Sound Classification Using Evolving Ensemble Models and Particle Swarm Optimization

Li Zhang[1], Chee Peng Lim[2], Yonghong Yu[3], and Ming Jiang[4]

[1]Department of Computer Science
Royal Holloway, University of London
Surrey, TW20 0EX, UK

[2]Institute for Intelligent Systems Research and Innovation
Deakin University
Waurn Ponds, VIC 3216, Australia

[3]College of Tongda
Nanjing University of Posts and Telecommunications
Nanjing, China

[4]School of Computer Science
Faculty of Technology
University of Sunderland, UK

Email: li.zhang@rhul.ac.uk; chee.lim@deakin.edu.au; yuyh@njupt.edu.cn; ming.jiang@sunderland.ac.uk

Abstract.
Automatic sound classification attracts increasing research attention owing to its vast applications, such as robot navigation, environmental sensing, musical instrument classification, medical diagnosis, and surveillance. In this research, we propose an ensemble convolutional bidirectional Long Short-Term Memory (CBiLSTM) network with optimal hyper-parameter selection for undertaking sound classification. We first transform each audio signal into a spectrogram representation using the Short-time Fourier transform (STFT). A Particle Swarm Optimization (PSO) variant is subsequently proposed to optimize the learning rate, weight decay, numbers of filters and hidden units in the convolutional and BiLSTM layers, respectively, in order to extract effective spatial-temporal characteristics from the spectrogram inputs. To tackle the issue of stagnation in optimization, the proposed algorithm incorporates local exploitation using secant and Newton-Raphson methods, promising leader generation using regular and irregular super-ellipse formulae, and three-dimensional spherical search coefficients. Moreover, it takes into account multiple fused elite signals in conjunction with numerical analysis based exploitation to balance between diversification and intensification. A variety of CBiLSTM networks with distinctive optimized settings are devised. An ensemble model is then constructed by incorporating a set of three yielded networks based on a majority voting scheme. Evaluated using several audio data sets, our ensemble CBiLSTM networks outperform those with default and optimal settings identified by other search methods, existing deep architectures and state-of-the-art related studies. In addition to sound classification tasks, the proposed PSO algorithm also outperforms a number of classical and advanced search methods for solving diverse unimodal and multimodal benchmark functions with statistical significance.

Keywords: Sound Classification, Evolutionary Algorithm, Deep Convolutional Bidirectional Long Short-Term Memory Network and Ensemble Classifier.

1. INTRODUCTION

Automatic sound classification has been widely adopted in a variety of real-life applications [1-5], such as health monitoring, security surveillance, environmental sensing, robot navigation, disaster identification and voice activity recognition. Sound classification tasks involve the extraction of acoustic characteristics from the audio signals and the subsequent identification of different sound classes. The broad range of sound classification deployments can be categorized into several disciplines, which include speech recognition, music instrument identification, environmental sound classification and abnormal sound identification for disease diagnosis. In comparison with speech and music sounds which possess proper high-level structures, the categories of diagnostic (such as respiratory and heart) sounds and environmental audio signals tend to be unstructured, which contain various clinical and natural acoustic noise. Furthermore, because of different sound production mechanisms (e.g. different body recording locations and equipment), sound classification with medical and environmental audio clips stands as a challenging problem [1-5].

It is thus important to generate effective audio representations to capture important discriminative characteristics and environmental acoustic cues to inform audio classification. In this respect, deep neural networks have demonstrated superior performances for signal processing tasks owing to their significant feature learning capabilities. Moreover, the hybridization of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) has gained increasing popularity because of the significant capabilities in spatial-temporal feature extraction [6, 7, 8]. In this research, we propose a convolutional bidirectional Long Short-Term Memory (CBiLSTM) network for sound classification, in view of the enhanced capabilities in audio representation. Since network structures and learning hyper-parameters play crucial roles in generating effective spatial-temporal features, we propose a new Particle Swarm Optimization (PSO) variant to optimize the learning parameters and configurations, which include the learning rate, weight decay, and convolutional and BiLSTM layer topologies, for performance enhancement.

Specifically, in this research, we propose evolving ensemble deep CBiLSTM networks with optimal hyper-parameter selection for undertaking sound classification. Firstly, the Short-time Fourier transform (STFT) is used to transform audio signals into spectrograms. In order to extract effective spatial-temporal dependencies from spectrogram inputs, a new PSO variant is devised to identify the optimal settings of the learning rate, weight decay, numbers of filters and hidden units in the convolutional and BiLSTM layers, respectively. The proposed PSO algorithm incorporates secant and Newton-Raphson algorithms for swarm leader enhancement, elite signal generation using regular and irregular super-ellipse adaptive operators, and three-dimensional (3D) spherical search coefficients. It exploits diverse hybrid elite indicators and numerical analysis based intensification to overcome stagnation. A set of optimized CBiLSTM networks with different layer structures and learning settings is produced using the proposed algorithm. An ensemble model is subsequently constructed by integrating three such optimized networks using a majority voting mechanism. Because of numerical analysis based intensification, elite signal generation operators, and parametric search coefficients, the proposed model shows great superiority in devising efficient hyper-parameter settings for yielding effective spatial-temporal dynamics. Our ensemble networks outperform those with optimal settings identified through classical and advanced search methods in diverse sound classification tasks. Figure 1 illustrates the system architecture.

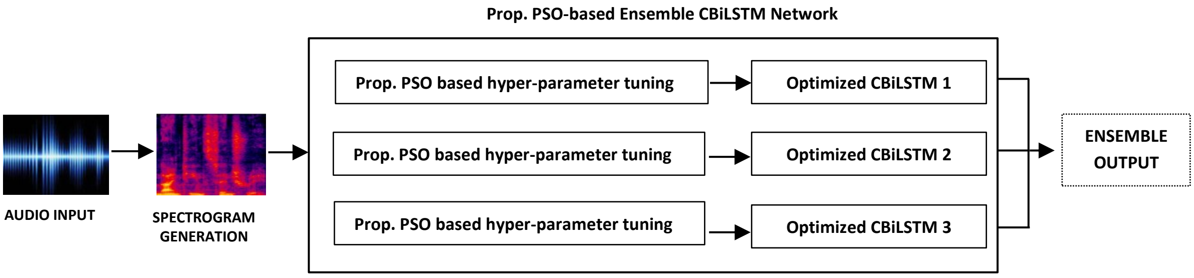

Prop. PSO-based Ensemble CBiLSTM Network
Prop. PSO based hyper-parameter tuning
Optimized CBiLSTM 1
Prop. PSO based hyper-parameter tuning
Optimized CBiLSTM 2
Prop. PSO based hyper-parameter tuning
Optimized CBiLSTM 3
AUDIO INPUT
SPECTROGRAM GENERATION
ENSEMBLE OUTPUT

Figure 1 The overall system architecture

The contributions of this research are summarized as follows.

1. We propose evolving ensemble CBiLSTM networks with optimal hyper-parameter selection using a PSO variant for undertaking sound classification. The proposed PSO model is used to optimize the associated network and learning configurations, such as the learning rate, weight decay, numbers of filters and hidden neurons in the convolutional and BiLSTM layers, respectively, in order to capture effective spatial and sequential cues of the spectrogram inputs. The yielded optimized networks with distinctive layer topologies and learning settings are able to equip the respective ensemble models with great diversity and complementary characteristics for enhancing classification performance.
2. The proposed PSO model incorporates secant and Newton-Raphson algorithms for swarm leader enhancement, regular and irregular elliptical formulae oriented crossover operators for hybrid

indicator generation, and diverse spherical search coefficients, to overcome stagnation. Firstly, after initializing the swarm, the secant and Newton-Raphson methods are used to enhance the global best solution. In comparison with stochastic random walk strategies such as Gaussian, Cauchy and Levy distributions for leader enhancement [9, 10], these root-finding numerical analysis algorithms provide *guided* local exploitation of promising solutions to overcome stagnation and accelerate convergence.

Secondly, to better balance the search between exploitation and exploration, a hybrid leader generation mechanism with adaptive weightings is proposed. Specifically, hybrid elite leader signals are devised using crossover operators based on regular and irregular adaptive super-ellipse formulae. In comparison with single-leader based search processes [11-20] as well as search operations led by multiple leaders with equal weightings [21-24], we leverage adaptive weighting factors yielded by super-ellipse formulae for hybrid signal generation to better balance between diversification and intensification. The enhanced swarm leader and the second-best solution are integrated using adaptive (increasing and decreasing) weights produced by regular and irregular 2D elliptical formulae. Such adaptive factors enable the search process to focus on global exploration by assigning a higher weighting to the second-best leader at the beginning of the search process, and gradually switch to local exploitation by emphasizing the impact of the global best solution towards the end of search iterations.

Thirdly, spherical parametric surfaces are exploited to produce search coefficients. Owing to the generation of distinctive spherical shapes in comparison with chaotic maps [16, 17] or Gaussian distribution [12] oriented search coefficients, the proposed spherical surfaces equip each particle with distinctive search scales, directions and trajectories in both social and cognitive components, in order to increase the chances of finding global optimality.

In short, the proposed model employs a variety of fused elite signals with diverse search trajectories as well as numerical-based intensification to overcome local optimum traps. The yielded elite leaders enable the search process to focus on local and global promising regions adaptively throughout the iterations. The empirical results indicate that our model shows superior efficiency in devising efficient layer structures and learning settings in CBiLSTM networks for generating effective spatial-temporal patterns to enhance performance.

3. Evaluated using challenging respiratory, heart and environmental sound data sets, i.e. ICBHI 2017 [25], the PhysioNet Computing in Cardiology Challenge 2016 (PhysioNet/CinC Challenge 2016) [26], and ESC-10 [27], the proposed ensemble CBiLSTM models outperform those with default and optimal settings identified by other search methods, existing deep networks and state-of-the-art related studies, significantly. In addition to sound classification tasks, the proposed PSO algorithm shows statistically significant superiority over a number of classical and advanced search methods for solving diverse unimodal and multimodal artificial landscapes.

The paper is organized as follows. In Section 2, we discuss a variety of related studies on sound classification for medical diagnosis and environmental sensing, as well as diverse recently proposed PSO variants. The proposed model with elite indicator generation and numerical analysis based exploitation is introduced in Section 3. Section 4 explains spectrogram generation and ensemble CBiLSTM networks with hyper-parameter selection. We present a comprehensive evaluation and introduce future directions in Sections 5 and 6, respectively.

2. RELATED WORK

In this section, we discuss state-of-the-art studies on respiratory, heart and environmental sound classification, as well as a variety of PSO variants and other search methods for solving diverse optimization problems.

2.1 Respiratory, Heart and Environmental Sound Classification

Automatic sound classification attracts significant research attention, in view of its broad range of real-life applications, such as autonomous navigation, environmental surveillance, and medical diagnosis. There are a variety of recent studies in the literature pertaining to heart and respiratory disease diagnosis using sound recordings [3, 28, 29, 30]. As an example, Wu et al. [29] proposed an ensemble CNN model with a Savitzky-Golay filter for heart sound anomaly detection. Their work first employed the Savitzky-Golay filter to denoise audio signals. STFT was subsequently used to transform the denoised audios into spectrograms.

Audio features such as mel-spectrograms and mel-frequency cepstral coefficients (MFCCs) were also extracted from on the spectrograms. The resulting features, i.e. spectrograms, mel-spectrograms and MFCCs, were used to train three VGG networks. An ensemble model was constructed based on these trained networks using a majority voting mechanism. Evaluated using the benchmark PhysioNet/CinC Challenge 2016 data set, their ensemble model outperformed several existing methods for abnormal heart sound classification. Zhang et al. [30] employed spectrogram and tensor decomposition for heart sound classification. After noise elimination, the heart cycles were estimated and segmented. STFT was used to transform each segmented cycle into a spectrogram. A bilinear interpolation method was used to rescale the spectrogram into a fixed size. Feature selection was then performed using a tensor decomposition method to extract the most significant structural pathological information from the spectrograms. A Support Vector Machine (SVM) classifier was used for heart abnormality identification. Evaluated using several heart sound data sets (e.g. the PASCAL challenge and PhysioNet/CinC Challenge 2016), their model achieved superior performances for abnormal heart signal detection.

Xiao et al. [31] conducted heart sound classification using a light-weight CNN model with an attention mechanism. Their proposed CNN architecture consisted of both clique and transition blocks. The transition block comprised batch normalization (BN), ReLu, an attention mechanism and an average pooling layer. The clique block adopted composite layers to connect feature maps bidirectionally. Instead of using the standard convolution, a separable convolution with the inverted bottleneck topology was used for feature extraction. Multi-scale feature fusion was also conducted by concatenating feature maps from different clique blocks. In comparison with related studies, their model illustrated superior capabilities in extracting discriminative features to enhance performance, along with a significant reduction on the number of model parameters. Kiranyaz et al. [32] conducted anomaly detection from heart sound signals using a 1D CNN model. A data purification scheme was incorporated in the back-propagation (BP) training stage with the intention to reduce the bias caused by the potential occurrence of normal heart beats embedded in the abnormal streams. Specifically, each heart beat was first segmented from the overall sound signal and subsequently normalized. Then the 1D CNN was used to classify each normalized beat as a normal or abnormal case. The majority voting rule was adopted to determine the final class label of the entire signal. In particular, during the training stage, their BP learning mechanism only considered beats classified as abnormal cases in the abnormal signal with a sufficient confidence level. Such a training scheme aimed to avoid the influence and distraction of the normal beats in the abnormal stream in the BP process. Their model achieved superior performance in comparison with those of existing methods.

Shuvo et al. [33] proposed a Convolutional Recurrent Neural Network (CRNN) model, namely CardioXNet, for classification of heart anomalies, including aortic stenosis, mitral stenosis, mitral regurgitation, and mitral valve prolapse. Three CNN streams were incorporated within their proposed network for hybrid feature learning, while BiLSTM layers were employed for extracting temporal dynamics. The three CNN components were frequency feature extractor (FFE), adaptive feature extractor (AFE), and pattern extractor (PE). The FFE and PE CNN modules both contained four 1D convolutional layers and two max-pooling layers for frequency and appearance pattern extraction, while the AFE CNN structure included 2D convolutional layers and squeeze-expansion layers for spatial feature extraction. The extracted features from these three CNN components were concatenated as the inputs for the BiLSTM layers. Moreover, the temporal information extracted by the BiLSTM component was also concatenated with the features extracted from the three CNN streams via a skip connection for final abnormality detection. Evaluated using the Github PCG and PhysioNet/CinC Challenge data sets, their CRNN model achieved promising performance.

There are also various recent studies for respiratory sound anomaly detection. Perna and Tagarelli [28] performed respiratory sound anomaly detection using LSTM networks. Using a benchmark ICBHI respiratory data set, the model detected the presence of crackles and/or wheezes in each respiratory cycle, as well as healthy, chronic and non-chronic lung conditions in each recording. Pre-processing procedures, such as respiratory cycle segmentation, MFCC feature extraction and Min-Max and Z-score normalization, were applied to yield the audio representations. The LSTM network in combination with seven window configurations (e.g. non-overlapping partitioning and 50% overlapping between consecutive windows) was used for anomaly detection. Their models achieved superior performance for crackle and/or wheeze detection as well as healthy/chronic/non-chronic signal classification. García-Ordás et al. [34] proposed the use of CNN in conjunction with variational autoencoders (VAE) for respiratory pathology classification. Mel-spectrograms were used to represent the audio signals. A CNN model was used for not only classifying healthy/chronic/non-chronic cases, but also identifying six specific pathology conditions (e.g. pneumonia, Chronic Obstructive Pulmonary Disease (COPD) and bronchiolitis). To tackle the class imbalance issues, several oversampling techniques, such as VAE, Synthetic Minority Oversampling Technique (SMOTE) and

Adaptive Synthetic Sampling Method (ASSM), were exploited to yield new labelled data for the minority classes in both classification tasks. The empirical results indicated that CNN with the VAE-based data augmentation method outperformed other oversampling techniques.

Zhao et al. [35] proposed a bidirectional gated recurrent unit (BiGRU)-Attention Network with XGBoost for respiratory sound classification. Their work first employed the time domain (e.g. short-term average energy and short-term average zero-crossing rate) and spectral features (e.g. spectral centroid, slope and contrast) for signal representations. Then, feature selection using the Gradient Boosting Decision Tree (GBDT) algorithm was conducted. The BiGRU-Attention model was used to extract temporal dependencies from the obtained optimal features. The ensemble classification XGBoost model was used to detect the presence of wheezes and crackles. Data augmentation techniques such as Griffifin-Lim and WORLD Vocoder were also employed to tackle the class imbalance problem. Their model (BiGRU-Attention-XGBoost with data augmentation) showed better performances than those of BiLSTM, BiGRU, BiLSTM-Attention, and BiGRU-Attention, for respiratory anomaly classification. Chen et al. [36] employed optimized S-transform (OST) and ResNets for identification of wheeze, crackle, and normal sounds, while Oletic and Bilas [37] performed wheezing detection from compressive sensing reconstructed spectra.

On the other hand, environmental sound classification has also attracted increasing research attention. Esmaeilpour et al. [38] proposed a Weighted Cycle-Consistent Generative Adversarial Network (WCCGAN) and unsupervised feature learning for environmental sound classification. Firstly, a discrete wavelet transform (DWT) was used to convert the audio clips into spectrogram representations. Motivated by the Cycle-Consistent Generative Adversarial Network (CCGAN), a Weighted CCGAN (WCCGAN) variant was proposed for high-level data augmentation of the extracted spectrograms. The WCCGAN model incorporated not only two identity mapping functions but also different architectures for both generator and discriminator with the intention to increase inter- and intra-class variations. The speeded up robust feature (SURF) (a variant of the scale invariant feature transform (SIFT)) descriptors and the spherical K-Means++ algorithm were used for feature extraction and feature clustering, respectively. A Random Forest (RF) classifier was adopted for classification of the audio clips. Evaluated using several environmental sound data sets, their model outperformed AlexNet and GoogLeNet as well as other related studies significantly. Zhang et al. [39] proposed a CRNN with an attention mechanism for environmental sound classification. Their CRNN model consisted of eight convolutional and two BiGRU layers. STFT was used to generate spectrograms. The obtained spectrograms were subsequently processed by a 128-band gammatone filter bank. The log gammatone spectrograms were obtained by converting the above outputs into the logarithmic scale. The delta information of the spectrograms was also derived, which was concatenated with the log gammatone spectrograms. These final concatenated features were used as the CRNN inputs. In addition, a sigmoid based attention mechanism was applied to the CNN and RNN layers, respectively. Data augmentation by mixing up spectrograms was also performed to enhance performance. Their work illustrated superior performance over those of existing studies when evaluated using several environmental sound data sets.

Medhat et al. [40] proposed a ConditionaL Neural Network (CLNN) and a Masked CLNN (MCLNN) pertaining to time-frequency spectrogram representations for sound classification. The CLNN model captured the temporal inter-frame relationships by using conditional connections as well as taking both the preceding and succeeding frames into account in the inference process. In order to better describe the spatial patterns from the time-frequency representation, the CLNN was extended into MCLNN, which employed a filter bank-like pattern to enable the exploration of different feature combinations. Their work showed impressive performances for the evaluation of music and environmental sound data sets. Besides environmental sound classification, other audio recognition tasks were available in the literature. As an example, Kuang et al. [41] conducted affective acoustic signal classification using an improved AlexNet with MFCC, STFT, and simplified inverse filter tracked features. Evaluated using emotional speech data sets, their work produced an impressive performance for identification of several basic emotion categories from affective acoustic signals.

2.2 Evolutionary and Swarm Intelligence Algorithms

Evolutionary algorithms offer effective search capabilities in solving a variety of optimization problems [42-46]. As one of the popular swarm intelligence algorithms, PSO [11] has been widely adopted in tackling discriminative feature selection, hyper-parameter fine-tuning, shallow and deep network generation, job scheduling and benchmark optimization. As indicated in Equations (1)-(2), the PSO algorithm employs the swarm leader $gbest$ and the personal best experience p_i to lead the search process.

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{1}$$

$$v_{id}^{t+1} = w \times v_{id}^{t} + c_1 \times r_1 \times (p_{id} - x_{id}^{t}) + c_2 \times r_2 \times (gbest_d - x_{id}^{t}) \quad (2)$$

where x_{id}^{t+1} and v_{id}^{t+1} denote the position and velocity for particle i in the $(t+1)^{th}$ iteration and the d^{th} dimension, respectively. The inertia weight, w, determines the impact of the previous velocity for generating the new velocity. The acceleration coefficients are denoted as c_1 and c_2, which are used to adjust the search steps in the cognitive and social components, respectively. In addition, r_1 and r_2 represent random vectors, where each element is uniformly distributed within [0, 1] in its respective dimension.

As illustrated in Equations (1)-(2), owing to the guidance of single global best solution, the search process of the PSO algorithm is likely to be trapped in local optima [47-50]. Diverse variants of the PSO model have been proposed in recent years to overcome the limitations. As an example, Li et al. [51] proposed an improved sticky binary PSO (ISBPSO) algorithm with new initialization and search space reduction strategies for discriminative feature selection. Their PSO variant initialized the swarm using feature weighting information provided by the entropy-based mutual information method to increase swarm quality. A masking strategy was also proposed to dynamically reduce the search space. A particular feature would be eliminated when it failed to be included in any personal best solution after a specified number of iterations. Such mask information was updated after a sufficient number of iterations to ensure an adequate exploration of the preceding feature space. Genetic operations (e.g. crossover and mutation operations) were also used to fine-tune and enhance the personal and global best solutions with the attempt to overcome stagnation. Evaluated using 12 UCI data sets, their model achieved improved performance with smaller selected feature subsets in most test cases, as compared with those from the sequential forward and backward selection strategies and six baseline PSO-based feature selection methods. Zhang et al. [52] proposed a PSO variant, namely Bayesian comprehensive learning PSO (BCLPSO), for solving diverse benchmark functions. Instead of relying on the swarm leader, their model produced a new exemplar using the Bayesian posterior probability to guide the search process. Specifically, a Bayesian iteration method was used for exemplar selection based on posterior probability. The historical prior information of particle swarm was used for posterior probability calculation. Evaluated using the CEC2017 test suite and practical quality control engineering optimization problems, their model achieved better performance than those of other advanced search methods. Xie et al. [53] developed a PSO variant for discriminative feature selection. The model incorporated an adaptive exemplar breeding mechanism, search coefficients yielded using trigonometric functions, and exponential exploitation and dispatching schemes for swarm leader and worse solution enhancement, respectively. A Logistic map was exploited to initialize the swarm. Multiple local and global best solutions were aggregated through adaptive weights with respect to exemplar generation. Nonlinear search parameters were produced using four formulae embedding sine, cosine, and hyperbolic tangent functions. Randomly selected personal best solutions as well as stochastic distributions were used to enhance three worst solutions in the swarm. An adaptive exponential function was also adopted for swarm leader enhancement. Evaluated using a total of 13 data sets, the model illustrated superior performance in comparison with those of other state-of-the-art PSO methods. Kılıç et al. [54] proposed a multi-population based PSO (MPPSO) algorithm for feature selection. Two swarms were initialized in the MPPSO model. One was randomly generated, while the other was yielded using the Relieff feature ranking method. Evaluated using 26 UCI and another 3 data sets from the feature selection repository of Arizona State University, their model identified the smallest feature subsets while achieving enhanced performance.

Molaei et al. [55] proposed a modified PSO algorithm with an enhanced Learning strategy and Crossover operator (PSOLC) for solving mathematically generated landscapes. To increase search diversity, the personal best experiences of all the particles were taken into account in the cognitive component for velocity updating. Besides employing an adaptively decreasing inertia weight factor, the cognitive acceleration coefficient was formulated as a proportion to the inverse of the fitness score pertaining to the personal best experience of the current particle. The largest cognitive search parameter among all the particles was also assigned as the social acceleration coefficient in each iteration. A crossover operator was used to diversify the swarm by applying a proportion of a randomly selected particle to the current particle. Evaluated using 29 benchmark functions, PSOLC outperformed the original and advanced PSO algorithms. Kan et al. [56] developed an adaptive PSO algorithm (APSO) for hyper-parameter identification in a 1D CNN model. An adaptive inertia weight scheme was proposed. The inertia weight was yielded based on the mean and minimum fitness scores of the overall population and that of the current particle. APSO was used to identify optimal settings of a set of 10 hyper-parameters with respect to network layer structure and learning configurations. Their devised CNN model outperformed a SVM classifier and deep networks with manual settings for Internet-of-Things (IoT) network intrusion detection, significantly. Li et al. [57] proposed a multipopulation cooperative PSO (MPCPSO) model with a mixed mutation strategy for solving diverse benchmark functions. The model divided the swarm into two populations in each iteration, i.e. elitist and general populations. For each particle in the general population, its dimension was split into several sub-sections by using a dynamic segment-based mean learning strategy (DSMLS). Exemplars were generated for each sub-section using particles from the elitist population via tournament selection. On the other hand, a multidimensional comprehensive learning strategy (MDCLS) was proposed to guide the search process of particles in the elitist population. A differential mutation operator was also used to increase search diversity. Evaluation using diverse benchmark functions, their model showed significant superiority over other methods in terms of performance and convergence speed.

Lawrence et al. [58] developed a modified PSO algorithm for evolving CNN architecture generation. A group-based encoding strategy was proposed to ensure that the frequency and position of the pooling operations could be adjusted in accordance with the image size. A new velocity updating scheme was also proposed by identifying the key network layer configuration variations between particles. The partial velocity was taken into account for position updating. Evaluated using eight well-known benchmark data sets (e.g. Rectangles, MNIST, and several MNIST variant data sets), their model outperformed several state-of-the-art deep architecture generation methods, such as psoCNN, in terms of accuracy and computational cost. Phung and Ha [59] proposed a spherical vector-based PSO model for navigating unmanned aerial vehicles (UAVs) in real-world complex environments, while a fuzzy hierarchical surrogate assisted probabilistic PSO was proposed by Chu et al. [60] for solving high-dimensional expensive optimization problems. Other PSO variants proposed in recent years include, e.g. a bi-objective PSO for incremental classifier generation in crime prediction [61], a micro-GA embedded PSO feature selection technique for facial expression recognition [47], a Bare-bones PSO variant for discriminative feature selection in lymphoblastic leukaemia diagnosis [21], a PSO variant with exemplar generation using sine, cosine and tanh formulae for deep ensemble network generation pertaining to video action classification [49], an adaptive learning PSO (ALPSO) model for hyper-parameter fine-tuning for skin lesion segmentation [62], a PSO variant with cosine oriented search coefficients for deep network generation with residual connections and dense connectivity [63], and a quantum behaved PSO algorithm with binary encoding for devising deep CNN architectures pertaining to image classification [64].

