Multimedia Tools and Applications (2023) 82:2437–2464
https://doi.org/10.1007/s11042-022-13378-x

An artificial intelligence driven facial emotion recognition system using hybrid deep belief rain optimization

Fakir Mashuque Alamgir[1] • Md. Shafiul Alam[1]

Received: 18 June 2021 / Revised: 24 January 2022 / Accepted: 7 June 2022 /
Published online: 27 June 2022

Abstract

Facial expression recognition is a process of identifying the different facial expressions of the individuals to categorize the mental health of the individual. This system is used in most of the fields but is vastly used in the medical field to identify the mental health issues. In this paper, a novel approach has been proposed to identify the facial expressions of the individuals and categorizing it into seven different emotions. Initially, the images collected from the dataset are subjected to pre-processing for de-noising. Then, the major geometric and appearance-based features are extracted from the images. The most relevant features are selected from the extracted feature set. Finally, based on the selected features, the classification is performed where the input images get labelled into seven different emotions. The classification is carried out with the use of the hybrid strategy called the Deep Belief Rain Optimization (DBRO) technique. The efficiency of the proposed model is proved through the simulations and it is identified to outperform the other existing approaches.

Keywords Facial expression recognition · Pre-processing · Feature extraction · Feature selection · Classification · Optimization

1 Introduction

Automatic recognition of facial expressions is increasingly becoming popular in recent times across various domains such as human-computer interactions, surveillance and health care. Recognizing facial expressions helps in identifying the mental health and mood swings of the individuals so that efficient measures can be taken abruptly [20]. Difficulties identified in the normal interactions and behaviours of the individuals are the major symptoms of abnormal

✉ Fakir Mashuque Alamgir
fma@ewubd.edu

[1] Department of Electrical & Electronic Engineering, University of Dhaka, Dhaka, Bangladesh

mental functioning [8]. Based on the social interactions and the ability of the individual in identifying others known to be mentalization will be degraded in many adults with diverse irregular mental stability [37]. Accurate determination of such symptoms in early stages helps in reducing the severity of the disease meanwhile improving the behaviour with proper social interactions. For this purpose, various researchers have formulated solutions with different techniques, the learning strategies outperformed the other techniques in accurate identification [23, 32].

The emotions of the individuals are identified from various factors such as audio, video, speech, images, EEG signals, etc. The learning techniques like machine and deep learning strategies help in identifying the facial expressions more accurately than any other techniques [35]. The accurate determination of the facial expressions through learning relies on the input features provided [30]. The input images subjected to facial emotion recognition, include diverse expressions of the individuals, and the classification system is able to diagnose the changes based on the muscle movements and the patterns extracted from the facial images. The patterns indicate the features extracted from the images [22]. Most important features indicating the facial expressions are extracted for better classification. Feature extraction phase contributes much in the classification system. Diverse features are required to be extracted to differentiate the expressions of the individuals [18]. There are various techniques employed in extracting the facial features such as Local Binary Pattern (LBP), Gabor Filters (GF), Active Appearance Model (AAM), etc. [5].

Identifying the most optimal features for classification ensures efficient classification results. Therefore, feature selection strategies are used to reduce the dimensionality of the features and to select the most relevant features based on the application [26]. Many techniques employed various strategies like heuristics, optimizations, etc. for selecting the optimal features. Out of all the available strategies, the optimization based strategies are reported to have efficient results with better accuracy [39]. Feature selection phase is responsible for reducing the dimensionality of the features and hence the useful information are conceded to the neural networks for classification. The meta-heuristic algorithms play a vital role in identifying the most relevant features from the extracted feature set [38]. Choice of better optimization strategies are appreciable to obtain efficient results with reduced computational complexity and cost. The meta-heuristic algorithms are highly capable of traversing the entire search space to identify the local and global optimal solutions. Therefore, all the best and optimal features can be taken for classification without the neglecting any feasible features from the feature set [2]. Recently, there are various meta-heuristics developed in literature for various purposes. Some of them include multi-objective self-organizing optimization for sparse span array (MOSSA) [15], graphics processing unit (GPU)-based parallel tabu search algorithm (GPTS) [11], biogeography-based optimization (BBO) [4], an efficient and robust fusion bat algorithm (ERFBA) [17] and hybrid optimization like whale optimization algorithm with differential evolution (WOA-DE) [16]. These algorithms are seen to be effective and can also be implemented for feature selection in near future.

