Received 13 November 2023, accepted 21 November 2023, date of publication 28 November 2023,
date of current version 1 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3336714

RESEARCH ARTICLE

Equilibrium Optimizer for Emotion Classification From English Speech Signals

LIYA YUE[1], PEI HU[2], SHU-CHUAN CHU[3], AND JENG-SHYANG PAN[3,4], (Senior Member, IEEE)
[1]Fanli Business School, Nanyang Institute of Technology, Nanyang, Henan 473004, China
[2]School of Computer and Software, Nanyang Institute of Technology, Nanyang, Henan 473004, China
[3]College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590, China
[4]Department of Information Management, Chaoyang University of Technology, Taichung 413310, Taiwan

Corresponding author: Jeng-Shyang Pan (jengshyangpan@gmail.com)

This work was supported in part by the Henan Provincial Philosophy and Social Science Planning Project under Grant 2022BJJ076, and in part by the Henan Province Key Research and Development and Promotion Special Project (Soft Science Research) under Grant 222400410105.

ABSTRACT Speech emotion recognition and its precise classification are challenging tasks that heavily depend on the quality of feature extraction and selection for speech signals. Many feature selection algorithms have been proposed to achieve recognition, however, their accuracy has not reached a satisfactory level. We introduce an improved equilibrium optimizer (iEO) algorithm and utilize mel frequency cepstral coefficients (MFCCs) and pitch features for emotion recognition. The transfer function is used to complete the binarization of iEO (BiEO), and the algorithm adopts multi-swarm and transfer functions to balance global search and local search. The performance of the proposed algorithm is verified using four English speech emotion datasets, eNTERFACE05, ryerson audio-visual database of emotional speech and song (RAVDESS), surrey audio-visual expressed emotion (SAVEE) and toronto emotional speech set (TESS). The experimental results illustrate that the proposed algorithm obtains an accuracy of 0.4923, 0.5581, 0,5575 and 0.9840 in eNTERFACE05, RAVDESS, SAVEE and TESS based on K-nearest neighbors, and an accuracy of 0.5279, 0.5862, 0.6752 and 0.9941 based on random forest.

INDEX TERMS Speech emotion recognition, feature selection, equilibrium optimizer, transfer function.

I. INTRODUCTION

Artificial intelligence has recently made significant progress in engineering applications [1], [2]. However, we are still unable to interact with machines naturally due to their inability to understand our emotional states or behaviors [3], [4].

Speech signals include information about speakers' age, gender, religion, origin, and emotional state. They are the best source for effective computing, and they are acquired more easily than other biological signals. There are many conditions for recognizing emotional individuals, such as human-robot interaction, entertainment, business applications, computer games, audio monitoring and call centers. As a result, most researchers place a high priority on speech emotion recognition (SER). The main goal of this research is to develop algorithms and models that can accurately identify and classify happiness, sadness, anger, fear, and more, based on acoustic features present in speech signals [5], [6].

The process of SER involves several steps. First, audio data containing human speech is collected and pre-processed to extract relevant acoustic features, such as pitch, intensity, and spectral information. Next, machine learning techniques build models that can learn and recognize patterns in these acoustic features. Once the models are trained, they are used to predict emotions.

Feature extraction converts raw speech signals into numerous features, and feature selection improves the efficiency and performance of emotion recognition systems. Feature

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

selection identifies the most relevant and discriminative acoustic features, while it is an NP hard task when dealing with a large feature space [7].

Metaheuristic algorithms generally tend to find better solutions through trial-and-error methods [8]. Although metaheuristics can usually discover near-optimal solutions with less computational effort than traditional optimization algorithms, they are not guaranteed to find the optimal solution [9], [10]. Metaheuristic algorithms efficiently explore the search space of feature subsets and reduce computational overhead [11], [12].

Equilibrium optimizer (EO) is a relatively new metaheuristic algorithm inspired by the behavior of individuals in ecosystems [13]. It was proposed by Shahryar Rahnamayan and Hamid R. Tizhoosh in 2018. EO searches for optimal solutions by simulating the interaction among individuals and their equilibrium behavior. Many efforts have been made to adapt EO for feature selection. Using principles from chaos theory, [14] addressed the issues of slow convergence and local optima trapping in the original EO algorithm. Chaotic maps have been added to the optimization process of EO to achieve an effective search. Additionally, eight S- and V-shaped transfer functions are employed to implement feature selection. Reference [15] utilized Entropy to enhance the performance of EO. Levy flight is employed to discover new solutions and improve the algorithm's global search, while local optimality is avoided by various jumps. Through using S- and V-shaped transfer functions, [16] brought binary EO as a powerful and efficient optimization technique for feature selection. Reference [17] combined the ReliefF guided approach with a novel binary EO that utilizes S- and V-shaped transfer functions for efficient feature selection in machine learning and data analysis tasks. Reference [18] advanced EO with Cauchy mutation, opposition-based learning (OBL), and a novel search approach. The population's diversity is enhanced by OBL, and the algorithm's exploration and exploitation are improved by the novel search and the Cauchy mutation. The solutions are converted into binary forms by time-varying S- and V-shaped transfer functions.

Although researchers have applied various algorithms to identify emotions, this task remains challenging. Because there is no theoretical basis to directly represent the features of human voice and associate them with different emotional states. We want to find the most useful emotional features from mel frequency cepstral coefficients (MFCCs) and pitch to improve recognition accuracy. Compared with particle swarm optimization (PSO) and gray wolf optimizer (GWO), EO presents great performance in CEC benchmark functions and engineering applications. It has fast convergence, and it is suitable for high-dimensional optimization. In this study, we investigate EO to recognize speech emotion through feature selection, and the primary contributions of this paper are summarized as follows:

(1) Propose a model for extracting speech emotion features.

(2) Propose an improved EO to implement feature selection.

(3) Propose a multi-swarm to balance the exploration and exploitation of EO. Each sub-swarm adopts a uniquely transfer function to implement searching, and the whole population shares the global information.

(4) Validate the accuracy, precision, recall and F1-Score of the proposed algorithm on four English speech emotion datasets.

The structure of this paper is organized as follows. Section II presents related works on SER. Section III introduces feature extraction, and the proposed feature selection algorithm. Section IV discusses the experimental results, and Section V provides the conclusion of the work.

II. RELATED WORKS

Metaheuristic algorithms have proven to be more reliable in classification tasks. Yogesh et al. developed a novel hybrid algorithm that combines PSO's search ability with BBO's diversity [19]. This hybrid optimization technique efficiently searches the large feature space of speech signals, and selects the most relevant features for emotion and stress recognition. Huang and Epps partitioned speech signals into smaller segments, and extracted acoustic features from them [20]. These partition-based features capture specific emotional cues present in different parts of speech, and provide a more detailed representation of emotional dynamics in continuous speech. By considering phonetic context, they are expected to encompass the influence of specific speech sounds and linguistic elements on emotional expression. Kalhor and Bakhtiari utilized the information from multiple speakers and emotions for classification [21]. Instead of treating each speaker and emotion separately, the model jointly optimize feature selection across multiple tasks, and it identifies the features that are relevant for emotion recognition of different speakers. Yogesh et al. used high order spectral analysis (HOSA) to extract BSF and BCF from speech [22]. By combining these features with the standard 2010 inter-speech features, the performance of a real-time SER system is further enhanced. The selected features generalize the model, and also reduce computational cost. Yildirim et al. modified the initial generation of metaheuristic algorithms [23], and evaluated the method on the nondominated sorting genetic algorithm II (NSGA-II) and the cuckoo search (CS) algorithm for SER.