Besides PSO methods, Grey Wolf Optimizer (GWO) [22-24] and its variants have been used in solving diverse optimization problems. As an example, Martin et al. [65] proposed a mixed GWO (mixedGWO) algorithm for joint denoising and unmixing of real-world multispectral images. The mixedGWO model was formed by combining an improved discrete GWO and a global continuous GWO, which was capable of searching parameter settings in both discrete and continuous search spaces. Firstly, an improved discrete GWO was proposed, where the position updating was conducted by randomly selecting a leader among a candidate group comprising three best wolf leaders and two randomly recruited individuals from the swarm. One out of these five candidate solutions was randomly selected as the elite signal to guide the search process during global exploration, while one out of the three global best leaders was randomly recruited to lead the swarm during local exploitation. Secondly, a global continuous GWO algorithm was devised, where the mean position of the aforementioned five candidate signals, i.e. three best wolf leaders and two randomly recruited individuals, was used to guide global exploration. In addition, the average position of the three best wolf leaders was adopted to lead intensification. The resulting search strategies in mixedGWO therefore can be used for discrete or continuous type of parameters. An adaptive variant of mixedGWO, i.e. amixedGWO, was also developed by implementing the exploration rate with higher power settings [65]. Evaluated using diverse continuous, discrete or mixed optimization tasks and real-world problems of simultaneous denoising and unmixing of multispectral images, the developed models showed enhanced performances as compared with those of the baseline methods.

3. THE PROPOSED PSO MODEL

In this research, we propose an ensemble CBiLSTM network with optimal hyper-parameter identification for undertaking audio classification. A new PSO algorithm is proposed to optimize the learning rate, weight decay, number of filters and hidden neurons in the convolutional and BiLSTM layers, respectively. The proposed PSO algorithm incorporates the secant and Newton-Raphson methods, crossover operators based on regular and irregular adaptive super-ellipse formulae, and spherical search coefficients to diversify the search process. Algorithm 1 illustrates the pseudo-code of the proposed algorithm.

Algorithm 1: Pseudo-Code of the Proposed PSO Algorithm

```
Start
Randomly initialize a particle swarm;
Conduct fitness evaluation of the population;
```

```
4.  Rank the particles based on the fitness scores and identify $gBest$ and the second-best solution;
5.  While (Stopping criterion is not satisfied)
6.  {
7.        Randomly select one of the following two operations for swarm leader enhancement;
8.        {//1. Enhance $gBest$ using the secant method;
9.          Assign $gBest$ and the second-best solution as the initial seeds for the secant method;
10.         Generate an offspring solution using Equation (3);
11.         Update $gBest$ if the new solution is fitter;
12.         Update the two seed solutions and repeat lines 10-11 until the termination criterion is met;
13.         //2. Enhance $gBest$ using the Newton's method;
14.         Assign $gBest$ as the initial seed for the Newton's method;
15.         Generate an offspring solution using Equation (4);
16.         Update $gBest$ if the new solution is fitter;
17.         Update the seed solution and repeat lines 15-16 until the termination criterion is met;
18.       }
19.       Sort the swarm and identify the second swarm leader;
20.       Produce the hybrid leader 1 as defined in Equations (5)-(8) using $gBest$ and the second swarm leader;
21.       Produce the hybrid leader 2 as defined in Equation (5) & (11)-(13) using $gBest$ and the second swarm leader;
22.       Produce the hybrid leader 3 as defined in Equation (5) & (14)-(16) using $gBest$ and the second swarm leader;
23.       Randomly select one of the above hybrid leaders for the following operation;
24.       For each individual in the overall swarm do {
25.            Generate the spherical search coefficient $\partial_1$ using Equations (18)-(22);
26.            Generate the spherical search coefficient $\partial_2$ using Equations (23)-(27);
27.            Conduct position updating as defined in Equations (1) and (17) using one of the randomly selected newly yielded hybrid leaders;
28.            Update $pBest$ and $gBest$;
29.       } End For
30.       Sort the overall swarm based on the fitness values and update $gBest$;
31. } Until (Stagnation)
32. Output $gBest$;
33. End
```

Referring to Algorithm 1, we first initialize a swarm randomly and rank the particles based on their fitness scores to identify $gbest$. We adopt either the secant algorithm or the Newton's method to further enhance the swarm leader. Specifically, we randomly select one of these numerical analysis methods for generating the offspring solutions of the swarm leader to increase search intensification. We assign $gbest$ as the initial seed for the Newton's method. Both $gbest$ and the second swarm leader are used as the seeds for the secant method for offspring solution generation. Both methods perform linear approximation to the global optima of the hyper-parameter optimization problems, and terminate when the pre-defined criterion (e.g. a precision score) is reached. Each of the new offspring solutions generated by either the Newton's or the secant methods is used to replace $gbest$ if it is fitter.

After obtaining the improved $gbest$ solution using one of the numerical analysis methods, the second swarm leader is identified. We generate enhanced offspring leaders using several distinctive crossover operators with $gbest$ and the second swarm leader as the parent chromosomes. To be specific, we produce three hybrid leader signals using crossover factors defined by three sets of adaptive regular and irregular elliptical functions. Such adaptive formulae assign larger weights to the second-best leader and smaller ones to $gbest$ at the beginning stage of the search process, in order to increase diversification. As the search progresses, smaller weights are produced for the second leader, while larger ones go to $gbest$ for increasing intensification. We then randomly select one of the resulting hybrid leaders for position updating. To equip the search process with different search scales and trajectories, two distinctive spherical outline surfaces are used to produce cognitive and social search coefficients, respectively. The proposed model incorporates a versatile search process with diverse hybrid leader indicators and distinctive elliptical search coefficients to overcome stagnation. The overall algorithm iterates until the termination criterion (i.e. the maximum number of function evaluations) is reached. We introduce each key search mechanism comprehensively in the following sub-sections.

3.1 Swarm Leader Enhancement Using the Secant Method

After initializing the swarm, we identify the swarm leader based on the ranking of the fitness scores. In order to increase local exploitation and avoid stagnation, we apply two root-finding algorithms, i.e. the secant and Newton-Raphson methods, to enhance the global best solution. In comparison with other stochastic processes such as Gaussian, Cauchy and Levy distributions based on random walks, these root-finding numerical analysis algorithms provide guided local exploitation of promising solutions during the course of finding global optimality.

We first introduce the secant method [66, 67] for swarm leader enhancement. It employs a succession of roots of secant lines to estimate the root of a specific function f. Equation (3) defines the operation.

$$x_n = \frac{x_{n-2}f(x_{n-1}) - x_{n-1}f(x_{n-2})}{f(x_{n-1}) - f(x_{n-2})} \tag{3}$$

where x_{n-1} and x_{n-2} represent two initial seed values which are assumed to be sufficiently close to the root, while x_n denotes the yielded better approximation of the root. Firstly a line is constructed using the two points, i.e. $(x_{n-2}, f(x_{n-2}))$ and $(x_{n-1}, f(x_{n-1}))$. The new approximation of the root, x_n, is calculated as the intersection of the *x*-axis and this newly formed line. We then use x_{n-1} and x_n as the subsequent seed values for the calculation of another new estimated root.

In this research, we assign the global best solution and the second swarm leader as the initial seed values, i.e. x_{n-1} and x_{n-2}, and use their fitness scores as $f(x_{n-1})$ and $f(x_{n-2})$, respectively. The newly generated offspring solution x_n is evaluated and used to replace $gbest$ if it is fitter. This process iterates until the maximum number of trials or the sufficient level of precision (i.e. a small difference between x_{n-1} and x_n) is reached.

3.2 Swarm Leader Enhancement Using the Newton-Raphson Method

Besides the secant method, the Newton-Raphson method (referred as Newton's method) [66, 67] is also employed to improve the $gbest$ solution. It employs one initial seed value to yield successive better approximations to the root of a specific function, assuming that the initial seed is sufficiently competent (i.e. close to the root). Equation (4) illustrates the operation.

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} \tag{4}$$

where x_{n-1} is the initial seed value for the root of function f, while x_n represents a better estimation of the root with f' referring to the derivative of f. As defined in Equation (4), the algorithm calculates the improved guess x_n as the intersection of the *x*-axis and the tangent of the graph of f at $(x_{n-1}, f(x_{n-1}))$. In this research, we assign x_{n-1} and $f(x_{n-1})$ as the global best solution and its fitness score, respectively. The new solution x_n is subsequently evaluated and used to replace the global best solution if it is fitter. The process iterates until the maximum number of trials or a sufficient precision score is reached.

Overall, the secant and the Newton's methods increase exploitation to overcome stagnation. We employ the improved $gbest$ solution produced by either of these numerical analysis methods for hybrid leader generation.

3.3 Hybrid Leader Generation Mechanisms

Motivated by GWO where three wolf leaders are used to guide the search process [22-24], in this research, we propose hybrid leader generation based on multiple leaders to overcome stagnation. Specifically, after determining the new $gbest$ solution using the secant or the Newton's algorithm, we sort the swarm and extract the second-best leader. Instead of using purely the $gbest$ solution to guide the search process, we propose an adaptive hybrid leader generation mechanism based on two swarm leaders, as defined in Equation (5), to avoid stagnation.

$$h_d^{t+1} = \alpha \times gbest_d^t + \beta \times sbest_d^t \tag{5}$$

where α and β represent the incremental and decremental coefficients for the swarm leader $gbest$ and the second-best solution $sbest$, respectively, with h denoting the generated hybrid leader. Owing to the adoption of increasing (α) and decreasing (β) coefficients with respect to generating the hybrid leader h, the impact of the second leader is comparatively more dominating at the beginning of the search process. As the search process iterates, the influence of the second leader is gradually reduced, while the guidance of the global best solution gains increasing effects.

In comparison with single leader-based search processes [11-20], hybrid leaders incorporating multiple elite signals are used in the proposed scheme to overcome stagnation. Moreover, in comparison with GWO [22-24] and a Bare-Bones PSO (BBPSO) variant [21] where multiple leaders with equal weightings are used to guide the search process, we adopt adaptive weighting factors yielded by super-ellipse formulae for hybrid signal generation to better balance between diversification and intensification. Therefore, the proposed hybrid leader generation mechanism offers enhanced capabilities in overcoming stagnation by emphasizing local and global optimal signals adaptively while following the optimal solutions simultaneously.

We elaborate the generation of adaptive parameters α and β as follows. In particular, to increase search diversity, three sets of regular and irregular elliptical 2D nonlinear formulae are used to produce adaptive coefficients α and β.

3.3.1 The first hybrid leader generation strategy

For the first crossover coefficient generation mechanism, a regular 2D elliptical curve [68] defined in Equations (6)-(8) for adaptive crossover factor generation is proposed, i.e.,

$$\varphi(\sigma) = |\, cos\left(\frac{\sigma}{2}\right)| + |\, sin\left(\frac{\sigma}{2}\right)| \qquad \sigma = [0{:}\,0.001{:}\,\pi] \tag{6}$$

$$x = \varphi \times cos(\sigma) \tag{7}$$

$$y = \varphi \times sin(\sigma) \tag{8}$$

where φ and σ denote the radius and angle, and x and y represent the coordinates of the generated points in x and y axes, respectively. Figure 2 illustrates the yielded elliptical curve using the above equations. An increasing elliptical sub-graph (denoted as the blue line) is generated when σ is assigned with values from $\pi/2$ to π with a step size of 0.001. On the other hand, a decreasing sub-graph (denoted as the orange line) is produced when σ is assigned with values from 0 to $\pi/2$ with a step size of 0.001. Each of the sub-curves consists of 1,571 unique points. These sub-curves are used to generate adaptive coefficients α and β, as defined in Equations (9) and (10).

$$\alpha = y[1{,}571 \times \frac{t}{max_{iterations}}] \qquad \sigma = [\pi/2{:}\,0.001{:}\,\pi] \qquad \text{(i.e. the blue line in Figure 2)} \tag{9}$$

$$\beta = y[1{,}571 \times \frac{t}{max_{iterations}}] \qquad \sigma = [0{:}\,0.001{:}\,\pi/2] \qquad \text{(i.e. the orange line in Figure 2)} \tag{10}$$

where t and $max_{iterations}$ indicate the current and maximum numbers of iterations, respectively.

As defined in Equations (9)-(10), we select $max_{iterations}$ number of values in an increasing order from the y-axis in the blue sub-graph with a step of $1{,}571/max_{iterations}$. Similarly, we select $max_{iterations}$ number of values in a decreasing order from the y-axis in the orange sub-graph with a step of $1{,}571/max_{iterations}$. Then we assign these two sets of increasing and decreasing values to α and β, respectively. In other words, as the iteration number t increases, both adaptive coefficients α and β are generated by assigning the increasing and decreasing values from the y-axis in the blue and orange elliptical sub-curves, respectively, with a step of $1{,}571/max_{iterations}$.

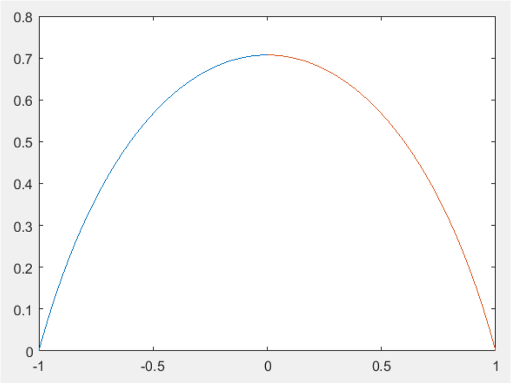


Figure 2 The elliptical curve generated using Equations (6)-(8) where the increasing (blue line) and decreasing (orange line) sub-graphs are used for the generation of adaptive coefficients α and β, respectively

Such increasing and decreasing crossover factors, α and β, are exploited to generate hybrid leaders, as defined in Equation (5). The resulting hybrid leaders in the early iterations are thus able to focus on global exploration by assigning larger values to β and smaller values to α. Then, local exploitation is gradually emphasized by assigning smaller values to β and larger values to α, towards the final iterations. Such a fused evolving leader signal is able to achieve a better balance between exploitation and exploration to overcome stagnation.

3.3.2 The second hybrid leader generation strategy

In comparison with the aforementioned regular elliptical function, a formula representing an irregular curve [68] is proposed as the second strategy for adaptive coefficient generation, as defined in Equations (11)-(13).

$$\omega(\sigma) = (|\frac{cos(22\times\sigma)}{300}| + \left|\frac{sin(16\times\sigma)}{300}\right|)^{\frac{1}{20}} \qquad \sigma = [0: 0.001: \pi] \tag{11}$$

$$x = \omega \times cos(\sigma) \tag{12}$$

$$y = \omega \times sin(\sigma) \tag{13}$$

where ω and σ denote the radius and angle, respectively. Figure 3 illustrates the elliptical graph produced using this new set of equations. The increasing (denoted by the blue line) and decreasing (denoted by the orange line) sub-graphs are generated with $\sigma = [\pi/2: 0.001: \pi]$ and $\sigma = [0: 0.001: \pi/2]$ respectively. There are 1,571 points in each sub-graph. We subsequently generate adaptive coefficients by assigning two sets of increasing and decreasing values extracted from the y-axis from the blue and orange sub-curves, respectively, with a step size of $1{,}571/max_{iterations}$. Therefore, two sets of $max_{iterations}$ number of values in increasing and decreasing orders are assigned to α and β, respectively. Besides following the general increasing and decreasing trends, the sub-curves yielded by the above functions embed many micro irregularities, which facilitate the generation of more distinctive hybrid leaders, in comparison with those produced using the aforementioned regular formulae, to increase search diversity.

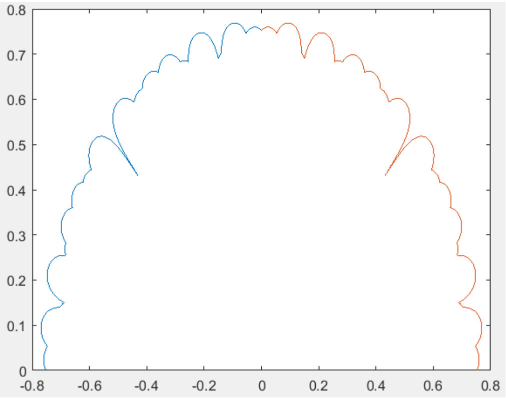


Figure 3 The elliptical curve generated using Equations (11)-(13) where the increasing (blue line) and decreasing (orange line) sub-graphs are used for the generation of adaptive coefficients α and β, respectively

3.3.3 The third hybrid leader generation strategy

Instead of using the formulae in Sections 3.3.1 and 3.3.2 separately, a combined scheme is proposed by applying the regular and irregular functions jointly for adaptive coefficient generation. Equation (14) defines this hybrid strategy for crossover factor generation.

$$\gamma(\sigma) = \begin{cases} \left|cos\left(\frac{\sigma}{2}\right)\right| + \left|sin\left(\frac{\sigma}{2}\right)\right|, & rand > 0.5 \\ (|\frac{cos(22\times\sigma)}{300}| + \left|\frac{sin(16\times\sigma)}{300}\right|)^{\frac{1}{20}}, & otherwise \end{cases} \tag{14}$$

$$x = \gamma \times cos(\sigma) \tag{15}$$

$$y = \gamma \times sin(\sigma) \tag{16}$$

where γ and $\sigma = [0{:}\,0.001{:}\,\pi]$ denote the radius and angle, respectively. In Equation (14), the upper formula is the same as that introduced in Equation (6). On top of it, we embed the lower formula defined in Equation (11) to diversify the production of weight factors. Figure 4 illustrates the resulting super-ellipse curves defined in Equation (14). The regular graph (the same as that illustrated in Figure 2) is produced using the first function, while the irregular contour (the same as that shown in Figure 3) is generated using the second formula. In each iteration, these regular and irregular formulae are used alternatively to generate the adaptive crossover factors α and β. Specifically, if a randomly assigned value is more than a threshold of 0.5, the first (regular) function is used for adaptive factor generation, otherwise the second (irregular) formula is applied to produce the adaptive coefficients.

As introduced previously, each of these regular and irregular ellipse graphs defined in Equation (14) consists of both increasing and decreasing sub-graphs, when $\sigma = [\pi/2{:}\,0.001{:}\,\pi\,]$ and $\sigma = [0{:}\,0.001{:}\,\pi/2\,]$, respectively. Therefore, by taking both formulae into account, there are two alternative sub-graphs with different shapes that can be used for generating each adaptive coefficient (i.e. the selection among $2 \times max_{iterations}$ number of increasing or decreasing values for generating α or β). In short, this new joint crossover generation mechanism is able to produce more diversified hybrid elite leaders, in comparison with those yielded using purely either the regular or irregular function, to overcome stagnation.

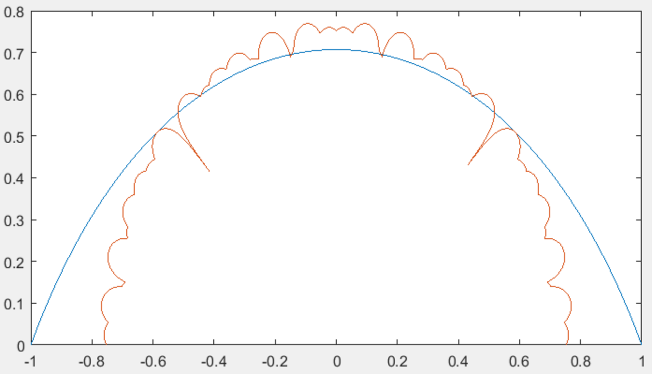


Figure 4 The two curves generated using Equation (14) where the regular graph (the same as that shown in Figure 2) is produced using the upper formula, while the irregular one (the same as that illustrated in Figure 3) is yielded using the lower formula

The aforementioned three leader generation operations using regular, irregular and combined formulae discussed in Sections 3.3.1-3.3.3 are randomly selected during the search process for producing diverse distinctive hybrid leaders. The aim is to avoid the search process from being trapped in local optima and mitigate premature convergence.

3.4 The Proposed PSO Operation

We now explain the proposed search operation for position updating. Equation (17) defines the new PSO action, where the generated hybrid leader is used to guide the search process.

$$v_{id}^{t+1} = w \times v_{id}^{t} + \partial_1 \times r_1 \times (h_d^t - x_{id}^t) + \partial_2 \times r_2 \times (p_{id}^t - x_{id}^t) \qquad (17)$$

where h_d^t represents the hybrid swarm leader produced using Equation (5), while ∂_1 and ∂_2 denote the dynamic search steps oriented from two spherical surfaces.

To increase search diversity, two distinctive geometrical surfaces are proposed for generating both social and cognitive coefficients, ∂_1 and ∂_2. We define the first spherical parametric surface in Equations (18)-(22) [68], with Figure 5 illustrating the yielded graph.

$$q(\vartheta) = (|cos(\tfrac{\vartheta}{2})|^{15} + |sin(\tfrac{\vartheta}{2})|^{32})^{-\frac{1}{10}} \quad \vartheta = [-\pi{:}\,0.05{:}\,\pi] \qquad (18)$$

$$\mu(\delta) = (|cos(\tfrac{\delta}{2})|^{15} + |sin(\tfrac{\delta}{2})|^{32})^{-\frac{1}{10}} \quad \delta = [-\pi/2{:}\,0.05{:}\,\pi/2] \qquad (19)$$

$$x = 1.5 \times q \times cos(\vartheta) \times \mu \times cos(\delta) \qquad (20)$$

$$y = 1.5 \times q \times sin(\vartheta) \times \mu \times cos(\delta) \qquad (21)$$

$$z = 1.5 \times \mu \times sin(\delta) \qquad (22)$$

where ϑ and δ represent the longitude and latitude with value within $[-\pi, \pi]$ and $[-\pi/2, \pi/2]$, while x, y and z denote the 3D coordinates of the generated points in x, y and z axes, respectively.

During each iteration of the search process, for each particle, a three-dimensional point is randomly selected from the generated spherical surface illustrated in Figure 5. The largest value among the absolute scores of x, y and z coordinates of this randomly selected point is selected as the search parameter ∂_1 in Equation (17). As shown in Figure 5, the generated spherical surface has a wide coverage of the three-dimensional space. Therefore the proposed search operation provides each particle with distinctive social search behaviours and large search steps to follow the hybrid elite signal.

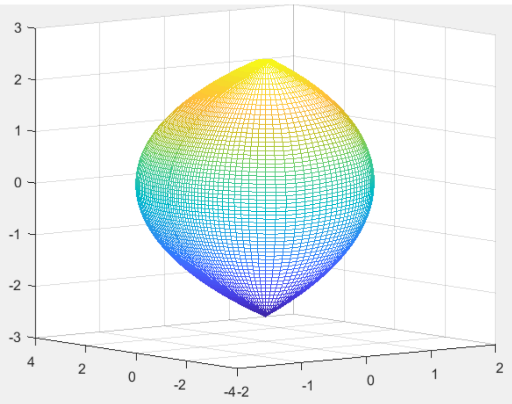


Figure 5 The spherical surface produced using Equations (18)-(22) for the production of the search coefficient ∂_1 in Equation (17)

Another geometrical curve is also proposed for generating the search coefficient ∂_2 in the modified PSO operation. Equations (23)-(27) define this new spherical parametric shape [68], with Figure 6 representing the generated contour. Similarly, in each iteration, for each particle, the maximum absolute value of the three coordinates of a randomly selected point is assigned as the search coefficient ∂_2 in Equation (17).

$$l(\vartheta) = (|cos(\vartheta)|^{\frac{1}{2}} + |sin(\vartheta)|^4)^{-2} \quad \vartheta = [-\pi{:}\,0.05{:}\,\pi] \tag{23}$$

$$k(\delta) = (|cos(\delta)|^{\frac{1}{2}} + |sin(\delta)|^4)^{-2} \quad \delta = [-\pi/2{:}\,0.05{:}\,\pi/2] \tag{24}$$

$$x = 3 \times l \times cos(\vartheta) \times k \times cos(\delta) \tag{25}$$

$$y = 3 \times l \times sin(\vartheta) \times k \times cos(\delta) \tag{26}$$

$$z = 3 \times k \times sin(\delta) \tag{27}$$

Owing to the generation of a distinctive spherical shape in comparison with that defined in Equations (18)-(22), different search behaviours are defined in the cognitive search operation. In comparison with the geometrical surface shown in Figure 5 with a comparatively wider coverage of the three-dimensional space, the newly yielded spherical contour in Figure 6 occupies principally the middle regions for coefficient generation. Therefore, the proposed search action enables each particle to follow the personal best solutions with comparatively smaller but more persistent search steps.

Overall, the two proposed spherical surfaces shown in Figures 5 and 6 equip each particle with distinctive search scales, directions and trajectories in the social and cognitive components, in order to increase the chances of finding global optimality.

In short, the proposed local and global search mechanisms, i.e. the secant and Newton's numerical analysis algorithms, adaptive crossover operators with weighting probabilities extracted from regular and irregular elliptical functions, and parametric spherical search coefficients, allow the search process to have a better trade-off between intensification and diversification to overcome premature convergence. We use the proposed model to optimize the hyper-parameters of the CBiLSTM network for the identification of heart and lung anomalies as well as environmental events using audio clips. We discuss hyper-parameter optimization in deep networks in detail in the next section.

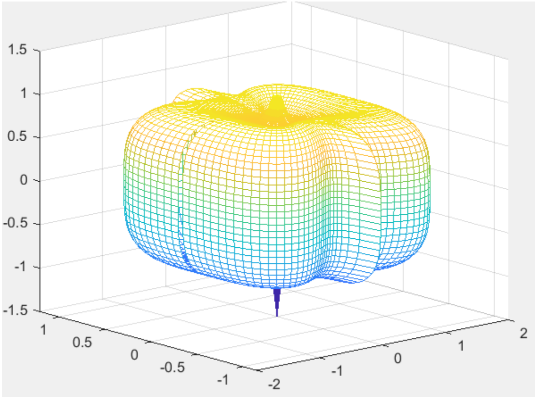


Figure 6 The spherical surface yielded using Equations (23)-(27) for the production of the search coefficient ∂_2 in Equation (17)

4. GENERATION OF EVOLVING DEEP CONVOLUTIONAL BILSTM NETWORKS

As a combination of multiple spectrums, spectrogram is used to describe changes in amplitude and frequency over a series of discrete times. In the pre-processing stage, we employ STFT for spectrogram generation, as defined in Equation (28).

$$\mathrm{STFT}\{s(t)\}(\tau, m) = X(\tau, m) = \int_{-\infty}^{\infty} s(t) w_f(t - \tau) e^{-jmt}\, dt \tag{28}$$

where $s(t)$ denotes the audio signal, and $w_f(\tau)$ represents the window function. Parameter m refers to the frequency, while τ indicates the time index which is considered as 'slow' time with t representing the high resolution time. In addition, $X(\tau, m)$ is the Fourier transform of $s(t)w_f(t - \tau)$, a function representing the phase and magnitude of the signal over time and frequency. The squared magnitude of the STFT produces the spectrogram representation of the function, as defined in Equation (29).

$$spectrogram\{s(t)\}(\tau, m) = |X(\tau, m)|^2 \tag{29}$$

The resulting spectrogram is used as the inputs to the CBiLSTM network. Existing studies [38, 39, 69] indicate that deep networks (e.g. AlexNet and GoogLeNet) based on such spectrogram inputs are able to outperform models that use 1D signal inputs.

