

# An evolutionary optimization method for selecting features for speech emotion recognition

Kesava Rao Bagadi<sup>1</sup>, Chandra Mohan Reddy Sivappagari<sup>2</sup>

<sup>1</sup>Department of Electronics and Communication Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, India

<sup>2</sup>Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University, Anantapur, Andhra Pradesh, India

---

## Article Info

### Article history:

Received Sep 01, 2021

Revised Nov 05, 2022

Accepted Nov 15, 2022

---

### Keywords:

Cuckoo search

Equilibrium optimization

Feature selection

Meta-heuristic optimization

MFCCs

Speech emotion recognition

---

## ABSTRACT

Human-computer interactions benefit greatly from emotion recognition from speech. To promote a contact-free environment in this coronavirus disease 2019 (COVID-19) pandemic situation, most digitally based systems used speech-based devices. Consequently, this emotion detection from speech has many beneficial applications for pathology. The vast majority of speech emotion recognition (SER) systems are designed based on machine learning or deep learning models. Therefore, need greater computing power and requirements. This issue was addressed by developing traditional algorithms for feature selection. Recent research has shown that nature-inspired or evolutionary algorithms such as equilibrium optimization (EO) and cuckoo search (CS) based meta-heuristic approaches are superior to the traditional feature selection (FS) models in terms of recognition performance. The purpose of this study is to investigate the impact of feature selection meta-heuristic approaches on emotion recognition from speech. To achieve this, we selected the rayerson audio-visual database of emotional speech and song (RAVDESS) database and obtained maximum recognition accuracy of 89.64% using the EO algorithm and 92.71% using the CS algorithm. For this final step, we plotted the associated precision and F1 score for each of the emotional classes.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



---

## Corresponding Author:

Kesava Rao Bagadi

Department of Electronics and Communication Engineering, Mahatma Gandhi Institute of Technology

C B Post, Gandipet, Hyderabad, Telangana, India-50075

Email: bkesavarao\_ece@mgit.ac.in

---

## 1. INTRODUCTION

There are a variety of sources of information we can use to detect emotions in people, such as speech, transcripts, facial expressions, brain signals (EEG), and a combination of two or more of these (multi-modal emotion recognition). Among these, emotional recognition from the speech is an essential element in the field of human-computer interaction. The process of speech emotion recognition involves using acoustic analysis to identify vocal changes caused by emotions and then determining which features to use to determine an emotion's presence [1]. However, many emotional databases contain either relevant or non-redundant information which can give low accuracy during classification. This issue can be addressed by applying effective feature selection (FS) methods to speech-based applications. Hence, it significantly improves the performance by the response time of the algorithm, which can turn to provide high classification accuracy. There are three main phases in the FS process. First, generate subset features from the whole set of databases, second is evaluation and finally validation [2].

As said, these traditional FS models required high computational requirements and time on speech emotional databases. There are many traditional feature selection algorithms developed for selecting relevant features for emotional classification from a speech signal. One among them is filter and wrapper approaches done based on the criterion of information gain [3], mutual information [4] and principal component analysis [5] and so on. Alternatively, in the wrapper approach, a classifier is used, such as the K-nearest neighbour (KNN) [6] and support vector machine (SVM) [7], among others, to assess the quality of the resulting subsets. At the time of the generation phase, selecting all possible features that are extracted, yields more computational efforts and computation time. Hence, the traditional FS methods are not that much impressive to speech emotion recognition (SER) tasks. Then, research is finding another way to solve this issue using a nature-inspired optimization algorithm called a meta-heuristic approach. These meta-heuristic algorithms are very intelligent search algorithms and already implemented many artificial intelligence problems [8]. Recently some researchers adopted nature-inspired meta-heuristic algorithms to improve the recognition accuracy along with fewer computational requirements. Some well-known meta-heuristic algorithms are genetic algorithm (GA), ant-colony, cuckoo search (CS), particle swarm optimization (PSO) and grey wolf optimization (GWO) employed to achieve optimal feature sub-set for speech based emotional tasks [9]. In this paper, we addressed the key concern i.e impact of feature selection models using meta-heuristic approaches for (speech emotion recognition) SER systems. An accurate classification model requires the appropriate generation of features, the selection of features, and the use of classification methods [10]. From this background, Figure 1 shows the role of feature selection methods for speech-based emotion recognition applications.

