificial intelligence (AI) and machine learning (ML) algorithms to images [5]. Extreme

✉ Musatafa Abbas Abbood Albadr
mustafa_abbas1988@yahoo.com

¹ CAIT, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

² School of Electrical Engineering, Department of Communication Engineering, Universiti Teknologi Malaysia, UTM Johor Bahru, Johor, Malaysia

learning machines (ELMs) have been demonstrated to outperform other ML algorithms, such as support vector machines (SVMs), convolutional neural networks (CNNs), and backpropagation-based neural networks (NNs), in such tasks [6]. For instance, the optimized genetic algorithm-extreme learning machine (OGA-ELM) has been shown to be able to detect COVID-19 from X-ray images with 100.00% accuracy [6]. In addition, the multiple kernels-ELM based on a deep neural network has been demonstrated to detect COVID-19 from computed tomography (CT) images with 98.36% accuracy [7].

Coughing is considered one of the most common COVID-19 symptoms as well as a symptom of more than a hundred other illnesses, which affects the respiratory system differently [8, 9]. For instance, lung illnesses can cause either obstruction or restriction in the airway, which can affect cough acoustics [10]. Additionally, it has been assumed that the behavior of the glottis acts differently under various pathological conditions [11, 12], which makes it possible to distinguish among coughs due to asthma, tuberculosis (TB), pertussis (whooping cough), and bronchitis [13–15].

Data on the respiratory system, such as speech, sneezing, eating behavior, coughing, and breathing, can be processed via ML algorithms to detect illnesses of the respiratory system [16]. Therefore, many recent studies have been conducted to detect COVID-19 by using voice data from the respiratory system. For example, the authors in [17] presented a new application for the detection of COVID-19 by implementing the short-time magnitude spectrogram features (SRMSF) and the ResNet18 classifier. The application was tested based on the Coswara dataset using cough voice data. The best result was obtained in terms of the area under the ROC curve (AUC), which reached up to 0.72. In addition, the authors in [18] proposed a system for the detection of COVID-19 by implementing the mel-filter bank features (MFBBF) and SVM classifier. The proposed system was evaluated based on speech voice data collected from YouTube. The experimental results revealed that the highest achieved result had an accuracy of 88.60%. Additionally, an application for the detection of COVID-19 was proposed in [19]. The proposed application was built by using the MFCC features and ResNet50 classifier. The application was tested based on cough voice data. The experimental results showed that the proposed application achieved the highest results with a specificity of 94.2%.

The authors in [20] proposed a COVID-19 detection system by implementing several handcrafted features and logistic regression (LR) classifiers. The proposed system was evaluated based on cough and breath voice data taken from a crowdsourced dataset. The experimental results showed that the highest precision was 80.00% for cough voice data and 69.00% for breath voice data. Moreover, a new COVID-19 detection system was proposed in [21]. The proposed

system uses mel-spectrogram features (MSF) and the CNN architecture. The system was evaluated based on the ESC-50 dataset and a collected dataset by recording COVID-19 samples using a mobile app. The evaluation was based on three different scenarios: speech, cough, and overall voice data. The highest results were achieved with an accuracy of 92.00%, 92.85%, and 88.00% for speech, cough, and overall, respectively. In addition, the authors in [22] proposed a COVID-19 detection system that consists of two phases. The first phase applies several steps to extract the needed features. These steps start with MFCC followed by variable Markov oracle (VMO) and recurrence plot (RP) and end with recurrence quantification analysis (RQA). The second phase feeds the extracted features into the eXtreme gradient boosting (XGBoost) classifier. The proposed system was evaluated based on cough and vowel “ah” voice data collected by Carnegie Mellon University (CMU). The experimental results showed that the proposed system achieved an accuracy of up to 97.00% and 99.00% for cough and vowel voice data, respectively. Furthermore, a new COVID-19 detection application was presented in [23]. The presented system was based on using MFCC features and the long short-term memory (LSTM) classifier. The application was tested by using three different scenarios: cough, breath, and speech voice data. The highest experimental results were achieved with an accuracy of 97.00% for cough, 98.20% for breath, and 88.20% for speech. Table 1 summarizes the previous works on COVID-19 detection by using respiratory system voice data.

Among the most common feature extraction methods used in the voice data processing field are linear predictive coding (LPC), cepstrum coefficients derived from LPC (LPCC), MFCC, and perceptual linear prediction (PLP) [24–26]. Out of all the aforesaid methods, MFCC is generally the most popular feature extraction approach in voice applications and has been cited to have the highest identification accuracy [27, 28].

Recently, the effectiveness of the ELM has been demonstrated in many domains, such as foreign accent identification [29], emotion recognition [30, 31], and language identification [32]. In addition, the ELM is preferred by researchers because it is superior to traditional SVMs [33–35] specifically in (1) thwarting overfitting, (2) its implementation on multi and binary classifications, and (3) its similar kernel-based capability SVM and working with an NN structure. These factors make the ELM more efficient in accomplishing a better learning performance.

Hence, overall, ELM demonstrates a speedier learning process, better generalization devoid of overtraining, and greater results based on the utilized input weights. Therefore, researchers have integrated PSO into ELM to obtain the best input weights and biases that are used in the hidden layer of the ELM. The effectiveness of this integration (PSO-ELM)

Table 1 Summary of the previous works on COVID-19 detection by using respiratory system voice data

Ref	Dataset	Features	Classifier	Parameters/modality	Results	Disadvantages
[17]	Coswara dataset	SRMSF	ResNet18	Cough	0.72 AUC	1. The output results are not encouraging and need more improvement
[18]	Collected from YouTube	MFBF	SVM	Speech	88.60% accuracy	2. Only one scenario has been conducted while other scenarios such as breath, counting fast, counting normal, and vowels (i.e., i, e, and o) have been ignored
[19]	Own collected dataset	MFCC	ResNet50	Cough	94.2% specificity	
[20]	Crowdsourced dataset	Several handcrafted features	LR	Cough and breath	80.00% precision for cough and 69.00% precision for breath	1. The output results are not encouraging and need more improvement 2. Only two scenarios have been conducted while other scenarios such as counting fast, counting normal, and vowels (i.e., i, e, and o) have been ignored
[21]	ESC-50 dataset and COVID-19	MSF	CNN	Speech, cough, and overall	Accuracy of 92.00%, 92.85%, and 88.00% for speech, cough, and overall, respectively	1. The output results are not encouraging and need more improvement 2. More scenarios such as counting fast, counting normal, and vowels (i.e., i, e, and o) have been ignored
[22]	Collected by CMU	MFCC-VMO-RP-RQA	XGBoost	Cough and vowel "ah"	97.00% accuracy for cough and 99.00% accuracy for vowel "ah"	More scenarios such as counting fast, counting normal, and vowels (i.e., i, e, and o) have been ignored
[23]	Own collected dataset	MFCC	LSTM	Cough, breath, and speech	Accuracy of 97.00% for cough, 98.20% for breath, and 88.20% for speech	

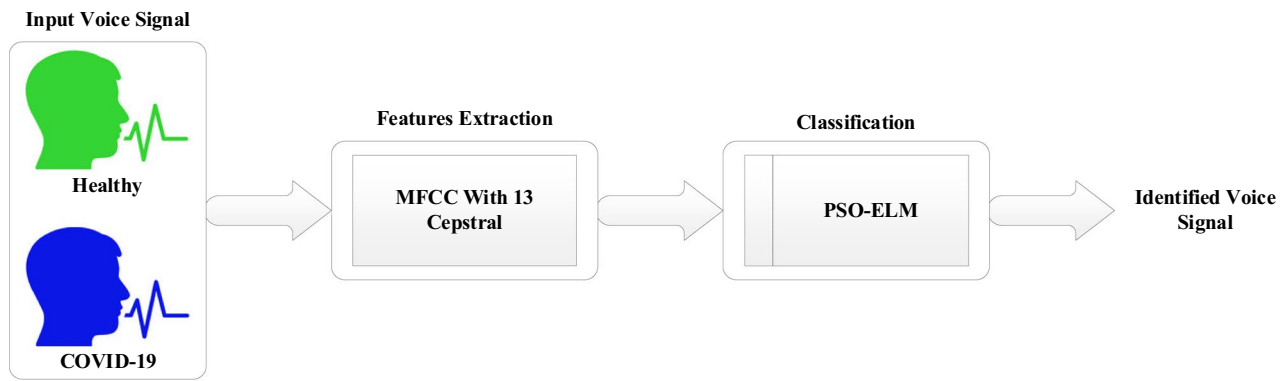


Fig. 1 Block diagram of the proposed COVID-19 detection system

has been demonstrated in many domains, including language identification [36], short-term temperature prediction [37], and breast cancer detection [38]. However, to the best of our knowledge, no research has used the PSO-ELM in the detection of COVID-19. Therefore, the aims of this study are as follows:

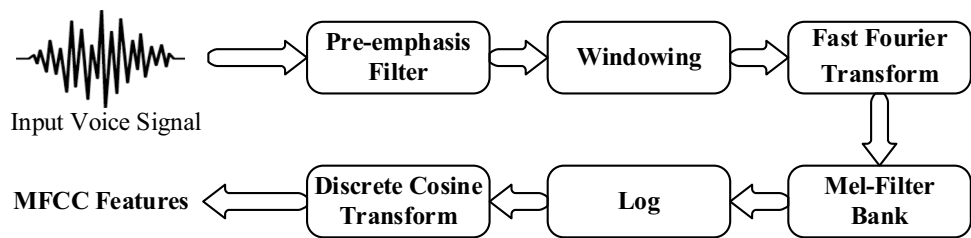
- Propose a new COVID-19 detection system based on the PSO-ELM classifier and MFCC features using different voice data of the respiratory system.
- In the proposed method, we used thirteen different scenarios: deep breath, shallow breath, all breaths, heavy cough, shallow cough, all coughs, count fast, count normal, all counts, vowel a, vowel e, vowel o, and all vowels.
- The NN, random forest (RF), and basic ELM classifiers are also implemented based on the MFCC features for the detection of COVID-19.
- Several evaluation measures, such as accuracy, recall, precision, specificity, F-measure, G-mean, and execution time, are used to evaluate the performance of the proposed system.

The remainder of this paper is organized as follows. “[Materials and Proposed Method](#)” section provides the materials and proposed method. “[Experimental Setup and Results](#)” section discusses the results of the experiments. Finally, “[Discussion](#)” section presents the conclusion of this paper.

Table 2 Description of the whole dataset

Deep breath		Shallow breath		All breath		Label
Class	Number of samples	Class	Number of samples	Class	Number of samples	
Healthy	39	Healthy	39	Healthy	78	1
COVID-19	39	COVID-19	39	COVID-19	78	2
Heavy cough		Shallow cough		All cough		Label
Class	Number of samples	Class	Number of samples	Class	Number of samples	
Healthy	45	Healthy	45	Healthy	90	1
COVID-19	45	COVID-19	45	COVID-19	90	2
Count fast		Count normal		All count		Label
Class	Number of samples	Class	Number of samples	Class	Number of samples	
Healthy	42	Healthy	45	Healthy	87	1
COVID-19	42	COVID-19	45	COVID-19	87	2
Vowel a		Vowel e				Label
Class	Number of samples	Class	Number of samples	Number of samples		
Healthy	43	Healthy	43	43		1
COVID-19	43	COVID-19	43	43		2
Vowel o		All vowels				Label
Class	Number of samples	Class	Number of samples	Number of samples		
Healthy	41	Healthy	127	127		1
COVID-19	41	COVID-19	127	127		2

Fig. 2 The process of MFCC feature extraction



Materials and Proposed Method

The general diagram of the proposed COVID-19 detection system using the PSO-ELM method is demonstrated in Fig. 1. The diagram consists of various stages that are used to create the COVID-19 detection system based on the voice signals. The first stage refers to the voice datasets (i.e., breathing, coughing, counting, and vowels) for healthy and COVID-19-infected people. In the second stage, the MFCC method is utilized to extract the needed features from the voice signals. Last, in the third stage, the MFCC extracted features are fed into the PSO-ELM classifier to detect COVID-19 based on the voice signal. These three stages of the proposed COVID-19 detection system are deliberated as subsections.

Dataset

In this study, the Corona Hack Respiratory Sound Dataset (CHRSD) is utilized to evaluate the proposed system. The CHRSD was downloaded from [39]. The dataset contains multiple categories of respiratory sounds, such as deep breath, shallow breath, heavy cough, shallow cough, count fast, count normal, vowel a, vowel e, and vowel o. The voice samples were recorded from healthy and COVID-19-infected people of both genders female and male. The duration of the voice samples is in the range of (1–25) seconds. Table 2 describes the dataset that was used in this study. It is worth mentioning that all the experiments in this study were conducted based on a ratio of 70% for training and 30% for testing. In addition, the data samples, in both training and testing, were shuffled.

Feature Extraction: MFCC

In this study, the MFCC features [40, 41] were extracted from the voice samples based on conducting several processing steps. These steps are preemphasis, windowing, fast Fourier transform (FFT), mel-filter bank, log, and discrete cosine transform (DCT). The explanation of the processing steps is provided as follows:

- Preemphasis represents the first step of MFCC feature extraction, which aims to boost energy at high frequencies.
- Windowing is the second step of the MFCC, and it segments the voice sample into frames.
- FFT denotes the third step of the MFCC, which aims to convert the voice signal from the time domain into a frequency domain because the voice data features exist in the frequency domain.
- The mel-filter bank is the fourth step of the MFCC, which aims to approximate how much energy appears in each area or point.
- Log indicates the fifth step of the MFCC, the goal of which is to assure that the low and high frequencies are separated, which simulates the human hearing system.
- DCT is the last step of the MFCC, which aims to convert the log mel spectrum back to time.

The whole process of extracting the MFCC features is demonstrated in Fig. 2. Table 3 depicts the variable values of the MFCC that have been utilized in this paper. The frame size and the frameshift size in the samples are calculated as depicted in Eqs. (1 and 2).

$$N_w = 10^{-3} \times T_w \times \text{Samplingrate} \tag{1}$$

N_w : the size of the frame in the samples.
 T_w : the duration of the frame in time.

$$N_s = 10^{-3} \times T_s \times \text{Samplingrate} \tag{2}$$

N_s : the size of the frameshift in the samples.
 T_s : the duration of the frameshift in time.

Table 3 The value of the MFCC variables that have been used in this study

Variable	Value
Sampling rate	44,100 Hz
T_w	25 ms
T_s	10 ms
N_w	1103
N_s	441
Number of MFCC features	13

Classification: PSO-ELM

The PSO-ELM is proposed based on the concept of the PSO algorithm. In the PSO-ELM, the values of both input weights and hidden layer biases are modified by updating the parameters of the PSO to achieve high accuracy. The PSO-ELM processing steps are provided in deep detail as follows alongside its flowchart, which is depicted in Fig. 3.

N is a group of featured samples (X_j, t_j) , where $X_j = [x_{j1}, x_{j2}, \dots, x_{jm}]^T \in R^n$, and $t_j = [t_{j1}, t_{j2}, \dots, t_{jm}]^T \in R^m$.

where:

X_j = the input, which is the extracted MFCC features.

t_j = the expected output (true value).

Step 1: Randomly initialize the particle swarm (P) in the range of $([0, 1]$ for the bias values of the hidden neurons and $[-1, 1]$ for the values of the input weights), where $P = \{p_1, p_2, \dots, p_z\}$ and z refer to the population size.

