Received: January 4, 2023. Revised: February 22, 2023.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Spiral Search Ant Colony Optimization Based AlexNet Model for Emotion Recognition

Pradipkumar Panchal[1]* **Hiren K. Mewada[2]**

[1]*Department of Electronics & Communication Engineering, Charotar University of Science and Technology, Anand, India*
[2]*Department of Electrical Engineering, Prince Mohammad Bin Fahd University, Dhahran 34754, Saudi Arabia*
* Corresponding author's Email: pradipeccse@gmail.com

Abstract: Emotion recognition is useful in understanding human recognition and is necessary for a variety of applications. Existing methods involve applying deep learning (DL) based techniques in emotion recognition and have the drawbacks of overfitting. This research involves applying the spiral search – Ant colony optimization (SS-ACO) for the optimal learning rate of the AlexNet model. The SS-ACO-AlexNet model is applied in the input face image for feature extraction. The extracted features are applied to long short term memory (LSTM) model for emotion classification. The spiral search technique is applied to increase the exploitation of the model and helps to escape the local optima trap. The SS-ACO method selects optimal learning rate and selective dropout to overcome overfitting problems in classification. On YALE datasets, the existing stochastic gradient descent linear collaborative discriminant regression classification (SGD-LCDRC) has 90.07% accuracy, which are lower when compared to proposed SS-ACO method, which has 99.2% accuracy.

Keywords: AlexNet, Ant colony optimization, Emotion recognition, Long short-term temory, Spiral search.

1. Introduction

Emotions have an important role in human well-being, which influences both physical and psychological health. The computer vision community views the task of identifying human emotion from facial expressions as challenging due to several issues, including individual variations in face shape, the challenge of identifying dynamic facial features, and the poor quality of digital images [1, 2]. The machine learning community finds it difficult to automatically recognize facial emotions in scenarios like partial occlusions, variable head positions, and lighting conditions. The key factor making it difficult to train an effective machine learning or deep learning model is the dearth of adequate samples with the aforementioned requirements in the baseline datasets [3]. The activities that convey our emotional state or attitude to others are known as emotional expressions and it occurs verbally and nonverbally. Studying physical characteristics from several modalities, particularly facial, vocal, and physical movements can help one understand complex human behavior [4]. Face emotion recognition technologies are used by many businesses that sell a wide variety of consumer goods to recognize client feedback and identify potential buyers. The field of machine learning gains additional value from these facial expressions. [5].

Researchers can accurately extract emotions from facial expressions using deep neural networks, specifically convolutional neural networks (CNN), recurrent neural networks (RNN), as well as LSTM models. For instance, numerous researchers showed that their deep learning-based approaches outperform competing algorithms on a variety of facial emotion identification data sets [6]. Consequently, a model for automatic emotion recognition includes a number of attention and applications. Applications of computer vision, as well as psychological research, include detecting depression in people as well as diagnosing developmental disorders in children by observing

their gaze and facial expressions during social interactions [7-9]. Other applications such as monitoring the conditions of the driver (such as the state of fatigue) as well as monitoring signs of attention are used to improve driver safety. Convolutional neural networks now have an increasing space ineffective feature extraction as well as recognition accuracy, notwithstanding their remarkable success in face emotion recognition algorithms [10]. The objective and the contribution of the model are discussed as follows:

1. The Spiral Search – Ant Colony Optimization (SS-ACO) model is proposed to increase the performance of emotion recognition. The SS-ACO model has the advantage of increasing the exploitation process around the best solution.
2. The SS-ACO model is applied for optimal learning rate and selective dropout in the AlexNet model to overcome the overfitting problem. The SS-ACO-AlexNet model extracts the relevant features for emotion recognition.
3. Extracted features were applied to the LSTM model for the classification of emotion from the input data. The SS-ACO model is highly effective than existing methods at categorization in emotion recognition.

This paper is structured as follows: recent research papers on emotion recognition are given in section 2 and the SS-ACO model explanation is given in section 3. The result analysis and comparison of the proposed SS-ACO model are demonstrated in section 4, and the conclusion of this research paper is given in section 5.

2. Literature review

Emotion recognition plays an important role in any interpersonal communication and this is considered as challenging task in computer vision. Some recent emotion recognition techniques are analyzed in this section.

Gabor wavelet transform (GWT) was recommended by Sulayman Ahmed [11] for feature extraction and feature reduction. The PSO approach was employed to determine the best feature. The deep learning method using six layers were utilised to recognise faces. To use the fitness function, that identifies the features which are the most similar to one another in quantity, the best and most optimal features are chosen by excluding the initial and final values of the characteristics. Discovered that the Gabor wavelet and deep learning cannot achieve the tests' findings for the recognition rate and accuracy because some volatility in the attribute values distorts the classification step. Consequently, there was a lot of diversity relative to Gabor wavelet characteristics.