The key contribution of this paper is summarized as studying the latest state-of-the-art meta-heuristic feature selection models for speech emotion recognition. Out of many heuristic approaches, analysis the impact of equilibrium optimization (EO) and CS algorithm for SER tasks. Finally, analyze the various performance metrics for the rayerson audio-visual database of emotional speech and song (RAVDESS) dataset towards speech emotion recognition. The rest of the paper is organized: section 2 provides the related work on SER using a meta-heuristic approach, materials such as speech emotional database used in this study describes in section 3, the methodology used for recognition of emotions from speech based on meta-heuristic focused in section 4, experimental results and analysis discussed in section 5, finally, section 6 gives conclusion and future perspective for this study.

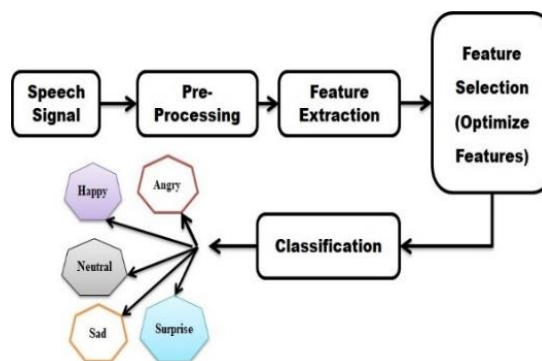


Figure 1. General framework of SER system

## 2. RELATED WORK

Many academics and research centres work on automatic speech emotion recognition and concentrate more on FS algorithms to avoid computational requirements. Initially, a modified multi-objective genetic feature selection algorithm was proposed for speech emotion recognition by Brester *et al.* [11] and achieved improvement on F1-score as 86.37% and 67.70% for the Berlin emotional speech database (EmoDB) and surrey audio-visual expressed emotion (SAVEE) databases respectively. Unlike content-based speech recognition systems, context-independent models use only signal parameters, classifiers consider these parameters as testing and training vectors [12]. The consistency of a feature selection algorithm is generated whenever new training samples are introduced or removed [13]. The selection of features that will identify important features is influenced by stability in knowledge discovery [14]. In [15], proposed a new approach of FS model using wrapper based PSO algorithm for SER tasks and achieve recognition rate up to 78.44% for SAVEE database. One more Kozodoi *et al.* [16] presented a new framework for scoring credit information using genetic algorithms. Another one proposed cuckoo search in [17] and this algorithm gives an impressive result for SER tasks. Dey *et al.* [18] on SAVEE and EMoDB, the hybrid-based meta-heuristic

optimization FS model was found to achieve an accuracy of 97.31% and 98.45%, respectively. Very recently, Daneshfar *et al.* [19] proposed a novel approach of quantum behaved particle swarm optimization (QPSO) algorithm for emotion recognition from the speech on various datasets. Zhang [20] attempted the SER using a weighted binary cuckoo search algorithm and achieved an F1-score of 83.80%. Another in [19] proposed particle swarm optimization (PSO) based on quantum behaviour for the dimensionality reduction of speech features. Compared to state-of-the-art algorithms, this method produced more accurate results. In all these works, researchers explored the various meta-heuristic optimization algorithms for SER tasks. Studies show that it is impossible to say that any feature selection method enables SER to improve or decrease performance. Features selection methods influence the success of SER depending on the classifier, the data, and the size reduction. With this literature analysis, we will address the impact of this FS methods on speech based emotion recognition. Finally, this work uses a public related speech emotional database i.e. RAVDESS for two different optimization algorithms such as EO and CS algorithm respectively.

### 3. MATERIALS

#### 3.1. Speech emotional database

The selection of a database is a crucial part of speech emotion recognition since the performance is determined by the naturalness of the database. In this paper, we have chosen a publicly available speech emotional database such as RAVDESS [21], which is in the English language. It contains various clipping profiles for both male and female speech samples of emotions such as anger, sadness, fear, excitement, happiness and neutral. A unique identification name is assigned to each sample in the dataset, and all samples are output as being either normal or strong in intensity. This study extracts features that contain the emotional information and selects the ones that are relevant for further processing and then classifies them using appropriate classifiers.

#### 3.2. Feature extraction

System performance and accuracy are dependent on the signal feature extraction. The salient features of speech signals need to be extracted to identify different emotional states and speech styles. Generally, speech features are classified as acoustic features and spectral features. To analyze the speech signal, acoustic characteristics such as pitch, energy, zero crossing rates, an average Mel frequency cepstral coefficient (MFCC) as well as a discrete wavelet transform are extracted. Even in traditional or some other feature selection methods based on SER tasks, MFCCs features are one of the most prominent features to recognize emotion from speech accurately. It provides a way to characterize the properties of the voice signal. It was found that MFCC was superior in terms of speech recognition, as it helped in creating human perception compassion that takes frequencies into account. Here 12 primary discrete cosines transform (DCT) coefficients for emotions were considered as a feature vector to recognise emotions. The process of extracting MFCCs is shown in Figure 2. In this work, we extracted MFCCs features using the openSMILE tool kit [22].