For any research article regarding the classification of facial emotions, the classification phase is highly crucial to obtain knowledge regarding the application. Selected features are subjected to classification to understand the problems with individuals possessing different behaviours [34]. Machine and deep learning strategies are known to produce efficient classification results than any other techniques available. Almost all the researchers have formulated solutions for classifications based on the learning strategies due to its efficiency in understanding the different behaviours of the features [31]. The psychological stress, mood swings, mental problems and social behaviours can be deeply understood by the learning strategies than any other techniques [29]. Emotions of the individuals are capable of demonstrating the different problems and mood swings in them. There are many studies formulated to provide knowledge about the facial emotions of the individuals and its applications in today's world [28]. Other than the available models, powerful classifiers such as ResNet [14], RetinaNet [36], YOLO [7], AlexNet [10] can also be implemented in this domain to obtain better results in future.

In this article, the facial expressions of individuals are monitored for identifying the victimized persons. The geometric and appearance based features play a vital role in accurate determination of diverse facial expressions. The deep learning strategy is used for the classification purposes and the proposed work uses the Deep Belief Network (DBN) for efficient classification. Though there are diverse models in literature for emotion recognition, there are various problems with those models. Most of the models fail to provide optimum results when there is a variation in the number of samples. Also, the models are not capable of producing even results when trained with different datasets. Apart from these, the results of the neural network is affected by the weight parameter and hence optimization of the weight parameter is more important. In this work, the exact weight parameter is identified using the Rain Optimization Algorithm (ROA). This algorithm is a recently developed optimization algorithm and can be used for various engineering applications and mostly for drilling applications. Due to its stability and better convergence ability, the algorithm has been used in the neural network for weight optimization. The lack of appropriate strategies to resolve the issues in identifying the facial expressions of individuals stayed a major motivation behind the proposed work. The major aim of the proposed work is to provide accurate classification results in identifying the facial expressions and to improve the research area in this field. The major contributions of the proposed work is as follows:

(i) Facial expression recognition is a major task that has wide range of applications. In this paper, a novel hybrid strategy called the Deep Belief Rain Optimization (DBRO) is proposed to identify the facial expressions of individuals.
(ii) Proposing the new Multi-Objective Seagull Optimization Algorithm (MOSOA) for the feature selection to reduce the dimensionality of the features by selecting the most useful and relevant features for classification.
(iii) Evaluating the performance of the proposed approach in terms of accuracy, precision, recall, f-measure and specificity against the other facial emotion recognition strategies.

The rest of the paper is organized as follows: Section 2 presents the literature review of the recent facial emotion recognition strategies, Section 3 presents the complete demonstration of the proposed methodology, Section 4 presents the simulation results along with comparisons and analysis, and finally Section 5 provides the conclusion of the proposed work.

2 Related work

Some of the most recent research works relevant to facial emotion recognition are summarized as follows Mehendale [19] introduced a technique called the facial emotion recognition using convolutional neural networks (CNN) known as FERC, to identify the facial expressions of individuals from the images. The technique included two parts where the first part relied on removing the background of the images and the second part relied on extracting the facial features. The expression vector (EV) was used in that technique for detecting the five common types of facial expressions. The FERC algorithm was capable of working with diverse orientations as a result of the 24 digit long EV matrix. The emotions from the images were accurately determined with the help of the background removal technique. The experimental results proved that the technique was able to detect the emotions of the individuals more accurately. The authors also suggested that the algorithm can be extended for lie detection, recognizing the moods of students and for many other applications.

