sciendo
ISSN 1335-8871

MEASUREMENT SCIENCE REVIEW

Journal homepage: https://content.sciendo.com

Comparative Performance Analysis of Metaheuristic Feature Selection Methods for Speech Emotion Recognition

Turgut Ozseven[1*], Mustafa Arpacioglu[2]

[1]*Department of Computer Engineering, Tokat Gaziosmanpaşa University, Tokat, 60100 Turkey,* turgut.ozseven@gop.edu.tr
[2]*Institute of Graduate Studies, Tokat Gaziosmanpasa University, Tokat, Turkey,* marpacioglu60@gmail.com

Abstract: Emotion recognition systems from speech signals are realized with the help of acoustic or spectral features. Acoustic analysis is the extraction of digital features from speech files using digital signal processing methods. Another method is the analysis of time-frequency images of speech using image processing. The size of the features obtained by acoustic analysis is in the thousands. Therefore, classification complexity increases and causes variation in classification accuracy. In feature selection, features unrelated to emotions are extracted from the feature space and are expected to contribute to the classifier performance. Traditional feature selection methods are mostly based on statistical analysis. Another feature selection method is the use of metaheuristic algorithms to detect and remove irrelevant features from the feature set. In this study, we compare the performance of metaheuristic feature selection algorithms for speech emotion recognition. For this purpose, a comparative analysis was performed on four different datasets, eight metaheuristics and three different classifiers. The results of the analysis show that the classification accuracy increases when the feature size is reduced. For all datasets, the highest accuracy was achieved with the support vector machine. The highest accuracy for the EMO-DB, EMOVA, eNTERFACE'05 and SAVEE datasets is 88.1%, 73.8%, 73.3% and 75.7%, respectively.

Keywords: Speech emotion recognition, metaheuristic, feature selection, acoustic analysis, feature optimization.

1. Introduction

Speech is formed by sounds that begin in the lungs, the so-called vocal tract, and end in the lips. This physiological structure of speech causes it to be easily influenced by emotional states. A person's emotional state is reflected in their voice and therefore in their speech [1]. According to the view presented by Damasio in 2000, an emotion sends commands to the body through the bloodstream and neurons, causing a general change in a person's state. This view has also been advocated by other authors such as Ekman [2], [3]. This change in speech and voice is used not only for emotion recognition but also for psychological diagnosis [1].

The main purpose of emotion recognition studies is to analyze speech using signal processing methods and to obtain features that distinguish emotions. For this purpose, acoustic, image and wavelet features are used. This study focuses on features based on acoustic analysis. Acoustic analysis is a field that has been studied for many years, and the statistical properties of acoustic parameters were first utilized [4], [5]. Acoustic features can be divided into three categories: continuous, qualitative and cepstral. In each feature, the speech signal is framed with a certain length and these features are extracted from each frame. Therefore, thousands of features can be extracted from each speech signal [6].

Speech Emotion Recognition (SER) essentially consists of pre-processing, feature extraction, post-processing and classification steps. In the pre-processing step, signal quality enhancement methods such as noise removal, resampling and pre-emphasis filtering are used. In the feature extraction step, acoustic parameters are extracted from each speech signal. The final processing step involves feature normalization or feature selection methods [7].

The feature set, which is an important step for the success of SER, has a direct impact on emotion recognition. Feature selection methods are used to achieve higher classification success with fewer features and to reduce classifier complexity [8]. In addition to traditional feature selection methods, metaheuristic-based methods are also used for this process. Traditional feature selection methods used for SER include Principal Component Analysis (PCA), Forward Feature Selection (FFS), Linear Discriminant Analysis (LDA), Backward Feature Selection (BFS), Sequential Floating Forward Selection (SFFS) and wrapper approach with forward selection [9]-[12].

In the investigation of metaheuristic methods, the Cuckoo Search algorithm (CS), the Harmony Search (HS) algorithm, the Non-dominated Sorting Genetic Algorithm (NSGA-II) and the Particle Swarm Optimization (PSO) methods were

DOI: 10.2478/msr-2024-0010

Corresponding author: turgut.ozseven@gop.edu.tr (Turgut Ozseven)

used for attribute optimization [13]-[19]. Reducing the feature size increased accuracy in some studies and decreased it in others [7]. In the study where two different PSO-based feature selection methods were proposed, the Extreme Learning Machines (ELM) classifier was used and a 10%-30% increase in SER performance was obtained when comparing three different datasets [19]. In the findings of the study in which the HS algorithm was used for feature selection, the feature size was reduced by 50%, while no significant change in classification accuracy was observed [14]. According to the results of the study in which the Biogeography-based PSO method was proposed, the feature size in the EMO-DB was reduced by approximately 83%, while the classification accuracy increased between 5% and 13%. In contrast, in the SAVEE dataset, the feature size was reduced by about 67%, while the classification accuracy increased between 4% and 16% [20]. In the study in which feature selection was performed using only the biogeography-based optimization method, a classification accuracy of 74.29% was achieved before optimization, while the classification accuracy after optimization was 90.13% [21]. According to the findings of another study, in which the Parallel Quantum Particle Swarm Optimization (pQPSO) method based on Quantum-behaved PSO (QPSO) was recommended, the proposed method increased the classification accuracy by 2% compared to the QPSO method [22]. According to the findings of the study, which included feature optimization using the semi-nonnegative matrix factorization method, the feature size was reduced from 65 to 56, while the classification accuracy increased from 55.36% to 90.12% [23]. Based on the average update method of PSO and Whale Optimization (WO), the model performed 1.02% better than the WO algorithm, 0.32% better than the Firefly Algorithm (FA), 23.45% better than PSO, and 23.41% better than Genetic Algorithm (GA) [24]. With the metaheuristic feature selection method using Golden Ratio Optimization (GRO) and Equilibrium Optimization (EO) as hybrids, a classification accuracy of 97.31% and 98.46% was achieved in the SAVEE and EMO-DB datasets, respectively [25]. In the study where CS and EO methods were used as hybrids, the classification accuracy increased by 8% compared to EO and 3% compared to CS [26]. Classification success increased by 3.39% in the study where the SER model based on GA and the Decision Tree (DT) fusion of deep and acoustic features were used [27]. In the study using a different metaheuristic method, the Enhanced Cat Swarm Optimization method achieved a 5% increase compared to PSO and a 7% increase compared to the Cat Swarm Optimization (CSO) [28]. In the study using the NSGA-II and CS methods, classification accuracy increased by about 10% depending on the dataset [17]. In the study where a weighted binary cuckoo search was used to determine the effective features for SER, the feature size was reduced by about 80%, while the classification success increased by about 16% [22]. In the study where a genetic algorithm was used for feature optimization and PCA for feature selection, the feature size was reduced from 1582 to 100, while the classification success increased by 5%-15% depending on the dataset [13]. In another study using PSO, one of the metaheuristic methods, a success rate of 72.96% was achieved for the SAVEE dataset [18]. In the study where a two-stage feature optimization method based on Feature Correlation Analysis (FCA) and the ReliefF algorithm is proposed to reduce the feature size and increase SER accuracy, the classification accuracy increased by about 30% [29]. In a hybrid model using a Clustering-based EO and Atom Search Optimization (ASO) algorithm, recognition accuracies of 98.01%, 98.72%, 84.62%, and 74.25% were achieved in the SAVEE, EMO-DB, RAVDESS, and IEMOCAP datasets, respectively [30]. To create a successful SER system, the classification phase is as important as feature selection and optimization. Although recent studies have focused more on deep learning approaches for feature selection and classification, most SER studies have used Support Vector Machine (SVM), decision trees, Neural Networks (NN), k-Nearest Neighbor (k-NN), and ELM as single, multiple, or hybrid classification algorithms [13].