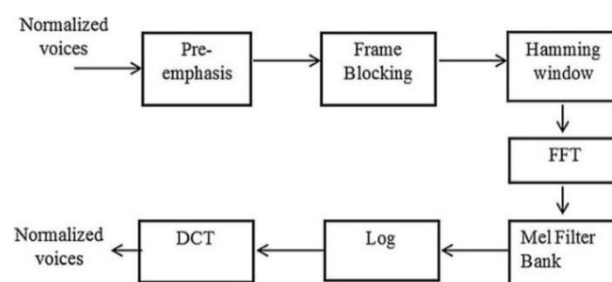


Figure 2. Process of MFCC s feature extraction from speech

#### 3.3. Feature selection

Over-fitting of machine learning algorithms occurs when the feature set dimension is large, resulting in low performance. For machine learning, FS has the objective of reducing the dimensionality of features and reducing the cost of classification. Unlike traditional feature selection methods; here we are selecting the optimal feature subset for emotion recognition based on meta-heuristic optimization algorithms i.e. EO and CS.

#### 3.4. Classifier

Classification involves applying a machine-learning algorithm to train a dataset as well as identifying or classifying new observations, or a test data set. In this work, we used to classify emotions using an SVM classifier. SVM is the easiest and most popular classifier. As far as classification is concerned,

it creates a hyperplane between different types of data, which is an optimal boundary [23]. The strength of this SVM is not to suffer any multiple local minima. Hence, in this work, we are selected SVM as a classifier to recognize emotions from speech.

#### 4. METHODOLOGY

Here, we attempted the impact of this meta-heuristic optimization algorithm like EO and CS on SER tasks. The framework of this FS model is shown in Figure 3. Initially, extract MFCCs features from the RAVDESS dataset using the openSMILE tool. Then, according to the principle of nature-inspired algorithms; first, generate the initial population of EO and CS algorithms. The purpose of this EO algorithm is to get both balancing and dynamic states from the control volume mass balance.

Considering exploration and exploitation simultaneously, it has the advantage of maintaining a good balance [24]. Dynamic mass balances of control volume systems are modelled by this algorithm. Describes the general mass balance equation in which the change in mass over time equals the mass entering a system plus the mass leaving it. A successful optimization method is cuckoo search. Yang and Deb developed the CS, one of the latest nature-inspired meta-heuristic algorithms, in 2010 [25], employing isotropic random walks, rather than by simple selection. According to recent studies, CS is potentially far more efficient than PSO. From a mathematical perspective, the success of this algorithm is to solve n-dimensional linear/non-linear optimization problems with low-level mathematics been developed in solving binary optimization problems.

Describes the general mass balance equation in which the change in mass over time equals the mass entering a system plus the mass leaving it. It is written as:

$$V \frac{dc}{dt} = QC_{eq} - QC + G \quad (1)$$

Whenever the control volume ( $V$ ) is filled with  $C$  concentration, there is a value.  $V \frac{dc}{dt}$  is the volumetric flow rate,  $Q$ , is the change in mass in the control volume,  $C_{eq}$  is the equilibrium concentration in the control volume under equilibrium condition without any generation, and is the mass generation rate inside the control volume. The initial population of an EO is also determined by the size and number of particles. A randomly generated initial population is represented by (2).

$$r_i^{initial} = r_{min} + \text{rand}_i (r_{max} - r_{min}) \quad (2)$$

Where  $r^{initial}$  represents initial vectors of  $i_{th}$  particle,  $r_{min}$  and  $r_{max}$  are optimal and maximal particle concentrations, and  $\text{rand}_i$  is between  $[0, 1]$  and  $n$  is the population size. Therefore, the equilibrium state concludes the optimization process since it optimizes globally.