Moghaddam et al. [27] presented a light field based technique to identify the facial expressions of humans based on deep learning. The facial features were extracted using the VGG 16 convolutional neural network and the detection was carried out using the Bi-directional Long Short Term Memory (Bi-LSTM). The learning process was regulated using the spatio-angular features extracted from the images. The architecture used the spatio-angular learning layer for exploring the angular relationships and analysing the input sequences in forward and backward directions. The attention layer was used in Bi-LSTM for advanced learning of the extracted features in the last layer. The classification was scored and the output was evaluated for identifying the facial emotions. Experimental results demonstrated that the technique was capable of identifying the facial expressions more accurately.

Arora and Kumar [3] presented an automatic hybrid technique known as Autofer for automatically identifying the facial expressions of humans. The technique hybridized the feature extraction and optimization techniques based on the Principal Component Analysis (PCA) and Particle Swarm Optimization (PSO) to achieve higher precision in detection. Initially, the pre-processing was implemented based on Gradient Filtering and PCA was called for extracting the major facial features where the eigen vectors were used to extract the eigen faces for differentiating the current and novel faces. The feature were selected on the basis of the meta-heuristic PSO algorithm, where the most relevant features were selected and the classification was carried out using the model. Upon experimentation, the technique proved to provide better classification results with higher precision than the compared techniques.

Hassan and Mohammed [9] presented a concept for recognizing facial emotions based on graph mining. The facial regions were represented a graph of nodes and edges where the emotions were recognized based on the sub-structures of the graph. The frequent substructures of the graphs were identified using the gSpan frequent sub-graphs mining algorithm. The overlap ratio metric was used to reduce the count of sub-graphs generated. Then the major sub-graphs were selected and the binary classification technique was used to classify the sub-graphs based on facial expressions into six different levels of classification. The technique introduced a meta-heuristic called the binary cat swarm optimization (CSO) in the classification layers to improve the overall classification accuracy. Upon simulations, the graph mining technique proved to outperform many other conventional techniques with 90% overall accuracy.

Alreshidi and Ullah [1] introduced a framework for emotion detection and classification using hybrid features. The framework employed two different machine learning algorithms for emotion detection and classification purposes. Initially, the faces of the individuals were detected from the images using the Adaboost cascade classifiers. Then the major appearance based facial features called the Neighbourhood Difference Features (NDF) were extracted. The extracted NDF features were based on the relationships between the neighbourhoods and with respect to the central region. Finally, the classification was carried out with the extracted features using the Random Forest Classifier (RFC). The RFC labelled the classes based on the votes gained into seven different classes. The experimental results demonstrated the effectiveness of the approach over the other approaches.

While using deep neural networks for class label prediction, the layers used in these networks post larger influence in the overall performance. An extended deep model for emotion recognition was introduced by Jain et al. [13] in which different convolution layers and residual blocks were installed to predict the appropriate class labels for facial images. The model was tested using two different datasets and it provided desired results on both the evaluations. But, that model suffered from computational complexity due to the complex structure. To attain higher classification accuracy, Wang et al. [33] introduced an optimized classification model in which the training was improved through JAYA algorithm. The feed forward neural network was utilized for classification and the random parameter was initialized with the use of that optimization. The model provided better classification outcomes than the compared approaches.

From the surveys taken, it has been identified that there are no efficient strategies to identify the facial emotions of individuals with improved accuracy. Most of the techniques used learning algorithms to classify the facial expressions and certain techniques relied on heuristics to categorize the emotions. Moreover, the available techniques for facial emotion recognition failed to achieve desired accuracy rate as a low accuracy rate might degrade the classification results leading to certain critical issues specifically in the medical sectors. The proposed framework works on automatically identifying the facial expressions of individuals, providing a better classification result with higher accuracy.