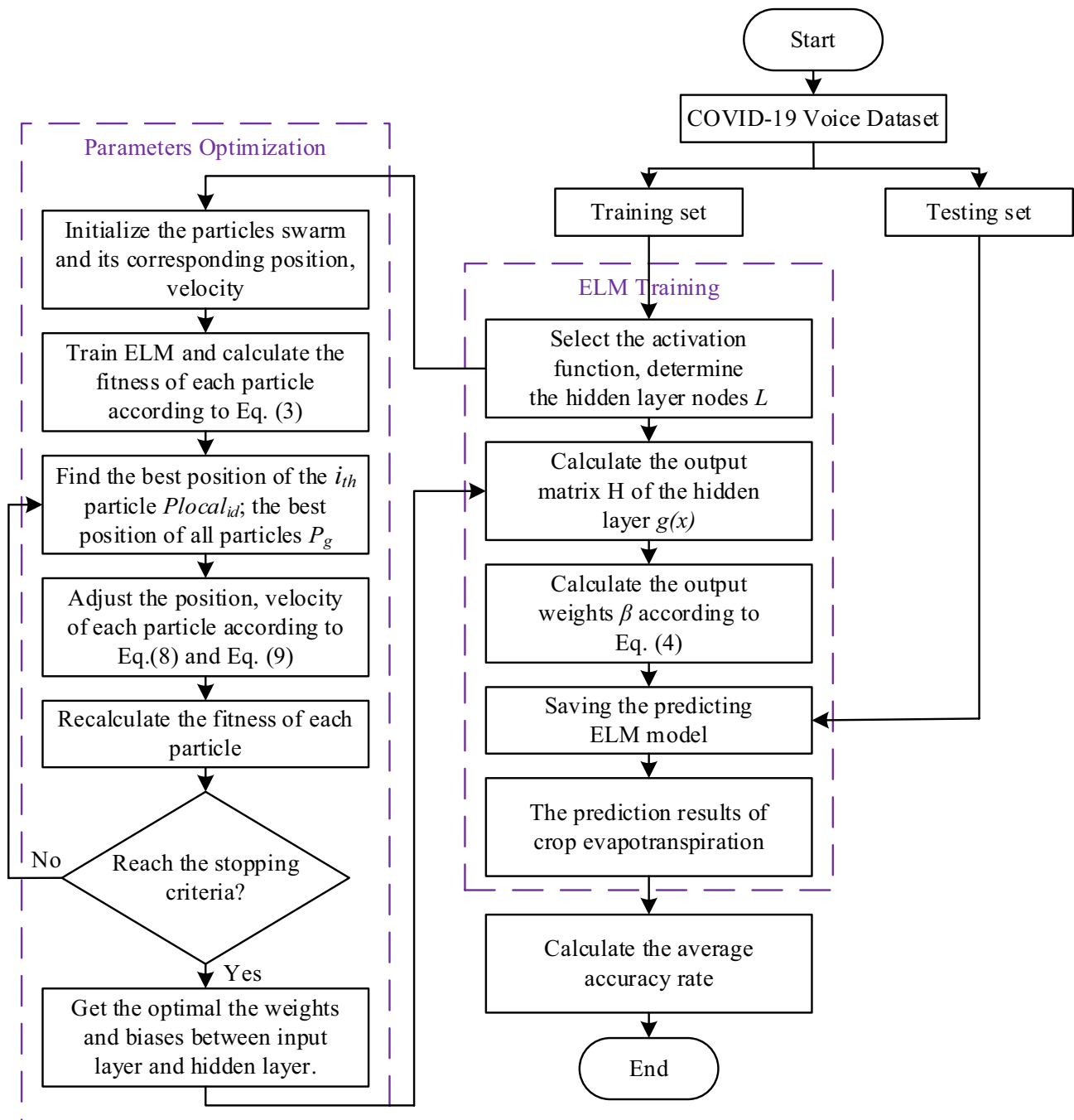


Fig. 3 PSO-ELM algorithm flowchart

The i th particle position shape is depicted as $p_i = \{w_{11}, w_{12}, \dots, w_{1n}, w_{21}, w_{22}, \dots, w_{2n}, w_{L1}, w_{L2}, \dots, w_{Ln}, b_1, \dots, b_L\}$. The i th particle velocity shape is denoted as $V_i = \{v_{i1}, v_{i2}, \dots, v_{iD}\}$, where $D = (1+n) \times L$. In addition, the velocity of each particle is limited to the range of $[-V_{max}, V_{max}]$. Determine the maximum iteration number k_{max} .

where:

$w_{ij} \in [-1, 1]$ denotes the values of the input weights that link the j th input neuron and the i th hidden neuron.

$b_i \in [0, 1]$ represents the bias of the i th hidden neuron. n refers to the number of input neurons.

L refers to the number of hidden neurons.

$(1+n) \times L$ denotes the dimension of the particle, which requires optimizing its parameters.

Step 2: Split the dataset into two subsets (training set and testing set).

Set the number of hidden layer neurons as L and select the appropriate ELM activation function $g(x)$ [42].

$$f(P) = \sqrt{\frac{\sum_j^N \|\sum_k^L \beta_k g(w_k x_j + b_k) - t_j\|_2^2}{N}} \quad (3)$$

where:

β is the output weights matrix.

t_j is the true value.

N is the number of training samples.

where:

$$\beta = HT \quad (4)$$

$$H = \begin{bmatrix} g(w_1.X_1 + b_1) & \dots & g(w_L.X_1 + b_L) \\ \vdots & \dots & \vdots \\ g(w_1.X_N + b_1) & \dots & g(w_L.X_N + b_L) \end{bmatrix}_{N \times L} \quad (5)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad \text{and} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}$$

H in Eq. (5) is the hidden layer output matrix of the ELM network; in H , the i th column is indicated to the i th hidden layer neuron on the input neurons. H is the Moore–Penrose generalized inverse of H . The activation function g is infinitely distinguishable when the desired number of hidden neurons is $L \leq N$.

Step 3: Calculate the fitness value of each particle (p) in the population (P) utilizing Eq. (3).

Step 4: Discover the best particle position $Plocal_{id}$ as well as the best position for all particles P_g . Subsequently, update them based on the following details:

$$Plocal_{id} = \begin{cases} p_i & p_i \text{ better than } Plocal_{id} \\ Plocal_{id} & \text{else} \end{cases} \quad (6)$$

$$P_g = \begin{cases} p_i & p_i \text{ better than } P_g \\ P_g & \text{else} \end{cases} \quad (7)$$

Step 5: Modify the position and velocity of each particle based on Eqs. (8 and 9) and recalculate the fitness value for each particle via Eq. (3).

$$v_{id}^{k+1} = \omega v_{id}^k + C_1 r_1 (Plocal_{id}^k - p_{id}^k) + C_2 r_2 (P_{gd}^k - p_{id}^k) \quad (8)$$

$$p_{id}^{k+1} = p_{id}^k + v_{id}^{k+1} \quad (9)$$

where C_1 and C_2 denote acceleration coefficients, which are nonnegative constants; k refers to the present iteration number; $i = 1, 2, \dots, z$; $d = 1, 2, \dots, D$; and ω denotes the inertial weight. r_1 and r_2 are two randomly generated numbers in the range of $[0, 1]$.

Step 6: If the stopping criteria are reached, then save the optimal input weights and biases of the input-hidden layers; otherwise, go to step 4.

Step 7: The outcome results of the PSO optimization are utilized as input weights and biases for the ELM algorithm, and the hidden layer output matrix (H) is calculated by using Eq. (5).

Step 8: Calculate the weights of the output (β) based on Eq. (4) and save the ELM prediction model for testing.

Table 4 illustrates the parameters of the ELM and PSO that have been used in this study.

Experimental Setup and Results

Several experiments were conducted based on four main scenarios: breath, cough, count, and vowels. In the breath scenario, three different scenarios were implemented: breath deep, breath shallow, and all breath. Additionally, three different scenarios (i.e., cough heavy, cough shallow, and all cough) were applied in the cough scenario. Moreover, in the count scenario, three different scenarios were performed, which were count fast, count normal, and all count. Last, four different scenarios (i.e., vowel a, vowel e, vowel o, and all vowels) were conducted in the vowel scenario. All the experiments were implemented based on using 70% of the dataset as a training dataset and 30% as a testing dataset.

The proposed PSO-ELM was applied in several experiments in each scenario based on the scenario’s dataset with a varying number of hidden neurons in the range of [100–600] with 50 increment steps, and each experiment had 100 iterations. It is worth mentioning that all the experiments have been implemented in MATLAB R2019a programming language over a PC Core i7 of 3.20 GHz with 16 GB RAM and SSD 1 TB (Windows 10). In this study, numerous evaluation measurements were utilized to evaluate the proposed PSO-ELM

Table 4 The ELM and PSO parameters

ELM		PSO	
Parameters	Values	Parameters	Values
P	Input weights and biases	Population (particles)	Contain positions and velocities
β	Output weights	Position	Generated randomly at the beginning, in the range of $[-1, 1]$ for input weights and $[0, 1]$ for biases
Input weight	In the range of $[-1, 1]$	Velocity	Start with zero values, and it is limited to the range of $[-2, 2]$
Biases values	In the range of $[0, 1]$	z	50
Input neurons number (n)	Input attributes	ω, C_1, C_2	0.7289, 1.496, 1.496
Hidden neurons number (L)	100–600, with 50 increment step	k_{max}	100
Output neurons	Class values	$P_{local_{id}}$	Best particle position
Activation function	Sigmoid	P_g	Best position of all particles

approach. The evaluation measurements rely on the ground truth, which entails the application of the model to expect the answer on the evaluation dataset followed by a comparison between the predicted target and the actual answer. The evaluation measurements were used to evaluate the proposed PSO-ELM approach regarding true positive (TP), true negative (TN), false positive (FP), false negative (FN), recall, accuracy, specificity, G-mean, precision, F-measure, and execution time. Equations (10–15) [43–45] depict these evaluation measurements.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \quad (10)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (11)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (12)$$

$$F - \text{Measure} = \frac{(2 \times \text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \quad (13)$$

$$G - \text{Mean} = \sqrt{\text{recall} \times \text{precision}} \quad (14)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (15)$$

The experimental results of the four main scenarios (breath, cough, count, and vowels) are provided and discussed separately in detail in the following subsections.