In addition to metaheuristic algorithms, deep learning techniques have recently been demonstrated their superiority over traditional machine learning methods in SER. Sheikhan et al. employed a hybrid approach to obtain the optimal weights setting and structure of a recurrent neural sentiment classifier with gravity search algorithm (GSA) and its binary version [24]. A rich feature set is constructed by speech signals related to affinity, speech quality and spectrum, and a fast feature selection method chooses more effective features. Mao et al. presented an innovative approach for SER by combining an improved decision tree algorithm, a layered feature selection strategy and a neural network [25]. The layered strategy involves a step-by-step process of feature evaluation and selection, including correlation analysis, mutual information, and other feature ranking methods. Amjad et al investigated the benefits of deep convolutional neural network (DCNN) [26]. They extract features from speech emotion databases using a pre-trained framework, and employ a feature selection approach to find the most influential features for SER. They classify seven emotions using K-nearest neighbors (KNN), support vector machine (SVM), decision tree (DT), random forest (RF), and multilayer perceptron classifier (MLP). SER has limitations in speech misunderstanding, data labeling, and other issues. Kumaran et al. extracted acustic features using MFCCs and gamma tone frequency cepstral coefficients (GFCCs) [27], and then utilized a deep convolutional current neural network to recognize emotions.

Although EO has demonstrated advantages in engineering applications, its research on SER is still limited. Dey et al. proposed a low-cost computational model for classification [28]. They utilize golden ratio optimization (GRO) and EO for feature selection. The input features are optimized for the XGBoost classifier where they are chosen from linear prediction cepstrum coefficient (LPCC) and linear predictive coding (LPC). To improve the accuracy of SER and reduce the burden of computational cost, Bagadi et al. proposed a robust metaheuristic feature selection model where CS and EO find the optimal features [29]. The hybrid algorithm achieves high emotion recognition in EMO-DB and RAVDESS.

III. MATERIALS AND METHOD

This section describes the proposed model, depicted in Figure 1. It includes pre-processing, feature extraction, feature selection and classifiers.

A. PRE-PROCESSING

The pre-processing of audio signals involves a series of steps to prepare and enhance audio data before it is used in speech recognition or music analysis. These steps are essential to improve the quality and reliability of audio data for subsequent analysis.

1) PRE-EMPHASIS

In audio signals, high-frequency components typically contain important information related to speech and other sound characteristics. However, during recording or transmission, low-frequency components tend to carry high energy and can dominate signals. As a result, high-frequency components may become less prominent and harder to detect, leading to the potential loss of valuable information. Pre-emphasis helps address this issue by applying a high-pass filter to amplify the amplitudes of high-frequency components in audio signals.

$$y(n) = x(n) - \alpha y(n-1) \tag{1}$$

where $y(n)$ and $x(n)$ imply the output and input signals at time index n, and α is a constant, which is usually set to 0.97.

2) FRAMING

Framing is a valuable tool for analyzing audio signals because they are non-stationary, meaning their characteristics change over time. In each frame, there is a fixed number of audio samples representing a small portion of the original audio signal. By analyzing these short segments through framing, we can capture temporal changes and extract essential features that are crucial for different audio analysis tasks.

$$N = (L - win\_size)/inc + 1 \tag{2}$$

where L is the length of a signal, win_size is the length of a frame, and inc is frame shift.

3) WINDOWING

Windowing reduces the abrupt changes in amplitude that would otherwise occur at the boundaries of segments. These abrupt changes introduce unwanted frequencies and cause spectral leakage.

Commonly used window functions consist of the Hamming window, Hanning window, Rectangular window, and Blackman window. Each window function has its own characteristics, and the choice of window depends on specific applications.

$$w(n) = 0.54 - 0.46 * cos(\frac{2\pi n}{N-1}) \tag{3}$$

where N is the length of a window, and n is a value within [0, N-1].

B. FEATURE EXTRACTION

Speech signals are divided into frames of 25ms with a 10ms overlap by the Hamming window, and then the fast fourier transform determines the power spectrum of each frame. We extract MFCCs and pitch features from raw audios. A total of 141 features are extracted, and Table 1 describes their details.

1) FAST FOURIER TRANSFORM (FTT)

Fast fourier transform is an algorithm that can efficiently compute discrete Fourier transform (DFT) and its inverses for given data points. FFT significantly reduces the computational complexity of DFT, and it is suitable for real-time applications.

$$X(k) = \sum_{n=0}^{O-1} w(n) * y(n) * e^{-i\frac{2\pi}{O}nk} \tag{4}$$

where $X(k)$ is the frequency spectrum at bin k, and O is the number of samples in a frame.

2) MEL-SCALE FILTER BANK

Mel-scale filter bank is designed to replicate the non-linear relationship between frequency and perceived pitch by

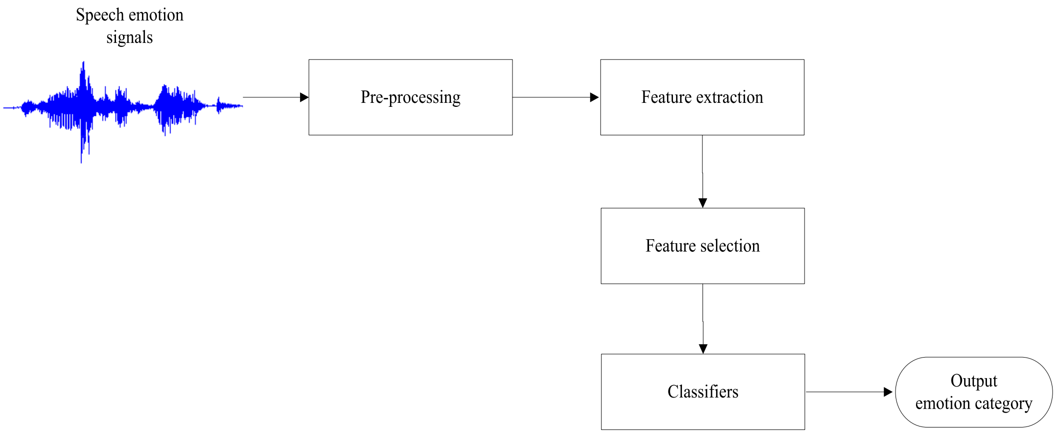


FIGURE 1. The flowchart of the proposed model.

TABLE 1. The details of features.

Features	Details
MFCC	the min, max, mean, median and variance value of each coefficient the min, max, mean, median and variance derivatives of each coefficient spurt length
Pitch	the min, max, mean, median and variance value of each pitch the min, max, mean, median and variance derivatives of each pitch

human ears. Mel-scale is mathematically defined based on psychoacoustic studies that analyze human auditory perception.

$$Mel(f) = 2595log_{10}(1 + \frac{f}{700}) \tag{5}$$

where f is frequency.

$$H_m(k) = \begin{cases} \frac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) <= k <= f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & f(m) <= k <= f(m+1) \\ 0 & else \end{cases} \tag{6}$$

3) LOG ENERGY

Log energy is a common feature in speech recognition, and it produces valuable information about the intensity and loudness of audio signals within each time frame. Combined with other features like MFCCs and delta features, log energy builds effective and robust audio signal processing systems.

$$s(m) = ln(\sum_{k=0}^{N-1} X(k)^2 H_m(k)), 0 <= m <= M \tag{7}$$

4) DISCRETE COSINE TRANSFORM (DCT)

Discrete cosine transform converts a sequence of data points, typically in the spatial or time domain, into cosine function coefficients in the frequency domain. The main advantage of DCT is that it can concentrate most of the signal energy into a few low-frequency coefficients, while involve less valuable messages in high-frequency coefficients. Through quantizing and discarding less significant coefficients, DCT achieves efficient data compression and preserves essential features for human perception.

$$C(n) = \sum_{m=0}^{N-1} s(m)cos(\frac{\pi(m-0.5)n}{N}), n = 1, 2, \ldots, L \tag{8}$$

where L determines the number of MFCCs, and it is set to 13 in this paper.