To enhance the effectiveness of facial recognition, T. Syed Akheel [12] proposed a novel collaborative approach called stochastic gradient descent linear collaborative discriminant regression classification (SGD-LCDRC). The different inputs of the data images were distinguished in the suggested method by using content and style losses. Furthermore, by optimising the proportion of WCRE and CBCRE, the best projection matrix was learned using learning algorithm with velocity. Since the data dimensionality in this study was greatly reduced using logistic regression model, the suggested SGD-LCDRC individual's cost was linear. Examining the outcomes, it can be seen that the recognition rate of the algorithms rises as the samples in the training set increases.

Sparse auto encoder (SAE) was recommended by Rabah Hammouche [13] to enhance the features recovered by the Gabor filter bank. The Gabor filter bank extracts the features first. The SAE was being used to enhance the traits that had been retrieved. The complexity of the features was then reduced using the PCA + LDA method to get improved accuracy. Eventually, the cosine Mahalanobis distance is used to meet the face image set. It was found that the system performs at its best on all 7 datasets when the SAE approach is applied to the Gabor filter bank, improving the outcomes. Every database has an equal amount of training and testing samples for every individual, except for the ones with odd numbers of samples, where the training dataset outnumber the testing samples by another. The SAE's hidden size was steadily increased by one since the optimal solution was achieved.

A unique feature representation for facial emotion identification was proposed by D. Lakshmi and R. Ponnusamy [14] utilizing modified histogram of oriented gradients (HOG) and local binary pattern (LBP) feature descriptors. First, Butterworth high pass filter has been used to improve the observed region in order to locate the eye, nose, or mouth regions that use the Viola-Jones approach. After Viola-Jones face detection has identified the facial images. The properties of the identified eye, nose, and tongue areas are then extracted using the suggested modified HOG and LBP feature descriptor. Multi-class Support Vector Machine would then be utilised for recognition and classification. By using longitudinal, horizontal, and orthogonal gradient computing, modified HOG features are extracted, which increases the reliability of FER.

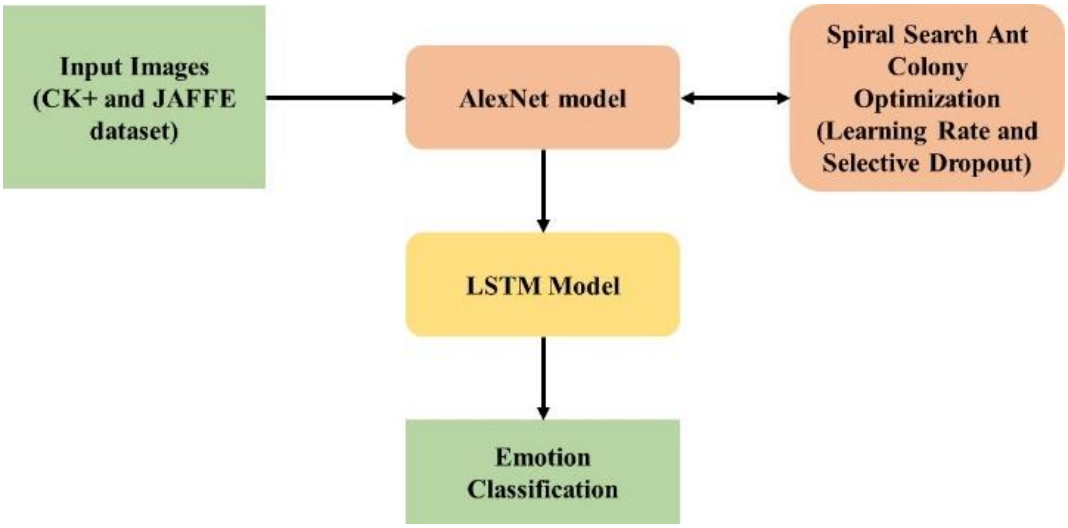

Figure. 1 Emotion Classification using SS-ACO model

In order to recognise eight facial emotions in static RGB images, Ahmed Hesham Mostafa [15] proposed the CNN Craft-net framework, which concatenates the output features from CNN, recurrent neural networks, and handmade features SIFT, SURF, and ORB calculated by a collection of visual words. To assess the developed framework, researchers employed the high-imbalanced AffectNet and FER2013 statistics. It should be observed that transfer learning aids in boosting the recommended accurateness, which rises steadily from None to ImageNet to AffectNet. When compared to many of the current state-of-the-art procedures, our solution for Fer2013 does not produce a favourable outcome for a variety of reasons.

A facial emotion identification process that uses deep belief networks (DBNs) and ouantum particle swarm optimization (QPSO) was proposed by Kamal A. El Dahshan [16]. The input image is first normalised by reducing the Region of Interest (ROI), allowing the targeted region to be obtained and any unnecessary portions to be removed. Second, several blocks are created from the ROI, and the best blocks are selected by using the structuring element. Third, to enhance system efficiency, the image down sampling method is modified to lower the size of the newly created sub image. Finally, the DBN is used to determine the emotion's class. The values of DBN variables are dynamically optimised using QPSO as opposed to changing the position. Even though adding more DBN layers takes longer, accuracy could be marginally improved. Thus, despite adding more layers, the effectiveness of the estimation results might degrade.