The CBiLSTM network used for audio classification consists of three convolutional blocks, two BiLSTM layers, and one dense block. Each convolutional block contains one 2D convolutional layer, which is succeeded by a batch normalization layer (i.e. a ReLU layer), a max pooling layer and a dropout layer. A kernel size of 3×3 is used in the convolutional layers. As discussed earlier, we employ a 2D representation of audio signals, i.e. STFT spectrograms, as the network input. Such 2D spectrograms can be treated as images. Therefore, as recommended by existing studies [38, 39, 69], the 2D convolution is adopted in this research for feature extraction, where the kernel in the convolutional layer slides over the 2D spectrogram input data, and performs an elementwise multiplication for feature extraction.

In addition, two BiLSTM layers are subsequently attached after the convolutional layers. Each BiLSTM layer is composed of two hidden LSTM layers of opposite directions. It learns bidirectional long-term dependencies between time steps in time series data. It employs the information from past (backwards) and future (forward) states simultaneously to inform the output layer. Specifically, in comparison with unidirectional LSTM where the future input information is not reachable from the current state, the BiLSTM layer adopts inputs from both forward and backwards directions to increase the contextual information available to the network. Existing studies also indicate the superior performance of BiLSTM networks over unidirectional LSTM models in time series forecasting [70, 71]. In this research, the BiLSTM layers embedded in the proposed network are used to extract temporal sequential cues from the spectrogram features yielded by the convolutional layers. Finally, an additional dense block is appended to the network which embeds a dropout layer, a batch normalization layer, and a linear layer. Figure 7 illustrates the detailed network architecture. We also provide the detailed architecture and parameter settings of the proposed CBiLSTM network as follows.

- Convolutional Block 1 – The first convolutional block is embedded with a 2D convolutional layer with a kernel size of 3×3 and a stride of 1. To effectively extract discriminative spatial features, the number of filters in this convolutional layer is optimized by the proposed PSO algorithm. A batch normalization layer (i.e. a ReLU layer) is subsequently attached to accelerate network training as well as reduce network sensitivity to initialization. This is followed by a max pooling layer with a kernel size of 3×3 and a stride of 3 to reduce the dimensions of feature maps. A dropout layer with a dropout probability of 0.1 is also used to avoid overfitting.
- Convolutional Blocks 2 & 3 – We set the numbers of filters in the second and third convolutional layers in these two subsequent stacked blocks to twice the size of the optimized number of filters in the first convolutional layer. A kernel size of 3×3 and a stride of 1 are used in these convolutional layers. A ReLU layer is subsequently attached after each convolutional layer to speed up the training process. The extracted spatial features are then down-sampled by a max pooling layer with a kernel size of 4×4 and a stride of 4, while a dropout layer with a dropout probability of 0.1 is appended in each block.
- BiLSTM Layers – Two BiLSTM layers are inserted to the network. Each BiLSTM layer is used to capture bidirectional temporal dependencies from the spatial spectrogram features by learning from both forward and backward states simultaneously at each time step. We optimize the number of hidden neurons in each BiLSTM layer with the proposed PSO algorithm. The optimized number of hidden units leads to an effective preservation of bidirectional discriminative sequential patterns to avoid overfitting and underfitting issues.
- A Dense Block – The final dense block comprises a dropout layer, a batch normalization layer and a linear layer. Specifically, it overcomes overfitting by using the dropout layer with a dropout probability of 0.3. The linear layer is assigned with the number of neurons equivalent to the number of target classes, which is used to produce the final network output.

In order to extract effective audio representations, as mentioned earlier, the proposed PSO algorithm is used to optimize the learning hyper-parameters, i.e. the learning rate and weight decay, as well as the network layer topologies, i.e. the numbers of filters in the three convolutional layers and the numbers of hidden neurons in the two BiLSTM layers.

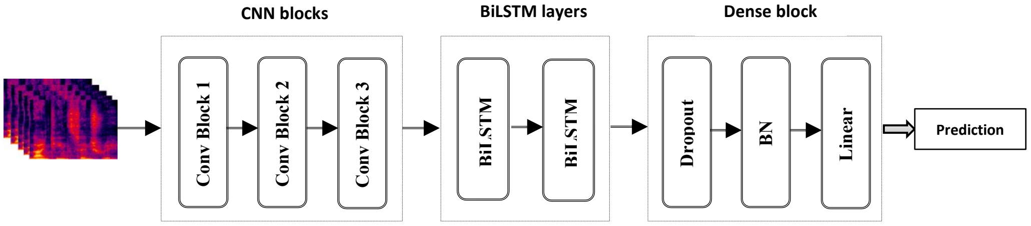


Figure 7 The proposed CBiLSTM network

The optimized learning and model settings play significant roles in affecting the network performance. Specifically, the numbers of filters in the convolutional layers determine spatial feature learning capabilities of the network. Filters act as feature detectors. Each filter generates a feature map, which learns the explanatory factors within the image. A sufficient number of filters offers the network competitive capabilities in capturing complex spatial details. In contrast, an inadequate number of filters discards important spatial patterns, while an exaggerated setting memorizes the exact details of input images. On the other hand, the numbers of hidden neurons in the BiLSTM layers define the network capabilities in extracting temporal dynamics from spatial features. A significantly large number of hidden neurons is likely to extract redundant temporal relationships. The BiLSTM layers with considerably small neuron settings result in constrained efficiency in discovering manifold sequential patterns. Moreover, the inclusion of a regularization term in the loss function is an effective way to reduce overfitting. The weight decay (i.e. the regularization factor) is important for determining the effects of this additional regularization term in the loss function, thus affecting convergence speed and model diversity. In addition, the learning rate setting determines the distinctive network learning behaviours. Overall, the combination of these learning parameters and the convolutional and BiLSTM layer settings greatly affects model performance. Therefore, we employ the proposed PSO algorithm to optimize these key elements for audio classification. Table 1 shows the detailed optimized hyper-parameters. Specifically, we optimize four hyper-parameters, i.e. the learning rate, weight decay, the number of filters in the first convolutional layer and the number of hidden neurons in each BiLSTM layer. As mentioned above, we generate the numbers of filters in the second and third convolutional layers by multiplying the optimized number of filters in the first convolutional layer by 2.

The optimal hyper-parameter selection is conducted as follows. A particle swarm is randomly initialized with each dimension of the particle positions in the range of [0, 1]. Each dimension of the particle represents one optimized factor. The proposed search actions, such as secant and Newton's methods for leader enhancement and the position updating operation with adaptive hybrid signals and spherical coefficients, are used to guide the search process. Each particle moves around in the search space with each dimension assigned with a continuous value. During fitness evaluation, each particle position is decoded into a set of network and learning configurations, which are used to construct a CBiLSTM network. This resulting network is trained with the training set, while the performance of the validation set is used as the fitness score for each particle.

Specifically, the Adam optimizer is used for training each devised CBiLSTM network. The negative log likelihood loss (provided by the PyTorch library) is adopted as the loss function. A batch size of 8 is employed. The network shuffles the training and validation data sets before each training and validation cycle. In addition, the learning rate is updated every 10 epochs by multiplying with a drop factor of 0.5. A smaller number of epochs of 5 is used for the hyper-parameter search during training. The most optimal settings encoded in the global best solution are employed to construct the output network, which is trained with the combined training and validation data sets and a comparatively larger number of epochs, i.e., 100. The resulting model is evaluated with the unseen test set for performance comparison.

In order to avoid overfitting, we conduct two types of data augmentation. Firstly, audio data augmentation is conducted during the training stage. We apply pitch and time shifting as well as noise addition for audio augmentation. As an example, we employ the pitch shifting range of [-2, 2] semitones, time shifting range of [-0.3, 0.3] seconds, and noise addition SNR range of [-20, 40] dB for generating new signals. Such augmentation setting provides a reasonable trade-off between performance and computational cost. Secondly, spectrogram augmentation at the image level, i.e., random flipping and image translation and rotation, is also conducted, to avoid overfitting.

According to Esmaeilpour et al. [38] and Zhang et al. [39], the above audio and image augmentation operations mainly focus on low-level transformation. As discussed earlier, the audio signals (such as respiratory, heart and environmental sounds) used in this research are unstructured with unpredictable factors in comparison with other speech or music signals. As an example, the audio signals sometimes consist of transient (e.g. person sneeze) and intermittent (e.g. baby cry and dog bark) data, resulting in discriminative salient features to be distributed in a few frames only. As such, when the signals and converted spectrograms do not embed many active areas or properties (e.g. a large number of silent periods or noisy frames) or contain complex temporal characteristics, the linear nature of low-level augmentation techniques may not be sufficient enough to cause a high impact on the classifier decision boundaries, despite their efficiency in tackling other image classification tasks using normal images. This observation is indicated in our experimental study and existing literature [34, 72]. Therefore, motivated by existing studies such as Wu et al. [29], besides data augmentation procedures, we employ data duplication to further strengthen the signals from minority classes. Detailed data duplication and augmentation processes are discussed in the evaluation section.

Table 1 The selected optimized hyper-parameters

Optimized hyper-parameters	**Search ranges**
Learning rate	[5e-4, 6e-3]
Weight decay	[0.01, 0.05]
No. of filters in the 1st convolutional layer	[24, 64]
No. of filters in the 2nd and 3rd convolutional layers	[48, 128] (multiplying the optimized number of filters in the 1st convolutional layer by 2)
No. of hidden units in each BiLSTM layer	[48, 128]

After obtaining a set of optimized CBiLSTM networks with different model and learning settings, we embed these optimized networks in a number of ensemble models. The majority voting mechanism is used for combining the outputs. Owing to the identified distinctive topologies, the optimized CBiLSTM networks illustrate superior discriminative and complementary characteristics to enhance the ensemble model performance.

Besides the optimized networks, a default CBiLSTM model with manually assigned configurations is employed for performance comparison. The default network uses 64 filters in the first convolutional layer, and 128 in the subsequent two convolutional layers, with 128 numbers of hidden units in each of the BiLSTM layers. We also use the default learning rate and weight decay of 0.002 and 0.01, respectively.

Several additional baseline deep learning models are employed for performance comparison, i.e. CNN, GoogLeNet, and ResNet18. These models produce competitive performances in diverse audio and image classification tasks. In our experiments, a variety of training and feature learning strategies are embedded into these baseline networks. Firstly, a CNN with a VGGNet architecture is used as the baseline method. It consists of 8 convolutional layers, with each convolutional layer succeeded by a batch normalization layer, a ReLU layer, an optional max or average pooling layer. We denote this CNN model as VGG8. This VGG8 network is trained from scratch in our experiments.

In addition, ImageNet pre-trained GoogLeNet and ResNet18 models are used for performance comparison. Specifically, we apply transfer learning to GoogLeNet for fine-tuning the network to tackle audio classification. In contrast, we extract deep spatial features directly from the global pooling layer (i.e. pool5) of ResNet18 without any training process, in order to take advantage of the representational power of the pre-trained networks. Such high-level features extracted from the deep layers in ResNet18 embed sufficient descriptive characteristics of the spectrograms. They are subsequently used to train a multiclass SVM classifier for audio classification. This baseline model is denoted as ResNet18+SVM in the result presentation. We elaborate the experimental details in the next section.

5. EXPERIMENTAL STUDIES

Three audio data sets, i.e. ICBHI 2017 [25], PhysioNet Challenge 2016 [26], and ESC-10 [27], are used in our experimental studies. The ICBHI 2017 data set contains a total of 920 audio clips with 6,898 respiratory cycles from 126 subjects. In addition, each recording contains several respiratory cycles, and has a length between 10 and 90 seconds. There are two types of annotations provided in ICBHI, i.e. (1) with respect to each respiratory cycle, whether or not crackles and/or wheezes are present, and (2) with respect to each subject, whether or not a particular disease condition is present. In this research, we focus on the latter, i.e. the identification of a particular disease condition for each subject using the entire audio signal, owing to the fact that it is a comparatively more challenging task. Moreover, in ICBHI, the audio clips are annotated with the following labels, i.e. healthy, asthma, Chronic Obstructive Pulmonary Disease (COPD), Lower Respiratory Tract Infection (LRTI), Upper Respiratory Tract Infection (URTI), bronchiectasis, pneumonia, and bronchiolitis. Owing to a severe imbalanced ratio of data samples in each class, we follow the existing studies [28, 34] to categorize all non-healthy cases into two groups, i.e. COPD, bronchiectasis, and asthma as chronic cases and URTI, LRTI, pneumonia, and bronchiolitis as non-chronic cases. Therefore we conduct 3-class classification at the pathology level to detect healthy, chronic and non-chronic conditions using this respiratory sound data set. In addition, owing to the collection of respiratory sound recordings using various devices at different chest locations as well as environmental noise, this respiratory sound data set presents very challenging characteristics for sound classification tasks.

A heart sound data set, i.e. PhysioNet/CinC Challenge 2016, is also employed to test model efficiency. It consists of five databases with a total of 3,240 heart sound clips, collected from clinical and non-clinical environments. There are 2,575 normal (denoted as -1) and 665 abnormal (denoted as 1) instances in the data set. They are contributed by both healthy subjects and pathological patients, with each subject donating 1-6 recordings. Each heart sound clip is resampled to 2,000 Hz, and has a length between 5 and 120 seconds. This data set is the largest public collection of phonocardiogram (PCG) recordings. Owing to the inclusion of samples recorded in a variety of uncontrolled environments, the data set embeds various challenging factors for sound classification.

A third data set containing environmental sound data samples is also used to evaluate the optimized CBiLSTM networks, i.e. ESC-10. It contains 400 environmental recordings from 10 classes (i.e. sea waves, baby cry, dog bark, rain, person sneeze, clock tick, chainsaw, rooster, helicopter, and fire crackling), with each class containing 40 samples. Each sound clip has a length of 5 seconds with 44.1 kHz sampling frequency. Owing to the overlap of different sound sources and various natural environmental noise, the data set poses great challenges to many sound classification methods.

Several baseline search methods are employed for performance comparison, i.e. the original PSO algorithm, a modified PSO (MPSO) [14] with an adaptive inertia weight factor and linear acceleration coefficients, and a PSO variant (PSOVA) [73] with dynamic coefficients and DE-based leader enhancement. Motivated by

CPSO [74], a PSO model with sine search coefficients (SPSO) is also implemented for performance comparison. We follow the parameter settings of these baseline methods from their original studies. We adopt adaptive 2D and 3D super-ellipse formulae for adaptive crossover factor and search coefficient generation, respectively, for the proposed PSO algorithm. Table 2 shows the detailed parameter settings of each search method.

Besides the above classical and advanced search methods, several deep learning methods are also employed for comparison, i.e. GoogLeNet (transfer learning), VGG8 (training from scratch), and ResNet18-based spatial feature extraction with SVM-based classification.

To evaluate the efficiency of the proposed PSO algorithm, the following optimization tasks are experimented, i.e. (1) optimal hyper-parameter selection in CBiLSTM networks, and (2) numerical optimization using unimodal and multimodal benchmark functions. The first optimization task combines both deep learning and evolutionary computation domains with the attempt to generate optimal deep networks for sound classification. The second optimization task is dedicated to solving diverse challenging artificial landscape functions to test model convergence speed and capabilities in attaining global optimality for high-dimensional optimization problems. In this experimental study, owing to complexities of the problem domains, different experimental settings have been employed.

In the first optimization task, i.e., optimal hyper-parameter selection, since there are four key hyper-parameters (i.e., the learning rate, weight decay, the number of filters in the first convolutional layer, and the number of hidden units in each BiLSTM layer) to be optimized, we set dimension=4. To ensure a fair comparison, all the search methods terminate when the maximum number of function evaluations, i.e., 15 (population) × 20 (the maximum number of iterations), is reached. We perform 30 trials for each search method in each experiment. The resulting 30 optimized networks are used to construct 10 ensemble models for performance comparison, where each ensemble classifier incorporates three optimized networks using a majority voting scheme.

To further assess model efficiency in solving high-dimensional optimization problems, in the second optimization task, two sets of benchmark functions are employed, i.e. (i) a set of 11 basic benchmark functions, and (ii) a suite of 13 artificial landscapes widely adopted in the literature [22, 65, 75-77]. We set dimension=30 and trial=30 in the experiments. Comparatively larger numbers of function evaluations are adopted owing to variations and complexities of the test functions, i.e., maximum number of function evaluations=population (30) × iterations (2000) and maximum number of function evaluations=population (30) × iterations (3000) for the numerical problems, respectively. Each search method terminates when these maximum numbers of function evaluations are reached. We discuss the details of the experimental studies in the following sub-sections.

Table 2 Parameter settings of each search method

Models	Parameter settings
Prop. PSO	maximum velocity=0.6, inertia weight=0.6, using 2D adaptive super-ellipse coefficients as crossover factors for hybrid leader generation, and 3D super-ellipse coefficients as search parameters
PSOVA [73]	maximum velocity=0.6, inertia weight=0.6, adaptive acceleration coefficients generated using nonlinear functions
MPSO [14]	an adaptive inertia weight factor and linear acceleration coefficients
PSO [11]	maximum velocity=0.6, inertia weight=0.5, acceleration constants c_1= c_2= 1.5
SPSO	adaptive acceleration coefficients generated using sine annealing functions

5.1 Evaluation Using the ICBHI Data Set

We first evaluate the proposed ensemble CBiLSTM networks using the ICBHI data set. As mentioned earlier, it has a total of 920 audio recordings [25]. We conduct the 3-class classification, i.e. healthy, chronic and non-chronic cases, for this respiratory sound data set. Each search method is used to optimize the learning rate, weight decay, the number of filters in the first convolutional layer and the number of hidden neurons in each BiLSTM layer. We employ a ratio of 80-20 for splitting the training and test sets. Each class is split consistently using this aspect ratio.

Since nearly 88% of the training samples belong to the chronic cases (648), we conduct data augmentation of the healthy (28) and non-chronic (60) instances, in order to ensure all the three class signals are balanced in the training set. Motivated by Wu et al. [29], we duplicate each healthy audio clip 23 times and each non-chronic signal 10 times, producing additional 644 healthy and 600 non-chronic samples. By adding these generated samples into the original training set, we have 648 chronic, 660 non-chronic and 672 healthy cases altogether. These samples are divided into training and validation sets using a ratio of 80-20 for hyper-parameter selection.

Besides duplicating the audio samples, standard audio augmentation procedures, such as pitch and time shifting and noise addition, are also conducted. Specifically, we employ the pitch shifting range of [-2, 2] semitones, time shifting range of [-0.3, 0.3] seconds, and noise addition SNR range of [-20, 40] dB, for generating the new signals. Moreover, spectrogram augmentation at the image level, i.e., random flipping and image translation and rotation, has also been conducted, to avoid overfitting. The most optimal learning configurations and model structures identified by each search method are subsequently decoded to construct the output CBiLSTM network. We then train this optimized network using the combined training and validation sets and a larger number (i.e. 100) of training epochs. Each trained network is then evaluated using the unseen audio clips in the test set. Note that we only conduct data duplication and augmentation processes at the training stage, and do not apply such processes to the test set.

To avoid accidental factors, we conduct 30 runs for hyper-parameter search. An ensemble model is subsequently constructed by incorporating three optimized CBiLSTM networks where the majority voting method is used to yield the final prediction. The average accuracy rate of the 10 ensemble models is used in performance comparison. In order to compare with related studies and better tackle class imbalance issues, we employ the mean score of sensitivity and specificity (also referred as a modified accuracy score), for performance comparison. The detailed evaluation metrics are defined in Equations (30)-(32).

$$Score = \frac{Sensitivity+Specificity}{2} \tag{30}$$

$$Sensitivity = \frac{C_{chronic}+C_{non-chronic}}{N_{chronic}+N_{non-chronic}} \tag{31}$$

$$Specificity = \frac{C_{healthy}}{N_{healthy}} \tag{32}$$

where C and N represent the correctly classified and total numbers of samples with respect to a particular class.

All the search methods terminate when the maximum number of function evaluations, i.e. 15 (population) × 20 (maximum number of iterations), is reached. Table 3 illustrates the mean score of each search method over 30 runs.

Table 3 The mean evaluation result of each search method over a set of 30 runs for the ICBHI data set

Models	Methodologies	Sensitivity	Specificity	Mean score
Prop. PSO-based Ensemble	Prop. PSO + ensemble CBiLSTM	0.9605	1	0.9803
PSOVA-based Ensemble	PSOVA + ensemble CBiLSTM	0.9266	0.7143	0.8205
SPSO-based Ensemble	SPSO + ensemble CBiLSTM	0.9435	0.7143	0.8289
MPSO-based Ensemble	MPSO + ensemble CBiLSTM	0.9209	0.7143	0.8176
PSO-based Ensemble	PSO + ensemble CBiLSTM	0.9209	1	0.9605
Ensemble model with default settings	Ensemble CBiLSTM with default settings	0.9153	0.7143	0.8148

As indicated in Table 3, the optimized ensemble models devised by the proposed PSO algorithm achieve a mean score of 98.03% and illustrate great superiority over the ensemble networks formed by the baseline search methods. The proposed search strategies, which include elite signal generation with regular and irregular adaptive weighting, diversified spherical coefficients, and secant and Newton's methods for leader enhancement, allow the search process to better balance between diversification and intensification and to overcome stagnation. In addition, the ensemble networks identified by the original PSO model achieve impressive performance with a mean score of 96.05%. Our proposed models and PSO-based deep networks show superior capabilities and robustness in identifying both the diseased and healthy cases, as indicated by the sensitivity and specificity results. The SPSO-based ensemble models with a mean score of 82.89% outperform PSOVA and MPSO-based ensemble networks. While these PSO-based networks are useful in classifying the chronic and non-chronic cases, they show less capabilities in identifying the healthy cases. The ensemble model with default base classifier settings illustrates limited diversity capabilities, in view of its lower performance as compared with those of ensemble networks formed by all the search methods.

Table 4 The identified mean optimal hyper-parameters for each search method over a set of 30 runs for the ICBHI data set

	Score	Learning rate	Weight decay	No. of filters	No. of hidden units
Prop. PSO	0.9803	0.003815	0.02759	37.67	85.33
PSO	0.9605	0.003638	0.02533	33.5	83.5
PSOVA	0.8205	0.002870	0.01859	44	112.75
SPSO	0.8289	0.003902	0.02413	32.5	96.5
MPSO	0.8176	0.004001	0.03473	40.67	106.67

Table 4 shows the identified mean optimal hyper-parameters for each search method over 30 runs. The proposed PSO model identifies comparatively moderate learning rates and weight decays, as well as moderate numbers of filters and hidden neurons, in comparison with those yielded by other search methods. Such moderate model and learning settings equip the networks with superior capabilities in extracting efficient discriminative spatial-temporal cues as well as prevent them from memorizing the exact details of the training images and sequential dynamics. The ensemble CBiLSTM models identified by the PSO algorithm show similar characteristics as those obtained by the proposed model, but with smaller numbers of filters and hidden units, which can potentially omit certain important descriptive indicators for spectrogram analysis. In contrast, the networks constructed by PSOVA and MPSO possess larger numbers of filters and hidden neurons. Therefore, they are more likely to suffer from overfitting by excessive extraction of spatial-temporal dependencies. The models formulated by SPSO have smaller mean numbers of filters but larger numbers of hidden units. Thus, they are likely to discard important spatial information, while over learning the sequential details.

We compare our optimized ensemble models with other deep networks in Table 5. The augmented data sets from our approach are used for training the baseline deep learning methods. As illustrated in Table 5, the proposed ensemble models with optimal hyper-parameter selection outperform all the baseline deep networks, i.e. VGG8, GoogLeNet with transfer learning and ResNet18-based deep feature extraction with SVM-based classification. Instead of purely extracting spatial features as in these baseline CNN-based methods, each of our optimized CBiLSTM networks takes both spatial and temporal dynamics into account for respiratory sound classification. Owing to the adoption of diverse optimized model structures, the proposed networks possess distinctive learning behaviours and significant diversity to enhance the performance. The confusion matrix of our ensemble networks is provided in Table 6.

Table 5 Performance comparison with other baseline deep networks for the ICBHI data set

Models	Methodologies	Sensitivity	Specificity	Mean score
Prop. PSO-based Ensemble	Prop. PSO + ensemble CBiLSTM	0.9605	1	0.9803
CNN	CNN with 8 convolutional layers	0.9096	0.7143	0.8120
GoogLeNet	Transfer learning	0.9266	0.7143	0.8205
ResNet18+SVM	Features extracted from deep layers+SVM	0.8531	0.4286	0.6409

Table 6 The confusion matrix of the proposed ensemble network for the ICBHI data set

	Non-chronic	**Chronic**	**Healthy**
Non-chronic	10	4	1
Chronic	2	160	0
Healthy	0	0	7

To indicate model efficiency, Table 7 shows a performance comparison between our yielded ensemble model and several existing studies. Since different evaluation strategies have been used, Table 7 depicts a rough estimation of model performance. In comparison with the existing studies using various CNN and LSTM networks in combination with a variety of data augmentation techniques, the proposed PSO-based ensemble network shows superior performance, which appears to be the top performer for respiratory anomaly detection.

Table 7 Comparison with existing studies for the ICBHI data set

Relate studies	Methodologies	Evaluation strategy	Sensitivity	Specificity	Score
Prop. PSO-based Ensemble	Prop. PSO + ensemble CBiLSTM	80-20 split	0.9605	1	0.9803
Perna [72]	CNN	80-20 split	0.89	0.76	0.83
Perna and Tagarelli [28]	LSTM with frame composition (50% overlapping between consecutive windows)	80-20 split	0.98	0.82	0.9
Perna and Tagarelli [28]	LSTM with non-overlapping partitioning	80-20 split	0.98	0.80	0.89
García-Ordás et al. [34]	CNN + Synthetic Minority Oversampling Technique	10-fold	0.950	0.167	0.558
García-Ordás et al. [34]	CNN + Adaptive Synthetic Sampling Method	10-fold	0.965	0.857	0.911
García-Ordás et al. [34]	CNN + dataset weighted	10-fold	0.953	0	0.476
García-Ordás et al. [34]	CNN + dataset unbalanced	10-fold	0.941	0	0.471
García-Ordás et al. [34]	CNN + VAE	10-fold	0.985	0.990	0.988

5.2 Evaluation Using the PhysioNet/CinC Challenge 2016 Data Set

The proposed ensemble CBiLSTM networks are evaluated using a heart sound data set, i.e. PhysioNet/CinC Challenge 2016. The data set has a total of 3,240 sound clips. A 5-fold cross validation is used, which is the same as in some existing studies [26, 78]. Specifically, a total of 2,592 audio samples (from randomly selected four folds) are used for training with the remaining fold of 648 unseen instances for testing. Such a process is performed 5 times so that all the 3,240 samples can be used as test data. For each validation (i.e. the evaluation of each fold), there are 532 positive and 2,060 negative instances in the training set. To balance the training set, as recommended by Wu et al. [29], we duplicate each training signal in the positive class three times. The generated signals are added into the pool of original positive recordings to produce a total of 2,128 positive instances. The updated training set thus consists of 2,128 abnormal and 2,060 normal signals.