5) MEL-FREQUENCY CEPSTRAL COEFFICIENTS

MFCCs are powerful features for audio and speech processing tasks because they effectively capture essential spectral information while reducing the dimensionality of feature space. They have been widely adopted in SER, speaker identification, music genre classification, and various other audio-related applications due to their effectiveness and robustness. MFCCs play a crucial role in advancing the field of audio signal processing, and they are a standard feature representation in audio processing systems. Figure 2 exhibits the model of MFCC.

C. IMPROVED EQUILIBRIUM OPTIMIZER FOR FEATURE SELECTION

To obtain the optimal features from MFCCs and pitch features, we need to calculate 2^{141} - 1 times, which is almost impossible. Therefore, we utilize the improved EO as a feature selector to acquire an approximate optimal solution within an acceptable time range.

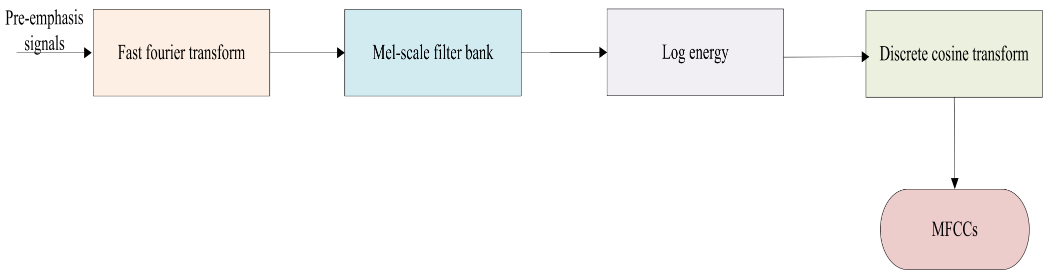


FIGURE 2. The model of MFCC.

1) EQUILIBRIUM OPTIMIZER

Equilibrium optimizer randomly initializes the positions of the population and its position update is defined as follows.

$$X_i(n+1)=X_{eq}(n)+(X_i(n)-X_{eq}(n))F(n)+\frac{G(n)}{\lambda}(1-F(n)) \tag{9}$$

where X_{eq} represents the equilibrium pool, and it is constructed by the positions of the first four optimal solutions and their average value. The algorithm randomly chooses one from X_{eq} for each run.

F controls the balance between exploration and exploitation, as shown in Eq. (11).

$$t(n)=(1-\frac{n}{Max\_iter})^{(2\frac{n}{Max\_iter})} \tag{10}$$

$$F(n)=sign(r-0.5)[e^{-\lambda t(n)}-1] \tag{11}$$

where Max_iter denotes the maximum iteration. λ and r are two random numbers within the rage [0,1]. *Sign* is the signum function of Matlab. G assists the algorithm in acquiring better performance, and it is computed as follows.

$$GCP=\begin{cases}0.5r_1 & if\,(r_2\geq GP)\\ 0 & else\end{cases} \tag{12}$$

$$G_0(n)=GCP*(X_{eq}(n)-X_i(n)) \tag{13}$$

$$G(n)=G_0(n)*F(n) \tag{14}$$

where r_1 and r_2 are two random values between [0,1].

2) IMPROVED EQUILIBRIUM OPTIMIZER

In EO, solutions are guided by the equilibrium pool. However, the presence of the four optimal solutions, which might be located at a local optimum, can potentially trap the while population. Alternatively, if these solutions are scattered in different search regions, they can hinder convergence.

Exploration and exploitation are two critical aspects used to evaluate the effectiveness of metaheuristic algorithms. Improved exploration enhances the EO's global search ability and aids in escaping from local optima. Additionally, better exploration empowers the algorithm with strong local search ability and promotes it to thoroughly exploit promising regions and discover the optimal solution. At the beginning of the proposed BiEO, the population is divided into three sub-swarms, as illustrated in Figure 3. The first sub-swarm focuses on exploration ability, and the second sub-swarm maintains population diversity and convergence. The last sub-swarm is responsible for exploitation ability, and the sub-swarms share valuable information.

The transfer function plays a significant role in binary metaheuristic algorithms, and it is responsible for converting continuous values into a binary string. A well-designed transfer function can help maintain diversity in the population, prevent premature convergence, and ensure effective exploration of search space. We utilize three transfer functions (as shown in Figure 4) and Eq. (15) to implement binarization.

$$X_i^j(n+1)=\begin{cases}X_i^j(n) & if\,(S(value)<rand)\\ 1-X_i^j(n) & else\end{cases} \tag{15}$$

$$value=(X_i(n)-X_{eq}(n))F(n)+\frac{G(n)}{\lambda}(1-F(n)) \tag{16}$$

where S represents the transfer function, and $X_i^j(n)$ means the position of individual i in the j-th dimension at the n-th iteration.

In the first sub-swarm, the equilibrium pool is constructed by the sub-swarm, and it adopts S_1 as its transfer function where S_1 quickly switches positions. This sub-swarm explores more space, and it has excellent global search ability. In the second sub-swarm, it adopts S_2 as its transfer function. It has the advantages of EO, and it well balances global search and local search. In the third sub-swarm, X_{eq} is constituted of the global optimal solution, and this sub-swarm adopts S_3 as its transfer function where S_3 slowly changes positions. The sub-swarm exploits optimal solutions, and it has great local search ability.

D. CLASSIFIERS

KNN and RF establish classification models, and K-fold cross validation evaluates the performance of the models.

1) K-NEAREST NEIGHBOR

KNN is a non-parametric algorithm that doesn't assume any specific data distribution. Instead, it relies on the similarity or distance among data points to make predictions. When predicting a new data point, KNN finds the K nearest data points in a training set based on a distance metric (like Euclidean distance), and then determines the majority class

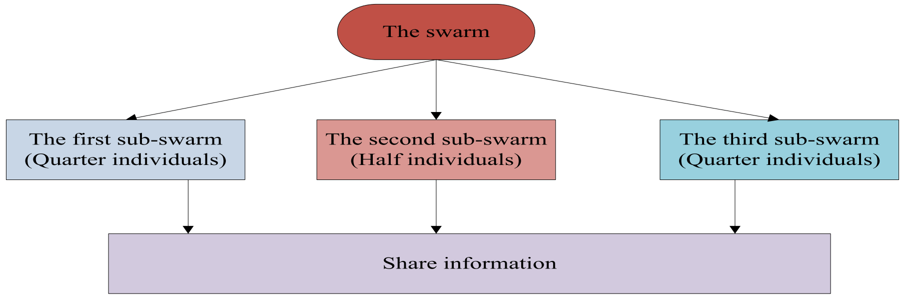


FIGURE 3. The division of multi-swarm.

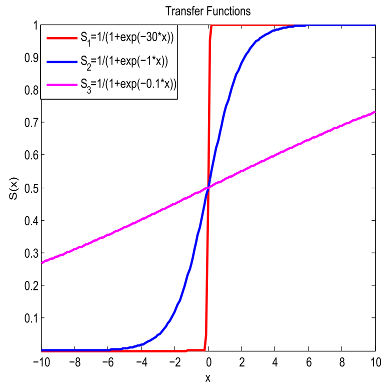


FIGURE 4. Transfer functions.

among those K neighbors.

$$d(x, x') = \sqrt[2]{\sum_{i=1}^{n}(x(i) - x'(i))^2} \tag{17}$$

where x and x' represent training and test data, and n is the number of features.