Convolutional neural network (CNN) integrated facial expression recognition using image filtering method was proposed by Ritanshi Agarwal [17]. Two convolutional layers and pooling layers make up our simple Classifier. It uses two different types of filtering to support the learning process: a Gaussian filter and a LoG filter. With Gaussian-CNN and LoG-CNN combinations, the chosen CNN design provides very good recognition accuracy. This is feasible as a result of the augmentation of facial images brought about by the application of the LoG or Gaussian filters. Conversely, since there are numerous LoG-CNN combinations that offer the highest level of accuracy, pooling layer seems to be more efficient than maximum pooling.

3. Proposed method

Emotion recognition is important in many applications and also it employs deep learning-based algorithms. The SS-ACO technique is applied for optimal selection of learning rate and selective dropout in the AlexNet model. The AlexNet extracted features were applied as input to the LSTM model and the LSTM model performs emotion classification. The flow of the SS-ACO model is shown in Fig. 1.

3.1 AlexNet

Deep architectures include more than just a secret layer. These concealed layers make it easier to improve and extract features. The effectiveness of deep networks for image classification is thus demonstrated to achieve a high categorization rate compared to other approaches, which stimulates everyone to adopt deep networks. A large network called AlexNet ranges between 650,000 and 60 million neurons. Multiple improvements were made to train these parameters by Krizhevsky [18].

The activation feature is the initial enhancement and the activation function for nonlinearity in the

traditional neural network is restricted to arctan, tanh, logistic function, etc. These activation functions will encounter the gradient vanishing problem because the gradient values they produce will only be important when the input is close to tiny range 0. To solve this issue, a rectified linear unit (ReLU), a new activation function, has been employed. The gradient of RELU is always 1 if the input is greater than 0. Additionally, it quickens the learning curve. RELU is defined by the following Eq. (1).

$$y = \max(0, x) \tag{1}$$

ReLU is used to calculate, and has a simple gradient. This deep network is made up of a number of little sub-networks. Since all sub-networks have the same loss function and are susceptible to overfitting, it would be beneficial to remove some of these layers. By removing many of these layers, the second modification sought to prevent overfitting. By removing the fully connected layers, this improvement can be implemented. For each iteration, a portion of the neurons are trained during the dropout. As a result of the neurons being compelled to cooperate by dropping out, the joint adaptation will be lowered, which will also increase and enhance the generalization. The overall network output is calculated as the average of the sub-networks. It is now obvious that dropout also strengthens and promotes toughness.

By using convolutional layers and a pooling layer, features are automatically extracted and minimized. For an image I with height h and width w, where m is the convolutional kernel with height b and width c, the convolution is described as follows in Eq. (2).

$$C(h,w) = (I \times m)(h,w) = \sum_b \sum_c I(h-b, w-c)m(b,c) \tag{2}$$

Convolution will enable the model to learn from visual features, and these parameters are shared to make the model simpler. The extracted features are cut down by using the layer pooling technique. A collection of nearby pixels from the feature map are used by the pooling layers to generate values for representation. In AlexNet, the feature map was condensed via max pooling. To create a 2x2 block with the maximum values, the max pooling uses a 4x4 block from the feature map.

Cross-channel normalization is a local normalization technique that enhances the feature generalization process. The feature maps must undergo normalization before being fed to the following layers. The result of cross-channel normalization is a sum of multiple nearby maps with the same positions. This mechanism has also been observed in actual neurons. In fully connected layers, the categorization was carried out.

3.2 Spiral search ant colony optimization

There are several rounds in the ACO algorithm [19, 20]. In each cycle, a number of ants build comprehensive answers based on heuristic data as well as the accumulated knowledge of earlier populations of ants. The pheromone trail that is left on a solution's parts serves as a representation of these accumulated experiences. According to the problem that needs to be solved, the pheromone might be placed on the connections or the components of a solution. The following describes the steps involved in the pheromone update rule.

3.2.1 Transition rule

In the ACO algorithm, a simple computational agent is an ant. It builds a remedy for the issue at hand iteratively. Each ant advances from a state r to a state s on each iteration of the method, which represents a more thorough intermediate solution. The k^{th} ant from state r to state s is chosen from the memorized unvisited states in J_k^r based on the Eq. (3).

$$s = \begin{matrix} \arg \\ u \in J_r^k \end{matrix} \max\left[\tau_i(r,u)^{\alpha}.\eta(r,u)^{\beta}\right] \quad if\ q \le q_0(Exploitation) \tag{3}$$

$\tau_i(r,u)^{\alpha}$ is stated as the amount of pheromone, $\eta(r,u)^{\beta}$ is stated as desirability of state transition. A posteriori evaluation of the effectiveness of the move is provided by the trail level. The trails are often updated once all ants have completed their responses, with the level of the trails altering to reflect moves that were either a part of "good" or "bad" answers.