Breath Scenario

This section provides and discusses the performance of the PSO-ELM in the breath scenario, where the breath scenario includes three different scenarios: breath deep, breath

shallow, and all breath. In the breath deep scenario, we used the voice samples that contain only breathing deep sounds. The shallow breath scenario was performed by using voice samples that contain only shallow breathing sounds. In the scenario of all breaths, we used the voice samples that contain both deep breathing and shallow breathing sounds. Table 5 describes the dataset that was used in each scenario.

The best experimental results of the proposed PSO-ELM in the deep breath, shallow breath, and all breath scenarios were obtained with overall accuracies of 95.83%, 91.67%, and 89.13% for the deep breath, shallow breath, and all breath scenarios, respectively. Figure 4 shows the ROC of the best results for the deep breath, shallow breath, and all breath scenarios.

Cough Scenario

This section provides and discusses the performance of the PSO-ELM in a cough scenario where the cough scenario includes three different scenarios: heavy cough, shallow

Table 5 The dataset that was used in deep breath, shallow breath, and all breath scenarios

Class	Number of all samples	Number of training samples	Number of testing samples	Label
Deep breath scenario				
Healthy	39	27	12	1
COVID-19	39	27	12	2
Shallow breath scenario				
Healthy	39	27	12	1
COVID-19	39	27	12	2
All breath scenario				
Healthy	78	55	23	1
COVID-19	78	55	23	2

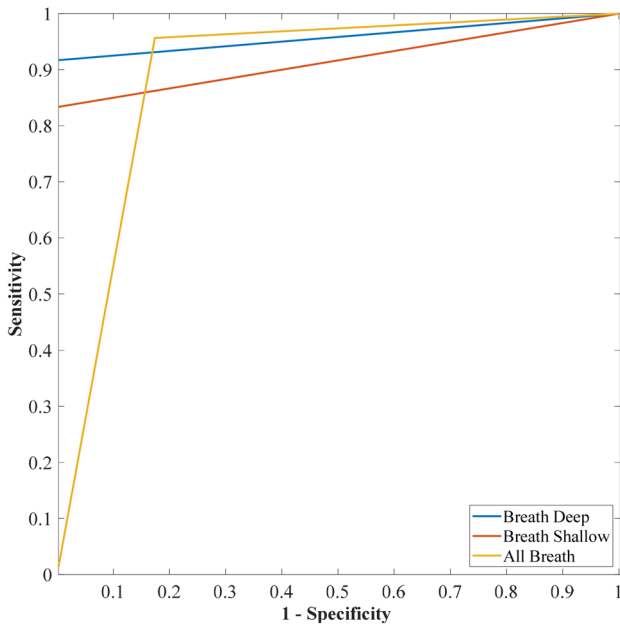


Fig. 4 The ROC of PSO-ELM best results for the deep breath, shallow breath, and all breath scenarios

cough, and all coughs. In the heavy cough scenario, we used voice samples that contain only heavy coughing sounds. The shallow cough scenario was performed by using the voice samples that contain only shallow coughing sounds. In the scenario of all coughs, we used voice samples that contain both heavy coughing and shallow coughing sounds. Table 6 describes the dataset that was utilized in each scenario.

The best experimental results of the proposed PSO-ELM in heavy cough, shallow cough, and all cough scenarios were obtained with overall accuracies of 96.43%, 92.86%, and 88.89% for heavy cough, shallow cough, and all cough scenarios, respectively. Figure 5 shows the ROC

Table 6 The dataset that was utilized in heavy cough, shallow cough, and all cough scenarios

Class	Number of all samples	Number of training samples	Number of testing samples	Label
Deep cough scenario				
Healthy	45	31	14	1
COVID-19	45	31	14	2
Shallow cough scenario				
Healthy	45	31	14	1
COVID-19	45	31	14	2
All cough scenario				
Healthy	90	63	27	1
COVID-19	90	63	27	2

of the best results for the heavy cough, shallow cough, and all cough scenarios.

Count Scenario

This section provides and discusses the performance of the PSO-ELM in a count scenario where the cough scenario contains three different scenarios: count fast, count normal, and all count. In the count fast scenario, we utilized the voice samples that contain only counting fast sounds. The count normal scenario was performed by utilizing the voice samples that contain only counting normal sounds. In the scenario of all counts, we utilized the voice samples that contain both counting fast and counting normal sounds. Table 7 describes the dataset that was used in each scenario.

The best experimental results of the proposed PSO-ELM in count fast, count normal, and all count scenarios were obtained with overall accuracies of 96.15%, 96.43%, and 88.46% for count fast, count normal, and all count scenarios, respectively. Figure 6 shows the ROC of the best results for the count fast, count normal, and all count scenarios.

Vowel Scenario

This section provides and discusses the performance of the PSO-ELM in a vowel scenario where the vowel scenario contains four different scenarios: vowel a, vowel e, vowel o, and all vowels. In the vowel a scenario, we used the voice

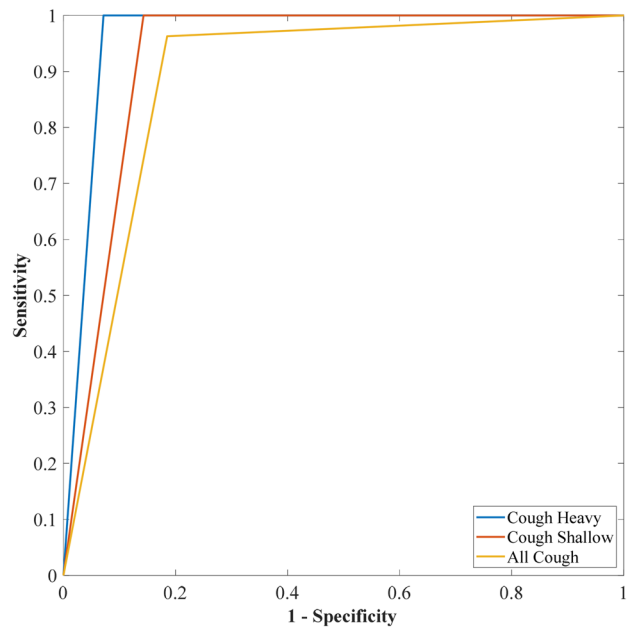


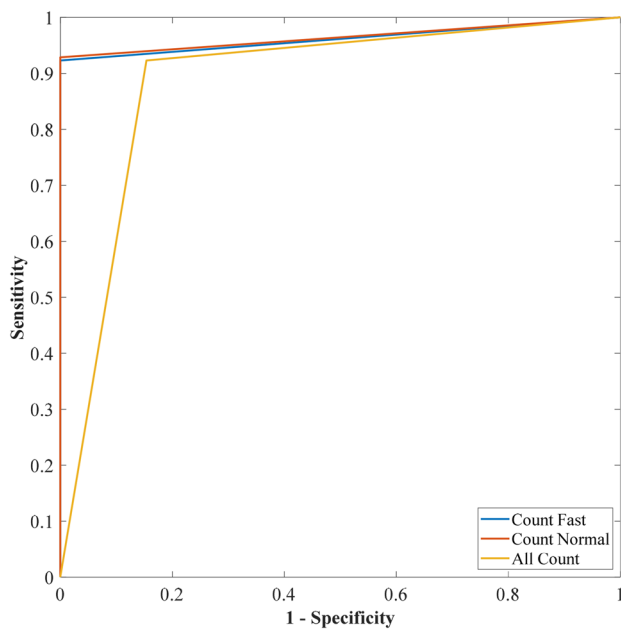
Fig. 5 The ROC of PSO-ELM best results for the heavy cough, shallow cough, and all cough scenarios

Table 7 The dataset that was used in count fast, count normal, and all count scenarios

Class	Number of all samples	Number of training samples	Number of testing samples	Label
Count fast scenario				
Healthy	42	29	13	1
COVID-19	42	29	13	2
Count normal scenario				
Healthy	45	31	14	1
COVID-19	45	31	14	2
All count scenario				
Healthy	87	61	26	1
COVID-19	87	61	26	2

samples that contain only vowel a sounds. The vowel e scenario was performed by using the voice samples that contain only vowel e sounds. In the vowel o scenario, we used the voice samples that contain only vowel o sounds. In the scenario of all vowels, we used the voice samples that contain vowel a, vowel e, and vowel o sounds. Table 8 describes the dataset that has been utilized in each scenario.