3 Proposed methodology

Recognition of facial expressions is vital to demonstrate the characteristics of an individual and to identify the medical conditions related to mental health. To classify the facial expressions of individuals, the article proposes a novel hybrid Deep Belief Rain Optimization (DBRO) technique. The proposed approach consists of four major phases such as pre-processing, feature extraction, feature selection and classification. Initially, the images collected from the dataset are pre-processed to remove the noises that is present in the images due to environmental conditions. After pre-processing, the major geometric and appearance based features are extracted using the Histogram of Oriented Gradients (HOG) and Gabor filter (GF). The extracted features are then passed to the feature selection phase, where the most relevant features are selected for classification. The feature selection is carried out using the Multi-Objective Seagull Optimization Algorithm (MOSOA). Finally, based on the selected features, the classification is carried out using the DBRO technique. This is a hybridized technique combining Deep Belief Network (DBN) and Rain Optimization Algorithm (ROA) to achieve higher accuracy in classification. The global architecture of the proposed framework is presented in Fig. 1.

3.1 Image pre-processing

Pre-processing is the initial phase of the proposed work carried out to improve the quality of the images by reducing the noises and distortions. The proposed work uses the joint bilateral filter (JBF) for pre-processing. This filter highly helps in preserving the edge information present in the images. The conventional bilateral filtering preserves the edge information based

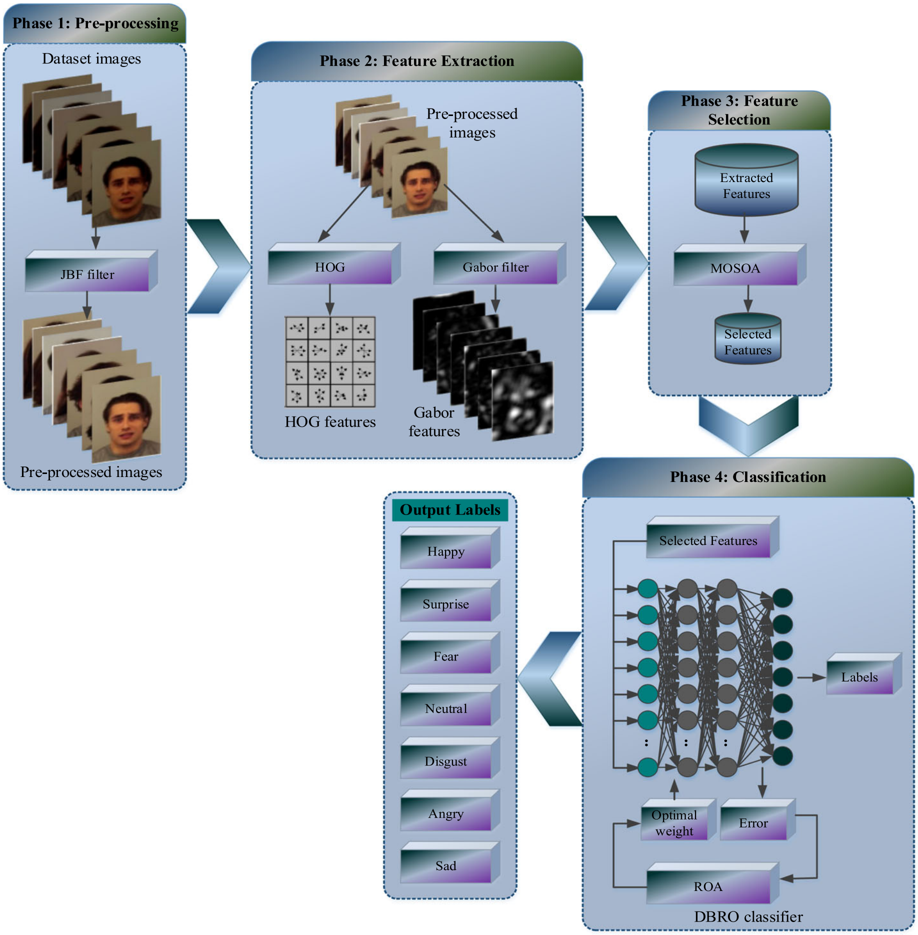


Fig. 1 Global architecture of the proposed framework for emotion classification

on the intensities of the images. But that filter is not capable of preserving the edge information of the images with lower intensity. Therefore, to overcome such disadvantages, the proposed work uses the JBF filtering [24] technique. To preserve the edge information more efficiently than the conventional filter, the first principal component is used as the guidance image replacing the range filtering kernel. The structural diagram of the pre-processing technique is depicted in Fig. 2.