2) RANDOM FOREST

RF creates multiple decision trees during a training phase and combines their predictions to produce the final prediction. Each decision tree is built using random data and features. This randomness introduces diversity among decision trees and improves the model's generalization.

3) K-FOLD CROSS VALIDATION

K-fold cross validation randomly divides a original dataset into K folds. A classification model is then trained and tested K times where each fold is used as a test set and the remaining K-1 folds are employed as a training set.

In this study, we use 5-NN and 10-fold cross validation, and the number of decision trees is set to 20.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATASETS DESCRIPTION

In this paper, eNTERFACE05, ryerson audio-visual database of emotional speech and song (RAVDESS), surrey audio-visual expressed emotion (SAVEE) and toronto emotional speech set (TESS) are employed to evaluate the emotion recognition of the proposed algorithm.

1) eNTERFACE05

eNTERFACE05 contains both audio and visual data. The dataset records a range of emotional expressions, from basic emotions like happiness, sadness, anger, and surprise to more complex emotional states. This multimodal nature allows researchers to investigate how emotions are conveyed through both speech and facial cues.

2) RYERSON AUDIO-VISUAL DATABASE OF EMOTIONAL SPEECH AND SONG

This dataset includes an array of audio and video recordings featuring performances by professional actors [29]. These skilled actors adeptly portray a diverse spectrum of emotions, including, but not limited to, sadness, anger, happiness, fear, calmness, surprise, and neutrality. These emotional expressions are thoughtfully captured at varying levels of intensity, and provide an extensive and nuanced resource for in-depth emotion analysis.

3) SURREY AUDIO-VISUAL EXPRESSED EMOTION

SAVEE contains recordings of sentences spoken in seven emotional expressions: anger, fear, disgust, sadness, happiness, surprise, and neutral [28]. Each emotion is portrayed by multiple actors, ensuring a diverse set of emotional variations. The database includes recordings from four male and four female actors, which balances the gender representation of emotional expression.

4) TORONTO EMOTIONAL SPEECH SET

TESS includes recordings of sentences spoken with happiness, sadness, anger, disgust, fear, surprise, and neutrality. One notable feature of TESS is the careful selection of actors to portray these emotions. Professional actors are chosen to ensure that the emotional expressions are portrayed convincingly and consistently across the dataset, and this method enhances the dataset's credibility and utility for emotion analysis research.

TABLE 2. The main parameters setting of the compared algorithms.

Algorithm	Main parameters
EO	a1=2; a2=1; GP=0.5;
BBO_PSO	KeepRate = 0.2; alpha = 0.9; pMutation = 0.1; w = 0.9; wdamp = 0.99; c1 = 2; c2 = 2;
MOBFA	N_{re} = 4; N_{ed} = 2; P_{ed} = 0.25; C_i = 0.05;
BiEO	a1=2; a2=1; GP=0.5;

TABLE 3. The experimental results of the compared algorithms.

Dataset	EO		BBO_PSO		MOBFA		BiEO	
	Accuracy	Length	Accuracy	Length	Accuracy	Length	Accuracy	Length
eNTERFACE05	0.4158	66.4	0.4042	66.15	0.3935	69	0.4923	46.2
RAVDESS	0.4674	62.8	0.4036	70.1	0.3497	66.65	0.5581	62.9
SAVEE	0.5230	70.1	0.5252	67.6	0.5113	70.45	0.5575	52.4
TESS	0.9507	66.15	0.8534	66.4	0.8620	71.7	0.9840	55.65
>/=/<	0/0/4		0/0/4		0/0/4		4/0/0	
Rank	2.25		3		3.75		1	
P-Value	0.0194							

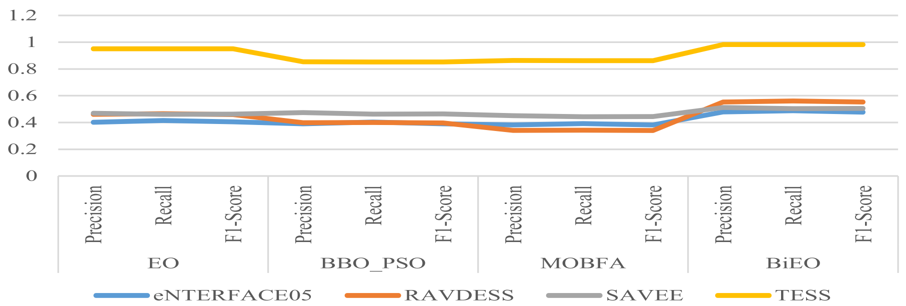


FIGURE 5. The precision, recall and F1-Score of the algorithms.

B. EXPERIMENTAL SETUP

To validate the superiority of the proposed BiEO, the classification performance is compared with EO [28], BBO_PSO [22] and MOBFA [30]. EO is utilized to test whether the proposed BiEO enhances the performance of classical EO. BBO_PSO and MOBFA are used for emotion classification, so they can test the effectiveness of the proposed emotion recognition model. Table 2 presents further details of the algorithms.

The maximum iterations of the algorithms is set to 100, while the population size is 20. The experimental data obtained is the results of repeating 20 times. To evaluate the statistical significance of experimental results, we employ Wilcoxon rank sum and Friedman test at a significance level of 0.05.

C. OBJECTIVE FUNCTION

Classification accuracy and the number of selected features are considered in feature selection [31], while in SER, classification accuracy is the main indicator for evaluating algorithms. Therefore, it is used as the objective function in the experiments, as shown in Eq. (18). We compare the algorithms in precision, recall, F1-Score (as shown in Eqs. (19-21)) [32], the number of selected features and running time.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

$$precision = \frac{TP}{TP + FP} \tag{19}$$

$$recall = \frac{TP}{TP + FN} \tag{20}$$

$$F1 - Score = \frac{2 * TP}{2 * TP + FP + FN} \tag{21}$$

D. EXPERIMENTAL ANALYSIS

1) SIMULATION RESULTS ON THE KNN CLASSIFIER

Table 3 presents the results using the KNN classifier where *Accuracy* represents prediction accuracy and *Length* means the number of selected features.

The recognition accuracy of BiEO in four datasets is 0.492, 0.558, 0.557 and 0.984, respectively, and BiEO is superior to EO. The proposed multi-swarm improves the classification accuracy of EO and can be applied to English SER. BBO_PSO obtains the accuracy of 0.404, 0.404, 0.525 and 0.853, and MOBFA acquires the accuracy of 0.394, 0.350, 0.511, 0.862. BiEO outperforms EO, BBO_PSO and MOBFA in eNTERFACE05, RAVDESS, SAVEE and TESS. The Wilcoxon rank sum reveals that the algorithms have no

eNTERFACE05	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	71%	7%	5%	2%	6%	8%
Disgust	19%	44%	9%	9%	11%	9%
Fear	15%	18%	31%	9%	13%	14%
Happiness	2%	9%	10%	66%	5%	8%
Sadness	15%	16%	14%	6%	34%	15%
Surprise	10%	8%	15%	9%	13%	46%

(a) eNTERFACE05

RAVDESS	Angry	Disgust	Fearful	Happy	Neutral	Sad	Surprised	Calm
Angry	67%	8%	8%	8%	1%	2%	4%	2%
Disgust	8%	50%	7%	7%	3%	9%	4%	12%
Fearful	6%	8%	59%	8%	3%	8%	3%	5%
Happy	9%	9%	15%	46%	4%	5%	8%	3%
Neutral	0%	3%	3%	5%	62%	7%	2%	18%
Sad	3%	8%	12%	7%	10%	42%	3%	15%
Surprised	7%	9%	8%	14%	6%	5%	48%	1%
Calm	0%	3%	2%	1%	10%	10%	0%	75%