According to the results of the studies in the literature, the metaheuristic methods CS, HS, PSO, WO, FA and GA were most frequently used. Among these, PSO is used as a single and hybrid method in many studies. In terms of classifiers, SVM, k-NN, artificial NN and ELM methods are mostly used, although deep learning-based approaches have come to the fore in recent years. Furthermore, the results of the studies have shown that classification success is directly related to the classifier used, the feature selection method, the feature set and the database. In this study, feature selection with eight different metaheuristic methods and SER performance with four different classifiers were evaluated. In this study, feature selection with eight different metaheuristic methods and SER performance with four different classifiers were evaluated. Metaheuristic methods used: PSO [31], Multi-Verse Optimization (MVO) [32], Grey Wolf Optimizer (GWO) [33], Moth-flame optimization (MFO) [34], WO [35], FA [36], Bat Algorithm (BAT) [37] and CS [38]. The classifiers used in our study are SVM, k-NN and Artificial Neural Network (ANN). Our main motivation for this study is that different metaheuristic methods for SER are not in the literature when we look at the studies in the literature. In contrast to the studies in the literature, our study aims to evaluate the performance of metaheuristic algorithms on different datasets and classification algorithms for SER. The contribution of this study to the literature can be summarized as follows: 1) Comparison of metaheuristic-based feature selection methods for SER. 2) Analysis of database-based performances. 3) Investigation of the relationship between classifier and metaheuristic method. 4) Analysis of the positive or negative effects of optimization on SER. 5) Creating a guide for metaheuristic-based feature selection for researchers conducting research in the field of SER. The flow chart that the study followed to reach the determined goals is shown in Fig. 1.

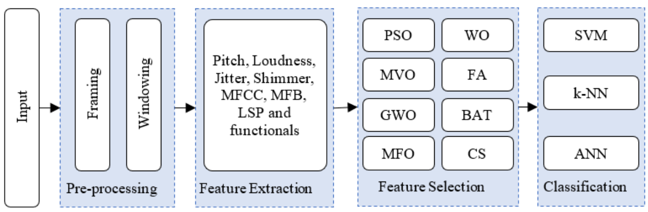

Fig. 1. Flow chart of this study.

The other parts of this study are organized as follows: Section 2 provides information about the methods, databases and metaheuristic algorithms used in our study for feature selection. The experimental results are highlighted in Section 3, the results obtained in Section 4 are compared with the results of other studies in the literature, the results are interpreted and future studies are indicated.

2. MATERIALS & METHODS

A. Database

The EMO-DB, eNTERFACE'05, EMOVO, and SAVEE datasets were used in this study. EMO-DB consists of data in which a total of 10 actors voiced 10 sentences used in daily communication. The audio recordings have a frequency of 16 kHz. [39]. eNTERFACE'05 is an audio-visual emotion dataset. The dataset contains 42 subjects. The audio sampling rate is 48 kHz in an uncompressed 16-bit stereo format [40]. EMOVO is a dataset consisting of the voices of up to 6 actors voicing 14 sentences simulating seven emotional states. The recordings were recorded in 16-bit stereo, wav format with a sampling frequency of 48 kHz [41]. The SAVEE dataset consists of the recording of 4 male actors expressing 7 different emotions (anger, disgust, fear, happiness, sadness, neutral, surprise) in a total of 480 sentences in British English. The sampling rate for this dataset is 44.1 kHz [12]. The distribution of the characteristics of the data used in the study can be found in Table 1.

Table 1. Emotion-based distribution of speech recordings used in the study.

Data Set	Anger	Disgust	Anxiety/Fear	Happiness	Sadness	Boredom	Surprise	Neutral	Total
EMO-DB	127	46	69	71	62	81	n/a	79	535
EMOVO	72	68	69	72	72	n/a	72	72	497
eNTERFACE’05	176	190	196	198	182	n/a	205	n/a	1147
SAVEE	60	60	60	60	60	n/a	60	120	480

B. Pre-processing

Pre-processing is used to extract better descriptive features from the input signals, to remove unwanted noise in the signal or to enhance the highlights in the signal [30]. The use of these pre-processing steps is not mandatory. However, in the SER process, the signal is not used while the acoustic features are extracted and the signal is divided into parts. The purpose of dividing the signal into parts is to keep the speech signals stable in small time intervals. This process is called framing. Overlapping is used to smooth the transition between frames and prevent data loss. Another method used in framing is windowing. Windowing is used to regulate interference due to spectral leakage and aliasing in the signal [1], [42]. In our study, the frame size was 25 ms, the overlap rate was 50% and the Hamming method was used for windowing.

C. Feature sets

Since the speech signal is stationary in short time intervals, the signal is not processed as a whole but divided into frames [7]. In our study, 1582 features were obtained by using openSMILE [43] for feature extraction. These are Mel-Frequency Cepstral Coefficients (MFCC), pitch, loudness, jitter, shimmer, Line Spectral Pair (LSP), log Mel-Frequency Bands (MFB). Pitch contains information about the thickness and thinness of the sound, and 164 features are extracted from each sound recording. Loudness is the feature related to the loudness of the sound and includes 42 features for each sound recording. Jitter covers the irregularities in the periods of the audio signal and 76 features are obtained. Shimmer is the periodic variation between amplitude peaks and 38 features are determined for each sound recording. The Mel unit is used for modeling the human hearing system. It performs a logarithmic conversion from the actual frequency value to the detected frequency value [44]. Cepstral coefficients extracted using this frequency scale are called MFCC and 966 features are obtained from each sound recording. LSP is the mathematical modeling of linear predictive coding. Linear predictive coding is the ability of each sound sample to be derived from a linear combination of previous sound samples. In our study, 336 LSP features were used.

D. Metaheuristic feature selection

Feature selection aims to remove redundant features and to select a feature subset from the high-dimensional feature set [45]. The main goal is to remove redundant features so that the classifier finds the optimal solution in terms of complexity and accuracy [17]. The Evolopy-Fs framework developed in the open-source Python programming language was used in this study [46]. PSO, CS, GWO, MVO, MFO, WO, BAT and FA were used in this framework [46]. The parameters used for the metaheuristic methods can be found in Table 2. NumOfRuns = 10, PopulationSize = 30, Iterations = 100 for all implemented methods.