The representation of the JBF filtering technique is as follows:

$$I_{JBF}(j,k,b) = \frac{1}{i(j,k)} \sum \Big\{ G_{\sigma_d}(j{-}p, k{-}q) \\ {}^*G_{\sigma_r}[I_{Pc}(j,k){-}I_{Pc}(p,q)] I(p,q,b) \Big\} \tag{1}$$

where, I_{Pc} denotes the input image at a pixel location (j, k), indicates the first principal component, σ_d *and* σ_r are the domain and range parameters, p *and* q are the pixel points

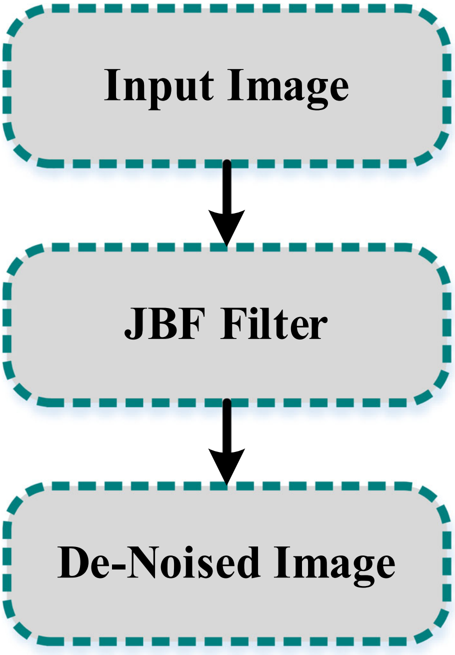


Fig. 2 Structural Diagram of Pre-Processing

of the image and b indicates the index of spectral band. The normalization factor of JBF is represented as follows:

$$i(j,k) = \sum\Big\{G_{\sigma_d}(j{-}p, k{-}q) \\ {*}G_{\sigma_r}(I_{Pc}(j,k){-}I_{Pc}(p.q))\Big\} \tag{2}$$

The parameter σ_r is known as the edge preserving factor and a larger value causes a Gaussian blurring, hence it is preferred to be small (around 0.1). The filter helps in smoothening the features making it suitable for feature extraction. By using this filter for pre-processing, the image quality has been enhanced and the edge information has been preserved. Therefore, the outcome from this phase supports in better feature extraction.

3.2 Feature extraction

The second phase of the proposed approach is feature extraction, where the important facial features are extracted. In the proposed work, the major geometric and appearance-based features are extracted using the HOG [40] and GF [25]. The facial components like eyes, nose, mouth, etc. are located and the features are extracted from these components.

(i). HOG for geometric feature extraction

HOG is a feature extraction method that helps in extracting the major facial components used in recognizing the expressions of the individual. The HOG descriptor is used in various fields and applications for different types of feature extractions and one of the major application is the facial emotion recognition technique. The main advantage of the HOG descriptor is that it is invariant to the photometric and geometric transformations and can extract the features accurately. The proposed technique uses the HOG descriptor to extract the geometric points of the face. The major facial components like eyes, nose and mouth are identified and the geometric points are drawn over those components to extract the geometric features. The HOG is computed based on the orientation and gradient information subjected to the first derivative of the image. For any image $I(p, \ q)$, the gradient about the image's arbitrary pixel point is a vector given as follows:

$$\nabla I(p,q) = \left[G_p, G_q\right]^T = \left[\frac{\partial I}{\partial p}, \frac{\partial I}{\partial q}\right] \tag{3}$$

where, G_p is the gradient in x-axis and G_q is the gradient in y-axis.