In general, the k^{th} ant moves from state r to state s with probability $p_k(r,s)$ as shown in Eq. (4),

$$p_k(r,s) = \begin{cases} \frac{\tau(r,s)^{\alpha}.\eta(r,s)^{\beta}}{\sum_{u \in J_r^k} \tau(r,u)^{\alpha}.\eta(r,u)^{\beta}}, & if\ s \in J_r^k \\ 0, & otherwise \end{cases} \tag{4}$$

In this case, the transition probability is denoted by $p_k(r,s)$, the pheromone concentration among the state r and the state u in the i^{th} population is denoted by $\tau(r,u)$ and the length of the trail between the two states r and u is denoted by $\eta(r,u)$. J_r^k is the collection of unexplored states for the kth ant in the ith population, α and β are control parameters, as well as q is a uniform probability [0, 1].

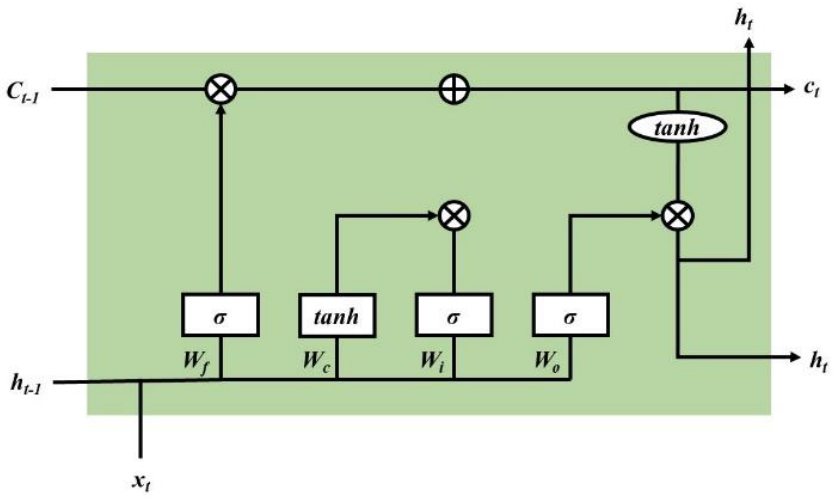

Figure. 2 LSTM cell unit consists of input, output and forget gate

3.2.2. The pheromone update rule

The pheromone trails need to be modified to raise the level of solution quality. Local, as well as global updates, are made to trails. The local trail update formula is explained in Eq. (5).

$$\tau(r,u) = (1-\rho)\tau(r,s) + \sum_{k=1}^{m} \Delta\tau_k(r,s) \quad (5)$$

In Eq. (3), $\rho(0 < \rho < 1)$ is the pheromone trial evaporating rate. $\Delta\tau_k(r,s)$ is the quantity of pheromone trail added to the edge (r,s) by ant k between time t and $t + \Delta t$ in the tour. It is described as shown in Eq. (6).

$$\Delta\tau_k(r,s) = \begin{cases} \frac{Q}{L_k} & (r,s) \in \pi_k \\ 0 & otherwise \end{cases} \quad (6)$$

Where the constant parameter is stated as$Q$, L_k is the distance of the sequence π_t toured by ant in Δt.

3.2.3 Spiral search technique

The second stage of the CSA was to produce a place in space at random. This kind of blind operation will reduce the algorithm's rate of convergence. We devised the spiral search process as a solution to this issue. The spiral search method gives crows more opportunities to take full use of their ideal locations, enhancing their capacity to find comprehensive solutions during the optimization process. By measuring the distance between the previous position and the crow's preferred position, the search space for the spiral structure was stated mathematically in Eq. (7).

$$x^{t+1} = D.\exp(b \times l).\cos(2\pi l) + gbest \quad (7)$$

Where $D = |gbest - x_i^t|$ Signified a constant for determining the shape of the logarithmic spiral, d indicated the distance between the current best position and the $i-th$ crow, and l stood for a random number in the range [1, 1].

3.3 Long short term memory

The LSTM network expands the RNN [21, 22]. The LSTM model is versatile in dealing with parameters with enormous dimensions and employs each layer's nonlinear activation patterns, allowing it to recognize nonlinear trends in data and retain knowledge from the past over an extended period. As a result, several time series issues have been successfully tackled with LSTM. The three different sorts of gates in the LSTM structure the input, forget, and output is an advantage. As shown in Fig. 2, the LSTM resolves the vanishing gradient issue with RNNs as well as permits long-term information storage.