(b) RAVDESS

SAVEE	Angry	Disgust	Fearful	Happy	Neutral	Sad	Surprised
Angry	46%	13%	4%	19%	14%	2%	4%
Disgust	3%	46%	3%	8%	29%	8%	4%
Fearful	6%	2%	33%	18%	1%	10%	29%
Happy	7%	3%	14%	47%	2%	12%	15%
Neutral	2%	6%	1%	0%	87%	4%	0%
Sad	1%	9%	3%	7%	23%	55%	1%
Surprised	4%	2%	30%	21%	0%	6%	38%

(c) SAVEE

TESS	Angry	Disgust	Fearful	Happy	Neutral	Sad	Surprised
Angry	100%	0%	0%	0%	0%	0%	0%
Disgust	0%	99%	0%	0%	0%	0%	0%
Fearful	0%	0%	100%	0%	0%	0%	0%
Happy	0%	0%	0%	98%	0%	0%	1%
Neutral	0%	0%	0%	0%	99%	1%	0%
Sad	0%	0%	0%	0%	1%	99%	0%
Surprised	0%	1%	0%	7%	0%	0%	92%

(d) TESS

FIGURE 6. **Confusion matrices.**

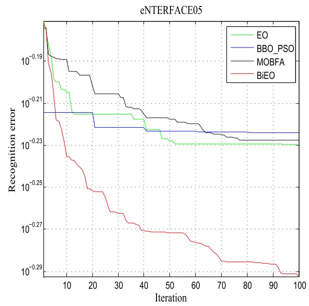
TABLE 4. **The average running time of the compared algorithms (Second).**

Dataset	EO	BBO_PSO	MOBFA	BiEO
eNTERFACE05	168.4900	170.8084	170.3729	**163.186**
RAVDESS	415.8207	418.4271	410.4288	**377.9914**
SAVEE	176.2913	177.9443	177.5689	**169.0756**
TESS	1174.5043	1189.2512	1204.1097	**997.4129**

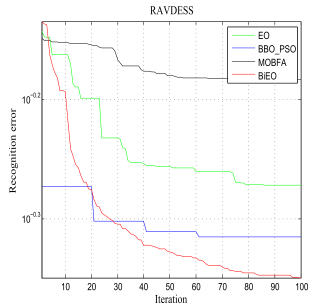
similar experimental data in the four datasets. The Friedman test presents that their average ranks are 2.25, 3, 3.75 and 1, respectively, with P-values less than 0.05. The superiority of BiDE is confirmed by the Wilcoxon rank sum and the Friedman test.

BiEO achieves emotion recognition with a minimal number of features. Interestingly, EO, BBO_PSO and MOBFA utilize more features than BiEO, but their classification accuracy doesn't surpass BiEO. This suggests that, in feature selection, using too more features may not conducive to improving classification performance.

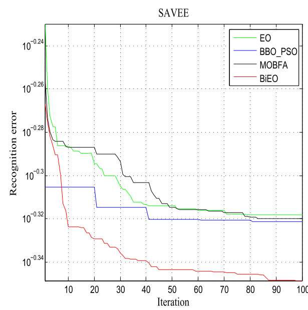
Figure 5 depicts the precision, recall and F1-Score of the algorithms. The results in TESS are better than in eNTERFACE05, RAVDESS and SAVEE. EO and BBO_PSO obtain the best data in TESS, followed by SAVEE, RAVDESS and eNTERFACE05. For MOBFA, the precision, recall and F1-Score in RAVDESS are better than in eNTERFACE05. For BiEO, the results in SAVEE are inferior to those in RAVDESS. Overall, BiEO exhibits the best performance in precision, recall and F1-Score.



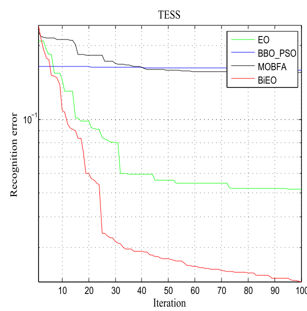
(a) eNTERFACE05



(b) RAVDESS



(c) SAVEE



(d) TESS

FIGURE 7. **Convergence curves.**

Figure 6 shows the confusion matrices of BiEO. A confusion matrix contains the information about the actual and predicted classification of a SER system. In eNTERFACE05, BiEO obtains a high recognition accuracy on Anger and Happiness, but a low accuracy on Fear and Sadness. In RAVDESS, BiEO has an accuracy of 75% on Clam, while its accuracy on Happy, Sad, and Surprised is less than 50%. In SAVEE, BiEO has higher classification accuracy on Neutral than other emotions. In TESS, BiEO achieves high emotional recognition.

TABLE 5. The experimental results of the compared algorithms.

Dataset	EO		BBO_PSO		MOBFA		BiEO	
	Accuracy	Length	Accuracy	Length	Accuracy	Length	Accuracy	Length
eNTERFACE05	0.5174	68	0.5248	62.6	0.5140	73	**0.5279**	56.8
RAVDESS	0.5458	74.4	0.5584	69	0.5420	70	**0.5862**	49.8
SAVEE	0.6465	71.4	0.6545	68.2	0.6487	70.6	**0.6752**	48.4
TESS	0.9908	68.6	0.9924	71	0.9915	69.2	**0.9941**	56.6
>/=/<	0/1/3		0/2/2		0/1/3		4/0/0	
Rank	3.5		2		3.5		1	
P-Value	0.0129							

TABLE 6. The average running time of the compared algorithms (Second).

Dataset	EO	BBO_PSO	MOBFA	BiEO
eNTERFACE05	6305.9544	6260.3155	6290.8553	**5990.6368**
RAVDESS	21050.7790	21124.4899	21175.2304	**19343.1022**
SAVEE	7135.4514	7031.1318	7116.9895	**6611.9405**
TESS	25160.5714	26345.8442	24241.2402	**22994.5546**

TABLE 7. The experimental results of the compared algorithms in EMO-DB.

Algorithm	Accuracy	Precision	Recall	F1-Score	Length	Time
EO	0.5554	0.5434	0.5268	0.5175	68.15	124.0763
BBO_PSO	0.4622	0.4447	0.4314	0.4203	68.4	117.2673
MOBFA	0.4602	0.4463	0.4298	0.42	69.25	114.783
BiEO	**0.6775**	**0.6633**	**0.6535**	**0.6476**	**54**	**112.5669**
EO	0.7503	0.7518	0.726	0.7194	69.6	10962.1237
BBO_PSO	0.7563	0.76	0.7315	0.7256	75.8	10951.0444
MOBFA	0.7525	0.7532	0.7253	0.7189	73.4	11081.0599
BiEO	**0.7636**	**0.763**	**0.7355**	**0.7276**	**66.4**	**10315.633**

Figure 7 displays the convergence curves of the algorithms. It can be observed that BiEO has the fastest convergence speed in eNTERFACE05, RAVDESS, SAVEE and TESS, and it is always searching for the optimal solution. BiEO has the ability to escape local traps in emotion recognition. BBO_PSO includes two operations, BBO and PSO, so it updates the global optimum every 20 iterations. In eNTERFACE05, RAVDESS and TESS, the convergence rates of EO are better than BBO_PSO and MOBFA, but not as excellent as BiEO, which shows that the proposed method improves the performance of EO.