There is no knowledge of the equilibrium state at the beginning of the optimization process, so only potential candidates can be determined. The equilibrium states of the algorithm are the highest quality and are the global optimum. Based on the results of complete optimization, these four are the best candidates. An additional particle, whose concentration equals the average of the four particles mentioned above, is based on numerous experiments under various types of case issues. For other optimization algorithms, the number of particles selected is arbitrary. A vector named the equilibrium pool is constructed by combining five selected objects listed in (3).

$$\bar{C}_{eq.pool} = \bar{C}'_{eq(1)}, \bar{C}'_{eq(2)}, \bar{C}'_{eq(3)}, \bar{C}'_{eq(4)} \quad (3)$$

The exponential term ( $F$ ) contributes to the main concentration updating rule in (4).

$$\bar{F} = e^{-\nu(t-t_0)} \quad (4)$$

In (5) time is defined as a function that decreases with an increase in the number of iterations ( $n_{Iter}$ ).

$$t = (1 - \frac{n_{Iter}}{Max_{Iter}})(a_2 \frac{n_{Iter}}{Max_{Iter}}) \quad (5)$$

Here,  $a_2$  is a variable that enables the exploitation skill to grow. As shown by (6), increasing exploration and exploitation abilities will allow us to easily achieve convergence by slowing down the search speed.

$$t_0 = \frac{1}{\phi} \ln(-k_1 \text{sign}(h - 0.5)[1 - e^{-\phi \cdot t}]) + t \quad (6)$$

Where,  $k_1$  denotes the ability of exploration,  $\text{sign}(h - 0.5)$  provides direction for exploration and exploitation. Here  $h$  lies between  $[0, 1]$ . The modified version of (4) is written as:

$$\vec{F} = k \cdot \text{sign}(h - 0.5)[e^{-\phi \cdot t} - 1] \tag{7}$$

In addition, the generation rate is a crucial step that helps to provide a good exploitation phase to provide an exact solution to the optimization problem. The well-known 1 – D space model is one of many models to calculate generation rate is in (8).

$$H_G = H_0 \cdot e^{\phi(t-t_0)} \tag{8}$$

Where  $H_0$  and  $\phi$  represents the initial value and the decay constant respectively. To produce a more symmetrical and controlled search output, and then equation can be rewritten in (9).

$$E_0 = GCP(C_{eq} - \phi C) \tag{9}$$

Here, generation rate control parameter (GCP) is the parameter of the control of the generation which represents the real probability of the update term. In conclusion, the (10) represents the following  $E_0$  updating rule:

$$C = c_{eq} + (P - p_{eq})M + \frac{F}{\phi V(1-M)} \tag{10}$$

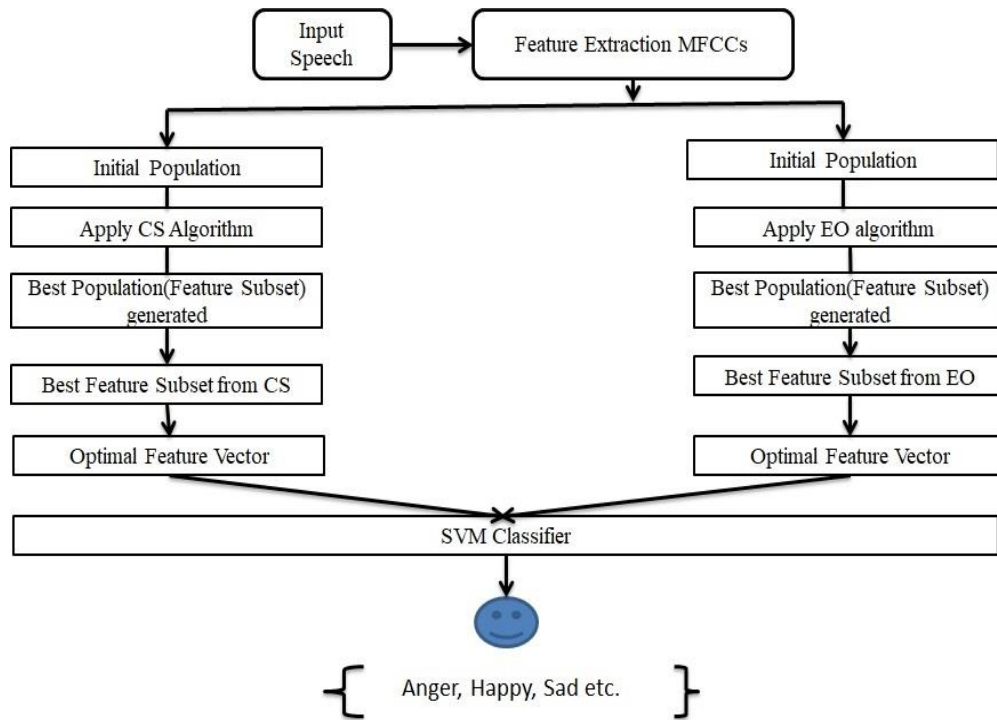


Figure 3. Experimental framework for speech emotion recognition

To perform good optimization cuckoo search follow the three basic rules: a) cuckoos lay one egg at a time, then dump it into a nest and try to choose at random; b) keeping healthy nests and passing down the best eggs to the next generations is the top priority; and c) the assumption is that the number of nests with available hosts is fixed and that the cuckoo’s eggs are discovered by the host birds with a probability of  $p_a$  and  $\phi$  (0, 1). Alternatively, the host bird can remove the egg from the nest or abandon the nest and build a new one to achieve a successful hatch. The nests are updated by random Lévy flights in the first stage of the algorithm. The two feature selection algorithms pseudocode is given below. Algorithm 1 gives the procedure to find the best optimal feature set for the speech recognition model from the above main feature set. One more popular nature-inspired algorithm i.e. CS used to find the optimal feature set and improve the recognition performance. The pseudo-code is described in Algorithm 2.

Algorithm 1. Pseudo code for EO based FS model for SER [24]

---

```

Input: generate initial population and feature space
Output: this is the final combination of features (best option)
1 Particle population is initialized as  $j = 1, 2, 3, \dots, n$ 
2 Give each equilibrium candidate a high fitness level
3 Parameters can be freely assigned  $k_1 = 2, k_2 = 1, GP = 0.5$ ;
4 While ( $i < \max[i]$ ) do // read all the sub folders in dataset in main folder
5   For  $i = 1, \dots, n$ ,  $n$  is number of particles do
6     Determine each  $i_{th}$  particle according to its fitness
7     If  $fit(p_i) < fit(p_{eq(1)})$  then
8       Replace  $p_{eq(1)}$  with  $p_i$  and  $fit(p_i)$  with  $fit(p_i)$ 
9     Else if  $fit(p_i) > fit(p_{eq(1)})$  and  $fit(p_{eq(2)})$  then
10      Replace  $p_{eq(2)}$  with  $p_i$  and  $fit(p_2)$  with  $fit(p_i)$ 
11     Else if  $fit(p_i) > fit(p_{eq(1)})$  and  $fit(p_i) > fit(p_{eq(1)})$  and  $fit(p_{eq(2)})$  and
12       $fit(p_i) > fit(p_{eq(3)})$  then
13       Replace  $p_{eq(3)}$  with  $p_i$  and  $fit(p_3)$  with the  $fit(p_i)$ 
14     Else if  $fit(p_i) > fit(p_{eq(1)})$  and  $(p_i) > fit(p_i) > fit(p_{eq(2)})$  and
15       $fit(p_i) > fit(p_{eq(3)})$  and  $fit(p_i) < fit(p_{eq(4)})$  then
16       Replace  $p_{eq(4)}$  with  $p_i$  and  $fit(p_4)$  with the  $fit(p_i)$ 
17     End if
18   End for
19    $EP_{avg} = (ep_{eq(1)} + epe_{eq(2)} + ep_{eq(3)} + ep_{eq(4)})/4$ 
20   Equilibrium pool  $E_{P_{eq.pool}} = (epeq(1), epeq(2), epeq(3), epeq(4), epeq(avg))$ 
21   Ensure the saving of memory if ( $i > 1$ )
22   Assign  $t = (1 - i/\max_i)(k_2 \cdot i/\max_i)$ 
23   For  $i = 1, \dots, n$ ,  $n$  is number of particles do
24     From the equilibrium pool, choose a random candidate
25     Generate random number  $\psi$  and  $mE = k_1 \times \text{sign}(h - 0.5) \times [\exp - \psi \cdot i - 1]$ 
26     Construct  $GCP = 0.5 \cdot h$  if  $h > GP$  else 0
27     Construct  $F0 = GCP(P_{eq} - \epsilon p)$ 
28     Update concentration  $P = p_{eq} + (P - p_{eq}) \cdot M + G/\psi V \times (1 - M)$ 
29   End for
30    $i = i + 1$ 
31 End while

```

---

Algorithm 2. Pseudo code for CS based FS model for SER [25]

---

```

Input: max number of iterations, maximum population size, and maximum number of features
Output: here is the final feature combination (best option)
1 Begin
2 Objective function  $f(x)$ , where  $x = (x_1, x_2, x_3, \dots, x_d)^T$ 
3 Initially populate  $n$  host nests  $x(i), i = 1, 2, \dots, n$ 
4 While ( $t < \text{Max\_Generation}$ ) or (stop criterion) do
5   Evaluate the quality or fitness of random cuckoos by Levy flights  $F_i$ 
6   Pick one nest out of the many  $n$  (say,  $j$ )
7   If ( $F_i < F_j$ ) then
8     Replace  $j$  by a new solution
9   End if
10  In ( $p_a$ ) of worst nests, a fraction is removed
11  Choose the most efficient solution (or keep nests with it)
12  Compare the current best solution to the solution ranked first
13  Results of the post-processing and visualization
14 End while
15 End