Mathematical formulations exist for the primary information flow of the LSTM cell (Fig. 2). In the paradigm, the symbols $\oplus$ and $\otimes$ stand for addition and multiplication, while the arrow represents the direction of information flow. It is possible to express the first layer of the memory gate as follows in the Eq. (8).

$$f_t = \sigma(W_f \times x_t + U_f \times h_{t-1} + b_f) \quad (8)$$

where f_t signifies forgetting threshold at time t, s signifies the sigmoid activation function, W_f as well as U_f signifies the weights, x_t signifies the input value, h_{t-1} signifies the output value at time $t-1$, as well as b_f denotes the bias term.

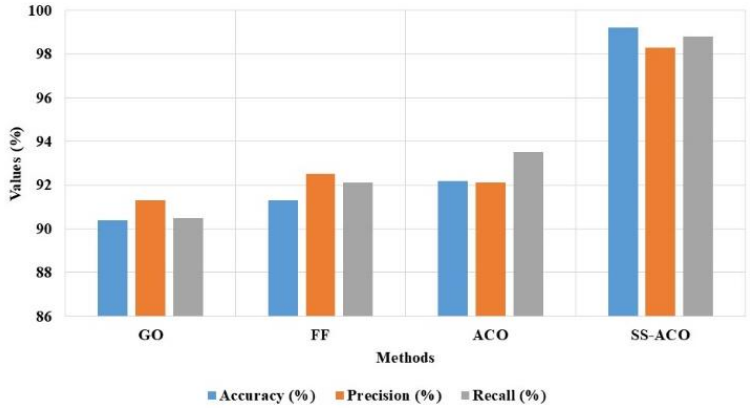

Figure. 3 The SS-ACO method compared with the existing feature selection technique

Table 1. The SS-ACO method with the existing feature selection technique on the PIE dataset

Methods	Accuracy (%)	Precision (%)	Recall (%)
GO	90.4	91.3	90.5
FF	91.3	92.5	92.1
ACO	92.2	92.1	93.5
SS-ACO	99.2	98.3	98.8

Table 2. The SS-ACO comparison with the existing classifier on PIE dataset

Methods	Accuracy (%)	Precision (%)	Recall (%)
RF	88.3	88.7	88.8
SVM	86.2	85.8	85.7
LSTM	90.4	91.2	91.5
CNN	92.3	91.4	91.7
SS-ACO	99.2	98.3	98.8

The second input gate selects the data from the current input vector that should be kept in the cell state. Decision i_t, is one example of this, as it modifies the value and $tanh$ layer to produce a fresh state value C_t. The specific values in equations are as follows (9 & 10).

$$i_t = \sigma(W_i + x_t + U_i \times h_{t-1} + b_i) \quad (9)$$

$$\bar{C}_t = \sigma(W_c \times x_t + U_c \times h_{t-1} + b_c) \quad (10)$$

where i_t signifies the input threshold at time t, W_i, U_i, W_c, and U_c are the weights, b_c and b_i are bias terms. To upgrade the cell's current time t, the phrase reads as follows in Eq. (11).

$$C_t = f_t \times C_{t-1} + i_t \times \bar{C}_t \quad (11)$$

The third layer is represented as follows and is generated as output data in the present time step in Eq. (12).

$$O_t = \sigma(W_o \times x_t + U_o \times h_{t-1} + b_0) \quad (12)$$

Where O_t signifies the output threshold at time t, W_o as well as U_o are the weights and b_o is the bias term. Then, the cell's final result can be stated in Eq. (13).

$$h_t = O_t \times \tanh(C_t) \quad (13)$$

Where h_t indicates the output value of the cell at time t, $tanh$ denotes the activation function, as well as C_t signifies the state of the cell at time t. Following the passage of the data via the three gates, the output is useful information as well as there is a loss of incorrect information.

4. Results

The SS-ACO method is applied in Alexnet-LSTM model for optimum learning rate and selective dropout. The SS-ACO method is tested on two datasets and compared with the existing technique for performance analysis.

Datasets: PIE dataset [23], FEI dataset [24], and Yale Face dataset [25] are employed to assess the generated model's performance.

The SS-ACO method is compared with contemporary feature selection techniques, as shown in Table 1 and Fig. 3. The SS-ACO method has the advantage of spiral search that involves in increase the exploitation of the model. Existing optimization techniques such as ACO, GO, and FF has limitations of local optima trap, lower exploitation, and poor convergence. The spiral search technique in the SS-ACO method involves improving the exploitation by searching around the best solution region. The SS-ACO method has 99.2 % accuracy and the ACO method has 92.2 % accuracy on the PIE dataset.