Table 4 illustrates the running time of the algorithms. Notably, BiEO demonstrates superior efficiency compared to other algorithms in eNTERFACE05, RAVDESS, SAVEE, and TESS. The number of selected features is a key factor affecting the efficiency of feature selection algorithms. BiEO stands out by utilizing a minimal number of features to accomplish classification. The algorithms exhibit high execution efficiency in eNTERFACE05 and SAVEE, but they run the slowest in TESS. Because TESS contains many samples, while eNTERFACE05 and SAVEE have a small amount of data.

From the above discussion, it's clear that BiDE excels in classification accuracy, precision, recall, F1-Score, the number of selected features and running time. Consequently, it is a highly suitable choice for English SER.

2) SIMULATION RESULTS ON THE RF CLASSIFIER

Table 5 presents recognition accuracy and the number of selected features through the RF classifier.

The accuracy obtained with RF is superior to that acquired by KNN. Table 5 illustrates that BiEO yields the best results in eNTERFACE05, RAVDESS, SAVEE and TESS. Its accuracy in the four emotion datasets is 0.5279, 0.5862, 0.6752 and 0.9941, and it outperforms EO, BBO_PSO and MBFOA. Through the Wilcoxon rank sum, EO, BBO_PSO, MBFOA and BiEO perform well in 1, 2, 1 and 4 datasets, respectively. In eNTERFACE05, their statistical data appears similar, and the Wilcoxon rank sum is unable to differentiate the experimental data of BBO_PSO and BiEO in TESS. The Friedman test reveals that the average ranks of EO, BBO_PSO, MBFOA and BiEO are 3.5, 2, 3.5 and 1, and the P-Value is 0.0129. EO is superior to MOBFA in eNTERFACE05 and RAVDESS, but MOBFA excels EO in the other databasets. BiEO also utilizes the fewest features to implement recognition. The features obtained by EO, BBO_PSO and MBOFA are higher than BiEO in the datasets. Table 5 provides the evidence of the superiority of MDE in English SER.

Figure 8 presents the precision, recall and F1-Score of the algorithms. The algorithms have the best precision, recall and F1-score in TESS, followed by SAVEE, RAVDESS and eNTERFACE05. BiEO demonstrates superior performance in eNTERFACE05, RAVDESS and SAVEE, and BBO_PSO

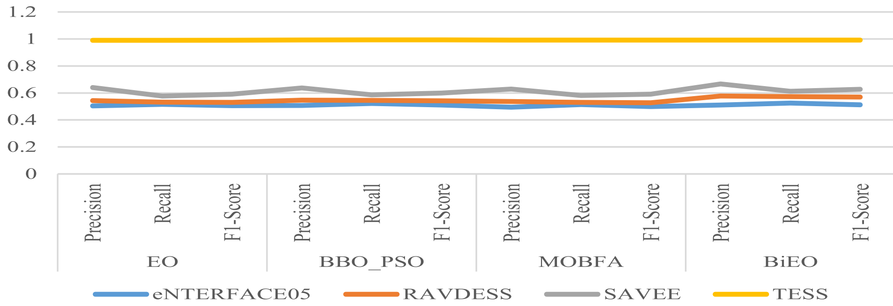


FIGURE 8. The precision, recall and F1-Score of the algorithms.

TABLE 8. Comparison the results with existing methods.

Reference	Dataset	classifier	Accuracy
Ref. [33]	eNTERFACE05	KNN	41%
Ref. [35]	RAVDESS	KNN	80%
Ref. [26]	RAVDESS	KNN	69%
Ref. [26]	RAVDESS	RF	76%
Ref. [36]	RAVDESS	KNN	61%
Ref. [33]	SAVEE	KNN	57 %
Ref. [26]	SAVEE	RF	62%
Ref. [36]	SAVEE	KNN	56%
Ref. [34]	TESS	KNN	93%

outperforms the other algorithms in TESS. Figure 8 illustrates that BiEO selects the most relevant features in the speech emotion datasets, and it provides a balance between precision and recall.

Figure 9 presents the confusion matrices of BiEO. In eNTERFACE05, the emotions of Disgust, Fear, and Sadness have higher confusion compared to Anger and Happiness. In RAVDESS, Calm has the highest recognition accuracy, while Neutral is the smallest. In SAVEE, Happy, and Neutral affect the recognition of Angry, and Neutral causes serious interference of Disgust. In TESS, BiEO achieves 100% recognition accuracy in Angry, Fearful and Sad emotions.

Figure 10 depicts the convergence cures of BiEO. In eNTERFACE05, BBO_PSO has fast convergence performance in the first 70 iterations, but BiEO performs better than other algorithms after that. BiEO not only has excellent convergence, but also has global optimization. In RAVDESS and SAVEE, the convergence curves of BiEO are better than EO, BBO_PSO and MOBFA, followed by BBO_PSO. In TESS, BiEO has the fast convergence rate, followed by BBO_PSO, MOBFA and EO.

Table 6 is the running time of the algorithms. The time of the algorithms on RF is significantly higher than that on KNN, because RF has higher time complexity than KNN. BiEO exhibits the fastest execution efficiency, and the time difference of the other algorithms is not significant. Among the algorithms, eNTERFACE05 and SAVEE require longer running time, whereas RAVDESS and TESS execute more quickly.

eNTERFACE05	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	75%	7%	3%	2%	7%	6%
Disgust	14%	40%	13%	13%	13%	7%
Fear	12%	14%	30%	19%	12%	13%
Happiness	1%	7%	10%	73%	4%	5%
Sadness	16%	14%	12%	5%	44%	10%
Surprise	10%	8%	10%	9%	11%	53%

(a) eNTERFACE05

RAVDESS	Angry	Disgust	Fearful	Happy	Neutral	Sad	Surprised	Calm
Angry	71%	9%	3%	6%	0%	3%	6%	1%
Disgust	8%	52%	5%	7%	2%	8%	8%	10%
Fearful	8%	8%	57%	9%	2%	8%	4%	6%
Happy	10%	10%	14%	42%	2%	5%	12%	4%
Neutral	1%	7%	3%	5%	41%	13%	2%	28%
Sad	1%	10%	11%	7%	4%	43%	6%	18%
Surprised	3%	6%	4%	10%	1%	5%	69%	2%
Calm	0%	3%	1%	1%	4%	7%	0%	83%

(b) RAVDESS

SAVEE	Angry	Disgust	Fearful	Happy	Neutral	Sad	Surprised
Angry	39%	7%	4%	23%	19%	5%	3%
Disgust	1%	49%	3%	4%	30%	9%	3%
Fearful	1%	4%	53%	8%	4%	13%	16%
Happy	4%	4%	9%	59%	2%	5%	17%
Neutral	0%	1%	0%	1%	93%	5%	0%
Sad	0%	4%	2%	5%	17%	72%	0%
Surprised	0%	0%	18%	15%	1%	3%	63%

(c) SAVEE

TESS	Angry	Disgust	Fearful	Happy	Neutral	Sad	Surprised
Angry	100%	0%	0%	0%	0%	0%	0%
Disgust	0%	99%	0%	0%	0%	0%	0%
Fearful	0%	0%	100%	0%	0%	0%	0%
Happy	0%	0%	0%	98%	0%	0%	1%
Neutral	0%	0%	0%	0%	99%	1%	0%
Sad	0%	0%	0%	0%	1%	99%	0%
Surprised	0%	1%	0%	7%	0%	0%	92%

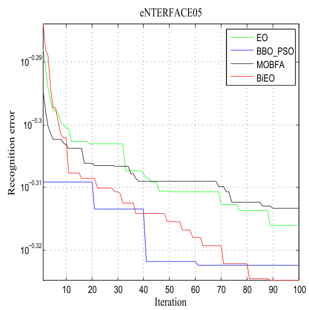
(d) TESS

FIGURE 9. Confusion matrices.