We further split this augmented training set into training and validation sets with a ratio of 80-20 for optimal hyper-parameter selection. The performance of the validation set is used as the fitness score. Besides duplicating the audio samples, standard audio augmentation procedures, such as pitch and time shifting and noise addition, as well as spectrogram augmentation at the image level, such as random flipping and image translation and rotation, have also been conducted, to avoid overfitting. The identified optimal settings are then used to construct a CBiLSTM network, which is subsequently trained using the combined training and validation sets with a larger number (i.e. 100) of epochs. The performance of the unseen test set (i.e. the remaining fold) is used for comparison. Note that the aforementioned data duplication and signal and image augmentation procedures are only conducted for the training set in each validation, without implementing them in the test set.

For each fold, we conduct 30 trials. Therefore a set of 30 optimized base classifiers is obtained, which are used to construct 10 ensemble models. The mean score of the 10 ensemble classifiers is produced for each fold. Such a process is conducted five times for 5 test folds where hyper-parameter search is performed 30 runs for each fold.

According to PhysioNet 2016, the modified accuracy score, i.e. the mean of sensitivity and specificity as defined in Equation (30), is employed for performance comparison. The new definitions of sensitivity and specificity with respect to the PhysioNet data set are provided below, i.e. sensitivity refers to the percentage of abnormal signals that are correctly classified as abnormal cases, as shown in Equation (33), while specificity indicates the percentage of normal phonocardiograms that are correctly identified as normal cases, as in Equation (34). Then a mean score of sensitivity and specificity is calculated using Equation (30).

$$Sensitivity = \frac{C_{abnormal}}{N_{abnormal}} \tag{33}$$

$$Specificity = \frac{C_{normal}}{N_{normal}} \tag{34}$$

where C and N denote the correctly classified and total numbers of samples with respect to a specific class.

Table 8 presents the mean result of 5-fold cross validation for each search method, where hyper-parameter search is conducted 30 trials in each fold.

Table 8 The mean evaluation results of the 5-fold cross validation for the PhysioNet data set with 30 trials in each fold

Models	Methodologies	Sensitivity	Specificity	Mean score
Prop. PSO-based Ensemble	Prop. PSO + ensemble CBiLSTM	0.9158	0.9227	0.9193
PSOVA-based Ensemble	PSOVA + ensemble CBiLSTM	0.8707	0.9060	0.8883
SPSO-based Ensemble	SPSO + ensemble CBiLSTM	0.8241	0.9364	0.8803
MPSO-based Ensemble	MPSO + ensemble CBiLSTM	0.8511	0.9273	0.8892
PSO-based Ensemble	PSO + ensemble CBiLSTM	0.8616	0.9122	0.8869
Ensemble model with default settings	Ensemble CBiLSTM with default settings	0.8236	0.9327	0.8782

As illustrated in Table 8, our optimized ensemble networks outperform those devised by the baseline methods, with mean accuracy score of 91.93%. Owing to the guidance of the secant and Newton's methods and diverse super-elliptical search courses for exploitation and exploration, the proposed PSO model is efficient in overcoming local optima traps for hyper-parameter selection. The sensitivity and specificity results indicate that our ensemble CBiLSTM networks illustrate competitive performance for classification of both positive and negative cases. MPSO, PSOVA and PSO-based ensemble networks obtain promising accuracy scores of 88.92%, 88.83% and 88.69%, respectively, with SPSO-based ensemble models obtaining a lower performance of 88.03%. As shown in Table 8, the ensemble networks devised by these PSO variants illustrate promising performance in identifying the healthy recordings, but with less competence in classifying the abnormal heart sound signals. The ensemble models yielded by all the search methods possess sufficient robustness and show better accuracy scores than those of the ensemble model with default settings.

Table 9 The identified mean optimal hyper-parameters for an example fold over a set of 30 runs for the PhysioNet data set

	Mean score	Learning rate	Weight decay	No. of filters	No. of hidden units
Prop. model	0.9193	0.003222	0.03238	45.5	95
PSOVA	0.8883	0.002419	0.01649	50	126
PSO	0.8869	0.003953	0.02559	47	81.67
MPSO	0.8892	0.003144	0.01827	42.33	116.33
SPSO	0.8803	0.004943	0.04108	33	90

Table 9 illustrates the identified mean optimal hyper-parameters for a particular fold over 30 runs. In comparison with those yielded by the baseline search methods, the proposed model identifies moderate settings of the learning rate and weight decay, moderate numbers of filters and hidden neurons. Such moderate settings of the optimized CBiLSTM networks are beneficial for extracting discriminative spatial-temporal information from the spectrogram inputs, leading to superior performance for classification of both positive and negative cases. On the other hand, PSOVA and MPSO identify comparatively larger numbers of hidden units in the BiLSTM layers, with larger and smaller numbers of filters in the convolutional layers, respectively. Therefore, they are more likely to capture excessive details of the sequential dependencies. The PSOVA-based networks potentially extract redundant spatial features. Moreover, PSO selects comparatively smaller numbers of hidden units and larger numbers of filters. Therefore they are capable of describing complex spatial patterns but have a constrained storage of the temporal relationships. SPSO extracts both comparatively smaller numbers of filters and hidden units. The resulting networks show limited capabilities in capturing complex spatial-temporal dynamics, thus a lower performance.

The moderate learning rates and weight decays from our proposed approach equip the networks with a good balance between diversity and convergence. Comparatively, PSO and SPSO obtain larger learning rates, which cause the networks to converge quickly to a suboptimal solution. MPSO and PSOVA yield smaller learning rates, resulting in slow convergence. With respect to the evaluation of other test folds, the devised hyper-parameter settings of our model show similar characteristics, i.e. moderate learning rates and weight decays and moderate numbers of filters and hidden units.

Table 10 shows a comparison between our devised ensemble models and other deep networks. A 5-fold cross validation is used to evaluate each baseline method. The baseline networks are trained using the augmented data set in each validation as those used in our experiment. As illustrated in Table 10, our optimized ensemble models outperform the baseline deep networks, i.e. VGG8, GoogLeNet with transfer learning and ResNet18-based spatial feature extraction with SVM-based classification, significantly. Among the baseline methods, GoogLeNet with transfer learning obtains a better performance, owing to its significant feature learning capabilities produced by both pre-training using ImageNet and transfer learning using the new spectrogram data set. Our devised ensemble CBiLSTM models incorporate a variety of networks with distinctive topologies, and show superior robustness in discriminative spatial-temporal feature extraction, therefore yielding a better performance. Table 11 illustrates the confusion matrix of our ensemble networks.

Table 10 Performance comparison with other baseline deep networks for the PhysioNet data set

Models	Methodologies	Sensitivity	Specificity	Mean score
Prop. PSO-based Ensemble	Prop. PSO + ensemble CBiLSTM	0.9158	0.9227	0.9193
CNN	CNN with 8 convolutional layers	0.7639	0.9351	0.8496
GoogLeNet	Transfer learning	0.8481	0.9122	0.8802
ResNet18+SVM	Features extracted from deep layers+SVM	0.7925	0.9076	0.85

Table 11 The confusion matrix of the proposed ensemble network for the PhysioNet data set

	Abnormal	**Normal**
Abnormal	609	56
Normal	199	2376

Table 12 illustrates a comparison between our devised ensemble model and existing studies. Since different evaluation strategies and sample sizes have been used in different publications, Table 12 serves as a rough performance comparison. Comparing with the existing methods, our obtained ensemble CBiLSTM network achieves superior performances for classification of both normal and abnormal cases. Therefore, our proposed model offers an alternative solution for abnormal heart sound classification.

Table 12 Comparison with existing studies for the PhysioNet data set

Relate studies	Methodologies	Evaluation strategy	Sample size	Sensitivity	Specificity	Score
Prop. PSO-based Ensemble	Prop. PSO + ensemble CBiLSTM	5-fold	3,240	0.9158	0.9227	0.9193
Wu et al. [29]	Ensemble VGGNet with Savitzky-Golay filter	80-20 split	3,240	0.8646	0.8563	0.8604
Wu et al. [29]	Ensemble VGGNet with Savitzky-Golay filter	10-fold	3,240	0.9173	0.8791	0.8981
Li et al. [78]	CNN	5-fold	3,153	0.87	0.866	0.868
Zhang et al. [30]	Scaled spectrogram and tensor decomposition	10-fold	3,240	0.88	0.92	0.9
Zhang et al. [30]	Scaled spectrogram and partial least squares regression	10-fold	3,240	0.82	0.95	0.88
Thomae and Dominik [79]	Deep gated RNN with a convolutional front end	85-15 split	3,153	0.96	0.83	0.89
Her and Chiu [80]	Time-frequency features	304 test samples	3,240	0.844	0.869	0.857
Potes et al. [81]	Ensemble of feature-based and deep learning-based classifiers	80-20 split	3,240	0.88	0.82	0.85
Homsi et al. [82]	A nested set of ensemble algorithms	10-fold	3,240	-	-	0.884
Xiao et al. [31]	1D CNN	10-fold	3,153	0.86	0.95	0.905

5.3 Evaluation Using the ESC-10 Data Set

The proposed ensemble networks are evaluated with a multi-class environmental sound data set, i.e. ESC-10 [27]. A total of 400 audio signals from 10 classes are available in the ESC-10 data set, where each class is composed of 40 instances. Following the guideline of the ESC-10 data set, we employ a 5-fold cross validation in our experiments. In each validation, we randomly select 320 samples (i.e. four folds) for hyper-parameter selection with the remaining 80 signals for testing. Owing to balanced sample sizes in each class, we do not duplicate any training instances, but only conduct standard signal and spectrogram augmentation procedures at the training stage in this evaluation.

We conduct 30 evaluations in each fold. Specifically, we divide the training set (i.e. randomly selected four folds) into training and validation sets with a ratio of 80-20. They are subsequently used for hyper-parameter selection where the validation set is used for fitness evaluation. The identified optimal CBiLSTM network from each search method is trained using the combined training and validation sets (i.e. all the data from the selected four folds) with a significantly larger number (i.e. 100) of epochs. The average result of the diagonal accuracy rates of the confusion matrix is used for performance comparison.

For each fold, as mentioned earlier, we conduct 30 trials. Therefore, a set of 30 optimized base classifiers is obtained for constructing 10 ensemble models. The mean result of the 10 ensemble models is recorded in each fold, which is used to compute the average result of 5-fold cross validation. Table 13 presents the detailed results for performance comparison.

Table 13 The mean evaluation results of 5-fold cross validation for the ESC-10 data set with 30 trials in each fold

Models	Methodologies	Mean accuracy
Prop. PSO-based Ensemble	Prop. PSO + ensemble CBiLSTM	0.93
PSOVA-based Ensemble	PSOVA + ensemble CBiLSTM	0.9023
SPSO-based Ensemble	SPSO + ensemble CBiLSTM	0.8967
MPSO-based Ensemble	MPSO + ensemble CBiLSTM	0.8984
PSO-based Ensemble	PSO + ensemble CBiLSTM	0.9142
Ensemble model with default settings	Ensemble CBiLSTM with default settings	0.8931

As indicated in Table 13, our ensemble networks achieve the highest mean accuracy rate of 93%, outperforming those ensemble networks devised by the original and other baseline PSO algorithms, significantly. PSO and PSOVA-based networks obtain competitive performances with mean accuracy rates of 91.42% and 90.23%, respectively. The ensemble networks yielded by all the search methods comprise base models with distinctive topologies and configurations, therefore illustrating better complementary and boosting characteristics than that of the ensemble network with default settings.

Table 14 The identified mean optimal hyper-parameters for an example fold over a set of 30 runs for the ESC-10 data set

	Accuracy	Learning rate	Weight decay	No. of filters	No. of hidden units
Prop. PSO	0.93	0.002624	0.02021	40.25	64.75
PSOVA	0.9023	0.003340	0.02610	47	57.67
MPSO	0.8984	0.003455	0.01786	45	48.33
PSO	0.9142	0.002005	0.03460	51.67	70.67
SPSO	0.8967	0.003895	0.03101	35.67	81.33

Table 14 illustrates the identified mean hyper-parameter settings for a specific fold over 30 runs for each search method. The proposed model identifies moderate settings of the learning rate and weight decay, as well as moderate numbers of filters and the hidden units, in comparison with those discovered by the baseline methods. Such network and learning configurations illustrate competitive robustness in extracting manifold discriminative spatial features as well as abundant temporal relationships. In addition, the original PSO model identifies comparatively larger mean numbers of filters and hidden units. The resulting model is susceptible to learning redundant spatial-temporal details, which affect its performance. In contrast, PSOVA and MPSO obtain larger numbers of filters and smaller numbers of hidden neurons. Such model settings have limited learning capabilities in discovering elaborated temporal relationships, despite the extraction of enriched spatial features. SPSO identifies a comparatively smaller mean number of filters and a significantly larger mean number of hidden units. This configuration results in the omission of some important discriminative spatial characteristics as well as an exaggerated storage of the sequential dynamics. In comparison with the moderate learning rates devised by the original and the proposed PSO algorithms, larger learning rates are identified by other PSO variants, which result in fast convergence to suboptimal solutions.

The empirical results also reveal that for the evaluation of other test folds, the identified optimized hyper-parameters illustrate similar characteristics. Clearly, the proposed PSO model is able to identify moderate learning rates and weight decays, as well as moderate numbers of filters and hidden units.

The theoretical justification pertaining to the proposed PSO variant and the baseline search methods is as follows. SPSO and MPSO employ adaptive sine and linear functions for search coefficient generation to adaptively adjust the search process. PSOVA adopts dynamic search coefficients, DE-based leader enhancement, and global optimal signals such as average and random leaders as well as g_{best} to overcome stagnation. However, all these baseline PSO variants (MPSO, SPSO and PSOVA) and the original PSO algorithm do not generate any hybrid leader signals and purely rely on one leader at a time to lead the search process, therefore less search diversity. On the other hand, the proposed PSO model first employs the secant and Newton's methods for swarm leader enhancement. It integrates the enhanced leader and the second-best solution using adaptive weights produced by regular, irregular and combined 2D elliptical formulae. Such improved leader signals in conjunction with diverse 3D spherical coefficients are used to balance the search between diversification and intensification. The proposed mechanisms offer the search process a variety of search directions and scales to mitigate premature convergence. Therefore, our yielded ensemble networks outperform those constructed by the baseline search methods for undertaking diverse sound classification tasks that possess various challenging factors.

Table 15 depicts a comparison between our optimized ensemble models and other deep networks. A 5-fold cross validation is applied with standard audio and image augmentation processes as used in our experiments. As illustrated in Table 15, the proposed ensemble models with optimal hyper-parameter selection outperform all the baseline deep networks, i.e. VGG8, GoogLeNet with transfer learning and ResNet18-based deep feature extraction with SVM-based classification. These baseline CNN-based models are prone to extracting spatial information with fixed learning and network settings. Comparatively, our CBiLSTM networks take spatial-temporal dependencies into account, and are equipped with diverse optimized network topologies with enhanced feature learning capabilities. The confusion matrix of our devised ensemble networks is illustrated in Figure 8.

Table 15 Performance comparison with other baseline deep networks for ESC-10 data set

Models	Methodologies	Mean accuracy
Prop. PSO-based Ensemble	Prop. PSO + ensemble CBiLSTM	0.93
CNN	CNN with 8 convolutional layers	0.885
GoogLeNet	Transfer learning	0.86
ResNet18+SVM	Features extracted from deep layers+SVM	0.85

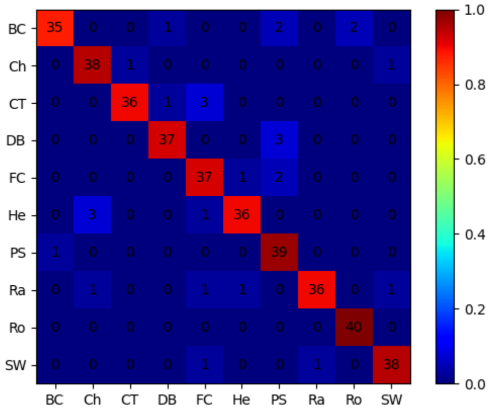


Figure 8 The confusion matrix of the proposed ensemble network for the ESC-10 data set where BC, Ch, CT, DB, FC, He, PS, Ra, Ro and SW denote baby cry, chainsaw, clock tick, dog bark, fire crackling, helicopter, person sneeze, rain, rooster, and sea waves, respectively.

Table 16 depicts a comparison between our ensemble CBiLSTM network and several state-of-the-art models. All the reported studies employed a 5-fold cross validation. As shown in Table 16, our evolving ensemble deep network with optimal hyper-parameter selection illustrates impressive performance, which is the top performer in this environmental sound classification problem.

Table 16 Comparison with existing studies for the ESC-10 data set

Relate studies	Methodologies	Mean accuracy
Prop. PSO-based Ensemble	Prop. PSO + ensemble CBiLSTM	0.93
Medhat et al. [40]	MCLNN	0.855
Esmaeilpour et al. [38]	Unsupervised feature learning and WCCGAN + RF as the classifier	0.87
Zhang et al. [39]	CRNN (8 convolutional layers+2 BiGRU layers)	0.93
Zhang et al. [69]	CNN + augmentation + mixed sound	0.917
Boddapati et al. [83]	GoogLeNet and AlexNet	0.86
Aytar et al. [84]	SoundNet	0.823
Su et al. [85]	A two-stream CNN	0.72
Piczak [86]	PiczakConvNets	0.805

daSilva et al. [87]	ANN, KNN + features cascading + optimization	0.78
Khamparia et al. [88]	CNN and tensor deep stacking network	0.56
Salamon and Bello [89]	CNN (3 conv layers + 2 fully connected layers)	0.77

5.4 Evaluation Using Benchmark Functions

To further evaluate the proposed PSO algorithm, we employ a set of 11 benchmark functions for performance assessment. They include multimodal functions, i.e. Ackley, Griewank, Rastrigin, and Powell, and unimodal artificial landscapes, i.e. Dixon-Price, Rotated Hyper-Ellipsoid (denoted as Rothyp), Rosenbrock, Sphere, Sum of Different Powers (denoted as Sumpow), Sum Squares (denoted as Sumsqu) and Zakharov. The unimodal benchmark functions have a single global minimum, while the multimodal landscapes illustrate multiple local optima. As indicated in [50, 90, 91], these mathematical functions constitute a challenging test suite with varied difficulties to test model efficiency.

In addition to the baseline methods, a set of additional classical and advanced search algorithms is employed for evaluation. These algorithms have shown significant superiority in solving diverse benchmark functions, i.e. FA [12], Genetic PSO (GPSO) [13], Dynamic Neighbourhood Learning PSO (DNLPSO) [15], ALPSO [62], Enhanced Leader PSO (ELPSO) [9], chaotic FA with Logistic map as random search parameters (CLFA) [16], chaotic FA with Gauss map as the attractiveness coefficients (CGFA) [17], FA with variable step sizes (VSSFA) [18], a modified FA (MFA) [19], and FA with neighbourhood attraction (NaFA) [20]. Besides the abovementioned PSO and FA-based methods, we employ ten other classical and advanced search methods for performance comparison, i.e. Moth-Flame Optimization (MFO) [75], Ant Lion Optimizer (ALO) [76], Dragonfly Algorithm (DA) [77], Cuckoo Search (CS) [92], Artificial Bee Colony (ABC) [93], Bat Algorithm (BA) [94], Whale Optimization Algorithm (WOA) [95], PSOGSA (the combination of PSO and Gravitational Search Algorithm (GSA)) [96], GWO [22] and amixedGWO [65].

Table 17 illustrates the experimental settings of these baseline methods, which are obtained from their original studies.

Table 17 Parameter settings of additional baseline methods

Methods	Parameter settings
MFO [75]	adaptive parameter settings
GPSO [13]	search coefficients c_1 = 2.6, c_2 = 1.5, inertia weight=0.9, maximum velocity=0.6.
DNLPSO [15]	refreshing gap=3, regrouping period=5, c_1 = c_2 =1.49445, inertia weight=0.9−(0.9−0.4)×(j−1)/(MaxGeneration−1), where j and MaxGeneration represent the current and maximum iteration numbers, respectively.
ALPSO [62]	random helix coefficients as the search parameters, maximum velocity=0.6, inertia weight=0.6.
ELPSO [9]	$c_1 = c_2$ =2, standard deviation of Gaussian mutation=1, scale parameter of Cauchy mutation=2, scale factor of DE-based mutation=1.2, inertia weight=0.9−(0.9−0.4)×(j−1)/(MaxGeneration−1).
FA [12]	Levy's index=1.5, randomization parameter=0.5, initial attractiveness=1.0, absorption coefficient=1.0.
CLFA [16]	Levy's index=1.5, randomization parameter=Logistic map, initial attractiveness=1.0, absorption coefficient=1.0.
CGFA [17]	attractiveness=Gauss map, absorption coefficient=1.0, Levy's index=1.5, randomization parameter=0.5.
VSSFA [18]	initial attractiveness=1.0, absorption coefficient=1.0, Levy's index=1.5, randomization parameter=0.4/(1+exp(0.015*(j-MaxGeneration)/3)).
MFA [19]	Levy's index=1.5, randomization parameter=0.5, initial attractiveness=1.0, absorption coefficient=1.0.
NaFA [20]	absorption coefficient=1.0, Levy's index=1.5, and randomization parameter $\alpha(j+1) = \left(\frac{1}{9000}\right)^{\frac{1}{j}} \times \alpha(j)$ where j is the current iteration.
GWO [22]	step size $A = (2 \times rand - 1) \times a,$ where a linearly decreases from 2 to 0, $rand \in (0, 1)$, search parameter $C = 2 \times rand$.
amixedGWO [65]	adopting parameter settings as in the vanilla GWO [22], and $\eta = 3$
ALO [76]	$I = 10^w \times \frac{j}{\text{MaxGeneration}}$, where $w = 2$ when $j > 0.1 \times \text{MaxGeneration}$, $w = 3$ when $j > 0.5 \times \text{MaxGeneration}$, $w = 4$ when $j > 0.75 \times \text{MaxGeneration}$, $w = 5$ when $j > 0.9 \times \text{MaxGeneration}$, and $w = 6$ when $j > 0.95 \times \text{MaxGeneration}$
DA [77]	separation weight = 0.1, alignment weight = 0.1, cohesion weight = 0.7, food factor = 1, enemy factor = 1, and inertia weight=0.9 – j×((0.9-0.4)/ MaxGeneration)
CS [92]	discovery probability = 0.25
ABC [93]	limit=dimension×population
BA [94]	loudness = 0.5, pulse rate = 0.5
WOA [95]	step size $A = (2 \times rand - 1) \times a,$ where a linearly decreases from 2 to 0, $rand \in$

	$(0, 1)$, search parameter $C = 2 \times rand$, $b = 1$, and $l = (a_2 - 1) \times rand + 1$, where a_2 linearly decreases from -1 to -2.
PSOGSA [96]	inertia weight = $rand$, acceleration constants c_1=0.5, c_2=1.5, initial gravitational constant G_0 = 1, and descending coefficient α = 20.

The following settings are used in our experiments, i.e. dimension=30, trial=30, and maximum number of function evaluations=population (30) × iterations (2000). The search process terminates when the maximum number of function evaluations is reached. Table 18 illustrates the mean, minimum, maximum, and standard deviation results over 30 runs for each search method. The best results are highlighted in bold. As indicated in Table 18, the proposed PSO algorithm achieves superior performance for solving diverse unimodal and multimodal benchmark functions. It outperforms 24 baseline methods in most test cases, except for the Ackley and Dixon-Price functions, where ALPSO and ABC achieve the best results, respectively. The proposed model achieves the best global minima (i.e., 0) for two multimodal functions, i.e. Griewank and Rastrigin, and five unimodal functions, i.e. Rotated Hyper-Ellipsoid, Sphere, Zakharov, Sum of Different Powers and Sum Squares. ALPSO, WOA, GWO and amixedGWO attain the global minimum solution for Griewank, while WOA and GWO also obtain the best minimum solution for both Rastrigin and Sum of Different Powers. ALPSO and WOA obtain the global minimum solution for Ackley and Zakharov, respectively.