Table 2. Parameters used in metaheuristic methods.

Method	Parameters
PSO	$c_1=2$, $c_2=2$, $w_{max}=0.9$, $w_{min}=0.2$, $v_{max}=6$
MVO	$WEP_{max}=1$, $WEP_{min}=0.2$, $p=6$
GWO	*No custom parameters*
MFO	*No custom parameters*
WO	$b=1$
FA	$alpha=0.5$, $beta_{min}=0.2$, $gamma=1$
BAT	$Q_{min}=0$, $Q_{max}=2$, $A=0.5$, $r=0.5$
CS	$p_a=0.25$

PSO is an algorithm based on the movements of animals moving in herds to satisfy their basic needs, e.g. to find food. Each individual that moves to find the solution is called a particle. The algorithm is started with a solution set that contains random particles. Based on this solution set, the particle position and best value are updated according to the best value found by the swarm to find the best solution at each iteration [47]. The velocity and position of each particle can be examined in [47] for details on how to calculate the best position and best value of each particle.

The MVO uses three concepts: white, black and wormhole. These concepts are used for exploration, usage, and local search, respectively. It is used to determine the search space of the white and black hole MVO. The wormhole helps to use the search space [32]. During the algorithm process, there is sometimes an exchange of objects between universes. High-inflation universes tend to send their objects to low-inflation universes, while low-inflation universes tend to receive objects from high-inflation universes [48]. In addition, each solution is assigned an inflation rate associated with the fitness function value [32]. Mathematical expressions for the MVO can be found in [32].

Gray wolves usually live in groups of 5 to 12 animals. There is a dominant social hierarchy in the groups. The GWO was developed based on the social and hunting behavior of gray wolves. Four types of gray wolves are used to express the social hierarchy. These are alpha, beta, delta and omega [33]. The alpha wolf has the highest level and is the leader of the entire team. Second-level beta wolves are secondary wolves and support the alpha wolves in decision-making. Delta wolves consist of wolves with functions such as scouts, sentries, hunters and caretakers. Deltas obey the alpha and beta wolves and dominate the omega wolves. Omega is the lowest-ranking wolf in the pack. They are the last wolves to eat. In some cases, the omega wolf is also a caretaker in the subgroup [49]. The hunting process of the gray wolf consists of three main elements: surrounding the prey, hunting and attacking (preying) the prey [49]. When besieging the prey, each wolf updates its position in relation to the approximate location of the prey and then circles it. When hunting, alpha, beta and delta have better information about the probable location of the prey. Other wolves update their position based on these three positions.

The MVO was developed based on the transverse orientation flight mechanism of moths [34]. Moths can fly long distances at a fixed angle to the moon, using moonlight to fly at night. The fact that the distance between the moth and the moon is quite large allows the moths to move in a straight path. In the algorithm, the proposed solutions of the moths and the position of the moths in the solution space are used as problem variables. Another important component in this algorithm are the flames. The flames are the matrix that indicates the best position for the moths. Therefore, the moths search around the flames and never lose their best position [50]. In other words, the flames can be considered as flags that the moths leave in the search space. Thus, each moth searches for a flag in all iterations and updates the existing flag when it finds a better solution. In this way, no moth ever loses the best solution.

WO is an algorithm inspired by the hunting behavior of humpback whales. Humpback whales have a unique spiral hunting technique. In this technique, the whales attack their prey with a spiral encirclement or position update. The algorithm considers the best available solution as the prey or the one closest to the prey. The whale's hunt manifests itself in recognizing the research space and solving the attack on its prey. More detailed information about the algorithm can be found in the citations [35], [51].

The FA is an optimization algorithm based on the glow characteristics of fireflies [36]. The most important feature of fireflies is that they have flashing lights. The firefly glow patterns generated by a bioluminescence method are unique to each of the 2000 firefly species alive today. This flashing has two main purposes, firstly to attract potential prey, and secondly to mate with mates. There are 3 idealized rules to simplify FA. These are:

1. The total number of fireflies is unisexual. As a result, you can be sexually attracted to them.
2. There is a positive correlation between the degree of attraction and brightness, i.e. the one that is less bright is more likely to be attracted to the one that has more light. Random attraction occurs when no one has the greatest amount of light.
3. The light intensity of a firefly is influenced by the appearance of the neutral function. The basis of attraction and the light intensity are important aspects. The light intensity is formulated as in [52].

BAT is a natural metaheuristic optimization algorithm inspired by the echo behavior of bats in nature. Bats use echolocation to hunt, distinguish between food and objects, and avoid obstacles [37]. The search space of the algorithm is the region where the food sources are located. The goal of the algorithm is to find the optimum of these resources by bats. Since the region of the food sources is unknown, the bat population is randomly distributed over the search space. The bats calculate the fitness value of their location in relation to the food source and store it in memory. This fitness value of the bats represents the next action plan of the population. The next action plan is to be positioned according to the solution value of the best individual [37], [53].

CS is an algorithm based on the breeding behavior of cuckoos and inspired by brood parasitism. Cuckoos are characterized by the fact that they lay their eggs in the nests of other creatures. In the nest of other creatures, the owner of the nest throws away the eggs with a high probability of hatching or leaves the nest. If the owner of the nest thinks that the eggs belong to him, he hatches them and brood parasitism occurs. When it is time to lay eggs, the cuckoos migrate to the area where their eggs resemble more closely to other eggs and have more food sources for their young. Different groups of cuckoos choose the location of the cuckoos with the best environment as a target. Migration continues until these groups have found the best environment. These behaviors form the basis for the CS algorithm. The algorithm is based on three principles: Each cuckoo lays one egg at a time in a randomly selected nest. If the egg is good, it is passed on to future generations. The owner of the host nest can recognize the laid egg with a probability $p_a \in [0, 1]$ [38]. For this study, the simplest approach was used, where each nest has only one egg. To study the algorithm in more detail, you can refer to reference [38].

E. Classifiers

In our study, three different classification algorithms, SVM, k-NN and ANN, were used after a metaheuristic feature selection for SER. SVM is widely used for emotion recognition as well as many machine learning problems [54].

SVM

Since the SVM classifier is a type of supervised learning, it must first be trained with labeled data. The main purpose is to find the most appropriate boundary between classes using the training data. For this purpose, the dataset is transformed into a high-dimensional space so that the classes can be better decomposed. The boundary that provides this separation is called the hyperplane. Kernel functions are used to move the dataset into the high-dimensional space. There are many kernel types (RBF, polynomial, linear, cubic, etc.) used in the literature and in applications [55]. In this study, the cubic kernel function is used for the SVM classifier. Success evaluation was done with 10-fold cross validation for all the classifiers we used. In addition, 70% of the dataset was used for training and 30% for testing.

k-NN

k-NN is a supervised machine learning model that makes predictions about the similarity of samples in the training data to each other. This classifier uses two basic parameters, *k* and distance. The k parameter specifies the number of nearest neighbors used for the calculation. This value has a direct effect on the classification result. Distance is the distance between the point to be estimated and other points. There are different methods that can be used for distance, such as Euclidean, Manhattan and Chebyshev [56]. In this study, the value for the number of neighbors (*k*) is 10 and the distance is Euclidean.