```

---

## 5. EXPERIMENTAL RESULTS AND DISCUSSIONS

We have relied on four prominent evaluating metrics like accuracy, F1 score, recall, and precision. These metrics are generated based on certain essential elementary measures contained in the confusion matrix. From the confusion matrix, we have calculated these parameters with the help of true positive, true negative, false positive and false-negative values. The two evolutionary algorithms above were developed with Python, openSMILE, and librosa tool kit. RAVDESS dataset contains a total of 400 speech samples of both males and females with different emotions in global language i.e. English. After applying the feature extraction to these data samples we got MFCCs and these dataset features are given to the SVM classifier to identify the emotion and estimate the accuracy of the model. In this work, to overcome the burden of classifier, the number of features is reduced using popular meta-heuristic approaches such as EO and CS. After applying the EO and CS algorithms individually, we achieved.

To evaluate the impact of this meta-heuristic approach precision and F1-score is the popular metrics used in the analysis of speech emotional classification. Here the Table 1 and Table 2 represents precision and F1-score of EO and CS-based FS model for SER tasks and corresponding graphical representation shown in Figure 4 and Figure 5. By analyzing the above results, we can say that meta-heuristic-based FS models have superior performance compared to the traditional feature selection methods which were discussed in section 2.

Finally, using this EO and CS algorithms-based FS model for speech emotion recognition accuracy is 89.64% and 92.71% respectively. Hence, most of the classification related problems like speech emotion recognition used these meta-heuristic optimizations and achieves impressive recognition rates. Table 3 shows the state of the art methods with our attempt using the meta-heuristic approach. In order to determine the impact of feature selection methods, the success rate obtained without any selection method is used as the reference value. From the above analysis, it is observed that, compared to traditional feature selection algorithms, the meta-heuristic approach is better accuracy for speech emotional intelligence.

Table 1. Precision and F1-score using EO based

FS model		
Emotional	Precision	F1-score
Angry	0.84	0.77
Happy	0.88	0.91
Sad	0.79	0.85
Disgust	0.86	0.79
Surprise	0.81	0.74
Neutral	0.82	0.75

Table 2. Precision and F1-score using CS based

FS model		
Emotional	Precision	F1-score
Angry	0.84	0.82
Happy	0.67	0.91
Sad	0.93	0.89
Disgust	0.95	0.91
Surprise	0.91	0.89
Neutral	0.87	0.95

Table 3. Comparison of existing work and proposed work

No.	Author	Features used	Classifier	Accuracy (%)
1	Ingale and Chaudhari [26]	MFCCs + LPCC	SVM	77
2	Shambhavi and Nitnaware [27]	MFCCs	SVM	84
3	El Ayadi, <i>et al.</i> [28]	MFCCs + prosodic	SVM	79
4	Mustaqeem and Kwon [29]	MFCCs	CNN	79.50
5	Issa <i>et al.</i> [30]	MFCCs	CNN	86.1
6	Proposed (EO and CS)	MFCCs	SVM	89.64 and 92.71

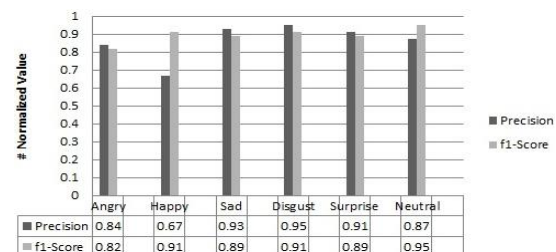
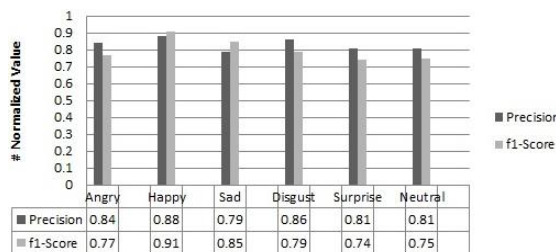


Figure 4. Precision and F1-score using EO-FS model      Figure 5. Precision and F1-score using CS-FS model

## 6. CONCLUSION

The main goal of this work is to achieve an impressive recognition rate with a smaller feature set. In real-time, the success rate was decreased due to the high dimensional feature set. To address this problem we attempted the FS model for SER using meta-heuristic approaches like EO and CS algorithm. During our experimentation, using normalization to reduce the feature set and increase the precision and F1-score.