Table 18 Evaluation results for benchmark functions with dimension=30 over a set of 30 runs

		Prop. PSO	**PSOVA**	**SPSO**	**MPSO**	**ALPSO**	**GPSO**	**DNLPSO**	**ELPSO**	**CLFA**	**CGFA**	**VSSFA**	**MFA**	**NaFA**
Ackley	mean	8.88E-16	1.76E+01	1.14E+01	1.69E+01	**0.00E+00**	1.77E+01	2.19E+00	1.51E+01	1.60E+01	1.46E+01	1.07E+01	2.03E+01	4.48E-03
	min	8.88E-16	1.62E+01	9.91E+00	1.46E+01	**0.00E+00**	1.69E+01	1.10E-03	1.40E+01	1.50E+01	1.39E+01	9.62E+00	2.03E+01	3.71E-03
	max	8.88E-16	1.87E+01	1.26E+01	1.92E+01	**0.00E+00**	1.83E+01	8.96E+00	1.58E+01	1.65E+01	1.51E+01	1.15E+01	2.03E+01	5.00E-03
	std	**0.00E+00**	7.33E-01	6.65E-01	1.22E+00	**0.00E+00**	4.65E-01	3.11E+00	5.82E-01	4.27E-01	3.77E-01	6.10E-01	**0.00E+00**	3.97E-04
Dixon	mean	**2.49E-01**	2.32E+05	1.39E+04	4.95E+05	2.84E+00	7.01E+05	3.96E+00	1.29E+05	1.50E+05	1.09E+05	1.28E+04	1.62E+06	7.07E-01
	min	**2.49E-01**	8.17E+04	8.63E+03	8.93E+03	7.06E-01	4.42E+05	6.67E-01	8.27E+01	1.04E+05	5.19E+04	7.40E+03	1.62E+06	6.67E-01
	max	**2.49E-01**	4.47E+05	1.96E+04	1.82E+06	6.00E+00	1.14E+06	1.17E+01	1.83E+05	1.97E+05	1.42E+05	1.56E+04	1.62E+06	8.65E-01
	std	1.87E-04	1.48E+05	3.51E+03	4.78E+05	1.51E+00	2.12E+05	3.54E+00	5.67E+04	2.97E+04	2.77E+04	2.45E+03	**1.74E-10**	7.75E-02
Griewank	mean	**0.00E+00**	1.66E+02	4.80E+01	1.90E+02	**0.00E+00**	3.15E+02	6.62E-01	1.68E+02	1.75E+02	1.51E+02	4.94E+01	6.08E+02	4.77E-03
	min	**0.00E+00**	8.85E+01	3.36E+01	3.94E+01	**0.00E+00**	1.67E+02	1.29E-02	1.17E+02	1.44E+02	1.14E+02	4.37E+01	6.08E+02	1.25E-03
	max	**0.00E+00**	3.38E+02	6.41E+01	3.85E+02	**0.00E+00**	3.60E+02	1.51E+00	2.18E+02	1.96E+02	1.70E+02	5.82E+01	6.08E+02	1.39E-02
	std	**0.00E+00**	7.18E+01	7.83E+00	1.04E+02	**0.00E+00**	5.47E+01	6.85E-01	3.67E+01	1.46E+01	1.60E+01	4.09E+00	1.20E-13	5.27E-03
Rastrigin	mean	**0.00E+00**	2.68E+02	2.18E+02	2.70E+02	1.82E+02	3.61E+02	7.63E+01	2.77E+02	2.61E+02	2.64E+02	2.22E+02	4.29E+02	5.14E+01
	min	**0.00E+00**	1.88E+02	1.81E+02	1.91E+02	1.26E+02	3.20E+02	3.99E+01	2.21E+02	2.22E+02	2.47E+02	2.04E+02	4.29E+02	3.08E+01
	max	**0.00E+00**	3.22E+02	2.52E+02	3.49E+02	2.12E+02	4.22E+02	1.33E+02	3.02E+02	2.84E+02	2.74E+02	2.40E+02	4.29E+02	9.05E+01
	std	**0.00E+00**	3.99E+01	1.57E+01	4.08E+01	1.95E+01	2.90E+01	3.07E+01	2.67E+01	1.81E+01	7.83E+00	1.23E+01	5.68E-14	1.66E+01
Rothyp	mean	**0.00E+00**	3.83E+04	3.15E+04	2.36E+05	9.30E+04	2.57E+05	8.93E+02	1.79E+02	9.85E+04	1.08E+05	3.23E+04	4.38E+05	1.67E-02
	min	**0.00E+00**	3.83E+04	1.84E+04	2.36E+05	6.75E+04	2.57E+05	4.74E-01	1.79E+02	9.85E+04	1.08E+05	2.84E+04	4.38E+05	4.31E-03
	max	**0.00E+00**	3.83E+04	4.79E+04	2.36E+05	1.34E+05	2.57E+05	3.60E+03	1.79E+02	9.85E+04	1.08E+05	3.70E+04	4.38E+05	3.20E-02
	std	**0.00E+00**	7.67E-12	6.07E+03	**0.00E+00**	1.51E+04	**0.00E+00**	1.35E+03	3.00E-14	1.53E-11	1.53E-11	3.28E+03	1.94E-11	1.06E-02
Rosenbrock	mean	**5.11E-04**	1.82E+05	1.50E+04	3.23E+05	1.42E+01	4.66E+05	6.64E+01	1.58E+05	1.34E+05	6.17E+04	8.87E+03	3.36E+06	5.71E+01
	min	**1.73E-09**	4.77E+04	5.50E+03	6.26E+04	3.14E-01	3.01E+05	2.25E+01	5.33E+04	9.81E+04	2.71E+04	6.04E+03	3.36E+06	2.44E+01
	max	**3.00E-03**	3.39E+05	2.30E+04	7.27E+05	7.09E+01	5.70E+05	3.02E+02	2.56E+05	1.75E+05	8.58E+04	1.14E+04	3.36E+06	2.10E+02
	std	7.76E-04	9.30E+04	4.34E+03	1.63E+05	1.89E+01	7.62E+04	8.53E+01	5.37E+04	2.62E+04	1.84E+04	1.84E+03	**4.66E-10**	6.26E+01
Sphere	mean	**0.00E+00**	4.92E+01	1.49E+01	4.92E+01	6.41E-01	9.11E+01	3.62E-01	4.48E+01	4.86E+01	4.14E+01	1.38E+01	1.77E+02	1.86E-06
	min	**0.00E+00**	2.36E+01	1.05E+01	8.87E+00	**0.00E+00**	6.93E+01	1.34E-10	3.46E+01	3.86E+01	2.52E+01	9.76E+00	1.77E+02	1.52E-06
	max	**0.00E+00**	8.55E+01	1.85E+01	1.09E+02	1.62E+00	1.20E+02	3.19E+00	5.98E+01	5.56E+01	4.90E+01	1.69E+01	1.77E+02	2.27E-06
	std	**0.00E+00**	2.10E+01	2.20E+00	2.41E+01	6.21E-01	1.72E+01	9.97E-01	6.71E+00	5.88E+00	6.79E+00	2.14E+00	3.42E-14	2.28E-07
Sumpow	mean	**0.00E+00**	1.70E-03	1.60E-04	3.67E-01	1.31E-03	2.60E-01	6.95E-10	2.79E-02	6.28E-03	4.63E-03	3.15E-04	5.82E-01	1.77E-08
	min	**0.00E+00**	1.05E-04	3.17E-05	4.58E-06	**0.00E+00**	6.11E-02	1.07E-74	1.22E-02	3.03E-03	1.31E-03	9.54E-05	5.82E-01	5.04E-09
	max	**0.00E+00**	7.37E-03	4.62E-04	2.00E+00	1.73E-02	5.74E-01	2.38E-09	6.49E-02	1.05E-02	1.14E-02	8.00E-04	5.82E-01	3.23E-08
	std	**0.00E+00**	2.33E-03	1.12E-04	5.56E-01	4.28E-03	1.79E-01	9.07E-10	1.48E-02	2.58E-03	3.05E-03	2.39E-04	1.33E-16	8.28E-09
Zakharov	mean	**0.00E+00**	4.70E+02	2.82E+02	3.52E+02	9.64E+01	4.60E+02	7.69E+01	3.61E+02	3.91E+02	3.18E+02	2.53E+02	9.34E+02	4.92E+01
	min	**0.00E+00**	3.87E+02	2.49E+02	2.57E+02	**0.00E+00**	4.25E+02	3.87E+01	3.24E+02	3.71E+02	2.93E+02	2.39E+02	9.34E+02	3.28E+01
	max	**0.00E+00**	5.67E+02	3.27E+02	5.46E+02	4.42E+02	4.82E+02	1.54E+02	3.95E+02	4.10E+02	3.41E+02	2.69E+02	9.34E+02	6.77E+01
	std	**0.00E+00**	6.52E+01	2.26E+01	6.97E+01	1.78E+02	1.81E+01	3.47E+01	2.54E+01	1.25E+01	1.62E+01	1.01E+01	6.56E-14	1.31E+01
Sumsqu	mean	**0.00E+00**	6.93E+02	1.83E+02	8.92E+02	4.07E-01	1.39E+03	6.17E-01	6.65E+02	6.40E+02	5.79E+02	1.90E+02	2.77E+03	1.34E-04
	min	**0.00E+00**	2.21E+02	1.08E+02	1.01E+02	2.25E-02	1.04E+03	2.85E-16	5.35E+02	5.26E+02	4.43E+02	1.56E+02	2.77E+03	3.37E-05
	max	**0.00E+00**	1.43E+03	2.41E+02	2.20E+03	1.38E+00	1.80E+03	5.58E+00	7.80E+02	7.42E+02	6.61E+02	2.30E+02	2.77E+03	3.26E-04
	std	**0.00E+00**	3.77E+02	2.91E+01	5.70E+02	3.59E-01	2.39E+02	1.75E+00	7.78E+01	7.35E+01	6.19E+01	2.30E+01	4.55E-13	1.05E-04
Powell	mean	**1.19E-270**	1.83E+03	2.73E+02	5.49E+03	1.14E+00	4.13E+03	1.56E+00	2.81E+03	1.61E+03	1.17E+03	3.59E+02	8.55E+03	3.42E-02
	min	**0.00E+00**	1.83E+03	1.60E+02	4.13E+02	2.30E-01	2.93E+03	4.65E-02	2.81E+03	1.01E+03	7.15E+02	2.54E+02	8.55E+03	1.38E-02
	max	**3.57E-269**	1.83E+03	3.98E+02	1.56E+04	2.49E+00	5.51E+03	7.89E+00	2.81E+03	2.15E+03	1.71E+03	5.21E+02	8.55E+03	5.50E-02
	std	**0.00E+00**	**0.00E+00**	6.28E+01	4.00E+03	6.04E-01	9.25E+02	2.40E+00	**0.00E+00**	3.61E+02	3.00E+02	7.85E+01	6.06E-13	1.52E-02

		Prop. PSO	**PSO**	**MFO**	**FA**	**ALO**	**DA**	**CS**	**ABC**	**BA**	**WOA**	**PSOGSA**	**GWO**	**amixedGWO**
Ackley	mean	**8.88E-16**	8.30E+00	1.40E+01	6.08E-03	1.90E+01	1.97E+01	9.56E-02	6.10E-06	1.86E+01	4.44E-15	1.34E+00	8.70E-15	**8.88E-16**
	min	**8.88E-16**	3.93E+00	8.83E+00	5.07E-03	1.90E+01	1.97E+01	3.39E-05	6.10E-06	1.74E+01	4.44E-15	1.34E+00	7.99E-15	**8.88E-16**
	max	**8.88E-16**	1.36E+01	1.80E+01	7.51E-03	1.90E+01	1.97E+01	9.51E-01	6.10E-06	1.98E+01	4.44E-15	1.34E+00	1.51E-14	**8.88E-16**
	std	**0.00E+00**	2.06E+00	3.08E+00	8.74E-04	**0.00E+00**	3.74E-15	2.86E-01	**0.00E+00**	9.03E-01	**0.00E+00**	2.34E-16	2.25E-15	**0.00E+00**
Dixon	mean	2.49E-01	1.10E+00	6.72E+04	9.89E-01	1.56E+06	3.29E+04	6.67E-01	**2.07E-03**	1.25E+02	6.67E-01	6.67E-01	6.67E-01	6.67E-01
	min	2.49E-01	6.67E-01	6.67E-01	6.67E-01	1.36E+06	3.29E+04	6.67E-01	**2.07E-03**	6.67E-01	6.67E-01	6.67E-01	6.67E-01	6.67E-01
	max	2.49E-01	7.40E+00	5.98E+05	1.64E+00	1.58E+06	3.29E+04	6.67E-01	**2.07E-03**	9.16E+02	6.67E-01	6.67E-01	6.67E-01	6.67E-01

	std	1.87E-04	1.32E+00	1.88E+05	3.60E-01	6.99E+04	7.67E-12	1.04E-04	4.57E-19	2.00E+02	1.17E-16	**0.00E+00**	8.69E-08	**0.00E+00**
Griewank	mean	**0.00E+00**	1.40E-01	9.07E+00	2.72E-03	6.22E+02	6.62E+00	1.40E-04	1.07E-09	5.76E+02	**0.00E+00**	9.03E+01	**0.00E+00**	**0.00E+00**
	min	**0.00E+00**	8.21E-09	4.99E-12	1.96E-03	5.11E+02	6.62E+00	2.20E-08	1.07E-09	3.42E+02	**0.00E+00**	9.03E+01	**0.00E+00**	**0.00E+00**
	max	**0.00E+00**	1.22E+00	9.02E+01	3.47E-03	6.35E+02	6.62E+00	2.34E-03	1.07E-09	6.87E+02	**0.00E+00**	9.03E+01	**0.00E+00**	**0.00E+00**
	std	**0.00E+00**	2.24E-01	2.85E+01	5.14E-04	3.92E+01	1.87E-15	4.38E-04	**0.00E+00**	8.04E+01	**0.00E+00**	1.50E-14	**0.00E+00**	**0.00E+00**
Rastrigin	mean	**0.00E+00**	6.09E+01	1.47E+02	3.39E+01	4.24E+02	1.14E+02	5.70E+01	2.42E-07	1.95E+02	**0.00E+00**	7.86E+01	**0.00E+00**	2.77E+01
	min	**0.00E+00**	3.58E+01	8.66E+01	1.59E+01	4.19E+02	1.14E+02	3.52E+01	2.42E-07	1.20E+02	**0.00E+00**	7.86E+01	**0.00E+00**	2.77E+01
	max	**0.00E+00**	8.86E+01	2.28E+02	4.97E+01	4.61E+02	1.14E+02	7.26E+01	2.42E-07	2.72E+02	**0.00E+00**	7.86E+01	**0.00E+00**	2.77E+01
	std	**0.00E+00**	1.33E+01	4.10E+01	1.22E+01	1.31E+01	1.50E-14	9.78E+00	**0.00E+00**	3.49E+01	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**
Rothyp	mean	**0.00E+00**	3.99E-01	1.61E+04	4.88E-02	4.10E+05	7.05E+03	2.63E-09	4.40E-16	2.78E+05	3.28E-313	1.88E-18	1.19E-122	3.93E-65
	min	**0.00E+00**	2.05E-11	3.20E-13	2.50E-02	4.06E+05	7.05E+03	2.36E-10	4.40E-16	1.95E+05	3.28E-313	1.88E-18	2.76E-123	3.93E-65
	max	**0.00E+00**	1.07E+01	8.87E+04	9.31E-02	4.50E+05	7.05E+03	5.98E-09	4.40E-16	3.41E+05	3.28E-313	1.88E-18	1.29E-122	3.93E-65
	std	**0.00E+00**	1.95E+00	2.72E+04	2.07E-02	1.40E+04	**0.00E+00**	1.50E-09	5.20E-32	4.19E+04	**0.00E+00**	4.06E-34	3.20E-123	8.89E-81
Rosenbrock	mean	**5.11E-04**	1.06E+02	8.70E+04	2.81E+01	1.91E+06	3.41E+02	1.73E+01	2.11E+00	9.47E+01	2.61E+01	2.48E+01	2.62E+01	2.67E+01
	min	**1.73E-09**	5.46E+00	4.22E+00	2.73E+01	1.86E+06	3.41E+02	1.86E-01	2.11E+00	2.29E+01	2.61E+01	2.48E+01	2.62E+01	2.67E+01
	max	**3.00E-03**	1.00E+03	2.91E+05	2.92E+01	1.92E+06	3.41E+02	7.50E+01	2.11E+00	4.61E+02	2.61E+01	2.48E+01	2.62E+01	2.67E+01
	std	7.76E-04	1.81E+02	8.28E+04	7.27E-01	1.74E+04	**0.00E+00**	1.76E+01	4.68E-16	9.69E+01	3.74E-15	3.74E-15	1.51E-02	3.74E-15
Sphere	mean	**0.00E+00**	2.06E-06	5.24E+00	3.39E-06	1.93E+02	1.90E+00	1.10E-12	6.88E-15	6.36E-54	2.59E-319	1.02E-19	2.62E-124	1.23E-69
	min	**0.00E+00**	1.41E-17	6.34E-18	2.03E-06	1.71E+02	1.90E+00	9.09E-14	6.88E-15	4.24E-54	2.59E-319	1.02E-19	2.32E-127	1.23E-69
	max	**0.00E+00**	2.91E-05	2.62E+01	4.93E-06	1.95E+02	1.90E+00	2.94E-12	6.88E-15	8.57E-54	2.59E-319	1.02E-19	2.91E-124	1.23E-69
	std	**0.00E+00**	6.90E-06	1.11E+01	8.35E-07	7.63E+00	**0.00E+00**	7.88E-13	8.32E-31	1.19E-54	**0.00E+00**	1.27E-35	9.21E-125	2.71E-85
Sumpow	mean	**0.00E+00**	1.72E-84	3.82E-35	2.78E-08	9.04E-01	3.41E-06	4.18E-24	3.57E-17	2.74E-07	**0.00E+00**	6.43E-14	**0.00E+00**	8.35E-214
	min	**0.00E+00**	2.79E-113	1.12E-42	1.46E-08	8.60E-01	3.41E-06	3.17E-34	3.57E-17	8.85E-08	**0.00E+00**	6.43E-14	**0.00E+00**	8.35E-214
	max	**0.00E+00**	5.10E-83	2.47E-34	5.26E-08	1.30E+00	3.41E-06	1.24E-22	3.57E-17	5.53E-07	**0.00E+00**	6.43E-14	**0.00E+00**	8.35E-214
	std	**0.00E+00**	9.31E-84	7.99E-35	1.17E-08	1.39E-01	4.46E-22	2.26E-23	**0.00E+00**	9.42E-08	**0.00E+00**	1.33E-29	**0.00E+00**	**0.00E+00**
Zakharov	mean	**0.00E+00**	1.40E+02	2.38E+02	3.13E+01	7.45E+02	3.25E+02	7.12E+01	8.04E-08	4.52E+02	**0.00E+00**	1.32E+02	5.12E-14	4.86E+00
	min	**0.00E+00**	9.55E+01	1.64E+02	1.99E+01	6.97E+02	3.25E+02	5.64E+01	8.04E-08	2.23E+02	**0.00E+00**	1.32E+02	**0.00E+00**	4.86E+00
	max	**0.00E+00**	1.86E+02	2.85E+02	6.47E+01	7.50E+02	3.25E+02	8.96E+01	8.04E-08	6.17E+02	**0.00E+00**	1.32E+02	5.68E-14	4.86E+00
	std	**0.00E+00**	2.84E+01	4.28E+01	1.34E+01	1.67E+01	5.99E-14	8.76E+00	**0.00E+00**	9.53E+01	**0.00E+00**	3.00E-14	1.80E-14	**0.00E+00**
Sumsqu	mean	**0.00E+00**	4.21E-05	9.70E+01	9.63E-03	2.62E+03	9.93E+00	1.77E-11	3.99E-12	7.60E-52	1.08E-321	3.43E-18	2.72E-125	1.32E-67
	min	**0.00E+00**	3.09E-15	1.02E-15	2.59E-04	2.60E+03	9.93E+00	4.61E-12	3.99E-12	3.93E-52	1.08E-321	3.43E-18	2.47E-125	1.32E-67
	max	**0.00E+00**	5.84E-04	3.15E+02	6.00E-02	2.75E+03	9.93E+00	5.73E-11	3.99E-12	1.24E-51	1.08E-321	3.43E-18	2.74E-125	1.32E-67
	std	**0.00E+00**	1.19E-04	1.03E+02	1.83E-02	4.63E+01	1.87E-15	1.36E-11	8.51E-28	2.27E-52	**0.00E+00**	4.06E-34	8.58E-127	1.74E-83
Powell	mean	**1.19E-270**	1.47E-01	6.37E+02	2.42E-01	1.03E+04	2.04E+01	1.08E-03	1.18E-02	2.16E-01	3.61E-27	1.78E-04	2.17E-08	4.70E-05
	min	**0.00E+00**	2.04E-03	2.23E-03	1.12E-01	1.03E+04	2.04E+01	1.48E-04	1.18E-02	3.23E-02	3.61E-27	1.78E-04	1.44E-08	4.70E-05
	max	**3.57E-269**	2.27E+00	3.57E+03	4.19E-01	1.08E+04	2.04E+01	5.06E-03	1.18E-02	8.14E-01	3.61E-27	1.78E-04	8.68E-08	4.70E-05
	std	**0.00E+00**	4.31E-01	1.14E+03	1.01E-01	1.53E+02	3.74E-15	9.92E-04	**0.00E+00**	2.07E-01	**0.00E+00**	**0.00E+00**	2.29E-08	7.14E-21

To further ascertain the results, the Wilcoxon rank sum test is conducted. Table 19 depicts the detailed statistical results. Since nearly all the *p*-values are lower than 0.05, the proposed model is statistically better than all baseline methods in nearly all the test cases. The exceptions are for Ackley and Dixon-Price, where ALPSO and ABC outperform the proposed model statistically. The proposed model and amixedGWO also show similar result distributions for Ackley. In addition, since the proposed model and some of the baseline methods, i.e. ALPSO, WOA, GWO and amixedGWO, achieve the global optimal solution (i.e. 0) for several test functions such as Griewank (ALPSO, WOA, GWO, and amixedGWO), Rastrigin (WOA and GWO), Sum of Different Powers (WOA and GWO) and Zakharov (WOA), our results achieve similar distributions to those of these methods in the respective test cases.

Table 19 The Wilcoxon rank sum test results for the test functions

	PSOVA	**SPSO**	**MPSO**	**ALPSO**	**GPSO**	**DNLPSO**	**ELPSO**	**CLFA**	**CGFA**	**VSSFA**	**MFA**	**NaFA**
Ackley	1.13E-11	1.21E-12	1.21E-12	1.69E-14	1.13E-11	1.13E-11	1.13E-11	1.13E-11	1.13E-11	1.13E-11	3.92E-12	1.13E-11
Dixon	3.12E-12	3.02E-11	3.02E-11	3.02E-11	3.13E-12	3.13E-12	3.13E-12	3.13E-12	3.12E-12	3.13E-12	1.07E-12	3.13E-12
Griewank	1.20E-12	1.21E-12	1.21E-12	1.00E+00	1.19E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	3.37E-13	1.20E-12
Rastrigin	1.20E-12	1.21E-12	1.21E-12	1.21E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	2.01E-13	1.20E-12
Rothyp	1.20E-12	2.37E-12	2.37E-12	2.37E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	3.50E-13	1.20E-12
Rosenbrock	3.00E-11	3.02E-11	3.02E-11	3.02E-11	3.00E-11	3.00E-11	3.00E-11	3.00E-11	3.00E-11	3.00E-11	1.17E-11	3.00E-11
Sphere	1.20E-12	2.37E-12	2.37E-12	1.20E-12	1.19E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	3.50E-13	1.20E-12
Sumpow	3.12E-12	1.21E-12	1.21E-12	3.12E-12	3.14E-12	3.14E-12	3.14E-12	3.13E-12	3.14E-12	3.13E-12	2.10E-12	3.14E-12
Zakharov	1.20E-12	1.21E-12	1.21E-12	1.21E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	7.67E-13	1.20E-12
Sumsqu	1.20E-12	1.72E-12	1.72E-12	1.72E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	1.20E-12	3.50E-13	1.20E-12
Powell	1.71E-12	1.72E-12	1.72E-12	1.72E-12	1.71E-12	1.70E-12	1.70E-12	1.71E-12	1.71E-12	1.70E-12	4.29E-14	1.70E-12

	PSO	**MFO**	**FA**	**ALO**	**DA**	**CS**	**ABC**	**BA**	**WOA**	**PSOGSA**	**GWO**	**amixedGWO**
Ackley	1.21E-12	1.12E-11	1.07E-11	3.92E-12	3.92E-12	1.21E-12	3.92E-12	1.21E-12	3.92E-12	3.92E-12	5.28E-12	1.00E+00
Dixon	3.02E-11	3.12E-12	3.13E-12	1.34E-12	9.62E-13	3.02E-11	5.28E-12	3.02E-11	5.28E-12	5.28E-12	7.06E-12	5.28E-12
Griewank	1.21E-12	1.20E-12	1.20E-12	4.82E-13	3.37E-13	1.21E-12	3.37E-13	1.21E-12	1.00E+00	3.37E-13	1.00E+00	1.00E+00
Rastrigin	1.21E-12	1.20E-12	1.20E-12	4.82E-13	3.37E-13	1.21E-12	3.37E-13	1.21E-12	1.00E+00	3.37E-13	1.00E+00	3.37E-13
Rothyp	1.21E-12	1.14E-12	1.20E-12	4.81E-13	3.37E-13	1.72E-12	3.37E-13	2.37E-12	1.69E-14	3.37E-13	4.82E-13	3.37E-13
Rosenbrock	3.02E-11	3.00E-11	3.00E-11	1.49E-11	1.14E-11	3.02E-11	1.14E-11	3.02E-11	1.14E-11	1.14E-11	1.49E-11	1.14E-11
Sphere	1.72E-12	1.17E-12	1.20E-12	4.82E-13	3.37E-13	1.72E-12	3.37E-13	3.16E-12	1.69E-14	3.37E-13	4.82E-13	3.37E-13
Sumpow	1.21E-12	3.14E-12	3.14E-12	1.35E-12	9.65E-13	2.37E-12	9.65E-13	2.37E-12	1.00E+00	9.65E-13	1.00E+00	2.75E-11
Zakharov	1.21E-12	1.20E-12	1.20E-12	6.13E-14	3.37E-13	1.21E-12	3.37E-13	1.21E-12	1.00E+00	3.37E-13	2.53E-11	3.37E-13
Sumsqu	1.21E-12	1.19E-12	1.20E-12	4.82E-13	3.37E-13	1.21E-12	3.37E-13	1.72E-12	1.69E-14	1.69E-14	4.82E-13	3.37E-13
Powell	1.72E-12	1.71E-12	1.71E-12	2.71E-14	4.95E-13	1.72E-12	8.84E-13	1.21E-12	8.84E-13	8.84E-13	1.26E-12	8.84E-13