ANN

ANN is a machine-learning method inspired by the learning function of the human brain. The brain's neural network mimics the abilities of learning, remembering and generalizing. ANN consists of a combination of more than one neuron. Each neuron consists of input, weighting, summation function, activation function and output. In ANN, information is given to the network in the input layer, which is processed in the middle layer, and the result is obtained in the output layer. These layers consist of more than one neuron [57]. In our study, ReLU was used for the activation function, 1 for the fully connected layer, 100 for the layer size, and 1000 for the number of training iterations.

3. Experimental results

The main purpose of this study is to compare the performance of metaheuristic feature selection methods for SER. For this purpose, the results of eight metaheuristic methods were compared on four datasets with SVM, k-NN, ANN and ELM classifiers. All analyses were performed on a laptop with an i5 3.0 GHz processor and 16 GB RAM. When performing the analyses, 70% of the dataset was used for training and 30% for testing. Summarized information on the classification accuracy achieved with the test dataset for all analyses can be found in Table 3.

According to the information in Table 3, feature selection does not always have a positive effect on classification accuracy. The classification accuracy increases or decreases depending on the method used. However, when examined as a classifier performance, feature selection makes a positive contribution. Details of the summarized information in Table 3 are explained below under the sub-headings.

Table 3. Classification accuracies (%) obtained using the test dataset.

Method	EMO-DB			EMOVA			eNTERFACE'05			SAVEE		
	SVM	k-NN	ANN	SVM	k-NN	ANN	SVM	k-NN	ANN	SVM	k-NN	ANN
No Opt.	87.5	72.5	84.4	73.8	60.4	71.8	67.4	39.8	68.3	73.6	59.0	67.4
PSO	85.6	68.1	84.4	72.5	55.0	67.8	71.2	44.8	71.2	73.6	49.3	75.0
MVO	87.5	78.1	85.0	72.5	57.0	60.4	70.1	41.9	66.0	75.7	57.6	70.1
GWO	85.6	74.4	83.1	70.5	51.7	61.1	71.8	45.9	69.5	74.3	57.6	73.6
MFO	83.1	76.2	81.9	72.5	56.4	67.1	73.3	42.4	69.8	70.8	55.6	63.9
WO	88.1	75.0	80.0	73.8	58.4	67.8	72.7	43.6	71.2	68.8	58.3	74.3
FA	88.1	77.5	85.6	65.1	50.3	64.4	68.6	42.2	69.8	71.5	55.6	66.0
BAT	87.5	70.6	85.6	73.2	56.4	61.7	69.2	48.5	66.6	74.3	50.0	75.0
CS	88.1	71.9	81.2	65.1	47.7	63.1	70.1	47.7	69.8	71.5	58.3	69.4

Table 4. SER accuracies without feature selection (%).

Classes	EMO-DB			EMOVA			eNTERFACE'05			SAVEE		
	SVM	k-NN	ANN	SVM	k-NN	ANN	SVM	k-NN	ANN	SVM	k-NN	ANN
Class1	93.3	91.0	92.1	72.0	64.0	78.0	88.7	85.5	80.6	76.2	47.6	66.7
Class2	85.7	49.0	89.8	67.3	55.1	55.1	57.6	21.2	59.8	64.3	42.9	54.8
Class3	87.5	80.4	89.3	72.9	60.4	58.3	65.2	37.0	61.6	59.5	35.7	69.0
Class4	81.2	81.2	78.1	52.9	37.3	47.1	73.4	25.2	64.0	50.0	38.1	57.1
Class5	70.0	38.0	64.0	80.0	70.0	72.0	75.6	59.8	73.2	95.2	85.7	92.9
Class6	92.7	69.1	87.3	74.0	64.0	78.0	74.1	38.5	69.2	71.4	54.8	71.4
Class7	90.9	93.2	93.2	58.0	48.0	50.0	-	-	-	64.3	52.4	64.3
Overall	87.5	72.5	84.4	73.8	60.4	71.8	67.4	39.8	68.3	73.6	59.0	67.4

A. SER results without feature selection

In this part of our study, the SER results were analyzed with 1582 features without feature selection. The accuracy rates obtained in each dataset are shown in Table 4.

According to the results in Table 4, the highest classification accuracy for SER was achieved with SVM and ANN classifiers. The classification accuracy of the k-NN classifier is much lower, especially in the eNTERFACE'05 database. In eNTERFACE'05, the highest success was achieved with ANN, while SVM was used in the other three datasets.

B. SER results with metaheuristic-based feature selection

The metaheuristic methods we used comprised 100 iterations using the parameters given in Table 2. The graphs of the accuracy rates obtained after iterations for each dataset are shown in Fig. 2.

As can be seen in Fig. 2, each method achieved different accuracies depending on the dataset. The PSO, MVO and GWO methods hardly changed depending on the iteration. For the other methods, changes in accuracy were observed depending on the iteration. The selected features in the iteration with the highest success after the metaheuristic feature selection are listed in Table 5.

After metaheuristic feature selection, the number of features changes depending on the method, and the included features also vary. In the datasets, the largest dimension reduction was achieved with BAT and the smallest dimension reduction with MFO. However, the lowest dimension reduction for eNTERFACE'05 was achieved with WO. With this selected feature, the classification was performed with SVM, k-NN and ANN, and the results are shown in Fig. 3.