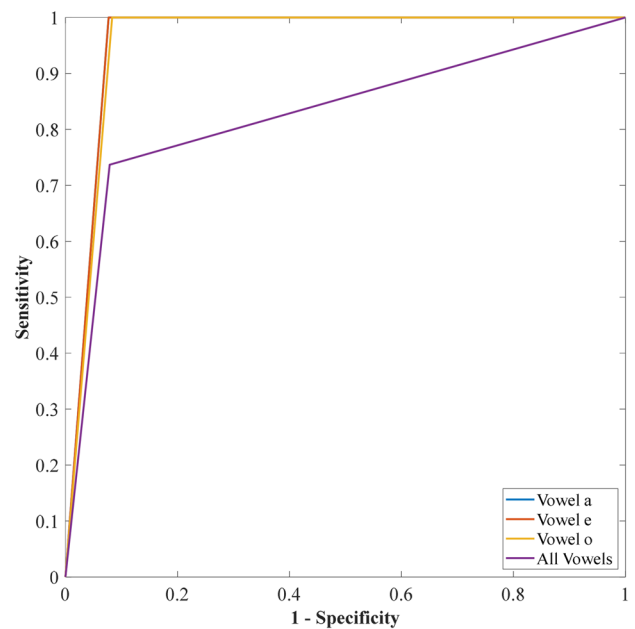
The best experimental results of the proposed PSO-ELM in vowel a, vowel e, vowel o, and all vowel scenarios have been obtained with overall accuracies of 96.15%, 96.15%, 95.83%, and 82.89% for vowel a, vowel e, vowel o, and all vowel scenarios, respectively. Figure 7 shows the ROC of the best results for vowel a, vowel e, vowel o, and all vowel

**Fig. 6** The ROC of PSO-ELM best results for the count fast, count normal, and all count scenarios**Table 8** The dataset that was utilized in vowel a, vowel e, vowel o, and all vowel scenarios

Class	Number of all samples	Number of training samples	Number of testing samples	Label
Vowel “a” scenario				
Healthy	43	30	13	1
COVID-19	43	30	13	2
Vowel “e” scenario				
Healthy	43	30	13	1
COVID-19	43	30	13	2
Vowel “o” scenario				
Healthy	41	29	12	1
COVID-19	41	29	12	2
All vowel scenario				
Healthy	127	89	38	1
COVID-19	127	89	38	2

scenarios. In addition, the best overall results of the evaluation measurements for all scenarios are shown in Table 9.

In addition, numerous experiments were performed based on the basic ELM, NN, and RF in thirteen different scenarios: deep breath, shallow breath, all breaths, heavy cough, shallow cough, all coughs, count fast, count normal, all count, vowel a, vowel e, vowel o, and all vowels. The ELM and NN approaches were implemented with a single hidden layer and the number of nodes in the range of [100–600] with 50 incremental steps. Note that because

**Fig. 7** The ROC of PSO-ELM best results for vowel a, vowel e, vowel o, and all vowel scenarios

the number of pages is limited, we reported only the highest performance of the ELM, NN, and RF in each scenario in terms of accuracy, recall, precision, specificity, G-mean, F-measure, TP, TN, FP, FN, and execution time. Table 10 provides all the experimental results of the basic ELM, NN, and RF techniques in the deep breath, shallow breath, all breaths, heavy cough, shallow cough, all coughs, count fast, count normal, all counts, vowel a, vowel e, vowel o, and all vowel scenarios. The best performance of the basic ELM was obtained with an accuracy of 66.67%, 70.83%, 65.22%, 75.00%, 64.29%, 64.81%, 69.23%, 67.86%, 63.46%, 65.38%, 61.54%, 62.50%, and 61.84% for deep breath, shallow breath, all breaths, heavy cough, shallow cough, all coughs, count fast, count normal, all counts, vowel a, vowel e, vowel o, and all vowel scenarios, respectively. The best performance of the NN had an accuracy of 62.50% deep breath, 66.67% shallow breath, 58.70% all breaths, 60.71% heavy cough, 60.71% shallow cough, 59.26% all coughs, 57.69% count fast, 71.43% count normal, 63.46% all count, 57.69% vowel a, 65.38% vowel e, 54.17% vowel o, and 56.58% all vowels. The best performance of the RF was achieved with an accuracy of 62.50%, 66.67%, 60.87%, 67.86%, 57.14%,

61.11%, 65.38%, 64.29%, 59.62%, 53.85%, 57.69%, 58.33%, and 60.53% for deep breath, shallow breath, all breaths, heavy cough, shallow cough, all coughs, count fast, count normal, all count, vowel a, vowel e, vowel o, and all vowel scenarios, respectively.

Furthermore, additional experiments have been conducted based on LSTM and XGBoost approaches in thirteen different scenarios: deep breath, shallow breath, all breaths, heavy cough, shallow cough, all coughs, count fast, count normal, all count, vowel a, vowel e, vowel o, and all vowels. Table 11 delivers all the experimental results of the LSTM and XGBoost approaches in the deep breath, shallow breath, all breaths, heavy cough, shallow cough, all coughs, count fast, count normal, all count, vowel a, vowel e, vowel o, and all vowel scenarios. The highest performance of the LSTM was achieved with an accuracy of 79.17%, 70.83%, 60.87%, 75.00%, 57.14%, 64.81%, 65.38%, 64.29%, 65.38%, 61.54%, 65.38%, 62.50%, and 60.53% for deep breath, shallow breath, all breaths, heavy cough, shallow cough, all coughs, count fast, count normal, all count, vowel a, vowel e, vowel o, and all vowel scenarios, respectively. The highest performance of XGBoost obtained an accuracy of 70.83%

Table 9 The best results of the proposed PSO-ELM in all scenarios

Hidden neurons	TP	TN	FP	FN	Accuracy	Precision	Recall	Specificity	F-measure	G-mean	Execution time (s)
Deep breath											
150	11	12	0	1	95.83	100.00	91.67	100.00	95.65	95.74	119.8567
Shallow breath											
300	10	12	0	2	91.67	100.00	83.33	100.00	90.91	91.29	119.5190
All breath											
250	22	19	4	1	89.13	84.62	95.65	82.61	89.80	89.96	132.7839
Heavy cough											
400	14	13	1	0	96.43	93.33	100.00	92.86	96.55	96.61	128.3579
Shallow cough											
550	14	12	2	0	92.86	87.50	100.00	85.71	93.33	93.54	127.9549
All cough											
200	26	22	5	1	88.89	83.87	96.30	81.48	89.66	89.87	176.8013
Count fast											
150	12	13	0	1	96.15	100.00	92.31	100.00	96.00	96.08	124.2523
Count normal											
350	13	14	0	1	96.43	100.00	92.86	100.00	96.30	96.36	125.7110
All count											
150	24	22	4	2	88.46	85.71	92.31	84.62	88.89	88.95	174.0120
Vowel a											
100	13	12	1	0	96.15	92.86	100.00	92.31	96.30	96.36	107.1898
Vowel e											
100	13	12	1	0	96.15	92.86	100.00	92.31	96.30	96.36	108.1326
Vowel o											
150	12	11	1	0	95.83	92.31	100.00	91.67	96.00	96.08	106.6515
All vowels											
200	28	35	3	10	82.89	90.32	73.68	92.11	81.16	81.58	198.1815

Table 10 The best experimental results of the basic ELM, NN, and RF techniques in all scenarios