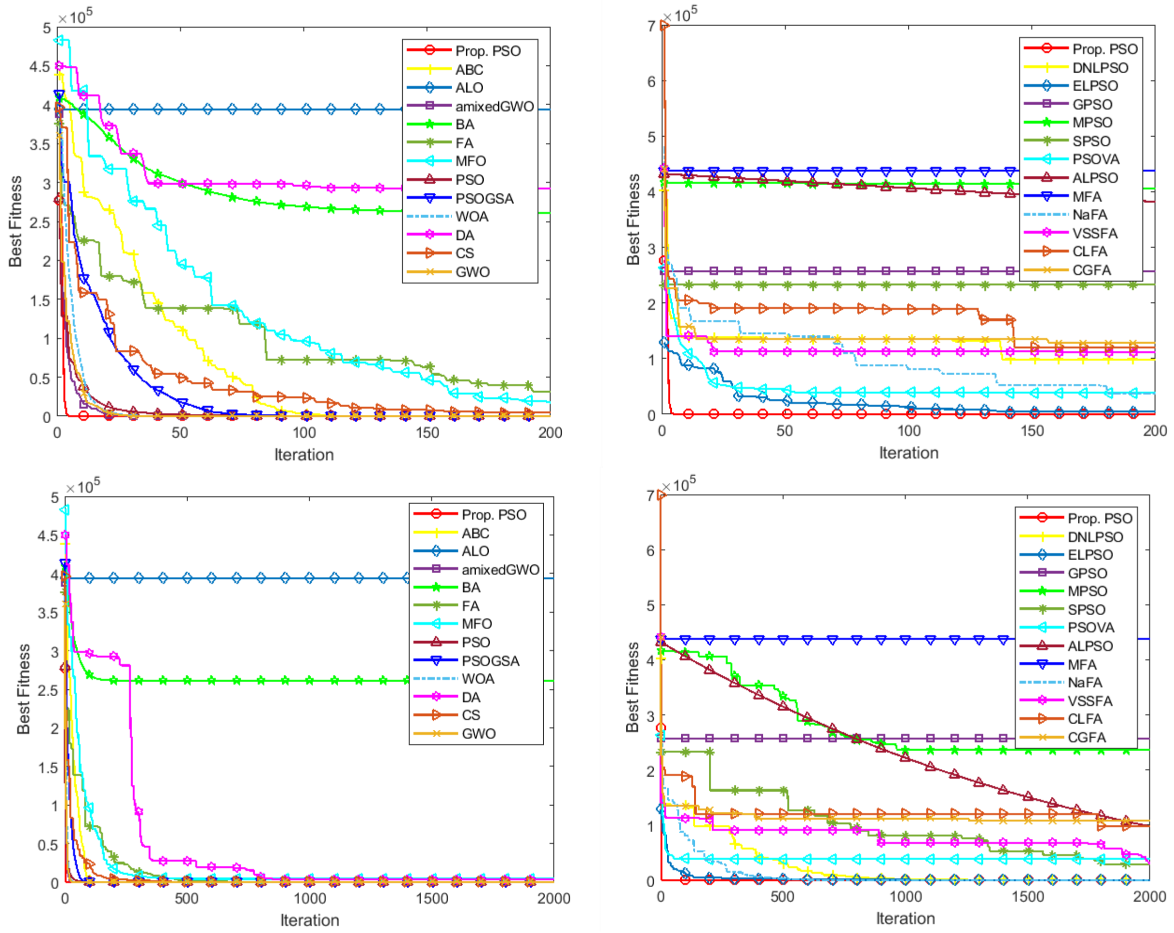
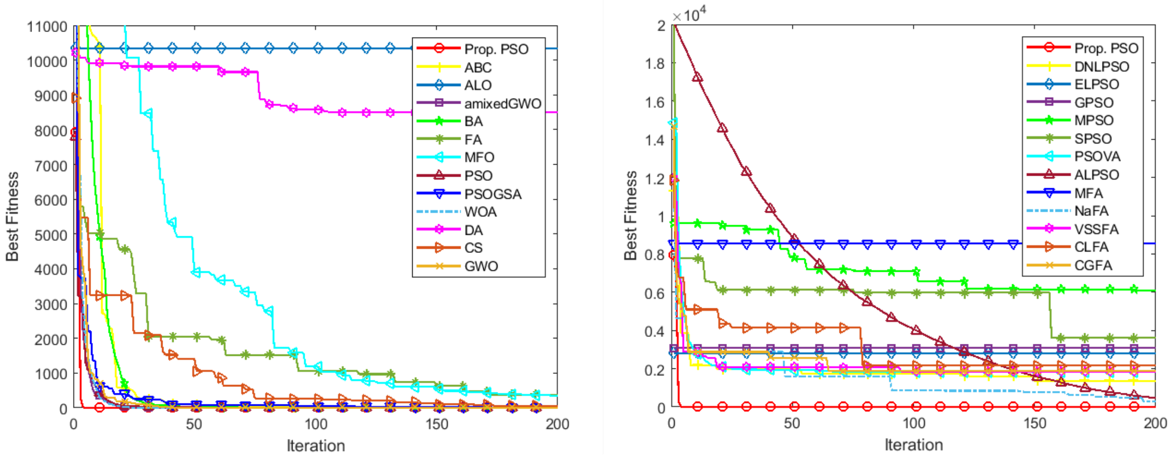


Figure 9 Convergence curves of the Rotated Hyper-Ellipsoid function over a set of 30 runs (Top row – 200 iterations, bottom row – 2000 iterations)

Figure 9 illustrates the convergence curves for both 200 (top row) and 2000 (bottom row) iterations pertaining to the Rotated Hyper-Ellipsoid function over a set of 30 runs. The proposed model depicts significantly faster convergence rates in comparison with those of the PSO variants and other search methods. It is able to achieve global minimum solution, i.e. 0, with a comparatively smaller number of iterations. As indicated in Figure 9 (top row), among the baseline methods, amixedGWO, GWO, WOA and PSO illustrate drastically faster convergence rates at the beginning of the search process, which are followed by those of PSOGSA, CS and ABC. In addition, ELPSO, PSOVA, NaFA and DNLPSO show faster convergence rates among the PSO and FA variants. Moreover, as indicated in Figure 9 (bottom row), as the iteration escalates, DA, MFO, FA, VSSFA, SPSO and ALPSO produce significant improvements.



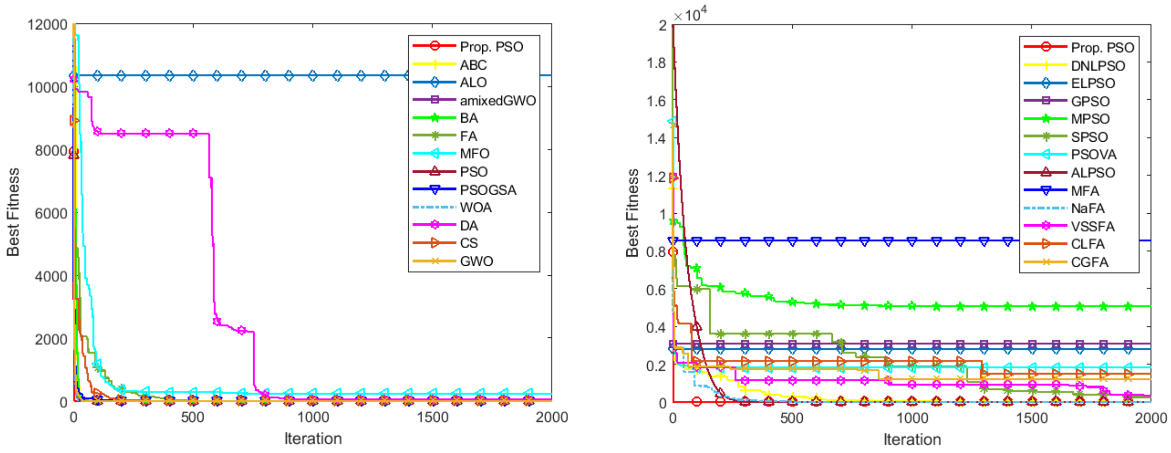


Figure 10 Convergence curves of the Powell function over a set of 30 runs (Top row – 200 iterations, bottom row – 2000 iterations)

In addition, the convergence curves of all search methods for both 200 (top row) and 2000 (bottom row) iterations pertaining to the multimodal function, i.e., Powell, over a set of 30 runs, are provided in Figure 10. Again, the proposed model yields the fastest convergence rates in comparison with those of all baseline methods throughout the search process. As indicated in Figure 10 (top row), among the baseline methods, WOA, GWO, amixedGWO, PSOGSA and PSO produce faster convergence speeds. This is followed by ABC and BA. NaFA and ALPSO show better convergence rates in comparison with those of other PSO and FA variants. As the search progresses, as shown in Figure 10 (bottom row), CS, FA, DA, DNLPSO, SPSO and VSSFA converge faster and produce significantly enhanced performances.

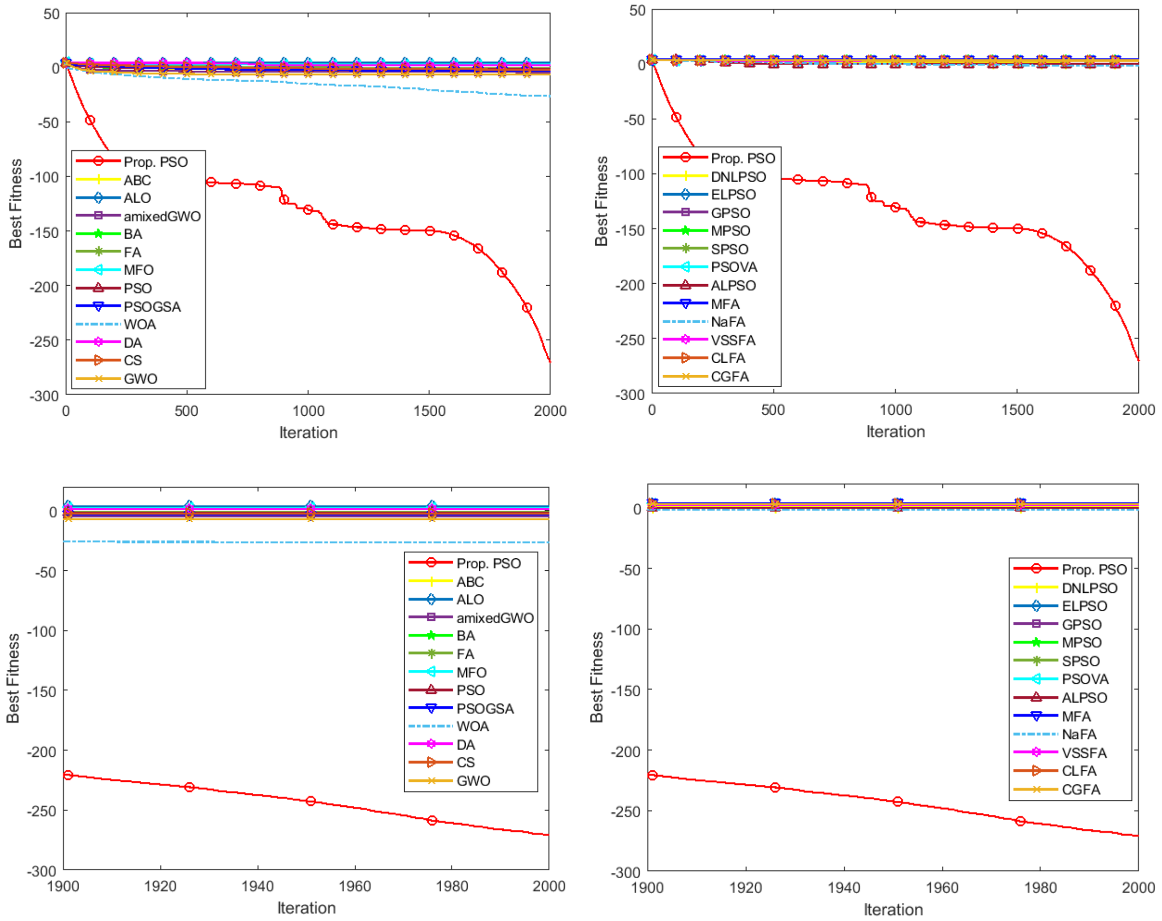


Figure 11 Convergence curves of the Powell function in the log scale over a set of 30 runs (Top row – 2000 iterations, bottom row – the last 100 iterations)

To further determine efficiency of the proposed PSO algorithm, we convert the convergence curves of the Powell function to the log scale with a base of 10. As illustrated in Figure 11, the top row shows the convergence curves for the overall 2000 iterations and the bottom row illustrates the last 100 iterations. The plots in Figure 11 depict a significantly faster convergence speed against those of all compared methods, therefore further ascertaining efficiency of the proposed algorithm. WOA and GWO show the fastest convergence rates among the baseline methods, while NaFA and ALPSO converge comparatively faster in comparison with those of other PSO and FA variants.

Moreover, the empirical results indicate that convergence curves of the proposed model for other test functions show similar characteristics and depict the fastest convergence rates in most test cases.

5.5 Evaluation Using Other Benchmark Functions

To further test model efficiency, we employ another suite of benchmark test functions widely adopted in existing studies [22, 65, 75-77] for performance comparison. It includes seven unimodal (F_1-F_7) and six multimodal (F_8-F_{13}) functions, as defined in Tables 20-21, respectively. These benchmark functions represent a variety of challenging mathematical artificial landscapes with varied difficulties.

Table 20 Unimodal benchmark functions used in existing studies [22, 65, 75-77]

Function	Dimension	Range	f_{min}
$F_1(x) = \sum_{i=1}^{n} x_i^2$	30	[-100, 100]	0
$F_2(x) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$	30	[-10, 10]	0
$F_3(x) = \sum_{i=1}^{n} (\sum_{j=1}^{i} x_j)^2$	30	[-100, 100]	0
$F_4(x) = \max_i\{\lvert x_i \rvert, 1 \le i \le n\}$	30	[-100, 100]	0
$F_5(x) = \sum_{i=1}^{n-1}[100\,(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	30	[-30, 30]	0
$F_6(x) = \sum_{i=1}^{n} ([x_i + 0.5])^2$	30	[-100, 100]	0
$F_7(x) = \sum_{i=1}^{n} i x_i^4 + random[0,1)$	30	[-1.28, 1.28]	0

Table 21 Multimodal benchmark functions used in existing studies [22, 65, 75-77]

Function	Dimension	Range	f_{min}
$F_8(x) = \sum_{i=1}^{n} -x_i sin(\sqrt{\lvert x_i \rvert})$	30	[-500, 500]	$-418.9829 \times Dim$
$F_9(x) = \sum_{i=1}^{n} [x_i^2 - 10\cos(2\pi x_i) + 10]$	30	[-5.12, 5.12]	0
$F_{10}(x) = -20\, exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_{\mathrm{i}})\right) + 20 + e$	30	[-32, 32]	0
$F_{11}(x) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	30	[-600, 600]	0
$F_{12}(x) = \frac{\pi}{n}\{10sin(\pi y_1) + \sum_{i=1}^{n-1}(y_i - 1)^2[1 + 10sin^2(\pi y_{i+1})] + (y_n - 1)^2\} + \sum_{i=1}^{n} u(x_i, 10, 100, 4)$ $y_i = 1 + \frac{x_i + 1}{4}$	30	[-50, 50]	0

$u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 & -a < x_i < a \\ k(-x_i - a)^m & x_i < -a \end{cases}$			
$F_{13}(x) = 0.1\{sin^2(3\pi x_1) + \sum_{i=1}^{n}(x_i - 1)^2[1 + sin^2(3\pi x_i + 1)] + (x_n - 1)^2[1 + sin^2(2\pi x_n)]\} + \sum_{i=1}^{n} u(x_i, 5, 100, 4)$	30	[-50, 50]	0

Owing to complexities of the test functions and model convergence rates, a comparatively larger experimental setting is used, i.e. maximum number of function evaluations=population (30) × iterations (3000), with dimension=30. All the search methods terminate when the maximum number of function evaluations is reached. The aforementioned 24 baseline search methods are employed for performance comparison. A set of 30 trials is conducted for each search method to eliminate bias and facilitate a statistical hypothesis test in performance comparison. Table 22 illustrates the detailed evaluation results. As shown in Table 22, the proposed PSO algorithm outperforms classical search methods and advanced PSO and FA variants in most test cases. The exceptions are F_7 and F_{10}, where ALPSO outperforms the proposed model, as well as F_6, F_{12} and F_{13}, where MFO, ABC and PSOGSA obtain the best results, respectively. ABC and PSOGSA also show better performances for F_6. Moreover, WOA, GWO and amixedGWO also illustrate impressive performances among the baseline methods. As an example, the proposed model, GWO and WOA achieve the global minimum solution (i.e. 0) for multimodal functions of F_9 and F_{11}. ABC also attains global optimum solution for F_9. Besides the above, the proposed model and WOA obtain the best minimum solution for F_1, while the proposed model and ALPSO attain the global optimum solution for F_{11}.

Table 23 shows the statistical test results for all test functions over a set of 30 runs. Since most of the *p*-values are lower than 0.05, they indicate that the proposed model outperforms baseline methods statistically in most test cases. The exceptions are as follows. ALPSO outperforms the proposed model statistically for F_7 and F_{10}. MFO, ABC and PSOGSA show statistical better results than those of the proposed model for F_6, while ABC and PSOGSA obtain significantly better result distributions for F_{13}. In addition, ABC achieves statistically better results for F_{12}. Moreover, our model and some baseline methods, i.e. ALPSO, ABC, WOA, GWO and amixedGWO, achieve the global optimum solution (i.e. 0) or the same results for several test functions such as F_1 (WOA), F_9 (ABC, WOA and GWO), F_{10} (amixedGWO) and F_{11} (ALPSO, WOA and GWO), similar distributions are produced by the proposed model and these baseline methods in such test cases. In short, for all test functions F_1-F_{13}, the proposed model shows statistically significant superiority over all baseline methods in most test cases.

Table 22 Evaluation results for the benchmark functions F_1-F_{13} with dimension=30 over a set of 30 runs

		Prop. PSO	PSOVA	SPSO	MPSO	ALPSO	GPSO	DNLPSO	ELPSO	CLFA	CGFA	VSSFA	MFA	NaFA
F_1	mean	**0.00E+00**	7.65E+03	5.06E+03	1.95E+04	1.92E+04	4.19E+04	1.00E+00	1.93E+04	1.53E+04	1.48E+04	5.92E+03	6.75E+04	7.76E-04
	min	**0.00E+00**	6.47E+03	3.77E+03	2.26E+03	1.05E+04	4.18E+04	5.54E-02	1.47E+04	1.52E+04	1.48E+04	5.13E+03	6.75E+04	5.25E-04
	max	**0.00E+00**	1.82E+04	5.88E+03	6.16E+04	2.71E+04	4.26E+04	9.55E+00	1.98E+04	1.65E+04	1.55E+04	6.01E+03	6.75E+04	8.04E-04
	std	**0.00E+00**	3.71E+03	7.82E+02	1.95E+04	4.66E+03	2.40E+02	3.00E+00	1.63E+03	4.11E+02	2.40E+02	2.78E+02	1.53E-11	8.82E-05
F_2	mean	**1.30E-322**	1.12E+02	3.08E+01	9.08E+01	7.40E+00	5.47E+05	2.19E-01	5.33E+01	6.59E+01	6.00E+01	2.78E+01	1.52E+14	1.05E-02
	min	1.20E-322	7.31E+01	2.88E+01	3.79E+01	**0.00E+00**	1.49E+05	1.66E-01	4.82E+01	6.35E+01	5.91E+01	2.70E+01	1.52E+14	1.04E-02
	max	**2.37E-322**	1.16E+02	3.34E+01	1.47E+02	1.89E+01	4.13E+06	6.94E-01	5.39E+01	8.72E+01	6.74E+01	3.50E+01	1.52E+14	1.07E-02
	std	**0.00E+00**	1.37E+01	1.74E+00	3.08E+01	8.10E+00	1.26E+06	1.67E-01	1.79E+00	7.47E+00	2.60E+00	2.56E+00	2.31E-01	9.25E-05
F_3	mean	**0.00E+00**	6.70E+04	1.05E+04	7.93E+04	1.80E+04	7.39E+04	6.78E+02	5.23E+04	2.28E+04	2.37E+04	1.53E+04	2.22E+05	7.40E-03
	min	**0.00E+00**	6.70E+04	7.27E+03	5.27E+04	1.10E+04	7.39E+04	7.38E+00	5.23E+04	2.25E+04	2.37E+04	1.31E+04	2.22E+05	7.40E-03
	max	**0.00E+00**	6.70E+04	1.51E+04	1.54E+05	2.34E+04	7.39E+04	7.52E+02	5.23E+04	2.54E+04	2.37E+04	1.56E+04	2.22E+05	7.40E-03
	std	**0.00E+00**	1.53E-11	2.35E+03	3.07E+04	3.73E+03	1.53E-11	2.36E+02	7.67E-12	9.18E+02	**0.00E+00**	8.04E+02	**0.00E+00**	1.83E-18
F_4	mean	**2.00E-323**	7.18E+01	2.89E+01	4.51E+01	5.36E-296	8.42E+01	3.12E-01	5.11E+01	7.89E+01	5.01E+01	2.47E+01	8.38E+01	1.29E-02
	min	1.00E-323	7.17E+01	2.71E+01	3.75E+01	**0.00E+00**	7.73E+01	1.21E-01	5.04E+01	6.94E+01	4.39E+01	2.43E+01	8.38E+01	1.25E-02
	max	**2.00E-323**	7.32E+01	3.10E+01	5.65E+01	5.36E-295	8.49E+01	2.03E+00	5.75E+01	8.00E+01	5.08E+01	2.81E+01	8.38E+01	1.61E-02
	std	**0.00E+00**	4.89E-01	1.20E+00	5.78E+00	**0.00E+00**	2.41E+00	6.03E-01	2.23E+00	3.35E+00	2.19E+00	1.19E+00	**0.00E+00**	1.13E-03
F_5	mean	**7.37E-04**	7.29E+07	1.77E+06	7.33E+07	1.17E+02	9.74E+07	4.34E+02	2.90E+07	1.53E+07	1.75E+07	1.97E+06	2.19E+08	2.83E+01
	min	**1.47E-08**	1.29E+07	1.37E+06	1.10E+06	4.89E+00	6.10E+07	4.24E+02	1.82E+07	1.52E+07	1.67E+07	1.75E+06	2.19E+08	2.64E+01
	max	**8.18E-04**	7.95E+07	2.68E+06	1.61E+08	2.09E+02	1.01E+08	5.23E+02	3.02E+07	1.66E+07	1.76E+07	1.99E+06	2.19E+08	2.85E+01
	std	2.59E-04	2.11E+07	3.68E+05	5.87E+07	6.35E+01	1.28E+07	3.14E+01	3.79E+06	4.57E+05	2.66E+05	7.58E+04	**0.00E+00**	6.71E-01
F_6	mean	**6.28E-06**	2.86E+04	4.94E+03	2.67E+04	1.79E+04	2.00E+04	3.79E+03	1.64E+03	1.86E+04	1.57E+04	5.32E+03	6.79E+04	6.14E-04
	min	**7.69E-08**	1.93E+04	3.58E+03	1.22E+04	1.17E+04	1.92E+04	3.78E-03	2.30E+02	1.69E+04	1.56E+04	5.31E+03	6.79E+04	5.98E-04
	max	**1.98E-05**	2.97E+04	6.31E+03	4.31E+04	2.36E+04	2.64E+04	4.21E+03	1.43E+04	1.87E+04	1.65E+04	5.43E+03	6.79E+04	7.54E-04
	std	6.86E-06	3.27E+03	9.33E+02	9.50E+03	3.54E+03	2.27E+03	1.33E+03	4.45E+03	5.82E+02	2.75E+02	3.83E+01	**1.53E-11**	4.92E-05
F_7	mean	9.08E-06	8.69E+00	1.04E+00	1.98E+01	**0.00E+00**	6.43E+01	9.28E-03	1.38E+01	5.41E+00	7.95E+00	1.60E+00	1.18E+02	2.53E-02
	min	8.45E-06	7.48E+00	7.00E-01	6.93E-01	**0.00E+00**	3.28E+01	7.57E-03	1.37E+01	5.28E+00	6.59E+00	1.06E+00	1.18E+02	1.59E-02

	max	1.47E-05	1.95E+01	1.47E+00	6.86E+01	**0.00E+00**	6.78E+01	9.47E-03	1.46E+01	6.60E+00	8.10E+00	1.66E+00	1.18E+02	1.10E-01
	std	1.99E-06	3.81E+00	2.05E-01	1.98E+01	**0.00E+00**	1.11E+01	6.00E-04	2.83E-01	4.19E-01	4.78E-01	1.91E-01	7.72E-03	2.97E-02
F_8	mean	**-6.48E+09**	-5.73E+03	-7.66E+03	-5.35E+03	-5.57E+03	-6.53E+03	-2.17E+04	-4.71E+03	-3.99E+03	-3.80E+03	-3.79E+03	-1.59E+03	-7.20E+03
	min	**-6.47E+10**	-6.73E+03	-8.46E+03	-6.00E+03	-6.36E+03	-6.72E+03	-2.32E+04	-4.78E+03	-4.04E+03	-3.95E+03	-3.81E+03	-1.59E+03	-7.40E+03
	max	-1.11E+04	-5.62E+03	-6.76E+03	-4.20E+03	-4.78E+03	-6.51E+03	**-2.16E+04**	-4.15E+03	-3.98E+03	-3.78E+03	-3.62E+03	-1.59E+03	-5.46E+03
	std	2.05E+10	3.49E+02	6.11E+02	5.94E+02	4.98E+02	6.64E+01	5.08E+02	1.99E+02	1.69E+01	5.43E+01	6.08E+01	**2.40E-13**	6.13E+02
F_9	mean	**0.00E+00**	2.63E+02	2.18E+02	3.15E+02	1.88E+02	2.91E+02	5.00E+01	2.43E+02	2.43E+02	2.48E+02	2.23E+02	4.29E+02	4.58E+01
	min	**0.00E+00**	2.63E+02	2.18E+02	3.15E+02	1.59E+02	2.91E+02	5.00E+01	2.43E+02	2.43E+02	2.48E+02	2.23E+02	4.29E+02	4.58E+01
	max	**0.00E+00**	2.63E+02	2.18E+02	3.15E+02	2.26E+02	2.91E+02	5.00E+01	2.43E+02	2.43E+02	2.48E+02	2.23E+02	4.29E+02	4.58E+01
	std	**0.00E+00**	5.99E-14	3.00E-14	5.99E-14	2.00E+01	5.99E-14	7.49E-15	**0.00E+00**	3.00E-14	**0.00E+00**	5.99E-14	5.99E-14	7.49E-15
F_{10}	mean	8.88E-16	2.01E+01	1.29E+01	1.75E+01	**0.00E+00**	1.94E+01	1.98E-01	1.69E+01	1.62E+01	1.65E+01	1.32E+01	2.06E+01	6.43E-03
	min	8.88E-16	1.90E+01	1.21E+01	9.90E+00	**0.00E+00**	1.94E+01	1.20E-01	1.64E+01	1.61E+01	1.65E+01	1.31E+01	2.06E+01	6.41E-03
	max	8.88E-16	2.02E+01	1.40E+01	2.01E+01	**0.00E+00**	1.99E+01	2.07E-01	1.70E+01	1.72E+01	1.72E+01	1.32E+01	2.06E+01	6.43E-03
	std	**0.00E+00**	3.78E-01	6.56E-01	4.01E+00	**0.00E+00**	1.77E-01	2.76E-02	1.73E-01	3.33E-01	2.20E-01	2.31E-02	**0.00E+00**	9.27E-06
F_{11}	mean	**0.00E+00**	8.52E+01	4.84E+01	1.90E+02	**0.00E+00**	3.03E+02	1.34E-01	1.82E+02	1.40E+02	1.49E+02	4.36E+01	6.08E+02	1.14E-03
	min	**0.00E+00**	7.39E+01	3.39E+01	3.24E+01	**0.00E+00**	2.61E+02	7.34E-02	1.37E+02	1.36E+02	1.40E+02	4.23E+01	6.08E+02	1.12E-03
	max	**0.00E+00**	1.87E+02	5.91E+01	3.06E+02	**0.00E+00**	3.08E+02	6.80E-01	1.87E+02	1.70E+02	1.50E+02	5.58E+01	6.08E+02	1.33E-03
	std	**0.00E+00**	3.58E+01	6.54E+00	1.05E+02	**0.00E+00**	1.49E+01	1.92E-01	1.59E+01	1.06E+01	3.31E+00	4.27E+00	1.20E-13	6.72E-05
F_{12}	mean	**5.46E-07**	4.34E+07	4.22E+04	1.03E+08	9.32E+03	1.11E+08	1.54E-01	3.52E+07	1.38E+07	9.41E+06	1.30E+04	4.25E+08	1.65E-06
	min	3.81E-10	3.90E+07	1.46E+03	4.45E+03	**0.00E+00**	7.18E+07	2.31E-04	3.37E+07	1.04E+07	9.14E+06	8.44E+03	4.25E+08	1.25E-06
	max	1.87E-06	8.35E+07	1.63E+05	5.12E+08	9.32E+04	1.15E+08	1.53E+00	3.53E+07	1.42E+07	1.18E+07	1.35E+04	4.25E+08	**1.69E-06**
	std	5.66E-07	1.41E+07	5.06E+04	1.79E+08	2.95E+04	1.37E+07	4.85E-01	5.16E+05	1.18E+06	8.42E+05	1.59E+03	**0.00E+00**	1.40E-07
F_{13}	mean	**1.44E-06**	3.68E+07	2.24E+06	2.47E+08	3.12E+05	4.85E+08	1.14E+04	6.41E+07	6.17E+07	3.78E+07	1.51E+06	9.10E+08	2.91E-05
	min	1.51E-07	3.25E+07	7.97E+05	5.45E+05	**0.00E+00**	4.13E+08	2.58E+00	5.51E+07	6.04E+07	3.70E+07	1.17E+06	9.10E+08	2.83E-05
	max	**1.58E-06**	3.73E+07	4.76E+06	8.23E+08	8.72E+05	4.93E+08	1.14E+05	1.45E+08	7.31E+07	3.79E+07	1.55E+06	9.10E+08	3.64E-05
	std	4.52E-07	1.51E+06	1.14E+06	3.46E+08	2.77E+05	2.53E+07	3.60E+04	2.86E+07	4.02E+06	2.76E+05	1.18E+05	**1.19E-07**	2.57E-06