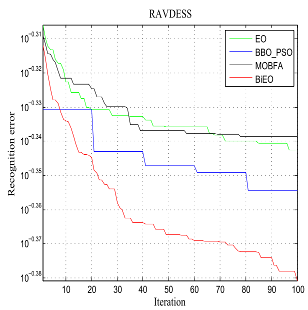
The experimental results of the algorithms on KNN and RF classifiers demonstrate that BiEO has the best performance, and the multi-swarm approach utilize fewer features to promote the search for optimal solutions. It improves population diversity and preserves the opportunity to find the optimal solution. The data from BiEO proves that it is suitable for English emotion recognition.

3) DISCUSSION

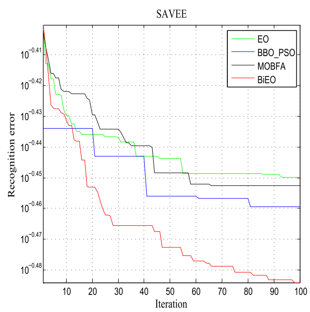
In order to further test the effectiveness of the proposed algorithm, we further analyze the performance of BiEO on other datasets and state-of-the-art algorithms. Table 7 describes the experimental results of BiEO in EMO-DB [28] where the first four rows are the results of KNN,



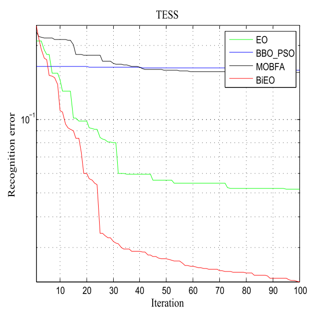
(a) eNTERFACE05



(b) RAVDESS



(c) SAVEE



(d) TESS

FIGURE 10. Convergence curves.

and the last four rows are the results of RF. The table reveals that the classification accuracy of BiEO is better than EO and BBO_PSO and MOBFA. The Wilcoxon rank sum indicates that BiEO and other algorithms have no data similarity. In precision, recall, F1-Score, the number of selected features and running time, BiEO also demonstrates excellent performance. This indicates that the proposed method is not only suitable for emotional recognition in English language, but also for recognition in other languages. The algorithm has the characteristic of robustness. Table 8 is the results with existing methods. From the Tables 3 and 5, we known that the proposed algorithm obtains an accuracy of 49%, 55%, 56% and 98% in eNTERFACE05, RAVDESS, SAVEE and TESS based on KNN, and an accuracy of 52%, 58%, 67% and 99% based on RF. The algorithm is superior to [26], [33], and [34] in eNTERFACE05 and TESS, but BiEO has poor performance in RAVDESS. The algorithm identifies emotional features that are more prominent or well-reported in the eNTERFACE05 and TESS datasets. The algorithm may not perform well when RAVDESS emphasizes diverse emotional cues.

V. CONCLUSION

SER research involves identifying emotional features using feature selection methods. However, it is infeasible to search all subsets. In this paper, we propose an improved equilibrium optimizer for feature selection to recognize speech emotion. Feature extraction is first performed on speech signals, and then BiEO determines the meaningful acoustic features. The efficacy of this algorithm lies in the abilities of multi-swarm to systematically search for emotional feature space and identify optimal features that contain intricate patterns for accurate emotion classification.

The performance of the proposed algorithm has been tested in eNTERFACE05, RAVDESS, SAVEE, and TESS with EO, BBO_PSO, and MOBFA. Based on the results of KNN and RF classifiers, BiEO exhibits excellent performance in accuracy, precision, recall, F1- Score, the number of selected features and running time. To further verify the robustness of the algorithm, we compare it with other algorithms and in EMO-DB database. The results reveal that the algorithm has excellent emotion recognition performance.

In the future, we can apply our proposed algorithm to real speech recognition scenarios and various corpus datasets. Although KNN runs fast, its performance is not as good as RF. We try to find a classifier suitable for emotion recognition and the proposed algorithm.

DATA AVAILABILITY

Data is available on request.

REFERENCES

[1] N. S. Murugan and G. U. Devi, "Detecting streaming of Twitter spam using hybrid method," *Wireless Pers. Commun.*, vol. 103, no. 2, pp. 1353–1374, Nov. 2018.

[2] N. S. Murugan and G. U. Devi, "Detecting spams in social networks using ML algorithms—A review," *Int. J. Environ. Waste Manage.*, vol. 21, no. 1, pp. 22–36, 2018.

[3] C. Zhang and L. Xue, "Autoencoder with emotion embedding for speech emotion recognition," *IEEE Access*, vol. 9, pp. 51231–51241, 2021.

[4] S. M. Nagarajan and U. D. Gandhi, "Classifying streaming of Twitter data based on sentiment analysis using hybridization," *Neural Comput. Appl.*, vol. 31, no. 5, pp. 1425–1433, May 2019.

[5] L.-M. Zhang, G. W. Ng, Y.-B. Leau, and H. Yan, "A parallel-model speech emotion recognition network based on feature clustering," *IEEE Access*, vol. 11, pp. 71224–71234, 2023.

[6] N. Liu, B. Zhang, B. Liu, J. Shi, L. Yang, Z. Li, and J. Zhu, ''Transfer subspace learning for unsupervised cross-corpus speech emotion recognition,'' *IEEE Access*, vol. 9, pp. 95925–95937, 2021.

[7] J. Deng, S. Frühholz, Z. Zhang, and B. Schuller, ''Recognizing emotions from whispered speech based on acoustic feature transfer learning,'' *IEEE Access*, vol. 5, pp. 5235–5246, 2017.

[8] J.-S. Pan, Q. Liang, S.-C. Chu, K.-K. Tseng, and J. Watada, ''A multi-strategy surrogate-assisted competitive swarm optimizer for expensive optimization problems,'' *Appl. Soft Comput.*, vol. 147, Nov. 2023, Art. no. 110733.

[9] J. Xi, Y. Chen, X. Liu, and X. Chen, ''Whale optimization algorithm based on nonlinear adjustment and random walk strategy,'' *J. Netw. Intell.*, vol. 7, no. 2, pp. 306–318, 2022.

[10] T.-T. Nguyen, T. Dong-Nguyen, and V.-T. Nguyen, ''An optimizing pulse coupled neural network based on golden eagle optimizer for automatic image segmentation,'' *J. Inf. Hiding Multimedia Signal Process.*, vol. 13, no. 3, pp. 155–164, 2022.

[11] E. Elhariri, N. El-Bendary, and S. A. Taie, ''Using hybrid filter-wrapper feature selection with multi-objective improved-salp optimization for crack severity recognition,'' *IEEE Access*, vol. 8, pp. 84290–84315, 2020.

[12] C.-Y. Lee, T.-A. Le, and Y.-T. Lin, ''A feature selection approach hybrid grey wolf and heap-based optimizer applied in bearing fault diagnosis,'' *IEEE Access*, vol. 10, pp. 56691–56705, 2022.

[13] J.-S. Pan, P. Hu, T.-S. Pan, and S.-C. Chu, ''Improved equilibrium optimizer for short-term traffic flow prediction,'' *J. Database Manage.*, vol. 34, no. 1, pp. 1–20, Apr. 2023.

[14] G. I. Sayed, G. Khoriba, and M. H. Haggag, ''A novel chaotic equilibrium optimizer algorithm with S-shaped and V-shaped transfer functions for feature selection,'' *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 6, pp. 3137–3162, Jun. 2022.