Technique	TP	TN	FP	FN	Accuracy	Precision	Recall	Specificity	F-measure	G-mean	Execution time (s)
Deep breath											
ELM	6	10	2	6	66.67	75.00	50.00	83.33	60.00	61.24	0.5236
NN	8	7	5	4	62.50	61.54	66.67	58.33	64.00	64.05	21.5222
RF	6	9	3	6	62.50	66.67	50.00	75.00	57.14	57.74	3.3139
Shallow breath											
ELM	7	10	2	5	70.83	77.78	58.33	83.33	66.67	67.36	0.5218
NN	10	6	6	2	66.67	62.50	83.33	50.00	71.43	72.17	24.5918
RF	9	7	5	3	66.67	64.29	75.00	58.33	69.23	69.44	3.3163
All breaths											
ELM	18	12	11	5	65.22	62.07	78.26	52.17	69.23	69.70	0.9417
NN	16	11	12	7	58.70	57.14	69.57	47.83	62.75	63.05	33.4789
RF	17	11	12	6	60.87	58.62	73.91	47.83	65.38	65.82	3.8154
Heavy cough											
ELM	11	10	4	3	75.00	73.33	78.57	71.43	75.86	75.91	0.6609
NN	11	6	8	3	60.71	57.89	78.57	42.86	66.67	67.45	22.4395
RF	10	9	5	4	67.86	66.67	71.43	64.29	68.97	69.01	3.6511
Shallow cough											
ELM	8	10	4	6	64.29	66.67	57.14	71.43	61.54	61.72	0.6421
NN	6	11	3	8	60.71	66.67	42.86	78.57	52.17	53.45	22.6169
RF	7	9	5	7	57.14	58.33	50.00	64.29	53.85	54.01	3.7370
All coughs											
ELM	19	16	11	8	64.81	63.33	70.37	59.26	66.67	66.76	1.0641
NN	18	14	13	9	59.26	58.06	66.67	51.85	62.07	62.22	34.6208
RF	19	14	13	8	61.11	59.38	70.37	51.85	64.41	64.64	4.1569
Count fast											
ELM	8	10	3	5	69.23	72.73	61.54	76.92	66.67	66.90	0.5920
NN	4	11	2	9	57.69	66.67	30.77	84.62	42.11	45.29	28.5440
RF	9	8	5	4	65.38	64.29	69.23	61.54	66.67	66.71	4.3687
Count normal											
ELM	11	8	6	3	67.86	64.71	78.57	57.14	70.97	71.30	0.6178
NN	14	6	8	0	71.43	63.64	100.00	42.86	77.78	79.77	25.1147
RF	10	8	6	4	64.29	62.50	71.43	57.14	66.67	66.82	4.4511
All count											
ELM	16	17	9	10	63.46	64.00	61.54	65.38	62.75	62.76	1.0476
NN	16	17	9	10	63.46	64.00	61.54	65.38	62.75	62.76	37.9394
RF	12	19	7	14	59.62	63.16	46.15	73.08	53.33	53.99	4.8791
Vowel a											
ELM	6	11	2	7	65.38	75.00	46.15	84.62	57.14	58.83	0.4631
NN	7	8	5	6	57.69	58.33	53.85	61.54	56.00	56.04	26.5335
RF	8	6	7	5	53.85	53.33	61.54	46.15	57.14	57.29	3.3466
Vowel e											
ELM	7	9	4	6	61.54	63.64	53.85	69.23	58.33	58.54	0.4873
NN	9	8	5	4	65.38	64.29	69.23	61.54	66.67	66.71	30.6789
RF	5	10	3	8	57.69	62.50	38.46	76.92	47.62	49.03	3.1672
Vowel o											
ELM	10	5	7	2	62.50	58.82	83.33	41.67	68.97	70.01	0.4789
NN	9	4	8	3	54.17	52.94	75.00	33.33	62.07	63.01	22.5454
RF	7	7	5	5	58.33	58.33	58.33	58.33	58.33	58.33	3.1421
All vowels											
ELM	25	22	16	13	61.84	60.98	65.79	57.89	63.29	63.34	1.5370

Table 10 (continued)

Technique	TP	TN	FP	FN	Accuracy	Precision	Recall	Specificity	F-measure	G-mean	Execution time (s)
NN	22	21	17	16	56.58	56.41	57.89	55.26	57.14	57.15	53.8420
RF	29	17	21	9	60.53	58.00	76.32	44.74	65.91	0.6653	4.5728

Table 11 The best experimental results of the LSTM and XGBoost approaches in all scenarios

Technique	TP	TN	FP	FN	Accuracy	Precision	Recall	Specificity	F-measure	G-mean	Execution time (s)
Deep breath											
LSTM	8	11	1	4	79.17	88.89	66.67	91.67	76.19	76.98	116.448
XGBoost	10	7	5	2	70.83	66.67	83.33	58.33	74.07	74.54	63.124
Shallow breath											
LSTM	8	9	3	4	70.83	72.73	66.67	75.00	69.57	69.63	113.005
XGBoost	8	9	3	4	70.83	72.73	66.67	75.00	69.57	69.63	65.930
All breaths											
LSTM	17	11	12	6	60.87	58.62	73.91	47.83	65.38	65.82	128.850
XGBoost	14	15	8	9	63.04	63.63	60.87	65.22	62.22	62.24	85.102
Heavy cough											
LSTM	10	11	3	4	75.00	76.92	71.43	78.57	74.07	74.12	112.323
XGBoost	10	8	6	4	64.29	62.50	71.43	57.14	66.67	66.82	55.048
Shallow cough											
LSTM	7	9	5	7	57.14	58.33	50.00	64.29	53.85	54.01	114.679
XGBoost	13	8	6	1	75.00	68.42	92.86	57.14	78.79	79.71	58.063
All coughs											
LSTM	17	18	9	10	64.81	65.38	62.96	66.67	64.15	64.16	127.551
XGBoost	16	18	9	11	62.96	64.00	59.26	66.67	61.54	61.58	82.043
Count fast											
LSTM	8	9	4	5	65.38	66.67	61.54	69.23	64.00	64.05	117.751
XGBoost	9	8	5	4	65.38	64.29	69.23	61.54	66.67	66.71	60.707
Count normal											
LSTM	10	8	6	4	64.29	62.50	71.43	57.14	66.67	66.82	115.631
XGBoost	8	10	4	6	64.29	66.67	57.14	71.43	61.54	61.72	58.609
All count											
LSTM	11	23	3	15	65.38	78.57	42.31	88.46	55.00	57.66	126.962
XGBoost	15	16	10	11	59.62	60.00	57.69	61.54	58.82	58.83	90.167
Vowel a											
LSTM	6	10	3	7	61.54	66.67	46.15	76.92	54.55	55.47	101.736
XGBoost	8	7	6	5	57.69	57.14	61.54	53.85	59.26	59.30	61.645
Vowel e											
LSTM	6	11	2	7	65.38	75.00	46.15	84.62	57.14	58.83	103.035
XGBoost	6	10	3	7	61.54	66.67	46.15	76.92	54.55	55.47	60.859
Vowel o											
LSTM	7	8	4	5	62.50	63.64	58.33	66.67	60.87	60.93	98.829
XGBoost	7	9	3	5	66.67	70.00	58.33	75.00	63.64	63.90	59.786
All vowels											
LSTM	26	20	18	12	60.53	59.09	68.42	52.63	63.41	63.59	133.452
XGBoost	22	24	14	16	60.53	61.11	57.89	63.16	59.46	59.48	120.183

deep breath, 70.83% shallow breath, 63.04% all breaths, 64.29% heavy cough, 75.00% shallow cough, 62.96% all coughs, 65.38% count fast, 64.29% count normal, 59.62% all count, 57.69% vowel a, 61.54% vowel e, 66.67% vowel o, and 60.53% all vowels.

Moreover, the proposed PSO-ELM is compared with some recent works in terms of accuracy based on various scenarios (i.e., deep breath, shallow breath, all breaths, heavy cough, shallow cough, all coughs, count fast, count normal, vowel a, vowel e, and vowel o scenarios). Table 12 demonstrates the comparison accuracy results of the proposed PSO-ELM and some other previous works.

Discussion

Based on the abovementioned experimental results in Table 9, we can conclude a critical observation. The PSO can create suitable weights and biases for the single hidden layer of the ELM to minimize classification errors.