		Prop. PSO	**PSO**	**MFO**	**FA**	**ALO**	**DA**	**CS**	**ABC**	**BA**	**WOA**	**PSOGSA**	**GWO**	**amixedGWO**
F_1	mean	**0.00E+00**	3.15E-01	2.12E-21	9.59E-04	6.96E+04	5.68E+02	1.08E-06	9.19E-16	5.49E+04	**0.00E+00**	1.52E-19	6.24E-185	3.23E-100
	min	**0.00E+00**	2.30E-01	2.10E-22	9.59E-04	5.50E+04	1.07E+02	3.56E-07	9.19E-16	4.77E+04	**0.00E+00**	1.52E-19	9.04E-190	4.19E-102
	max	**0.00E+00**	1.08E+00	2.33E-21	9.59E-04	7.69E+04	1.83E+03	3.85E-06	9.19E-16	6.16E+04	**0.00E+00**	1.52E-19	5.63E-184	1.54E-99
	std	**0.00E+00**	2.68E-01	6.70E-22	2.29E-19	6.51E+03	5.40E+02	7.22E-07	**0.00E+00**	4.67E+03	**0.00E+00**	2.54E-35	**0.00E+00**	6.34E-100
F_2	mean	**1.30E-322**	7.52E+00	5.60E+01	1.98E-02	2.62E+12	1.61E+01	7.15E-03	1.38E-10	1.23E+02	6.48E-317	1.49E-09	8.22E-107	6.33E-56
	min	1.20E-322	3.66E+00	2.00E+01	1.98E-02	3.51E+06	**0.00E+00**	2.26E-03	1.38E-10	9.76E+01	4.7E-322	1.49E-09	2.92E-108	1.24E-57
	max	**2.37E-322**	7.95E+00	6.00E+01	1.98E-02	7.54E+12	4.07E+01	1.89E-02	1.38E-10	1.48E+02	6.48E-316	1.49E-09	2.59E-106	1.46E-55
	std	**0.00E+00**	1.35E+00	1.26E+01	**0.00E+00**	2.92E+12	1.33E+01	3.93E-03	2.72E-26	1.92E+01	**0.00E+00**	2.18E-25	7.47E-107	4.67E-56
F_3	mean	**0.00E+00**	8.04E+02	1.50E+04	3.71E-01	2.14E+05	1.65E+04	1.23E+00	4.33E+03	8.93E+04	2.58E+02	1.50E+04	3.35E-61	1.27E-28
	min	**0.00E+00**	8.04E+02	1.50E+04	3.71E-01	1.44E+05	1.65E+04	1.23E+00	4.33E+03	8.93E+04	2.58E+02	1.50E+04	3.35E-61	6.48E-41
	max	**0.00E+00**	8.04E+02	1.50E+04	3.71E-01	2.21E+05	1.65E+04	1.23E+00	4.33E+03	8.93E+04	2.58E+02	1.50E+04	3.35E-61	8.49E-28
	std	**0.00E+00**	**0.00E+00**	1.92E-12	5.85E-17	2.46E+04	3.83E-12	2.34E-16	9.59E-13	**0.00E+00**	**0.00E+00**	1.92E-12	**0.00E+00**	2.86E-28
F_4	mean	**2.00E-323**	1.73E+01	6.58E+01	3.87E-02	8.93E+01	1.12E+01	3.44E+00	1.02E+01	7.81E+01	2.26E+01	1.43E+01	2.19E-45	2.22E-10
	min	**1.00E-323**	6.62E+00	5.96E+01	3.87E-02	8.56E+01	1.34E-01	1.03E+00	1.02E+01	6.82E+01	7.92E-02	1.43E+01	1.13E-47	2.34E-11
	max	**2.00E-323**	1.85E+01	6.65E+01	3.87E-02	9.23E+01	2.26E+01	5.38E+00	1.02E+01	8.24E+01	7.58E+01	1.43E+01	1.22E-44	7.31E-10
	std	**0.00E+00**	3.76E+00	2.19E+00	7.31E-18	1.94E+00	7.40E+00	1.15E+00	**0.00E+00**	4.43E+00	2.21E+01	3.74E-15	3.91E-45	2.68E-10
F_5	mean	**7.37E-04**	5.58E+01	6.79E+01	2.92E+01	2.69E+08	5.40E+04	3.31E+01	1.73E-02	3.49E+07	2.58E+01	2.42E+01	2.62E+01	2.67E+01
	min	**1.47E-08**	2.89E+01	2.82E+01	2.92E+01	1.62E+08	3.63E+02	2.00E+01	1.73E-02	1.83E+07	2.54E+01	2.42E+01	2.52E+01	2.52E+01
	max	**8.18E-04**	2.98E+02	7.23E+01	2.92E+01	3.23E+08	1.55E+05	8.40E+01	1.73E-02	4.56E+07	2.62E+01	2.42E+01	2.72E+01	2.88E+01
	std	2.59E-04	8.52E+01	1.39E+01	**0.00E+00**	4.92E+07	5.02E+04	1.97E+01	3.66E-18	9.13E+06	1.95E-01	3.74E-15	7.58E-01	1.09E+00
F_6	mean	6.28E-06	3.49E-05	**1.97E-22**	6.54E-04	6.81E+04	5.45E+02	9.32E-06	7.18E-16	5.16E+04	6.68E-04	1.08E-19	5.26E-01	1.83E+00
	min	7.69E-08	4.26E-13	**1.08E-23**	6.54E-04	5.10E+04	1.55E+02	9.22E-06	7.18E-16	3.22E+04	4.45E-04	1.08E-19	2.32E-06	1.01E+00
	max	1.98E-05	3.46E-04	**2.18E-22**	6.54E-04	7.60E+04	1.00E+03	1.02E-05	7.18E-16	5.78E+04	9.15E-04	1.08E-19	1.24E+00	2.65E+00
	std	6.86E-06	1.09E-04	6.55E-23	1.14E-19	7.79E+03	2.96E+02	3.07E-07	**0.00E+00**	7.25E+03	1.93E-04	**0.00E+00**	3.61E-01	5.11E-01
F_7	mean	**9.08E-06**	6.38E-01	8.04E-02	3.22E-02	1.24E+02	1.82E-01	3.88E-02	3.46E-02	1.18E-01	9.38E-04	1.76E-02	2.28E-04	9.89E-04
	min	8.45E-06	5.39E-01	4.20E-02	3.22E-02	9.48E+01	1.42E-02	1.64E-02	3.46E-02	7.12E-02	**7.99E-06**	1.76E-02	6.03E-05	4.94E-04
	max	**1.47E-05**	6.49E-01	8.46E-02	3.22E-02	1.59E+02	4.79E-01	1.28E-01	3.46E-02	1.92E-01	4.24E-03	1.76E-02	4.11E-04	1.54E-03
	std	1.99E-06	3.49E-02	1.35E-02	7.31E-18	1.99E+01	1.34E-01	2.19E-02	7.31E-18	3.29E-02	1.24E-03	**3.66E-18**	1.20E-04	3.73E-04
F_8	mean	**-6.48E+09**	-6.82E+03	-8.49E+03	-6.53E+03	-5.42E+03	-5.44E+03	-9.01E+03	-1.26E+04	-3.11E+03	-1.24E+04	-6.84E+03	-6.13E+03	-5.35E+03
	min	**-6.47E+10**	-7.00E+03	-8.60E+03	-6.53E+03	-5.42E+03	-6.22E+03	-9.92E+03	-1.26E+04	-3.47E+03	-1.26E+04	-6.84E+03	-7.18E+03	-6.44E+03
	max	-1.11E+04	-5.14E+03	-7.43E+03	-6.53E+03	-5.42E+03	-3.98E+03	-8.38E+03	**-1.26E+04**	-2.51E+03	-1.12E+04	-6.84E+03	-4.74E+03	-3.21E+03
	std	2.05E+10	5.87E+02	3.72E+02	9.59E-13	9.59E-13	6.81E+02	3.25E+02	**0.00E+00**	2.71E+02	4.41E+02	**0.00E+00**	6.85E+02	9.99E+02
F_9	mean	**0.00E+00**	7.46E+01	1.25E+02	5.27E+01	4.58E+02	1.16E+02	1.51E+01	**0.00E+00**	1.88E+02	**0.00E+00**	1.34E+02	**0.00E+00**	9.30E+00
	min	**0.00E+00**	3.18E+01	1.25E+02	5.27E+01	4.14E+02	1.16E+02	1.51E+01	**0.00E+00**	1.88E+02	**0.00E+00**	1.34E+02	**0.00E+00**	9.30E+00
	max	**0.00E+00**	1.15E+02	1.25E+02	5.27E+01	4.63E+02	1.16E+02	1.51E+01	**0.00E+00**	1.88E+02	**0.00E+00**	1.34E+02	**0.00E+00**	9.30E+00
	std	**0.00E+00**	2.25E+01	3.00E-14	1.50E-14	1.54E+01	3.00E-14	**0.00E+00**	**0.00E+00**	3.00E-14	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**
F_{10}	mean	**8.88E-16**	7.09E+00	1.98E+01	7.60E-03	2.00E+01	1.94E+01	1.44E+00	3.22E-09	2.00E+01	4.09E-15	1.91E+01	7.99E-15	**8.88E-16**
	min	**8.88E-16**	7.06E+00	1.84E+01	7.60E-03	2.00E+01	1.93E+01	5.88E-03	3.22E-09	2.00E+01	**8.88E-16**	1.91E+01	7.99E-15	**8.88E-16**
	max	**8.88E-16**	7.34E+00	1.99E+01	7.60E-03	2.00E+01	1.96E+01	5.15E+00	3.22E-09	2.00E+01	7.99E-15	1.91E+01	7.99E-15	**8.88E-16**
	std	**0.00E+00**	8.90E-02	4.78E-01	**0.00E+00**	**0.00E+00**	1.04E-01	1.09E+00	**0.00E+00**	1.44E-03	2.62E-15	3.74E-15	**0.00E+00**	**0.00E+00**
F_{11}	mean	**0.00E+00**	6.44E-02	1.75E-02	2.30E-03	5.89E+02	5.04E+00	1.80E-03	2.22E-16	5.78E+02	**0.00E+00**	1.48E-02	**0.00E+00**	1.89E-03
	min	**0.00E+00**	4.02E-13	1.23E-02	2.30E-03	5.26E+02	**0.00E+00**	1.34E-06	2.22E-16	4.56E+02	**0.00E+00**	1.48E-02	**0.00E+00**	**0.00E+00**
	max	**0.00E+00**	1.62E-01	6.37E-02	2.30E-03	7.20E+02	1.15E+01	9.40E-03	2.22E-16	6.56E+02	**0.00E+00**	1.48E-02	**0.00E+00**	1.08E-02
	std	**0.00E+00**	5.13E-02	1.63E-02	**0.00E+00**	6.21E+01	3.51E+00	2.33E-03	**0.00E+00**	7.32E+01	**0.00E+00**	**0.00E+00**	**0.00E+00**	4.03E-03
F_{12}	mean	5.46E-07	8.27E+00	8.95E-01	3.53E-06	6.19E+08	4.05E+04	6.99E-01	**7.16E-16**	1.12E+08	8.22E-05	1.53E+01	3.30E-02	1.20E-01
	min	3.81E-10	7.75E+00	4.15E-01	3.53E-06	5.19E+08	8.18E-01	1.77E-03	**7.16E-16**	4.10E+07	5.11E-05	1.53E+01	1.29E-02	4.04E-02
	max	1.87E-06	1.30E+01	5.22E+00	3.53E-06	7.41E+08	4.05E+05	1.78E+00	**7.16E-16**	1.54E+08	1.19E-04	1.53E+01	8.54E-02	2.11E-01
	std	5.66E-07	1.65E+00	1.52E+00	**0.00E+00**	7.37E+07	1.28E+05	5.50E-01	**0.00E+00**	3.36E+07	2.12E-05	**0.00E+00**	2.08E-02	5.94E-02
F_{13}	mean	1.44E-06	2.45E+01	4.39E-03	5.00E-05	1.16E+09	1.40E+05	8.05E-04	6.39E-16	2.51E+08	1.64E-02	**1.62E-20**	4.81E-01	1.28E+00
	min	1.51E-07	2.39E+01	**2.40E-21**	5.00E-05	7.65E+08	1.51E+01	5.49E-06	6.39E-16	1.17E+08	5.16E-04	1.62E-20	2.75E-01	8.11E-01
	max	1.58E-06	3.01E+01	4.39E-02	5.00E-05	1.47E+09	9.61E+05	8.83E-03	6.39E-16	4.04E+08	1.22E-01	**1.62E-20**	8.03E-01	1.66E+00
	std	4.52E-07	1.98E+00	1.39E-02	1.43E-20	2.18E+08	3.06E+05	1.66E-03	1.04E-31	1.03E+08	3.74E-02	**3.17E-36**	1.84E-01	2.97E-01

Table 23 The Wilcoxon rank sum test results for F_1-F_{13}

	PSOVA	SPSO	MPSO	ALPSO	GPSO	DNLPSO	ELPSO	CLFA	CGFA	VSSFA	MFA	NaFA
F_1	4.82E-13	1.20E-12	1.20E-12	1.20E-12	4.82E-13	4.82E-13	4.82E-13	4.82E-13	4.82E-13	4.82E-13	3.37E-13	6.13E-14
F_2	1.34E-12	3.11E-12	3.13E-12	3.13E-12	1.34E-12	1.34E-12	1.34E-12	1.34E-12	1.34E-12	1.34E-12	9.99E-13	1.34E-12
F_3	4.82E-13	1.20E-12	1.20E-12	1.20E-12	4.82E-13	4.82E-13	4.82E-13	4.82E-13	4.82E-13	4.82E-13	3.37E-13	4.82E-13
F_4	1.34E-12	3.13E-12	3.13E-12	7.57E-07	1.34E-12	1.34E-12	1.34E-12	1.34E-12	1.34E-12	1.34E-12	9.62E-13	1.34E-12
F_5	7.06E-12	3.13E-12	3.13E-12	2.98E-11	7.06E-12	7.06E-12	7.06E-12	7.06E-12	1.34E-12	7.06E-12	5.28E-12	7.06E-12
F_6	7.06E-12	2.95E-11	2.97E-11	2.98E-11	7.06E-12	1.34E-12	7.06E-12	7.06E-12	7.06E-12	7.06E-12	5.28E-12	7.06E-12
F_7	1.34E-12	3.12E-12	3.13E-12	1.20E-12	1.34E-12	1.34E-12	1.34E-12	1.34E-12	1.34E-12	1.34E-12	1.34E-12	1.34E-12
F_8	7.06E-12	2.98E-11	2.98E-11	2.98E-11	7.06E-12	4.17E-08	7.06E-12	7.06E-12	7.06E-12	7.06E-12	5.28E-12	1.34E-12
F_9	4.82E-13	1.20E-12	1.20E-12	1.20E-12	4.82E-13	4.82E-13	4.82E-13	4.82E-13	4.82E-13	4.82E-13	3.37E-13	6.13E-14
F_{10}	5.28E-12	1.13E-11	1.10E-11	3.37E-13	5.28E-12	5.28E-12	5.28E-12	5.28E-12	9.62E-13	5.28E-12	3.92E-12	5.28E-12
F_{11}	4.82E-13	1.20E-12	1.20E-12	1.00E+00	4.81E-13	4.82E-13	4.82E-13	4.82E-13	4.82E-13	4.82E-13	3.37E-13	4.82E-13
F_{12}	7.06E-12	2.97E-11	2.98E-11	2.47E-08	7.06E-12	7.06E-12	7.06E-12	7.06E-12	7.06E-12	7.06E-12	5.28E-12	7.06E-12
F_{13}	7.06E-12	2.97E-11	2.98E-11	1.06E-07	7.06E-12	7.06E-12	7.06E-12	1.34E-12	7.06E-12	7.06E-12	7.02E-13	7.06E-12

	PSO	MFO	FA	ALO	DA	CS	ABC	BA	WOA	PSOGSA	GWO	amixedGWO
F_1	4.82E-13	4.82E-13	3.37E-13	1.20E-12	1.20E-12	1.21E-12	3.37E-13	1.20E-12	1.00E+00	3.37E-13	1.20E-12	1.20E-12
F_2	1.34E-12	2.00E-13	9.62E-13	3.12E-12	1.34E-12	3.02E-11	9.62E-13	3.13E-12	3.06E-12	6.12E-14	3.13E-12	3.13E-12
F_3	3.37E-13	3.37E-13	3.37E-13	4.82E-13	1.69E-14	3.37E-13	3.37E-13	3.37E-13	3.37E-13	3.37E-13	3.37E-13	3.37E-13
F_4	1.34E-12	2.00E-13	9.62E-13	3.13E-12	3.13E-12	2.55E-11	9.62E-13	3.12E-12	3.12E-12	9.62E-13	3.12E-12	3.13E-12
F_5	7.04E-12	1.34E-12	4.82E-13	3.00E-11	3.00E-11	3.02E-11	5.28E-12	3.12E-12	3.00E-11	5.28E-12	3.00E-11	3.00E-11
F_6	7.06E-12	7.06E-12	5.28E-12	3.00E-11	3.00E-11	3.99E-04	5.28E-12	2.97E-11	3.00E-11	5.28E-12	1.84E-08	3.00E-11
F_7	1.34E-12	1.34E-12	6.13E-14	3.00E-11	3.00E-11	3.02E-11	9.65E-13	3.13E-12	1.69E-09	9.65E-13	3.00E-11	3.00E-11
F_8	7.04E-12	1.34E-12	5.28E-12	1.11E-11	2.92E-11	2.98E-11	3.45E-08	2.98E-11	1.05E-07	5.28E-12	2.93E-11	2.92E-11
F_9	4.82E-13	4.82E-13	3.37E-13	1.20E-12	1.20E-12	1.21E-12	1.00E+00	1.20E-12	1.00E+00	3.37E-13	1.00E+00	1.20E-12
F_{10}	5.28E-12	5.28E-12	3.92E-12	3.37E-13	1.12E-12	1.21E-12	3.37E-13	1.12E-11	1.71E-09	3.92E-12	3.37E-13	1.00E+00
F_{11}	4.82E-13	4.82E-13	3.37E-13	1.20E-12	5.73E-11	1.21E-12	3.37E-13	1.20E-12	1.00E+00	3.37E-13	1.00E+00	8.82E-07
F_{12}	7.06E-12	7.06E-12	5.28E-12	3.00E-11	3.00E-11	3.02E-11	5.28E-12	2.97E-11	3.00E-11	5.28E-12	3.00E-11	3.00E-11
F_{13}	7.04E-12	4.17E-08	5.28E-12	3.00E-11	3.00E-11	3.02E-11	5.28E-12	2.98E-11	3.00E-11	5.28E-12	3.00E-11	3.00E-11

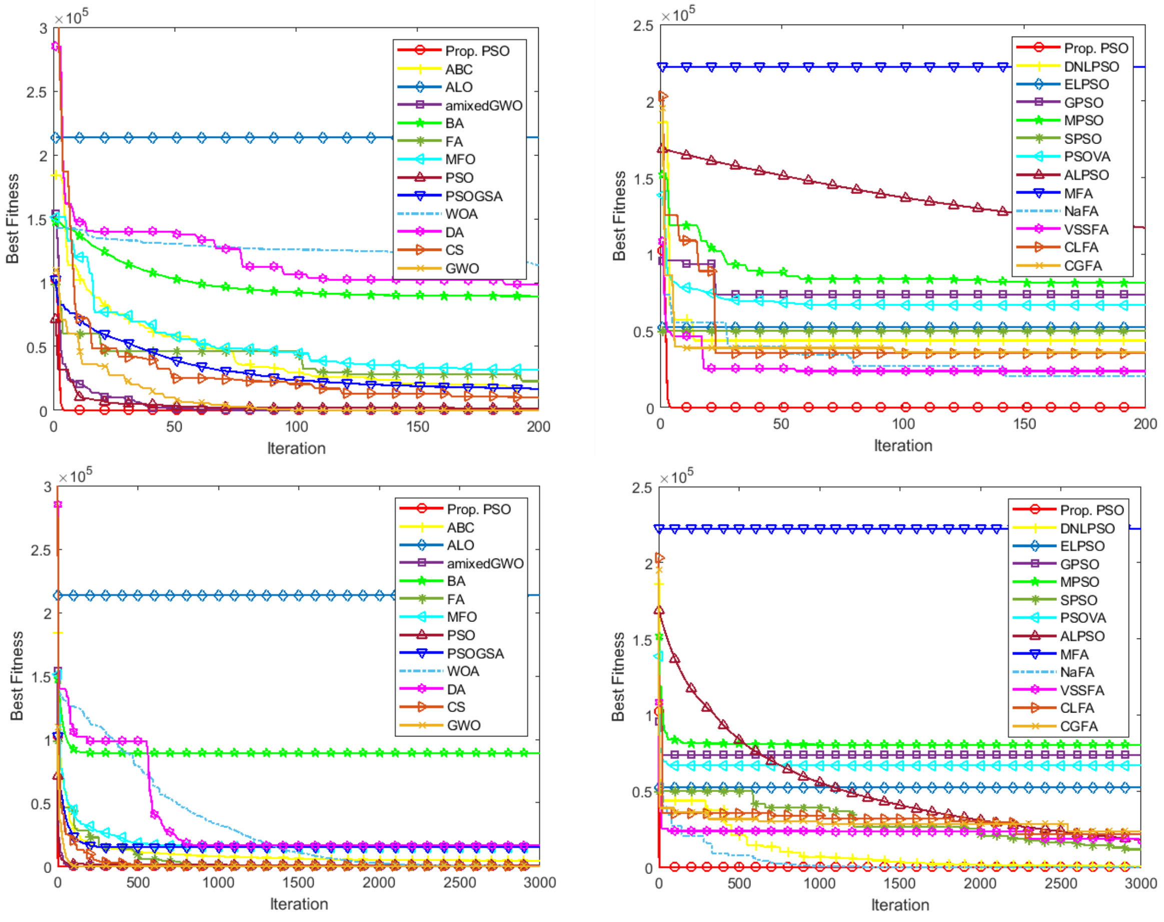


Figure 12 Convergence curves of F_3 over a set of 30 runs (Top row – 200 iterations, bottom row – 3000 iterations)

To indicate model convergence speed, Figure 12 illustrates the convergence curves for both 200 (top row) and 3000 (bottom row) iterations for the unimodal function, F_3, over a set of 30 runs. The proposed algorithm achieves significantly faster convergence rates in comparison with those of PSO variants and other search methods. As indicated in Figure 12 (top row), PSO, amixedGWO, GWO, CS and FA illustrate faster convergence rates among the baseline methods and achieve impressive performance with smaller numbers of iterations. NaFA and DNLPSO also show drastic improvements over smaller numbers of iterations in comparison with other PSO and FA variants. As the search intensifies, as shown in Figure 12 (bottom row), WOA, ABC, DA, SPSO, VSSFA and ALPSO are able to achieve improved convergence.

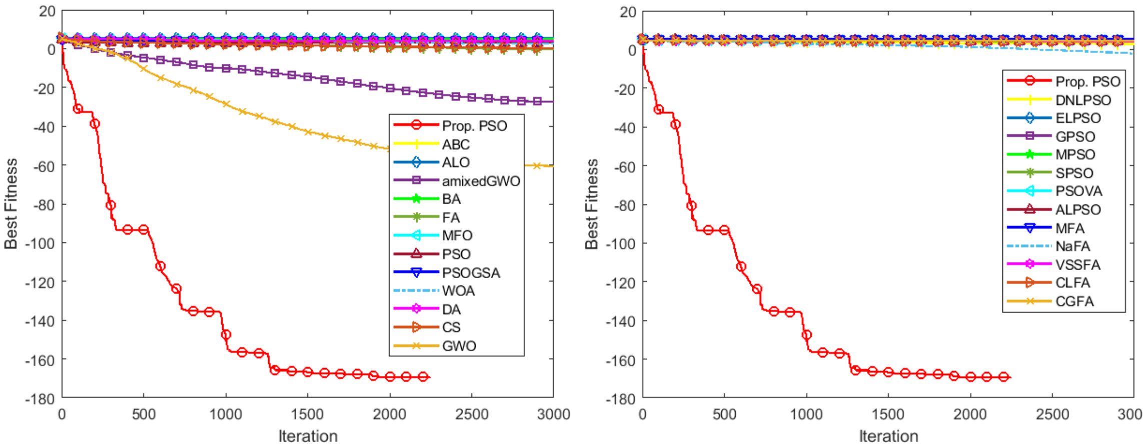
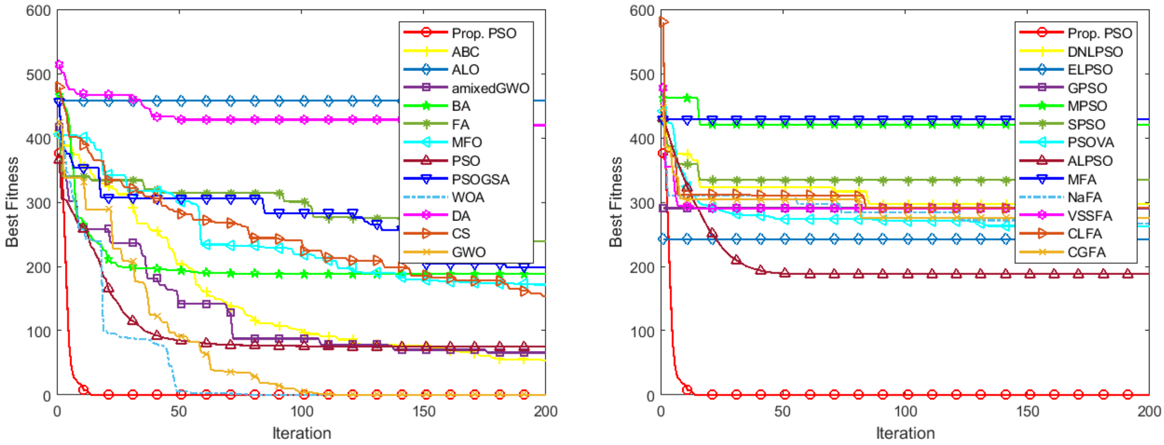


Figure 13 Convergence curves of F_3 in the log scale over a set of 30 runs

We also convert the convergence curves of F_3 to the log scale with a base of 10, as depicted in Figure 13. Our model achieves the global minimum, i.e. 0, after 2250 iterations, based on the mean average over 30 runs. Since $log_{10}(0) = -\infty$ (which cannot be displayed as a specific value), the convergence curve of our model in the log scale is illustrated until 2249 iterations, providing a clear overview of the model convergence speed. As shown in Figure 13, the proposed algorithm illustrates a significantly faster convergence rate than those of all compared methods. GWO and amixedGWO converge comparatively faster against all other baseline search methods, while NaFA illustrates the fastest convergence rate among the PSO and FA variants.