However, during this work, we have faced some challenges related to several iterations to achieve the best fitness. One more challenge is to perform manual feature engineering instead of automatic feature engineering. Hence, there will be room for applying these meta-heuristic approaches based on automatic feature engineering like deep learning.

## ACKNOWLEDGEMENTS

We thank some of the associated research scholars for their comments that greatly improved the manuscript. I am also thanks my supervisor for his continuous support during this experimentation and some mathematical insights.

## REFERENCES





- [1] T. Özseven, "A novel feature selection method for speech emotion recognition," *Applied Acoustics*, vol. 146, pp. 320–326, 2019, doi: 10.1016/j.apacoust.2018.11.028.
- [2] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, 2005, doi: 10.1109/TKDE.2005.66.
- [3] Y. Yang and J. O. Pedersen, "A Comparative Study on Feature Selection in Text Categorization," in *Proc. of the Fourteenth International Conference on Machine Learning*, 1997 pp. 412–420. [Online]. Available: <http://surdeanu.cs.arizona.edu/mihai/teaching/ista555-spring15/readings/yang97comparative.pdf>
- [4] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005, doi: 10.1109/TPAMI.2005.159.
- [5] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. Wyman Street, Waltham, USA: Elsevier, 2011. [Online]. Available: <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
- [6] C. Yan, J. Ma, H. Luo, and A. Patel, "Hybrid binary coral reefs optimization algorithm with simulated annealing for feature selection in high-dimensional biomedical datasets," *Chemometrics and Intelligent Laboratory Systems*, vol. 184, pp. 102–111, 2019, doi: 10.1016/j.chemolab.2018.11.010.
- [7] Y. Mingshun, Y. Zijiang, and J. Guoli, "Partial maximum correlation information: A new feature selection method for microarray data classification," *Neurocomputing*, vol. 323, pp. 231–243, 2019, doi: 10.1016/j.neucom.2018.09.084.
- [8] T. Stützle and M. L. Ibáñez, "Automated Design of Metaheuristic Algorithms," *Handbook of Metaheuristics*, pp. 541–579, 2018, doi: 10.1007/978-3-319-91086-4\_17.
- [9] F. Glover and G. A. Kochenberger, *Handbook of metaheuristics*, New York: Springer, 2006, vol. 57, doi: 10.1007/b101874.
- [10] T. Tuncer, S. Dogan, and U. R. Acharya, "Automated accurate speech emotion recognition system using twine shuffle pattern and iterative neighbourhood component analysis techniques," *Knowledge-Based Systems*, vol. 211, 2021, doi: 10.1016/j.knsys.2020.106547.
- [11] C. Brester, E. Semenkin, and M. Sidorov, "Multi-objective heuristic feature selection for speech-based multilingual emotion recognition," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 6, no. 4, pp. 243–253, 2016. [Online]. Available: <https://www.sciendo.com/article/10.1515/jaiscr-2016-0018>
- [12] S. G. Koolagudi and K. S. Rao, "Emotion recognition from speech using source, system, and prosodic features," *International Journal of Speech Technology*, vol. 15, pp. 265–289, 2012, doi: 10.1007/s10772-012-9139-3.
- [13] B. Xin, L. Hu, Y. Wang, and W. Gao, "Stable feature selection from brain MRI," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, vol. 29, no. 1, doi: 10.1609/aaai.v29i1.9477.
- [14] G. V. S. George and V. C. Raj, "Accurate and stable feature selection powered by iterative backward selection and a cumulative ranking score of features," *Indian Journal of Science and Technology*, vol. 8, no. 11, 2015, doi: 10.17485/ijst/2015/v8i11/71766.
- [15] Yogesh C. K. et al., "A new hybrid PSO assisted biogeography-based optimization for emotion and stress recognition from the speech signal," *Expert Systems with Applications*, vol. 69, pp. 149–158, 2017, doi: 10.1016/j.eswa.2016.10.035.
- [16] N. Kozodoi, S. Lessmann, K. Papakonstantinou, Y. Gatsoulis, and B. Baesens, "A multi-objective approach for profit-driven feature selection in credit scoring," *Decision Support Systems*, vol. 120, pp. 106–117, 2019, doi: 10.1016/j.dss.2019.03.011.
- [17] X. -S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *2009 World Congress on Nature and Biologically Inspired Computing (NaBIC)*, 2009, pp. 210–214, doi: 10.1109/NABIC.2009.5393690.
- [18] 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, doi: 10.1109/ACCESS.2020.3035531.
- [19] F. Daneshfar, S. J. Kabudian, and A. Neekabadi, "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*, vol. 166, 2020, doi: 10.1016/j.apacoust.2020.107360.
- [20] Z. Zhang, "Speech feature selection and emotion recognition based on weighted binary cuckoo search," *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 1499–1507, 2021, doi: 10.1016/j.aej.2020.11.004.
- [21] S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," *PLoS One*, vol. 13, no. 5, 2018, doi: 10.5281/zenodo.1188976.
- [22] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the Munich versatile and fast open-source audio feature extractor," in *MM '10: Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1459–1462, doi: 10.1145/1873951.1874246.
- [23] L. Chen, X. Mao, Y. Xue, and L. L. Cheng, "Speech emotion recognition: Features and classification models," *Digital Signal Processing*, vol. 22, no. 6, pp. 1154–1160, 2012, doi: 10.1016/j.dsp.2012.05.007.
- [24] A. Faramarzi, M. Heidarinejad, B. Stephens, and S. Mirjalili, "Equilibrium optimizer: A novel optimization algorithm," *Knowledge-Based Systems*, vol. 191, 2020, doi: 10.1016/j.knsys.2019.105190.
- [25] X. -S. Yang and S. Deb, "Engineering optimisation by cuckoo search," *International Journal Mathematical Modelling and Numerical Optimisation*, vol. 1, no. 4, pp. 330–343, 2010, doi: 10.48550/arXiv.1005.2908.