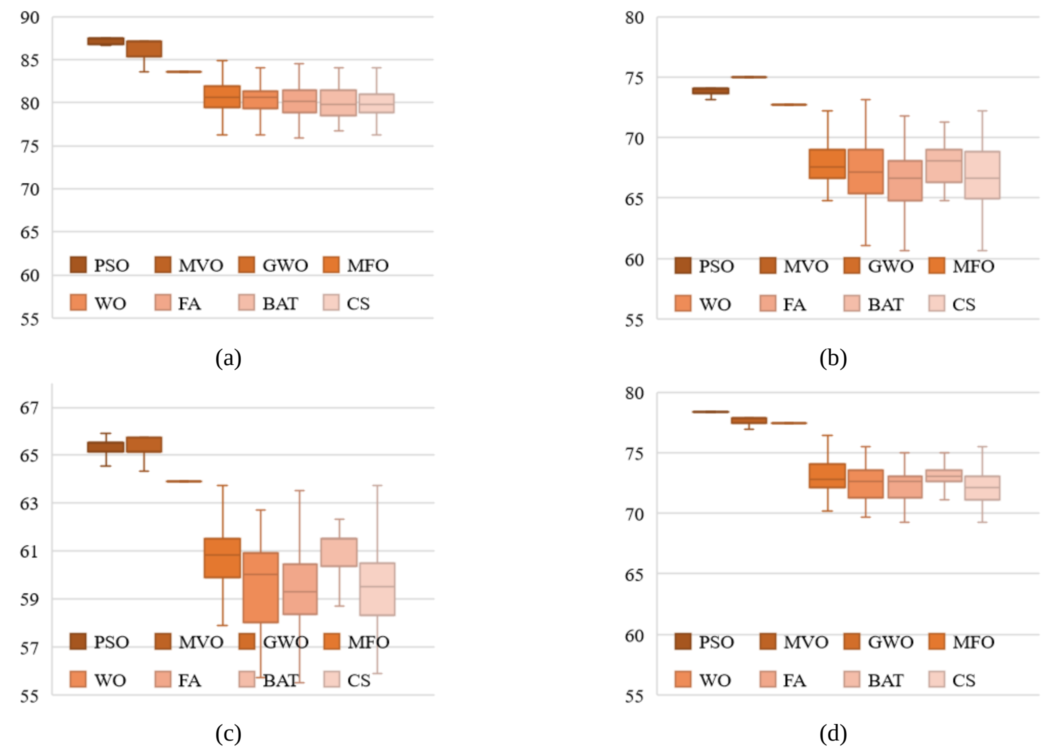

Fig. 2. Iteration results for metaheuristic feature selection (a) EMO-DB, (b) EMOVA, (c) eNTERFACE’05, (d) SAVEE.

Table 5. Number of features in data sets after metaheuristic feature selection.

Dataset	PSO	MVO	GWO	MFO	WO	FA	BAT	CS
EMO-DB	1158	956	781	1188	775	870	615	766
EMOVA	1178	959	797	1180	778	871	636	811
eNTERFACE’05	1143	966	804	1182	1203	931	726	774
SAVEE	1153	939	774	1178	782	883	622	802

EMO-DB

After feature selection for EMO-DB, SVM classifier and PSO, the classification accuracy of GWO and MFO decreased. For the k-NN classifier, the classification accuracy decreased for PSO, BAT and CS. For the ANN classifier, the classification accuracy decreased for GWO, MFO, WO and CS. However, this reduction in classification accuracy ranges from 2% to 4%, while the reduction in feature size is about 25%. The highest classification accuracy in EMO-DB is 88.1% for WO, FA and CS. Considering the number of features, the most successful method for EMO-DB is SVM+CS. With CS feature selection, the feature size was reduced by 51%, while the classification accuracy increased by 0.6%.

EMOVA

After feature selection for EMOVA, the classification accuracy did not change with only WO in SVM, and other methods caused a decrease in classification accuracy. The decrease in classification accuracy ranges from 1.3% to 8.7%, and the decrease in feature size is about 25%. All metaheuristic methods with k-NN and ANN classifiers caused a decrease in classification accuracy in EMOVA. The highest classification accuracy in EMOVA is 73.8% with SVM+WO. Although the feature size was reduced by 51%, the classification accuracy did not change.

eNTERFACE'05

In eNTERFACE'05, all feature selection methods increased classification accuracy. ANN showed high success before feature selection, and SVM showed higher success after feature selection. The most successful metaheuristic methods are MFO and WO. In contrast to other datasets, the WO method resulted in lower size reduction in eNTERFACE05. The highest classification accuracy in eNTERFACE05 is 73.3% with SVM+MFO. The feature size was reduced by 25.3%, while the classification accuracy was increased by 5.9%.

SAVEE

After feature selection for SAVEE, SVM classifier and MFO, the classification accuracy of WO, FA and CS decreased. All metaheuristic methods for the k-NN classifier decreased the classification accuracy. For the ANN classifier, the classification accuracy decreased for MFO and FA. The highest classification accuracy before feature selection was 73.6% for SVM, and after feature selection for MVO, the classification accuracy increased by 2.1% and the feature size decreased by 40.6%.

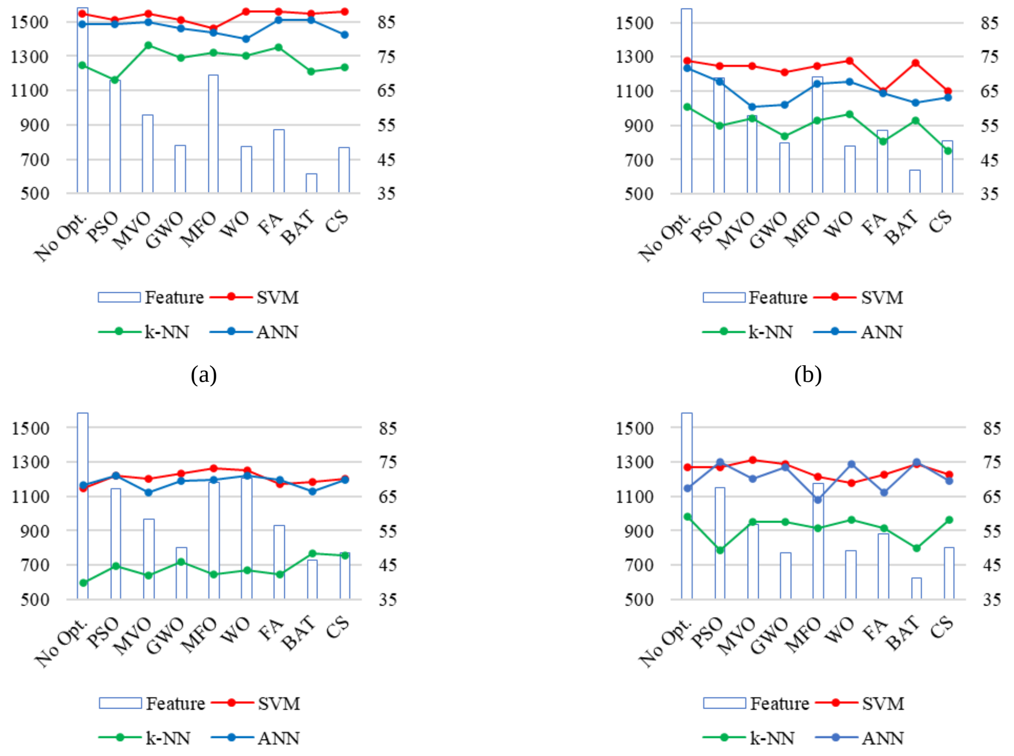

Fig. 3. Classification accuracies and feature size after feature selection (a) EMO-DB, (b) EMOVA, (c) eNTERFACE'05, (d) SAVEE.

Table 6. Class based accuracies after metaheuristic feature selection (%).