Table 12 The comparison of accuracy between the proposed PSO-ELM and other previous works

Technique	Accuracy	Technique	Accuracy
All breaths		All coughs	
PSO-ELM	89.13%	PSO-ELM	88.89%
SVM in [46]	81.50%	SVM [46]	85.70%
RF [47]	75.17%	RF [47]	70.69%
CNN [48]	70.37%	CNN [49]	88.48%
RF [50]	86.79%	Ensemble model [51]	77.10%
Technique	Accuracy	Technique	Accuracy
Deep breath		Shallow breath	
PSO-ELM	95.83%	PSO-ELM	91.67%
SVM [46]	62.30%	SVM in [46]	62.20%
BI-ATGRU [52]	94.50%		
Technique	Accuracy	Technique	Accuracy
Heavy cough		Shallow cough	
PSO-ELM	96.43%	PSO-ELM	92.86%
SVM [46]	72.30%	SVM [46]	74.10%
Technique	Accuracy	Technique	Accuracy
Count fast		Count normal	
PSO-ELM	96.15%	PSO-ELM	96.43%
SVM [46]	73.50%	SVM [46]	72.50%
Technique	Accuracy	Technique	Accuracy
Vowel a		Vowel e	
PSO-ELM	96.15%	PSO-ELM	96.15%
SVM [46]	59.30%	SVM [46]	68.20%
Technique		Accuracy	
Vowel o			
PSO-ELM		95.83%	
SVM [46]		69.20%	

Avoiding unsuitable weights and biases prevents the ELM from becoming stuck in the local maxima of weights and biases. Consequently, the performance of the PSO-ELM was impressive in thirteen different scenarios, with accuracies of 95.83%, 91.67%, 89.13%, 96.43%, 92.86%, 88.89%, 96.15%, 96.43%, 88.46%, 96.15%, 96.15%, 95.83%, and 82.89% for deep breath, shallow breath, all breath, heavy cough, shallow cough, all cough, count fast, count normal, all count, vowel a, vowel e, vowel o, and all vowel scenarios, respectively.

Furthermore, the additional observations can be made based on the results in Table 9. (a) Using a single sound of the respiratory system (e.g., deep breath, shallow breath, heavy cough, shallow cough, count fast, count normal, vowel a, vowel e, or vowel o) provides better results than combining two or more sounds of the respiratory system. (b) In both breath and cough, deepness and heaviness provide higher results than shallowness. (c) The speed (e.g., fast and normal) in count and the heaviness in cough deliver almost the same accuracy results.

Moreover, based on the results in Tables 9, 10, and 11, the PSO-ELM outperformed the basic ELM, NN, RF, LSTM, and XGBoost in thirteen different scenarios: deep breath, shallow breath, all breaths, heavy cough, shallow cough, all cough, count fast, count normal, all count, vowel a, vowel e, vowel o, and all vowels. This demonstrates that generating suitable weights and biases of the ELM leads to minimized classification errors. However, they outperform the proposed PSO-ELM in terms of the execution time because the PSO-ELM is based on PSO, which requires more time to obtain the best values of input weights and biases.

Lastly, based on all the results in Table 12, the PSO-ELM outperformed all the other previous works in the deep breath, shallow breath, all breaths, heavy cough, shallow cough, all coughs, count fast, count normal, vowel a, vowel e, and vowel o scenarios. This finding offers a promise that the proposed PSO-ELM is a reliable technique for the detection of COVID-19 by using different voice data of the respiratory system. Although the proposed method has shown good performance, there are some limitations, which are provided as follows:

- The voice datasets of the respiratory system that were used in this study for training and testing purposes are small.
- This study concentrated on classifying the voice data of the respiratory system into two classes (healthy/COVID-19) only, and other lung diseases were ignored.

Conclusion

In this study, we proposed a COVID-19 detection system based on the conventional MFCC features and PSO-ELM classifier. The PSO-ELM underwent thirteen different evaluation scenarios: deep breath, shallow breath, all breaths,

heavy cough, shallow cough, all coughs, count fast, count normal, all count, vowel a, vowel e, vowel o, and all vowels using the CHRSD dataset. The outcome indicated the advantage of the PSO-ELM over some previous works (see Table 12) and the basic ELM, NN, RF, LSTM, and XGBoost (see Table 9, 10, and 11) in the thirteen different scenarios. The performance of the PSO-ELM was impressive in all scenarios, with accuracies of 95.83%, 91.67%, 89.13%, 96.43%, 92.86%, 88.89%, 96.15%, 96.43%, 88.46%, 96.15%, 96.15%, 95.83%, and 82.89% for the deep breath, shallow breath, all breaths, heavy cough, shallow cough, all coughs, count fast, count normal, all count, vowel a, vowel e, vowel o, and all vowel scenarios, respectively. However, the current study has been evaluated on only a small dataset. Therefore, the future work of this study is to implement the proposed PSO-ELM on a larger dataset. Additionally, other optimization approaches for ELM will be further explored to generate the most suitable weights and biases for the ELM, which leads to the minimization of the classification errors.

Funding This study was funded by the Universiti Kebangsaan Malaysia (Research code: FRGS/1/2020/ICT02/UKM/02/1).

Data Availability The Corona Hack Respiratory Sound Dataset is available on <https://www.kaggle.com/praveengovi/coronahack-respiratory-sound-dataset>.

Declarations

Ethics Approval This article does not contain any study involving human or animal test subjects.

Conflict of Interest The authors declare no competing interests.

References

- Miyata R, Tanuma N, Hayashi M, Imamura T, Takanashi J-I, Nagata R, Okumura A, Kashii H, Tomita S, Kumada S. Oxidative stress in patients with clinically mild encephalitis/encephalopathy with a reversible splenic lesion (MERS). *Brain Develop.* 2012;34(2):124–7.
- Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA.* 2020;323(11):1061–9.
- Carfi A, Bernabei R, Landi F. Persistent symptoms in patients after acute COVID-19. *JAMA.* 2020;324(6):603–5.
- WHO coronavirus (COVID-19) dashboard. https://covid19.who.int/?gclid=CjwKCAjw64eJBhAGEiwABr9o2Mg-MBCGh7FvnEbtuQsG8icZ2fouAAAYrWPEOaUCrPvT49gytPdPZhoCuu4QAvD_BwE. Accessed 20 Aug 2021.
- Ieracitano C, Mammone N, Versaci M, Varone G, Ali A-R, Armentano A, Calabrese G, Ferrarelli A, Turano L, Tebala C. A fuzzy-enhanced deep learning approach for early detection of COVID-19 pneumonia from portable chest X-ray images. *Neurocomputing.* 2022.
- Albadr MAA, Tiun S, Ayob M, Al-Dhief FT, Omar K, Hamzah FA. Optimised genetic algorithm-extreme learning machine approach for automatic COVID-19 detection. *PLoS One.* 2020;15(12):e0242899.
- Turkoglu M. COVID-19 detection system using chest CT images and multiple kernels-extreme learning machine based on deep neural network. *IRBM.* 2021.
- Chang AB, Redding G, Everard M. Chronic wet cough: protracted bronchitis, chronic suppurative lung disease and bronchiectasis. *Pediatr Pulmonol.* 2008;43(6):519–31.
- Higenbottam T. Chronic cough and the cough reflex in common lung diseases. *Pulm Pharmacol Ther.* 2002;15(3):241–7.
- Chung KF, Pavord ID. Prevalence, pathogenesis, and causes of chronic cough. *The Lancet.* 2008;371(9621):1364–74.
- Knocikova J, Korpas J, Vrabec M, Javorka M. Wavelet analysis of voluntary cough sound in patients with respiratory diseases. *J Physiol Pharmacol.* 2008;59(Suppl 6):331–40.
- Korpáš J, Sadloňová J, Vrabec M. Analysis of the cough sound: an overview. *Pulm Pharmacol.* 1996;9(5–6):261–8.
- Al-Khassaweneh M, Bani Abdelrahman RE. A signal processing approach for the diagnosis of asthma from cough sounds. *J Med Eng Technol.* 2013;37(3):165–171.
- Pahar M, Miranda I, Diacon A, Niesler T. Deep neural network based cough detection using bed-mounted accelerometer measurements. In *ICASSP 2021–2021 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2021;8002–8006.* IEEE.
- Rudraraju G, Palreddy S, Mamidgi B, Sripada NR, Sai YP, Vodnala NK, Haranath SP. Cough sound analysis and objective correlation with spirometry and clinical diagnosis. *Informatics in Medicine Unlocked.* 2020;19: 100319.
- Belkacem AN, Ouhbi S, Lakas A, Benkhelifa E, Chen C. End-to-end AI-based point-of-care diagnosis system for classifying respiratory illnesses and early detection of COVID-19: a theoretical framework. *Front Med.* 2021;8:372.
- Bagad P, Dalmia A, Doshi J, Nagrani A, Bhamare P, Mahale A, Rane S, Agarwal N, Panicker R. Cough against COVID: evidence of COVID-19 signature in cough sounds. 2020. *arXiv preprint arXiv:2009.08790.*
- Ritwik KVS, Kalluri SB, Vijayasenani D. COVID-19 patient detection from telephone quality speech data. 2020. *arXiv preprint arXiv:2011.04299.*
- Laguarta J, Hueto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open Journal of Engineering in Medicine and Biology.* 2020;1:275–81.
- Brown C, Chauhan J, Grammenos A, Han J, Hasthanasombat A, Spathis D, Xia T, Cicuta P, Mascolo C. Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data. 2020. *arXiv preprint arXiv:2006.05919.*
- Imran A, Posokhova I, Qureshi HN, Masood U, Riaz MS, Ali K, John CN, Hussain MI, Nabeel M. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked.* 2020;20: 100378.
- Mouawad P, Dubnov T, Dubnov S. Robust detection of COVID-19 in cough sounds. *SN Computer Science.* 2021;2(1):1–13.
- Hassan A, Shahin I, Alsabek MB. COVID-19 detection system using recurrent neural networks. In *2020 international conference on communications, computing, cybersecurity, and informatics (CCCI).* 2020;1–5. IEEE.
- Basu S, Chakraborty J, Bag A, Aftabuddin M. A review on emotion recognition using speech. In *2017 international conference on inventive communication and computational technologies (ICICCT), 2017;109–114.* IEEE.
- Mar LL, Pa WP. Depression detection from speech emotion recognition: Seventeenth international conference on computer applications (ICCA). 2019.
- Tripathi A, Singh U, Bansal G, Gupta R, Singh AK. A review on emotion detection and classification using speech. in *Proceedings*