- [26] A. B. Ingale and D. S. Chaudhari, "Speech emotion recognition," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 2, no. 1, pp. 235–238, 2012. [Online]. Available: <https://www.ijscce.org/wp-content/uploads/papers/v2i1/A0425022112.pdf>
- [27] Shambhavi S. S. and V. N. Nitnaware, "Emotion speech recognition using MFCC and SVM," *International Journal of Engineering Research and Technology*, vol. 4, no. 6, pp. 1067-1070, 2015. [Online]. Available: <https://www.ijert.org/research/emotion-speech-recognition-using-mfcc-and-svm-IJERTV4IS060932.pdf>
- [28] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011, doi: 10.1016/j.patcog.2010.09.020.
- [29] Mustaqeem and S. Kwon, "Optimal feature selection based speech emotion recognition using two-stream deep convolutional neural network," *International Journal of Intelligent Systems*, vol. 36, no. 9, pp. 5116-5135, 2021, doi: 10.1002/int.22505.
- [30] D. Issa, M. F. Demirci, and A. Yazici, "Speech emotion recognition with deep convolutional neural networks," *Biomedical Signal Processing and Control*, vol. 59, 2020, doi: 10.1016/j.bspc.2020.101894.

## BIOGRAPHIES OF AUTHORS



**Kesava Rao Bagadi**     received the B. Tech degree in Electronics and Communication Engineering from Jawaharlal Nehru Technological University, Hyderabad, India in 2006 and post-graduation in Systems and Signal Processing from Osmania University, India, in 2011. He is currently pursuing Ph.D with the department of ECE, Jawaharlal Nehru Technological University, and Anantapur, India. He is currently an Assistant Professor with the department of ECE, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, India. He has published various research articles in national and international journals and conferences. His main research interests areas of speech signal processing, emotional intelligence, Internet of Things and Optimization Methods. His research areas include feature engineering, feature selection methods, emotion recognition, machine learning, meta-heuristic algorithms and soft computing techniques. He can be contacted at email: [bkesavarao\\_ece@mgit.ac.in](mailto:bkesavarao_ece@mgit.ac.in).



**Chandra Mohan Reddy Sivappagari**     received the B.Tech degree in Electronics and Communication Engineering from Acharya Nagarjuna University, Guntur, Andhra Pradesh, India in 2001 and post-graduation in Digital Techniques and Instrumentation from Rajiv Gandhi Proudhyigiki Viswavidyalaya (RGPV), Bhopal, India in 2003. He obtained his PhD from JNTU Anantapur in the field of Embedded Systems in 2014. He is currently working as an Associate Professor with the Department of ECE, JNTUA, Anantapur, India. He published various international national and national journals and conferences. He wrote various books on Embedded systems from his research work. He acts as revivers for various journals. His research interest includes signal processing for communications, embedded systems and Internet of things. He can be contacted at email: [cmr.ece@jntua.ac.in](mailto:cmr.ece@jntua.ac.in).