The SS-ACO method is compared with standard classifiers such as RF, SVM, LSTM as well as CNN, as in Table 2 and Fig. 4. The SS-ACO method increases the exploitation of the model using a spiral

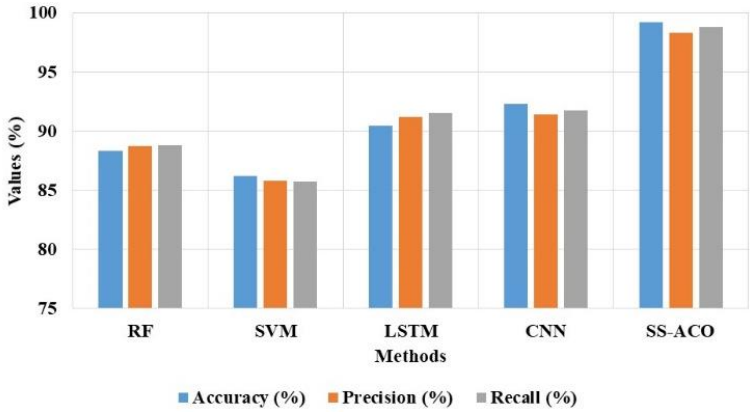

Figure. 4 The SS-ACO method is compared with the existing classifier on the PIE dataset

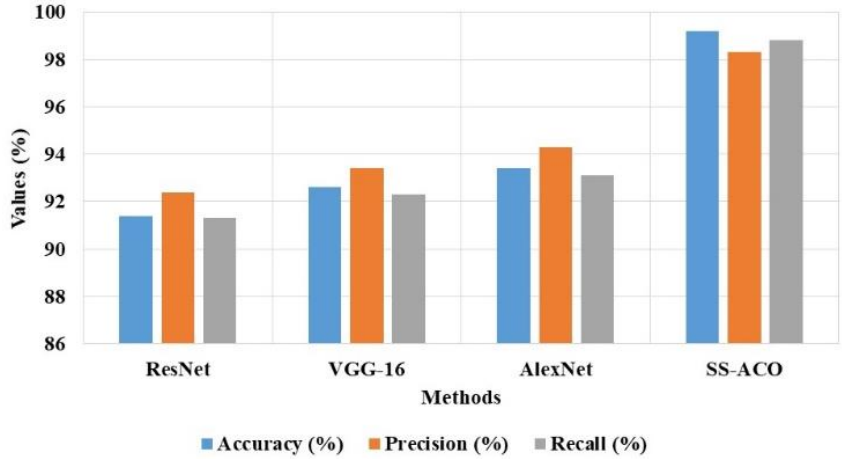

Figure. 5 The SS-ACO method is compared with the classifier on FEI dataset

Table 3. The SS-ACO method comparison with the deep learning method on FEI dataset

Methods	Accuracy (%)	Precision (%)	Recall (%)
ResNet	91.4	92.4	91.3
VGG-16	92.6	93.4	92.3
AlexNet	93.4	94.3	93.1
SS-ACO	99.2	98.3	98.8

Table 4. The SS-ACO method in comparison with the deep learning method on FEI dataset

Methods	Accuracy (%)	Precision (%)	Recall (%)
ResNet	93.2	93.7	93.4
VGG-16	94.8	95.2	95.3
AlexNet	95.1	94.6	94.2
SS-ACO	99.1	98.6	98.5

search for the best solution that helps to increase the performance of emotion recognition. The CNN model has an overfitting problem, and the LSTM model has to vanish gradient problem. The SVM model has an issue with imbalanced data, whereas the LSTM model has an issue with vanishing gradients. The SS-ACO method has accuracy of 99.2 % and the CNN model has 92.3 % of accuracy on emotion recognition.

The SS-ACO method is compared with deep learning techniques such as AlexNet, VGG-16, and ResNet model, as depicted in Table 3 and Fig. 5 and the models have drawbacks of overfitting in the classifier. The SS-ACO model applies spiral search to increase the exploitation in the optimal weight value to overcome the overfitting problem. The SS-ACO model has 99.2 % accuracy, AlexNet has 93.4 % accuracy, and VGG-16 has 92.6 % accuracy.

The SS-ACO method is compared with the deep learning technique on FEI dataset, as shown in Table 4 and Fig. 6. The SS-ACO method applies a spiral search technique to increase the exploitation for optimal learning rate for AlexNet and selective dropout. This technique helps to improve the classification performance and overcome the overfitting problem. The SS-ACO method has 99.1 % accuracy, and AlexNet has 95.1 % of accuracy on the JAFFE dataset. The SS-ACO model is tested on two datasets such as YALE, ORL, JAFFE and CK+ datasets as compared with existing in Table 5, 6, 7 and 8 respectively. From the overall comparisons