Classes	EMO-DB		EMOVA		eNTERFACE'05		SAVEE	
	SVM (No Opt.)	SVM+CS	SVM (No Opt.)	SVM+WO	ANN (No Opt.)	SVM+MFO	SVM (No Opt.)	SVM+MVO
Class1	93.3	93.3	72.0	70.0	80.6	87.9	76.2	69.0
Class2	85.7	81.2	67.3	59.2	59.8	51.9	64.3	52.4
Class3	87.5	86.0	72.9	62.5	61.6	62.8	59.5	52.4
Class4	81.2	87.5	52.9	51.0	64.0	68.1	50.0	61.9
Class5	70.0	65.3	80.0	80.0	73.2	71.1	95.2	94.0
Class6	92.7	85.7	74.0	78.0	69.2	67.8	71.4	59.5
Class7	90.9	93.2	58.0	52.0	-	-	64.3	71.4
Overall	87.5	88.1	73.8	73.8	68.3	73.3	73.6	75.7

The class-based accuracies of the analyses with the highest classification accuracy after metaheuristic feature selection can be found in Table 6.

If the class-based accuracies in Table 6 are compared with the results before feature selection, the class-based accuracies vary depending on the dataset, classifier and metaheuristic method. These changes do not constitute a clear opinion on the classification accuracy of the metaheuristic methods. This is because there are increasing and decreasing accuracies depending on the class. However, it can be said that metaheuristic methods do not cause a change in class-based accuracies that would affect the balance of overall accuracy.

4. DISCUSSION & CONCLUSION

In this study, we conducted a feature selection study for SER using metaheuristic methods. Our experimental study was conducted using the EMO-DB, eNTERFACE05, EMOVO and SAVEE datasets. The features in these datasets were extracted using the openSMILE software. The Evolopy-FS framework was used for feature optimization [46]. Feature selection was performed using eight metaheuristic algorithms (PSO, CS, GWO, MVO, MFO, WOA, BAT, FFA) within this framework. The results obtained were tested with SVM, k-NN and ANN classifiers and their accuracy rates were tested. When testing the results, the highest value was obtained with SVM+CS for EMO-DB, SVM+WO for EMOVA, SVM+MFO for eNTERFACE05 and SVM+MVO for SAVEE. The comparison of our results with those in the literature is shown in Table 7.

The EMO-DB dataset is mostly used for SER in the literature. The highest classification accuracy of 99.5% was achieved with this dataset. Hybrid optimization methods have been used in studies with higher accuracy than the result we obtained. In addition, it has been used for SER in deep attributes in recent years. A similar situation is true for SAVEE, and hybrid methods have proven to be very successful. No study was found in the literature that included meta-heuristic feature selection for eNTERFACE'05 and EMOVA. Furthermore, the studies in the literature aim to increase the success of SER and include only a limited number of comparisons of metaheuristic methods.

Table 7. Comparison of the results obtained in this study with literature.

Dataset	Classifier	Feature Selection	Feature Size	Accuracy	Reference
EMO-DB	ELM	WPSO	57	99.5	[19]
EMO-DB	SVM	CEOAS	132	98.7	[30]
EMO-DB	XGBoost	GRO+EO	98	98.5	[25]
EMO-DB	ELM	OGA	n/a	93.3	[58]
EMO-DB	SVM	CSEO	n/a	92.5	[26]
EMO-DB	ELM	PSOBBO	177	90.3	[20]
EMO-DB	SVM	PSO	n/a	90.1	[59]
EMO-DB	SVM	Semi-NMF	56	90.1	[23]
EMO-DB	SVM	BBO	n/a	90.1	[21]
EMO-DB	SVM	DGA	n/a	89.7	[13]
EMO-DB	k-NN	Semi-NMF	60	89.3	[23]
EMO-DB	SVM	CS	766	88.1	Our result
EMO-DB	SVM	MCS	255	87.7	[17]
EMO-DB	SVM	Fisher	n/a	86.9	[16]
EMO-DB	DT-SVM	GA	205	85.9	[27]
EMO-DB	LR	WCS	n/a	83.7	[60]
EMO-DB	GEBF-ANN	pQPSO	n/a	79.9	[22]
SAVEE	SVM	CEOAS	95	98.0	[30]
SAVEE	XGBoost	GRO+EO	87	97.3	[25]
SAVEE	SVM	MVO	939	75.7	Our result
SAVEE	ELM	WPSO	68	75.4	[19]
SAVEE	ELM	PSOBBO	336	62.5	[20]
SAVEE	LR	WCS	n/a	60.2	[60]
SAVEE	GEBF-ANN	pQPSO	n/a	59.4	[22]

Note:
AS: Atom Search optimization,
BBO: Biogeography-Based Optimization,
CEOAS: Clustering-based EO and AS,
CS: Cuckoo Search,
CSEO: CS-based EO,
DGA: Density-based spatial clustering of application with GA,
EO: Equilibrium Optimization,
GA: Genetic Algorithm,
GEBF-ANN: Gaussian Elliptical Basis Function type ANN,
GRO: Golden Ratio Optimization,
LR: Logistic Regression,
MCS: Modified-CS,
MVO: Multi-Verse Optimization,
OGA: Optimized GA,
PSO: Particle Swarm Optimization,
pQPSO: Parallel Quantum-behaved PSO,
Semi-NMF: Semi Non-negative Matrix Factorization,
WCS: Weighted binary CS,
WPSO: Wrapper-based PSO,
XGBoost: eXtreme Gradient Boosting.

When the results of this study are evaluated in general, the method to be used for feature selection varies depending on the dataset and feature set. One of the main purposes of feature selection is to increase the classification accuracy. The other goal is to reduce the classification complexity. The results obtained in our study have increased the classification accuracy, while most of the meta-heuristic feature selection methods have reduced the feature size. Another factor for classification accuracy is the classifier. In this study where SVM, k-NN and ANN were used, SVM provided more successful results in most of the analyses.

The performance comparison of different meta-heuristic methods, the use of multiple datasets and classifiers in this study are the strengths of this study. The limitation of this study is that hybrid meta-heuristics and data imbalance were not considered. Future studies can be conducted to improve SER performance by optimizing classifier parameters, using hybrid metaheuristics, correcting data imbalance, and different feature sets.

References

[1] Das, A., Guha, S., Singh, P. K., Ahmadian, A., Senu, N., Sarkar, R. (2020). A hybrid meta-heuristic feature selection method for identification of Indian spoken languages from audio signals. *IEEE Access*, 8, 181432-181449. https://doi.org/10.1109/ACCESS.2020.3028241

[2] Damasio, A. R. (2000). A second chance for emotion. In *Cognitive Neuroscience of Emotion*. Oxford University Press, 12-23. ISBN 9780195155921.

[3] Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science*, 3 (1), 34-38. https://doi.org/10.1111/j.1467-9280.1992.tb00253.x

[4] Ververidis, D., Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48 (9), 1162-1181. https://doi.org/10.1016/j.specom.2006.04.003

[5] Lee, C. M., Narayanan, S. S. (2005). Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing*, 13 (2), 293-303. https://doi.org/10.1109/TSA.2004.838534

[6] Özseven, T. (2022). A review of infant cry recognition and classification based on computer-aided diagnoses. In *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. IEEE. https://doi.org/10.1109/HORA55278.2022.9800038

[7] Özseven, T. (2019). A novel feature selection method for speech emotion recognition. *Applied Acoustics*, 146, 320-326. https://doi.org/10.1016/j.apacoust.2018.11.028

[8] Bandela, S. R., Kumar, T. K. (2020). Speech emotion recognition using unsupervised feature selection algorithms. *Radioengineering*, 29 (2), 353-364. http://dx.doi.org/10.13164/re.2020.0353

[9] Pao, T.-L., Chen, Y.-T., Yeh, J.-H., Chang, Y.-H. (2005). Emotion recognition and evaluation of Mandarin speech using weighted D-KNN classification. In *Proceedings of the 17th Conference on Computational Linguistics and Speech Processing*. The Association for Computational Linguistics and Chinese Language Processing.