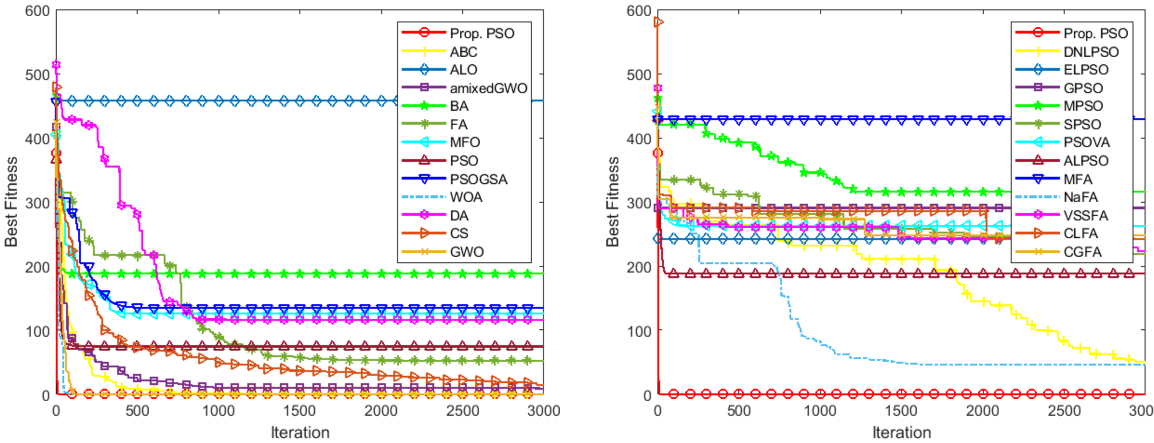


Figure 14 Convergence curves of F_9 over a set of 30 runs (Top row – 200 iterations, bottom row – 3000 iterations)

Figure 14 illustrate the convergence curves for both 200 (top row) and 3000 (bottom row) iterations for multimodal function F_9 over a set of 30 runs. The proposed model achieves the global minimum solution, i.e. 0, in a comparatively smallest number of iterations against those of all baseline methods. In addition, as illustrated in Figure 14 (top row), WOA, PSO, GWO, amixedGWO, ABC, and ALPSO show faster convergence rates in comparison with those of other baseline methods and achieve impressive performance with smaller numbers of iterations. As the search progresses, as shown in Figure 14 (bottom row), CS, FA, NaFA and DNLPSO yield drastic improvements. The convergence analysis using other benchmark functions depicts similar characteristics, where the proposed model illustrates the fastest convergence rates in most test cases.

Since the majority of these classical and advanced search methods employ single leaders at a time to guide the search process, they are more likely to be trapped in local optima. GWO employs multiple leaders to guide its search process. However, since equal weightings are assigned to the leader signals, its search capability in local exploitation and global exploration is compromised. The proposed PSO algorithm incorporates hybrid leaders generated by adaptive weightings based on regular and irregular adaptive crossover operators to better balance between diversification and intensification in search. Such enhanced elite signals in conjunction with the secant and Newton-Raphson-based intensification and spherical parametric search coefficients are useful to guide the search process to overcome local optimum traps and achieve global optimality. It outperforms all baseline search methods statistically in most evaluations in solving diverse unimodal and multimodal functions. The proposed PSO algorithm is also capable of devising efficient hyper-parameters in the CBiLSTM networks to achieve superior performance in sound classification.

6. CONCLUSIONS

In this research, we have proposed an ensemble CBiLSTM network with optimal hyper-parameter selection using the proposed PSO algorithm for undertaking audio classification tasks. To effectively overcome the stagnation problem, the proposed model incorporates secant and Newton's methods, spherical search coefficients, and regular and irregular adaptive elliptical formulae for generating hybrid leaders. It employs multiple elite fused signals in conjunction with diversified search steps and trajectories to overcome the limitations of the original PSO algorithm. The empirical results indicate its superiority in identifying effective network settings and yielding efficient spatial-temporal dynamics to improve sound classification performance. Evaluated using diverse sound data sets, the proposed ensemble CBiLSTM models outperform counterparts with optimal settings identified by other search methods, as well as several existing deep networks and state-of-the-art methods, significantly. The proposed PSO algorithm also produces statistically better results as compared with those from 24 classical and advanced search methods for solving diverse unimodal and multimodal benchmark functions. The empirical results clearly support the effectiveness of the formulated strategies, which include hybrid elite signal generation, numerical analysis based leader enhancement and super-elliptical curves and surfaces oriented search coefficients, in yielding significant superiority of the proposed model.

In future work, we will incorporate other strategies (e.g. Muller's method) [67, 97] for leader enhancement. In comparison with the secant and Newton's methods, such strategies are able to approximate the function quadratically to increase search diversity. Other oversampling techniques (such as autoencoder) will also be employed to tackle the class imbalance problem. We will further evaluate the proposed PSO algorithm for generating deep learning models [43, 58, 63] with respect to other signal and vision processing tasks such as voice activity detection [98], speech emotion recognition [41], visual saliency detection [19, 62], and image description generation [6, 7, 8, 99].

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] S. Li, F. Li, S. Tang and W. Xiong. (2020). A review of computer-aided heart sound detection techniques. *BioMed Research International*, Volume 2020, Article ID 5846191.

[2] I. Ozer, Z. Ozer and O. Findik. (2018). Noise robust sound event classification with convolutional neural network. *Neurocomputing*, 272, pp.505-512.

[3] C. Wall, L. Zhang, Y. Yu and L. Mistry. (2021). Deep Recurrent Neural Networks with Attention Mechanisms for Respiratory Anomaly Classification. In *Proceedings of International Joint Conference on Neural Networks (IJCNN)*.

[4] L. Nanni, G. Maguolo and M. Paci. (2020). Data augmentation approaches for improving animal audio classification. *Ecological Informatics*, 57, p.101084.

[5] K.J. Piczak. (2015). Environmental sound classification with convolutional neural networks. In *Proceedings of Proceedings of IEEE 25th International Workshop on Machine Learning for Signal Processing*, pp. 1–6.

[6] P. Kinghorn, L. Zhang and L. Shao. (2018). A region-based image caption generator with refined descriptions. *Neurocomputing*. 272 (2018) 416-424.

[7] P. Kinghorn, L. Zhang and L. Shao. (2019). A Hierarchical and Regional Deep Learning Architecture for Image Description Generation. *Pattern Recognition Letters*. 119 (2019) 77-85.

[8] J. Donahue, L. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko and T. Darrell. (2015). Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*

[9] A.R. Jordehi. (2015). Enhanced leader PSO (ELPSO): a new PSO variant for solving global optimisation problems. *Applied Soft Computing*. 26 (2015) 401–417.

[10] Y. Zhang, L. Zhang, S.C. Neoh, K. Mistry and A. Hossain. (2015) affect regression for bodily expressions using hybrid particle swarm optimization and adaptive ensembles. *Expert Systems with Applications*, 42 (22) 8678-8697. 2015.

[11] J. Kennedy and R. Eberhart. (1995). Particle Swarm Optimization. In *Proceedings of IEEE Int. Conf. Neural Networks*, vol. 4, pp.1942-1948.

[12] X.S. Yang. (2010). Firefly Algorithm, Levy Flights and Global Optimization. *Research and Development in Intelligent Systems.* 26 (2010) 209–218.

[13] Q. Chen, Y. Chen and W. Jiang. (2016). Genetic particle swarm optimization–based feature selection for very-high-resolution remotely sensed imagery object change detection. *Sensors*, 16 (8) 1204.

[14] D.R. Nayak, R. Dash and B. Majhi. (2018). Discrete ripplet-II transform and modified PSO based improved evolutionary extreme learning machine for pathological brain detection. *Neurocomputing*. 282 (2018) 232–247.

[15] M. Nasir, S. Das, D. Maity, S. Sengupta, U. Halder and P.N. Suganthan. (2012). A dynamic neighborhood learning based particle swarm optimizer for global numerical optimization. *Information Sciences*. 209 (2012) 16–36.

[16] A. Kazem, E. Sharifi, F.K. Hussain, M. Saberic and O.K. Hussain. (2013). Support vector regression with chaos-based firefly algorithm for stock market price forecasting, *Applied Soft Computing*. 13 (2) 947–958.

[17] A.H. Gandomi, X.S. Yang, S. Talatahari and A.H. Alavi. (2013). Firefly algorithm with chaos, *Communications in Nonlinear Science and Numerical Simulation*, 18, 89–98.

[18] S.H. Yu, S.L. Zhu, Y. Ma and D.M. Mao. (2015). A variable step size firefly algorithm for numerical optimization. *Applied Mathematics and Computation*. 263, 214–220.

[19] L. He and S. Huang. (2017). Modified firefly algorithm based multilevel thresholding for colour image segmentation. *Neurocomputing,* 240 (2017) 152-174.

[20] H. Wang, W. Wang, X. Zhou, H. Sun, J. Zhao, X. Yu and Z. Cui. (2017). Firefly algorithm with neighborhood attraction. *Information Sciences*. 382–383 (2017) 374–387.

[21] W. Srisukkham, L. Zhang, S.C. Neoh, S. Todryk and C.P. Lim. (2017). Intelligent Leukaemia Diagnosis with Bare-Bones PSO based Feature Optimization. *Applied Soft Computing*, 56 (2017) 405-419.

[22] S. Mirjalili, S.M. Mirjalili and A. Lewis. (2014). Grey wolf optimizer. *Advances in Engineering Software*, 69, pp.46-61.

[23] S. Mirjalili, S. Saremi, S.M. Mirjalili and L.D.S. Coelho. (2016). Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization. *Expert Systems with Applications*, 47, pp.106-119.

[24] H. Faris, I. Aljarah, M.A. Al-Betar and S. Mirjalili. (2018). Grey wolf optimizer: a review of recent variants and applications. *Neural Computing and Applications*, 30(2), pp.413-435.

[25] B.M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques and R.P. Paiva. (2017). A respiratory sound database for the development of automated classification. In *Proceedings of International Conference on Biomedical and Health Informatics* (pp. 33-37). Springer, Singapore.

[26] C. Liu , D. Springer , Q. Li , B. Moody , R.A. Juan , F.J. Chorro , F. Castells , J.M. Roig , I. Silva , A.E.W. Johnson , Z. Syed , S.E. Schmidt , C.D. Papadaniil , L. Hadjileontiadis , H. Naseri , A. Moukadem , A. Dieterlen , C. Brandt , H. Tang , M. Samieinasab , M.R. Samieinasab , R. Sameni , R.G. Mark , G.D. Clifford, (2016). *An open access database for the evaluation of heart sound algorithms*, *Physiol. Meas.* 37 (2016) 2181–2213 .

[27] K.J. Piczak. (2015). ESC: Dataset for Environmental Sound Classification. In *Proceedings of 23rd ACM Int. Conf. Multimedia*, pp. 1015–1018.

[28] D. Perna and A. Tagarelli. (2019). Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks. In *Proceedings of IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 50-55). IEEE.

[29] J.M.T. Wu, M.H. Tsai, Y.Z. Huang, S.H. Islam, M.M. Hassan, A. Alelaiwi and G. Fortino. (2019). Applying an ensemble convolutional neural network with Savitzky–Golay filter to construct a phonocardiogram prediction model. *Applied Soft Computing*, *78*, pp.29-40.

[30] W. Zhang, J. Han and S. Deng. (2017). Heart sound classification based on scaled spectrogram and tensor decomposition. *Expert Systems with Applications*, *84*, pp.220-231.

[31] B. Xiao, Y. Xu, X. Bi, J. Zhang and X. Ma. (2020). Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption. *Neurocomputing*, 392, pp.153-159.

[32] S. Kiranyaz, M. Zabihi, A.B. Rad, T. Ince, R. Hamila and M. Gabbouj. (2020). Real-time phonocardiogram anomaly detection by adaptive 1D convolutional neural networks. *Neurocomputing*, *411*, pp.291-301.

[33] S.B. Shuvo, S.N. Ali, S.I. Swapnil, M.S. Al-Rakhami and A. Gumaei. (2021). CardioXNet: A Novel Lightweight Deep Learning Framework for Cardiovascular Disease Classification Using Heart Sound Recordings. *IEEE Access*.

[34] M.T. García-Ordás, J.A. Benítez-Andrades, I. García-Rodríguez, C. Benavides and H. Alaiz-Moretón. (2020). Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data. *Sensors*, *20*(4), p.1214.

[35] X. Zhao, Y. Shao, J. Mai, A. Yin and S. Xu. (2020). Respiratory Sound Classification Based on BiGRU-Attention Network with XGBoost. In *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 915-920). IEEE.

[36] H. Chen, X. Yuan, Z. Pei, M. Li and J. Li. (2019). Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. *IEEE Access*, 7, pp.32845-32852.

[37] D. Oletic and V. Bilas. (2017). Asthmatic wheeze detection from compressively sensed respiratory sound spectra. *IEEE Journal of Biomedical and Health Informatics*, 22(5), pp.1406-1414.

[38] M. Esmaeilpour, P. Cardinal and A.L. Koerich. (2020). Unsupervised feature learning for environmental sound classification using weighted cycle-consistent generative adversarial network. *Applied Soft Computing*, *86*, p.105912.

[39] Z. Zhang, S. Xu, S. Zhang, T. Qiao and S. Cao. (2020). Attention based convolutional recurrent neural network for environmental sound classification. *Neurocomputing*.

[40] F. Medhat, D. Chesmore and J. Robinson. (2020). Masked Conditional Neural Networks for sound classification. *Applied Soft Computing*, *90*, p.106073.

[41] Y. Kuang, Q. Wu, Y. Wang, N. Dey, F. Shi, R.G. Crespo and R.S. Sherratt. (2020). Simplified inverse filter tracked affective acoustic signals classification incorporating deep convolutional neural networks. *Applied Soft Computing*, *97*, p.106775.

[42] S.C. Neoh, L. Zhang, K. Mistry, A.M. Hossain, C.P. Lim, N. Aslam and P. Kinghorn. (2015). Intelligent Facial Emotion Recognition Using a Layered Encoding Cascade Optimization Model. *Applied Soft Computing*. Volume 34, 72–93.

[43] B. Fielding, T. Lawrence and L. Zhang. (2019). Evolving and Ensembling Deep CNN Architectures for Image Classification. In *Proceedings of International Joint Conference on Neural Networks*.

[44] H. Xie, L. Zhang, C.P. Lim., Y. Yu, C. Liu, H. Liu and J. Walters. (2019). Improving K-means clustering with enhanced firefly algorithms. *Applied Soft Computing,* 84, p.105763.

[45] H. Xie, L. Zhang and C.P. Lim. (2020). Evolving CNN-LSTM models for time series prediction using enhanced grey wolf optimizer. *IEEE Access*, 8, pp.161519-161541.

[46] M. Willis, L. Zhang, H. Liu, H. Xie and K. Mistry. (2020). Object Recognition Using Enhanced Particle Swarm Optimization. In *Proceedings of International Conference on Machine Learning and Cybernetics (ICMLC)* (pp. 241-246). IEEE.

[47] K. Mistry, L. Zhang, S.C. Neoh, C.P. Lim and B. Fielding. (2017). A micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition. *IEEE Transactions on Cybernetics*. 47 (6) 1496–1509.

[48] T.Y. Tan, L. Zhang and C.P. Lim. (2020). Adaptive melanoma diagnosis using evolving clustering, ensemble and deep neural networks. *Knowledge-Based Systems*, 187, p.104807.

[49] L. Zhang, C.P. Lim and Y. Yu, Y. (2021). Intelligent human action recognition using an ensemble model of evolving deep networks with swarm-based optimization. *Knowledge-Based Systems*, *220*, p.106918.

[50] T.Y. Tan, L. Zhang, L., S.C. Neoh and C.P. Lim. (2018). Intelligent skin cancer detection using enhanced particle swarm optimization. *Knowledge-based systems*, 158, pp.118-135.

[51] A.D. Li, B. Xue and M. Zhang. (2021). Improved binary particle swarm optimization for feature selection with new initialization and search space reduction strategies. *Applied Soft Computing*, *106*, p.107302.

[52] X. Zhang, W. Sun, M. Xue and A. Lin. (2021). Probability-optimal leader comprehensive learning particle swarm optimization with Bayesian iteration. *Applied Soft Computing*, *103*, p.107132.

[53] H. Xie, L. Zhang, C.P. Lim, Y. Yu and H. Liu. (2021). Feature Selection Using Enhanced Particle Swarm Optimisation for Classification Models. *Sensors*, *21*(5), p.1816.

[54] F. Kılıç, Y. Kaya and S. Yildirim. (2021). A novel multi population based particle swarm optimization for feature selection. *Knowledge-Based Systems*, *219*, p.106894.

[55] S. Molaei, H. Moazen, S. Najjar-Ghabel and L. Farzinvash. (2021). Particle swarm optimization with an enhanced learning strategy and crossover operator. *Knowledge-Based Systems*, *215*, p.106768.

[56] X. Kan, Y. Fan, Z. Fang, L. Cao, N.N. Xiong, D. Yang and X. Li. (2021). A Novel IoT Network Intrusion Detection Approach Based on Adaptive Particle Swarm Optimization Convolutional Neural Network. *Information Sciences*. 568, 147-162, 2021.

[57] W. Li, X. Meng, Y. Huang and Z.H. Fu. (2020). Multipopulation cooperative particle swarm optimization with a mixed mutation strategy. *Information Sciences*, 529, pp.179-196.

[58] T. Lawrence, L. Zhang, C.P. Lim and E.J. Phillips. (2021). Particle Swarm Optimization for Automatically Evolving Convolutional Neural Networks for Image Classification. *IEEE Access*, *9*, pp.14369-14386.

[59] M.D. Phung and Q.P. Ha. (2021). Safety-enhanced UAV path planning with spherical vector-based particle swarm optimization. *Applied Soft Computing*, *107*, p.107376.

[60] S.C. Chu, Z.G. Du, Y.J. Peng and J.S. Pan. (2021). Fuzzy Hierarchical Surrogate Assists Probabilistic Particle Swarm Optimization for expensive high dimensional problem. *Knowledge-Based Systems*, 220, p.106939.

[61] P. Das, A.K. Das, J. Nayak, D. Pelusi and W. Ding. (2021). Incremental classifier in crime prediction using bi-objective Particle Swarm Optimization. *Information Sciences*, *562*, pp.279-303.

[62] T.Y. Tan, L. Zhang, C.P. Lim, B. Fielding, Y. Yu and E. Anderson. (2019). Evolving Ensemble Models for Image Segmentation Using Enhanced Particle Swarm Optimization. *IEEE Access*. 7, 34004-34019.

[63] B. Fielding and L. Zhang. (2020). Evolving Deep DenseBlock Architecture Ensembles for Image Classification. *Electronics*, *9*(11), p.1880.

[64] Y. Li, J. Xiao, Y. Chen and L. Jiao. (2019). Evolving deep convolutional neural networks by quantum behaved particle swarm optimization with binary encoding for image classification. *Neurocomputing*, *362*, pp.156-165.

[65] B. Martin, J. Marot and S. Bourennane. (2019). Mixed grey wolf optimizer for the joint denoising and unmixing of multispectral images. *Applied Soft Computing*, 74, pp.385-410.

[66] A. Gil, J. Segura and N.M. Temme. (2007). Numerical methods for special functions. Society for Industrial and Applied Mathematics.

[67] M. Avriel. (2003). Nonlinear programming: analysis and methods. Courier Corporation.

[68] J. Gielis. (2003). A generic geometric transformation that unifies a wide range of natural and abstract shapes. *American Journal of Botany*, 90(3), pp.333-338.

[69] Z. Zhang, S. Xu, S. Cao and S. Zhang. (2018). Deep convolutional neural network with mixup for environmental sound classification. In *Proceedings of Chinese conference on pattern recognition and computer vision (prcv)* (pp. 356-367). Springer, Cham.

[70] V. Makarenkov, L. Rokach and B. Shapira. (2019). Choosing the right word: Using bidirectional LSTM tagger for writing support systems. *Engineering Applications of Artificial Intelligence*, 84, pp.1-10.

[71] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad and S.W. Baik. (2017). Action recognition in video sequences using deep bi-directional LSTM with CNN features, *IEEE Access*, 6, 1155–1166.

[72] D. Perna. (2018). Convolutional neural networks learning from respiratory data. In *Proceedings of 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2109-2113). IEEE.

[73] L. Zhang and C.P. Lim. (2020). Intelligent optic disc segmentation using improved particle swarm optimization and evolving ensemble models. *Applied Soft Computing*, *92*, p.106328.

[74] B. Fielding and L. Zhang. (2018). Evolving Image Classification Architectures with Enhanced Particle Swarm Optimisation. *IEEE Access*. Vol 6. 68560–68575.

[75] S. Mirjalili. (2015). Moth-Flame optimization algorithm: A novel nature-inspired heuristic paradigm, *Knowledge-Based Systems*. 89 (2015) 228–249.

[76] S. Mirjalili. (2015). The ant lion optimizer. *Advances in Engineering Software*, 83, pp.80-98.

[77] S. Mirjalili. (2016). Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural Computing and Applications*, 27(4), pp.1053-1073.

[78] F. Li, H. Tang, S. Shang, K. Mathiak and F. Cong. (2020). Classification of heart sounds using convolutional neural network. *Applied Sciences*, 10(11), p.3956.

[79] C. Thomae and A. Dominik. (2016). Using deep gated RNN with a convolutional front end for end-to-end classification of heart sound. In *Proceedings of Computing in Cardiology Conference (CinC)* (pp. 625-628). IEEE.

[80] H.L. Her and H.W. Chiu. (2016). Using time-frequency features to recognize abnormal heart sounds. In *Proceedings of Computing in Cardiology Conference (CinC)* (pp. 1145-1147). IEEE.

[81] C. Potes, S. Parvaneh, A. Rahman and B. Conroy. (2016). Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In *Proceedings of computing in cardiology conference (CinC)* (pp. 621-624). IEEE.

[82] M.N. Homsi, N. Medina, M. Hernandez, N. Quintero, G. Perpiñan, A. Quintana and P. Warrick. (2016). Automatic heart sound recording classification using a nested set of ensemble algorithms. In *Proceedings of Computing in Cardiology Conference (CinC)* (pp. 817-820). IEEE.

[83] V. Boddapati, A. Petef, J. Rasmusson and L. Lundberg. (2017). Classifying environmental sounds using image recognition networks. *Procedia computer science*, *112*, pp.2048-2056.

[84] Y. Aytar, C. Vondrick and A. Torralba. (2016). Soundnet: Learning sound representations from unlabeled video, In *Proceedings of Neural Information Processing Systems*.

[85] Y. Su, K. Zhang, J. Wang and K. Madani. (2019). Environment sound classification using a two-stream CNN based on decision-level fusion. *Sensors*, 19(7), p.1733.

[86] K.J. Piczak. (2015). Environmental sound classification with convolutional neural networks. In *Proceedings of IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE.

[87] B. daSilva, A.W. Happi, A. Braeken and A. Touhafi. (2019). Evaluation of classical Machine Learning techniques towards urban sound recognition on embedded systems. *Applied Sciences*, 2019; 9(18):1–27.

[88] A. Khamparia, D. Gupta, N.G. Nguyen, A. Khanna, B. Pandey and P. Tiwari. (2019). Sound classification using convolutional neural network and tensor deep stacking network. *IEEE Access*, 7, pp.7717-7727.

[89] J. Salamon and J.P. Bello. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24(3), pp.279-283.

[90] L. Zhang, W. Srisukkham, S.C. Neoh, C.P. Lim and D. Pandit. (2018). Classifier ensemble reduction using a modified firefly algorithm: An empirical evaluation. *Expert Systems with Applications*. 93 (2018) 395-422.

[91] D. Pandit, L. Zhang, S. Chattopadhyay, C.P. Lim and C. Liu. (2018). A Scattering and Repulsive Swarm Intelligence Algorithm for Solving Global Optimization Problems. *Knowledge-Based Systems*.

[92] X.S. Yang and S. Deb. (2009). Cuckoo search via Lévy flights. In *Proceedings of World Congress on Nature & Biologically Inspired Computing (NaBIC)* (pp. 210-214). IEEE.

[93] Karaboga, D. and Basturk, B., 2008. On the performance of artificial bee colony (ABC) algorithm. *Applied Soft Computing*, 8(1), pp.687-697.

[94] X.S. Yang. (2010). A new metaheuristic bat-inspired algorithm. In *Proceedings of Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)* (pp. 65-74). Springer, Berlin, Heidelberg.

[95] S. Mirjalili and A. Lewis. (2016). The whale optimization algorithm. *Advances in Engineering Software*, 95, pp.51-67.

[96] S. Mirjalili and S.Z.M. Hashim. (2010). A new hybrid PSOGSA algorithm for function optimization. In *Proceedings of International Conference on Computer and Information Application* (pp. 374-377). IEEE.

[97] J.D. Hoffman and S. Frankel. (2018). Numerical methods for engineers and scientists. CRC press.

[98] I. Ariav, D. Dov and I. Cohen. (2018). A deep architecture for audio-visual voice activity detection in the presence of transients. *Signal Processing*, *142*, pp.69-74.

[99] P. Kinghorn, L. Zhang and L. Shao. (2017). Deep learning based image description generation. In *Proceedings of International Joint Conference on Neural Networks (IJCNN)*. pp.919-926.