- of the international conference on innovative computing & communications (ICICC). 2020.
27. Albadr MAA, Tiun S, Ayob M, Mohammed M, AL-Dhief FT. Mel-frequency cepstral coefficient features based on standard deviation and principal component analysis for language identification systems. *Cogn Comput*. 2021;1–18.
 28. Trang H, Loc TH, Nam HBH. Proposed combination of PCA and MFCC feature extraction in speech recognition system. In 2014 international conference on advanced technologies for communications (ATC 2014). 2014;697–702. IEEE.
 29. Kashif K, Wu Y, Michael A. Consonant phoneme based extreme learning machine (ELM) recognition model for foreign accent identification. in Proceedings of the 2019 the world symposium on software engineering. 2019;68–72.
 30. Abu F, Hagiescu D, Vladutu L, Puica M-A. Neural network approaches for children's emotion recognition in intelligent learning applications. In EDULEARN15 7th Annu Int Conf Educ New Learn Technol Barcelona, Spain, 6th-8th. 2015.
 31. Albadr MAA, Tiun S, Ayob M, AL-Dhief FT, Omar K, Maen MK. Speech emotion recognition using optimized genetic algorithm-extreme learning machine. *Multimed Tools Appl*. 2022;1–27.
 32. Albadr MAA, Tiun S, Ayob M, AL-Dhief FT, Abdali T-AN, Abbas AF. Extreme learning machine for automatic language identification utilizing emotion speech data. In 2021 international conference on electrical, communication, and computer engineering (ICECCE). 2021;1–6. IEEE.
 33. Albadr MAA, Tiun S, AL-Dhief FT, Sammour MA. Spoken language identification based on the enhanced self-adjusting extreme learning machine approach. *PLoS ONE*. 2018;13(4): e0194770.
 34. Albadr MAA, Tiun S, Ayob M, AL-Dhief FT. Spoken language identification based on optimised genetic algorithm-extreme learning machine approach. *Int J Speech Technol*. 2019;22(3):711–727.
 35. Albadra MAA, Tiuna S. Extreme learning machine: a review. *Int J Appl Eng Res*. 2017;12(14):4610–23.
 36. Albadr MAA, Tiun S. Spoken language identification based on particle swarm optimisation-extreme learning machine approach. *Circuits Systems Signal Process*. 2020;39(9).
 37. Kumar S, Pal SK, Singh R. A novel hybrid model based on particle swarm optimisation and extreme learning machine for short-term temperature prediction using ambient sensors. *Sustain Cities Soc*. 2019;49: 101601.
 38. Muduli D, Dash R, Majhi B. Fast discrete curvelet transform and modified PSO based improved evolutionary extreme learning machine for breast cancer detection. *Biomed Signal Process Control*. 2021;70: 102919.
 39. Corona Hack Respiratory Sound Dataset. Kaggle dataset. <https://www.kaggle.com/praveengovi/coronahack-respiratory-sound-dataset>. Accessed 01 Jun 2021.
 40. Al-Dhief FT, Latiff NMAA, Malik NNNA, Salim NS, Baki MM, Albadr MAA, Mohammed MA. A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms. *IEEE Access*. 2020;8:64514–64533.
 41. Dogra A, Kaul A, Sharma R. Automatic recognition of dialects of Himachal Pradesh using MFCC & GMM. In 2019 5th international conference on signal processing, computing and control (ISPCC). 2019;134–137. IEEE.
 42. Albadr MA, Tiun S, Ayob M, AL-Dhief F. Genetic algorithm based on natural selection theory for optimization problems. *Symmetry*. 2020;12(11):1758.
 43. AL-Dhief FT, Baki MM, Latiff NMAA, Malik NNNA, Salim NS, Albader MAA, Mahyuddin NM, Mohammed MA. Voice pathology detection and classification by adopting online sequential extreme learning machine. *IEEE Access*. 2021;9:77293–77306.
 44. AL-Dhief FT, Latiff NMAA, Baki MM, Malik NNNA, Sabri N, Albadr MAA. Voice pathology detection using support vector machine based on different number of voice signals. In 2021 26th IEEE Asia-Pacific conference on communications (APCC), 2021;1–6. IEEE.
 45. AL-Dhief FT, Latiff NMAA, Malik NNNA, Sabri N, Baki MM, Albadr MAA, Abbas AF, Hussein YM, Mohammed MA. Voice pathology detection using machine learning technique. In 2020 IEEE 5th international symposium on telecommunication technologies (ISTT). 2020;99–104. IEEE.
 46. Dash TK, Mishra S, Panda G, Satapathy SC. Detection of COVID-19 from speech signal using bio-inspired based cepstral features. *Pattern Recogn*. 2021;117: 107999.
 47. Muguli A, Pinto L, Sharma N, Krishnan P, Ghosh PK, Kumar R, Bhat S, Chetupalli SR, Ganapathy S, Ramoji S. DiCOVA challenge: dataset, task, and baseline system for COVID-19 diagnosis using acoustics. 2021. arXiv preprint [arXiv:2103.09148](https://arxiv.org/abs/2103.09148).
 48. Chowdhury ME, Ibtehad N, Rahman T, Mekki YMS, Qibalwey Y, Mahmud S, Ezeddin M, Zughaier S, Al-Maadeed SAS. QUCough-Scope: an artificially intelligent mobile application to detect asymptomatic COVID-19 patients using cough and breathing sounds. 2021. arXiv preprint [arXiv:2103.12063](https://arxiv.org/abs/2103.12063).
 49. Lella KK, Pja A. Automatic COVID-19 disease diagnosis using 1D convolutional neural network and augmentation with human respiratory sound based on parameters: cough, breath, and voice. *AIMS Public Health*. 2021;8(2):240.
 50. Meister JA, Nguyen KA, Luo Z. Audio feature ranking for sound-based COVID-19 patient detection. 2021. arXiv preprint [arXiv:2104.07128](https://arxiv.org/abs/2104.07128).
 51. Chaudhari G, Jiang X, Fakhry A, Han A, Xiao J, Shen S, Khanzada A. Virufy: global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough. 2020. arXiv preprint [arXiv:2011.13320](https://arxiv.org/abs/2011.13320).
 52. Wang Y, Hu M, Li Q, Zhang X-P, Zhai G, Yao N. Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner. 2020. arXiv preprint [arXiv:2002.05534](https://arxiv.org/abs/2002.05534).

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.