[10] Ververidis, D., Kotropoulos, C. (2006). Fast sequential floating forward selection applied to emotional speech features estimated on DES and SUSAS data collections. In *2006 14th European Signal Processing Conference*. IEEE.

[11] Sidorova, J. (2009). Speech emotion recognition with TGI+.2 classifier. In *Proceedings of the EACL 2009 Student Research Workshop*. Association for Computational Linguistics (ACL), 54-60.

[12] Haq, S., Jackson, P. J. B., Edge, J. D. (2008). Audio-visual feature selection and reduction for emotion classification. In *Proceedings of International Conference on Auditory-Visual Speech Processing (AVSP 2008)*. AVISA, 185-190. ISBN 978-0-646-49504-0.

[13] Kanwal, S., Asghar, S. (2021). Speech emotion recognition using clustering based GA-optimized feature set. *IEEE Access*, 9, 125830-125842. https://doi.org/10.1109/ACCESS.2021.3111659

[14] Tao, Y., Wang, K., Yang, J., An, N., Li, L. (2015). Harmony search for feature selection in speech emotion recognition. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 362-367. https://doi.org/10.1109/ACII.2015.7344596

[15] Liu, Z.-T., Wu, M., Cao, W.-H., Mao, J.-W., Xu, J.-P., Tan, G.-Z. (2018). Speech emotion recognition based on feature selection and extreme learning machine decision tree. *Neurocomputing*, 273, 271-280. https://doi.org/10.1016/j.neucom.2017.07.050

[16] Sun, L., Fu, S., Wang, F. (2019). Decision tree SVM model with Fisher feature selection for speech emotion recognition. *EURASIP Journal on Audio, Speech, and Music Processing*, 2019, 2. https://doi.org/10.1186/s13636-018-0145-5

[17] Yildirim, S., Kaya, Y., Kılıç, F. (2021). A modified feature selection method based on metaheuristic algorithms for speech emotion recognition. *Applied Acoustics*, 173, 107721. https://doi.org/10.1016/j.apacoust.2020.107721

[18] Panigrahi, S. N., Palo, H. K. (2021). Emotional speech recognition using particle swarm optimization algorithm. In *2021 International Conference in Advances in Power, Signal, and Information Technology (APSIT)*. IEEE. https://doi.org/10.1109/APSIT52773.2021.9641247

[19] Muthusamy, H., Polat, K., Yaacob, S. (2015). Particle swarm optimization based feature enhancement and feature selection for improved emotion recognition in speech and glottal signals. *PLoS ONE*, 10 (3), e0120344. https://doi.org/10.1371/journal.pone.0120344

[20] Yogesh, C. K., Hariharan, M., Ngadiran, R., Adom, A. H., Yaacob, S., Berkai, C., Polat, K. (2017). A new hybrid PSO assisted biogeography-based optimization for emotion and stress recognition from speech signal. *Expert Systems with Applications*, 69, 149-158. https://doi.org/10.1016/j.eswa.2016.10.035

[21] Ding, N., Ye, N., Huang, H., Wang, R., Malekian, R. (2018). Speech emotion features selection based on BBO-SVM. In *2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*. IEEE, 210-216. https://doi.org/10.1109/ICACI.2018.8377608

[22] Daneshfar, F., Kabudian, S. J., Neekabadi, A. (2020). Speech emotion recognition using hybrid spectral-prosodic features of speech signal/glottal waveform, metaheuristic-based dimensionality reduction, and Gaussian elliptical basis function network classifier. *Applied Acoustics*, 166, 107360. https://doi.org/10.1016/j.apacoust.2020.107360

[23] Bandela, S. R., Kumar, T. K. (2019). Speech emotion recognition using semi-NMF feature optimization. *Turkish Journal of Electrical Engineering and Computer Sciences*, 27 (5), 3741-3757. https://doi.org/10.3906/elk-1903-121

[24] Rajasekhar, B., Kamaraju, M., Sumalatha, V. (2020). A novel speech emotion recognition model using mean update of particle swarm and whale optimization-based deep belief network. *Data Technologies and Applications*, 54 (3), 297-322. https://doi.org/10.1108/DTA-07-2019-0120

[25] Dey, A., Chattopadhyay, S., Singh, P. K., Ahmadian, A., Ferrara, M., Sarkar, R. (2020). A hybrid meta-heuristic feature selection method using golden ratio and equilibrium optimization algorithms for speech emotion recognition. *IEEE Access*, 8, 200953-200970. https://doi.org/10.1109/ACCESS.2020.3035531

[26] Bagadi, K. R., Sivappagari, C. M. R. (2024). A robust feature selection method based on meta-heuristic optimization for speech emotion recognition. *Evolutionary Intelligence*, 17, 993-1004. https://doi.org/10.1007/s12065-022-00772-5

[27] Sun, L., Li, Q., Fu, S., Li, P. (2022). Speech emotion recognition based on genetic algorithm–decision tree fusion of deep and acoustic features. *ETRI Journal*, 44 (3), 462-475. https://doi.org/10.4218/etrij.2020-0458

[28] Gomathy, M. (2021). Optimal feature selection for speech emotion recognition using enhanced cat swarm optimization algorithm. *International Journal of Speech Technology*, 24 (1), 155-163. https://doi.org/10.1007/s10772-020-09776-x

[29] Pan, L., Wang, S., Yin, Z., Song, A. (2022). Recognition of human inner emotion based on two-stage FCA-reliefF feature optimization. *Information Technology and Control*, 51 (1), 32-47. https://doi.org/10.5755/j01.itc.51.1.29430

[30] Chattopadhyay, S., Dey, A., Singh, P. K., Ahmadian, A., Sarkar, R. (2023). A feature selection model for speech emotion recognition using clustering-based population generation with hybrid of equilibrium optimizer and atom search optimization algorithm. *Multimedia Tools and Applications*, 82, 9693-9726. https://doi.org/10.1007/s11042-021-11839-3

[31] Kennedy, J., Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*. IEEE. https://doi.org/10.1109/ICNN.1995.488968

[32] Mirjalili, S., Mirjalili, S. M., Hatamlou, A. (2016). Multi-Verse Optimizer: A nature-inspired algorithm for global optimization. *Neural Computing and Applications*, 27 (2), 495-513. https://doi.org/10.1007/s00521-015-1870-7

[33] Mirjalili, S., Mirjalili, S. M., Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software*, 69, 46-61. https://doi.org/10.1016/j.advengsoft.2013.12.007