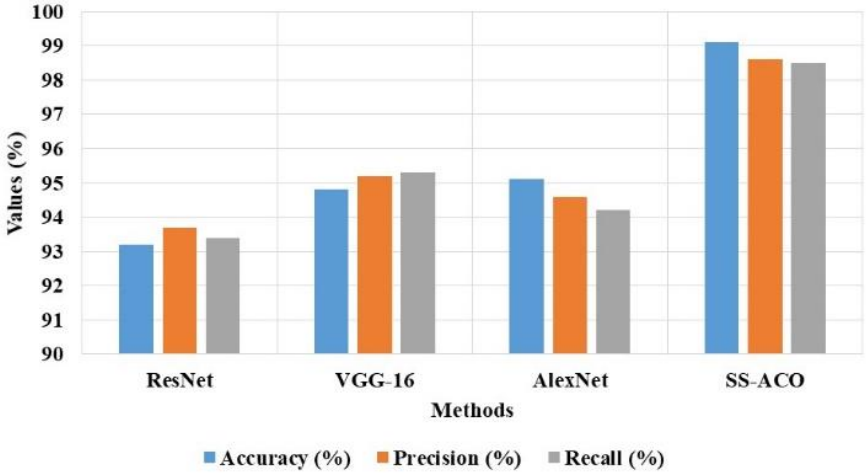

Figure. 6 The SS-ACO method is compared with the deep learning method on FEI dataset

Table 5. Comparative Analysis on YALE dataset

Methods	Datasets	Accuracy (%)
GWT-PSO [11]	YALE	94.66
SGD-LCDRC [12]	YALE	90.07
SS-ACO	YALE	99.20

Table 6. Comparison on ORL datasets

Methods	Datasets	Accuracy (%)
SGD-LCDRC [12]	ORL	93.34
SS-ACO	ORL	98.73

Table 7. Comparison on JAFFE datasets

Methods	Datasets	Accuracy (%)
SAE [13]	JAFFE	100
SS-ACO	JAFFE	100

Table 8. Comparison on CK+ datasets

Methods	Datasets	Accuracy (%)
Modified HOG and LBP [14]	CK+	97.66
SS-ACO	CK+	99.01

(YALE, ORL, JAFFE and CK+), the proposed SS-ACO achieved betters accuracy which is displayed in above tables. The SS-ACO method applies spiral search to increase the exploitation of the model. The SS-ACO method performs selection of optimal learning rate and selective dropout to overcome the overfitting problem and increase the performance. Existing methods such as GWT-PSO [11], SGD-LCDRC [12], SAE [13], Modified HOG and LBP [14] models have drawbacks of overfitting that affects the efficiency of the model. The SS-ACO method has 99.2 % of accuracy which is better when compared with existing GWT-PSO [11] and SGD-LCDRC [12], which has 94.66 % and 90.07 % of accuracy on YALE datasets.

5. Conclusion

Emotion recognition is necessary for many applications, and several new approaches for emotion recognition have been developed to analyze human behavior. Existing methods in emotion recognition have limitations of overfitting issues caused by feature selection. This research proposes the SS-ACO-AlexNet model to improve emotion recognition performance. The SS-ACO method has the advantage of applying a spiral search technique to increase the exploitation process. The SS-ACO method is applied for optimal selection of learning rate and selective dropout in the AlexNet model. The YALE, ORL, JAFFE and CK+ datasets are applied to evaluate the performance of the SS-ACO model. When compared to the existing GWT-PSO and SGD-LCDRC, which have 94.66% and 90.07% accuracy on YALE datasets, the SS-ACO approach has 99.2% accuracy, which is better. The model's future work will involve using a concentration layer in the emotion identification process.

Notations

Notation	Description
$\rho(0 < \rho < 1)$	Pheromone trial evaporating rate
$\eta(r,u).$	Length of the trail between the two states r and u
$\tau(r,u)$	Pheromone concentration among the state r and the state u in the i^{th} population
J_r^k	Collection of unexplored states
α and β	Control parameters
q	Uniform probability
$p_k(r,s)$	Transition probability

$\Delta\tau_k(r,s)$	Quantity of pheromone trail added to the edge (r,s)
L_k	Distance of the sequence π_t visited by ant in Δt
Q	Constant parameter
$\otimes$	Multiplication,
$\oplus$	Addition
x_t	Input value
h_{t-1}	Output value
s	Sigmoid activation function
f_t	Forgetting threshold
h_t	Output value of the cell
t	Time
$tanh$	The activation function
C_t	State of the cell
O_t	Output threshold
b_c and b_i	Bias
W_o and U_o	Weights
i_t	Input threshold
$p_k(r,s)$	k^{th} ant moves from state r to state s with probability
I	Image
$C(h,w)$	Convolution
w	Width
m	Convolutional kernel
h	Height
$\tau_i(r,u)^\alpha$	Amount of pheromone,
$\eta(r,u)^\beta$	Desirability of state transition.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

For this research work all authors' have equally contributed in Conceptualization, methodology, validation, resources, writing—original draft preparation, writing—review and editing.

References

[1] Y. Said, and M. Barr, "Human Emotion Recognition Based on Facial Expressions via Deep Learning on High-Resolution Images", *Multimedia Tools and Applications*, Vol. 80, No. 16, pp. 25241-25253, 2021.

[2] E. S. Salama, R.A. E. Khoribi, M. E. Shoman, and M. A. W. Shalaby, "A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition", *Egyptian Informatics Journal*, Vol. 22, No. 2, pp. 167-176, 2021.