[15] Z. A. Varzaneh, S. Hossein, S. Ebrahimi Mood, and M. M. Javidi, ''A new hybrid feature selection based on improved equilibrium optimization,'' *Chemometric Intell. Lab. Syst.*, vol. 228, Sep. 2022, Art. no. 104618.

[16] Y. Gao, Y. Zhou, and Q. Luo, ''An efficient binary equilibrium optimizer algorithm for feature selection,'' *IEEE Access*, vol. 8, pp. 140936–140963, 2020.

[17] M. Zhang, J.-S. Wang, J.-N. Hou, H.-M. Song, X.-D. Li, and F.-J. Guo, ''RG-NBEO: A ReliefF guided novel binary equilibrium optimizer with opposition-based S-shaped and V-shaped transfer functions for feature selection,'' *Artif. Intell. Rev.*, vol. 56, no. 7, pp. 6509–6556, Jul. 2023.

[18] A. M. Vommi and T. K. Battula, ''A hybrid filter-wrapper feature selection using fuzzy KNN based on Bonferroni mean for medical datasets classification: A COVID-19 case study,'' *Expert Syst. Appl.*, vol. 218, May 2023, Art. no. 119612.

[19] M. Hariharan, R. Ngadiran, A. H. Adom, S. Yaacob, C. Berkai, and K. Polat, ''A new hybrid PSO assisted biogeography-based optimization for emotion and stress recognition from speech signal,'' *Expert Syst. Appl.*, vol. 69, pp. 149–158, Mar. 2017.

[20] Z. Huang and J. Epps, ''An investigation of partition-based and phonetically-aware acoustic features for continuous emotion prediction from speech,'' *IEEE Trans. Affect. Comput.*, vol. 11, no. 4, pp. 653–668, Oct. 2020.

[21] E. Kalhor and B. Bakhtiari, ''Speaker independent feature selection for speech emotion recognition: A multi-task approach,'' *Multimedia Tools Appl.*, vol. 80, no. 6, pp. 8127–8146, Mar. 2021.

[22] M. Hariharan, R. Ngadiran, A. H. Adom, S. Yaacob, and K. Polat, ''Hybrid BBO_PSO and higher order spectral features for emotion and stress recognition from natural speech,'' *Appl. Soft Comput.*, vol. 56, pp. 217–232, Jul. 2017.

[23] S. Yildirim, Y. Kaya, and F. Kılıç, ''A modified feature selection method based on metaheuristic algorithms for speech emotion recognition,'' *Appl. Acoust.*, vol. 173, Feb. 2021, Art. no. 107721.

[24] M. Sheikhan, M. Abbasnezhad Arabi, and D. Gharavian, ''Structure and weights optimisation of a modified Elman network emotion classifier using hybrid computational intelligence algorithms: A comparative study,'' *Connection Sci.*, vol. 27, no. 4, pp. 340–357, Oct. 2015.

[25] Q. Mao, X. Wang, and Y. Zhan, ''Speech emotion recognition method based on improved decision tree and layered feature selection,'' *Int. J. Humanoid Robot.*, vol. 7, no. 2, pp. 245–261, Jun. 2010.

[26] A. Amjad, L. Khan, and H.-T. Chang, ''Effect on speech emotion classification of a feature selection approach using a convolutional neural network,'' *PeerJ Comput. Sci.*, vol. 7, p. e766, Nov. 2021.

[27] U. Kumaran, S. R. Rammohan, S. M. Nagarajan, and A. Prathik, ''Fusion of mel and gammatone frequency cepstral coefficients for speech emotion recognition using deep C-RNN,'' *Int. J. Speech Technol.*, vol. 24, no. 2, pp. 303–314, Jun. 2021.

[28] A. Dey, S. Chattopadhyay, P. K. Singh, A. Ahmadian, M. Ferrara, and R. Sarkar, ''A hybrid meta-heuristic feature selection method using golden ratio and equilibrium optimization algorithms for speech emotion recognition,'' *IEEE Access*, vol. 8, pp. 200953–200970, 2020.

[29] K. R. Bagadi and C. M. R. Sivappagari, ''A robust feature selection method based on meta-heuristic optimization for speech emotion recognition,'' *Evol. Intell.*, vol. 2022, pp. 1–12, Sep. 2022.

[30] M. Arumugam and M. Kaliappan, ''Feature selection based on MBFOA for audio signal classification under consideration of Gaussian white noise,'' *IET Signal Process.*, vol. 12, no. 6, pp. 777–785, Aug. 2018.

[31] F. Kılıç, Y. Kaya, and S. Yildirim, ''A novel multi population based particle swarm optimization for feature selection,'' *Knowl.-Based Syst.*, vol. 219, May 2021, Art. no. 106894. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095070512100157X

[32] G. G. Devarajan, S. M. Nagarajan, S. I. Amanullah, S. A. S. A. Mary, and A. K. Bashir, ''AI-assisted deep NLP-based approach for prediction of fake news from social media users,'' *IEEE Trans. Computat. Social Syst.*, early access, May 29, 2023, doi: 10.1109/TCSS.2023.3259480.

[33] T. Özseven, ''A novel feature selection method for speech emotion recognition,'' *Appl. Acoust.*, vol. 146, pp. 320–326, Mar. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X18309915

[34] P. T. Krishnan, A. N. Joseph Raj, and V. Rajangam, ''Emotion classification from speech signal based on empirical mode decomposition and non-linear features: Speech emotion recognition,'' *Complex Intell. Syst.*, vol. 7, no. 4, pp. 1919–1934, Aug. 2021.

[35] I. Shahin, O. A. Alomari, A. B. Nassif, I. Afyouni, I. A. Hashem, and A. Elnagar, ''An efficient feature selection method for Arabic and English speech emotion recognition using grey wolf optimizer,'' *Appl. Acoust.*, vol. 205, Mar. 2023, Art. no. 109279.

[36] M. Farooq, F. Hussain, N. K. Baloch, F. R. Raja, H. Yu, and Y. B. Zikria, ''Impact of feature selection algorithm on speech emotion recognition using deep convolutional neural network,'' *Sensors*, vol. 20, no. 21, p. 6008, Oct. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/21/6008

LIYA YUE received the M.A. degree in foreign language and literature from Henan University, Kaifeng, China, in 2015. She is mainly engaged in research on English teaching, computer-assisted English, and feature selection.

PEI HU received the M.Eng. degree in computer application technology from Central South University, Changsha, China, in 2009, and the Ph.D. degree in computer application technology from the Shandong University of Science and Technology, Qingdao, China, in 2023. His research interests include evolutionary computation, feature selection, and machine learning.

SHU-CHUAN CHU received the Ph.D. degree from the School of Computer Science, Engineering and Mathematics, Flinders University, Australia, in 2004. In December 2009, she joined Flinders University, after nine years with Cheng Shiu University, Taiwan. She has been the Research Fellow with the College of Science and Engineering, Flinders University, since December 2009. She has been a Research Fellow (with the Ph.D. advisor) with the College of Computer Science and Engineering, Shandong University of Science and Technology, since September 2019. Her research interests include swarm intelligence, intelligent computing, and data mining.

JENG-SHYANG PAN (Senior Member, IEEE) received the B.S. degree in electronic engineering from the National Taiwan University of Science and Technology, Taiwan, in 1986, the M.S. degree in communication engineering from the National Chiao Tung University, Taiwan, in 1988, and the Ph.D. degree in electrical engineering from the University of Edinburgh, U.K, in 1996. He is currently a Professor with the College of Computer Science and Engineering, Shandong University of Science and Technology. His current research interests include soft computing, information security, and signal processing. He joined the Editorial Board of *LNCS Transactions on Data Hiding and Multimedia Security*, the *Journal of Computers*, and the *Chinese Journal of Electronics*.