[34] Mirjalili, S. (2015). Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-Based Systems*, 89, 228-249. https://doi.org/10.1016/j.knosys.2015.07.006

[35] Mirjalili, S., Lewis, A. (2016). The Whale Optimization Algorithm. *Advances in Engineering Software*, 95, 51-67. https://doi.org/10.1016/j.advengsoft.2016.01.008

[36] Yang, X.-S. (2010). Firefly algorithm, stochastic test functions and design optimisation. *International Journal of Bio-Inspired Computation*, 2 (2), 78-84. https://doi.org/10.1504/IJBIC.2010.032124

[37] Yang, X.-S. (2010). A new metaheuristic bat-inspired algorithm. In *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*. Springer, SCI 284, 65-74. https://doi.org/10.1007/978-3-642-12538-6_6

[38] Yang, X.-S., Deb, S. (2009). Cuckoo Search via Lévy flights. In *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*. IEEE, 210-214. https://doi.org/10.1109/NABIC.2009.5393690

[39] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., Weiss, B. (2005). A database of German emotional speech. In *INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology*. ISCA, 1517-1520. https://doi.org/10.21437/Interspeech.2005-446

[40] Martin, O., Kotsia, I., Macq, B., Pitas, I. (2006). The The eNTERFACE' 05 audio-visual emotion database. In *22nd International Conference on Data Engineering Workshops (ICDEW'06)*. IEEE. https://doi.org/10.1109/ICDEW.2006.145

[41] Costantini, G., Iadarola, I., Paoloni, A., Todisco, M. (2014). EMOVO Corpus: an Italian emotional speech database. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. ELRA, 3501-3504.

[42] Rabiner, L. R. (1968). Digital-formant synthesizer for speech-synthesis studies. *The Journal of the Acoustical Society of America*, 43 (4), 822-828. https://doi.org/10.1121/1.1910901

[43] Eyben, F., Weninger, F., Gross, F., Schuller, B. (2013). Recent developments in openSMILE, the munich open-source multimedia feature extractor. In *MM '13: Proceedings of the 21st ACM International Conference on Multimedia*. ACM, 835-838. https://doi.org/10.1145/2502081.2502224

[44] Özseven, T., Düğenci, M. (2018). SPeech ACoustic (SPAC): A novel tool for speech feature extraction and classification. *Applied Acoustics*, 136, 1-8. https://doi.org/10.1016/j.apacoust.2018.02.009

[45] Song, P., Zheng, W., Yu, Y., Ou, S. (2021). Speech emotion recognition based on robust discriminative sparse regression. *IEEE Transactions on Cognitive and Developmental Systems*, 13 (2), 343-353. https://doi.org/10.1109/TCDS.2020.2990928

[46] Khurma, R. A., Aljarah, I., Sharieh, A., Mirjalili, S. (2020). EvoloPy-FS: An open-source nature-inspired optimization framework in Python for feature selection. In *Evolutionary Machine Learning Techniques: Algorithms and Applications*. Spinger, 131-173. https://doi.org/10.1007/978-981-32-9990-0_8

[47] Guangyou, Y. (2007). A modified particle swarm optimizer algorithm. In *2007 8th International Conference on Electronic Measurement and Instruments*. IEEE. https://doi.org/10.1109/ICEMI.2007.4350772

[48] Yılmaz, Ö., Altun, A. A., Köklü, M. (2022). Optimizing the learning process of multi-layer perceptrons using a hybrid algorithm based on MVO and SA. *International Journal of Industrial Engineering Computations*, 13 (4), 617-640. https://doi.org/10.5267/j.ijiec.2022.5.003

[49] Ma, C., Huang, H., Fan, Q., Wei, J., Du, Y., Gao, W. (2022). Grey wolf optimizer based on Aquila exploration method. *Expert Systems with Applications*, 205, 117629. https://doi.org/10.1016/j.eswa.2022.117629

[50] Nadimi-Shahraki, M. H., Banaie-Dezfouli, M., Zamani, H., Taghian, S., Mirjalili, S. (2021). B-MFO: A binary moth-flame optimization for feature selection from medical datasets. *Computers*, 10 (11), 136. https://doi.org/10.3390/computers10110136

[51] Sharawi, M., Zawbaa, H. M., Emary, E. (2017). Feature selection approach based on whale optimization algorithm. In *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)*. IEEE, 163-168. https://doi.org/10.1109/ICACI.2017.7974502

[52] Xu, H., Yu, S., Chen, J., Zuo, X. (2018). An improved firefly algorithm for feature selection in classification. *Wireless Personal Communications*, 102 (4), 2823-2834. https://doi.org/10.1007/s11277-018-5309-1

[53] Nakamura, R. Y. M., Pereira, L. A. M., Costa, K. A., Rodrigues, D., Papa, J. P., Yang, X.-S. (2012). BBA: A binary bat algorithm for feature selection. In *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*. IEEE. https://doi.org/10.1109/SIBGRAPI.2012.47

[54] Huang, S., Dang, H., Jiang, R., Hao, Y., Xue, C., Gu, W. (2021). Multi-layer hybrid fuzzy classification based on SVM and improved PSO for speech emotion recognition. *Electronics*, 10 (23), 2891. https://doi.org/10.3390/electronics10232891

[55] Wang, L. (ed.) (2005). *Support Vector Machines: Theory and Applications*. Springer, STUDFUZZ 177. https://doi.org/10.1007/b95439

[56] Al Dujaili, M. J., Ebrahimi-Moghadam, A., Fatlawi, A. (2021). Speech emotion recognition based on SVM and KNN classifications fusion. *International Journal of Electrical and Computer Engineering (IJECE)*, 11 (2), 1259. http://doi.org/10.11591/ijece.v11i2.pp1259-1264

[57] Challita, N., Khalil, M., Beauseroy, P. (2016). New feature selection method based on neural network and machine learning. In *2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*. IEEE, 81-85. https://doi.org/10.1109/IMCET.2016.7777431

[58] Albadr, M. A. A., Tiun, S., Ayob, M., AL-Dhief, F. T., Omar, K., Maen, M. K. (2022). Speech emotion recognition using optimized genetic algorithm-extreme learning machine. *Multimedia Tools and Applications*, 81 (17), 23963-23989. https://doi.org/10.1007/s11042-022-12747-w

[59] Li, C.-Z., Liu, F.-K., Wang, Y.-T., Wang, H., Zhang, Q. (2017). Speech emotion recognition based on PSO-optimized SVM. In *2nd International Conference on Software, Multimedia and Communication Engineering (SMCE 2017)*. DEStech Publications. https://doi.org/10.12783/dtcse/smce2017/12465

[60] Zhang, Z. (2021). Speech feature selection and emotion recognition based on weighted binary cuckoo search. *Alexandria Engineering Journal*, 60 (1), 1499-1507. https://doi.org/10.1016/j.aej.2020.11.004

Received April 19, 2023
Accepted March 12, 2024