[3] S. Kuruvayil, and S. Palaniswamy, "Emotion recognition from facial images with simultaneous occlusion, pose and illumination variations using meta-learning", *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, No. 9, pp. 7271-7282, 2022.

[4] L. Schoneveld, A. Othmani, and H. Abdelkawy, "Leveraging recent advances in deep learning for audio-visual emotion recognition", *Pattern Recognition Letters*, Vol. 146, pp. 1-7, 2021.

[5] P. Naga, S. D. Marri, and R. Borreo, "Facial emotion recognition methods, datasets and technologies: A literature survey", *Materials Today: Proceedings*, 2021.

[6] J. Rößler, J. Sun, and P. Gloor, "Reducing videoconferencing fatigue through facial emotion recognition", *Future Internet*, Vol. 13, No. 5, p. 126, 2021.

[7] L. N. Do, H. J. Yang, H. D. Nguyen, S. H. Kim, G. S. Lee, and I. S. Na, "Deep neural network-based fusion model for emotion recognition using visual data", *The Journal of Supercomputing*, Vol. 77, No. 10, pp. 10773-10790, 2021.

[8] D. Lakshmi, and R. Ponnusamy, "Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders", *Microprocessors and Microsystems*, Vol. 82, p. 103834, 2021.

[9] R. Magherini, E. Mussi, M. Servi, and Y. Volpe, "Emotion recognition in the times of COVID19: Coping with face masks", *Intelligent Systems with Applications*, Vol. 15, p. 200094, 2022.

[10] Y. Zhong, L. Sun, C. Ge, and H. Fan, "HOG-ESRs Face Emotion Recognition Algorithm Based on HOG Feature and ESRs Method", *Symmetry*, Vol. 13, No. 2, p. 228, 2021.

[11] S. Ahmed, M. Frikha, T. D. H. Hussein, and J. Rahebi, "Optimum feature selection with particle swarm optimization to face recognition system using Gabor wavelet transform and deep learning", *BioMed Research International*, pp. 1-13, 2021.

[12] T. S. Akheel, V. U. Shree, and S. A. Mastani, "Stochastic gradient descent linear collaborative discriminant regression classification-based face recognition", *Evolutionary Intelligence*, pp. 1-15, 2021.

[13] R. Hammouche, A. Attia, S. Akhrouf, and Z. Akhtar, "Gabor filter bank with deep autoencoder based face recognition system", *Expert Systems with Applications*, p. 116743, 2022.

[14] D. Lakshmi, and R. Ponnusamy, "Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders", *Microprocessors and Microsystems*, Vol. 82, p. 103834, 2021.
[15] A. Mostafa, H. E. Sayed, and M. Belal, "Facial Expressions Recognition Via CNNCraft-net for Static RGB Images", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 4, pp. 410-420, 2021, doi: 10.22266/ijies2021.0831.36.
[16] K. A. E. Dahshan, E. K. Elsayed, A. Aboshoha, and E. A. Ebeid, "Recognition of Facial Emotions Relying on Deep Belief Networks and Quantum Particle Swarm Optimization", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 4, pp. 90-101, 2020, doi: 10.22266/ijies2020.0831.09.
[17] R. Agarwal, N. Mittal, and H. Madasu, "Convolutional Neural Network Based Facial Expression Recognition Using Image Filtering Techniques", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 5, pp. 410-421, 2021, doi: 10.22266/ijies2021.1031.08.
[18] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size", *arXiv preprint arXiv:1602.07360*, 2016.
[19] W. Deng, J. Xu, and H. Zhao, "An Improved Ant Colony Optimization Algorithm Based on Hybrid Strategies for Scheduling Problem", *IEEE Access*, Vol. 7, pp. 20281-20292, 2019.
[20] D. Zhao, L. Liu, F. Yu, A. A. Heidari, M. Wang, D. Oliva, and H. Chen, "Ant colony optimization with horizontal and vertical crossover search: Fundamental visions for multi-threshold image segmentation", *Expert Systems with Applications*, Vol. 167, p. 11412, 2021.
[21] R. Swaminathan, S. Mishra, and A. A. Routray, "CNN-LSTM-based fault classifier and locator for underground cables", *Neural Comput & Applic*, Vol. 33, pp. 15293–15304, 2021.
[22] G. Kłosowski, T. Rymarczyk, K. Niderla, M. Rzemieniak, A. Dmowski, and M. Maj, "Comparison of Machine Learning Methods for Image Reconstruction Using the LSTM Classifier in Industrial Electrical Tomography", *Energies*, Vol. 14, No. 21, p. 7269, 2021.
[23] https://www.cs.cmu.edu/afs/cs/project/pie/multipie/multi-pie/home.html
[24] https://fei.edu.br/~cet/facedatabase.html
[25] http://vision.ucsd.edu/content/yale-face-database