The magnitude and direction of the gradient are given by the following equations:

$$|\nabla I(p,q) = \sqrt{G_p^2 + G_q^2}| \tag{4}$$

$$\theta(p,q) = arc\tan\frac{G_q}{G_p} \tag{5}$$

The pixels present in the images are divided into cells and the adjacent cells are formed into blocks with overlapping nature. The gradient and orientation information of the images are extracted through the application of the HOG descriptor over the image blocks. The extracted orientation information of similar cells are gathered and stored in the histogram bins. Later, these information are sorted and arranged into a histogram. The overall features generated through the HOG descriptor can be computed using the following equation:

$$N_{HOG} = A_s{*}A_j{*}N_b \tag{6}$$

where, A_s indicates the size of a block, A_j is the count of blocks present in the considered image sample and N_b indicates the count of histogram bins used. Using HOG, the geometric features of the facial images of individuals are extracted.

(ii). GF for appearance-based feature extraction

GF is highly preferable for feature extraction because of the spatial frequency localization property based on the uncertainty principle. The advantages include its robustness against illumination changes and image noises along with invariance to scale, rotation and transition. The sinusoidally modulated kernel function in spatial domain for a 2-d GF can be represented as follows:

$$G(p,q) = \frac{f^2}{\pi\gamma\phi}\exp\left(-\frac{p^{'2} + \gamma^2 q^{'2}}{2\sigma^2}\right){*}\exp\left(j.2\pi f q^{'} + \theta\right) \tag{7}$$

$$p^{'} = p\cos\varphi + q\sin\varphi \tag{8}$$

$$q^{'} = -p\sin\varphi + q\cos\varphi \tag{9}$$

where, σ indicates the standard deviation, γ indicates the ellipticity of the Gabor function, ϕ indicates the spatial aspect ratio, θ indicates the phase offset, j indicates an imaginary number, π indicates the orientation of the Gabor function and f indicates the sinusoidal frequency.

The image samples of the face used in the proposed work are convolved using different orientation set and spatial resolution of the 2-d GF bank. For any image $I(p,\ q)$, with filtering kernel $\psi_{u,\ v}(p,\ q)$, the image characterization of the image can be represented as follows:

$$O_{u,v}(p,q) = I(p,q).\psi_{u,v}(p,q) \tag{10}$$

The feature vectors generated through the GF are reduced through down sampling using the sampling factor and are normalized to unit variance. The proposed work uses the GF for extracting the appearance based features from the facial image samples. Both the geometric and appearance-based features are crucial to discover the different types of emotions from the face images. The extracted features are provided to the next phase to improve the overall classification.

3.3 Feature selection

Feature selection is one of the major phase followed in the proposed work to optimally select the required features for accurate classification. This phase can maximize the classification accuracy by eliminating the unwanted features and selecting only the needed features. The proposed work uses an optimization algorithm for automatic selection of the most required features from the extracted feature set. The feature selection process is formulated as a multi-objective optimization problem and a novel Multi-Objective Seagull Optimization Algorithm (MOSOA) [6] is proposed for this application. This algorithm helps in automatically selecting the most relevant features from the given feature set appropriate for prediction.

The intelligent behaviour of seagulls in migration and attacking of the prey are imitated in the MOSOA for formulating the search. The migrating behaviour is known to the exploration and the attacking behaviour is termed as exploitation of the searching process. The behaviour for the seagulls can be described in three steps as follows:

(i) Seagulls migrate in groups and maintain different initial positions to avoid collisions.
(ii) The group of seagulls will follow the best fittest seagull found in the group while exploration.
(iii) Based on the position of the best seagull in the group, the remaining seagulls will update their position.

The above steps are the behaviours while exploring the search space and the exploitation is seen when the seagulls attack other birds migrating over the sea. These behaviours can be formulated into a multi-objective optimization problem for appropriate feature selection.

3.3.1 MOSOA for optimal feature selection

The MOSOA algorithm can be imitated for the feature selection process, where the seagulls are the search agents (features). The migration process is carried out for exploring the search space well and to obtain the optimal features out of the available set of features. The objective functions for the proposed feature selection process involves the minimization of the classification error and the minimization of the features taken as input. The overall fitness function can be formulated as follows:

$$Min\ F_t = \delta * \Psi + (1 - \delta) * \frac{|f|}{|F|} \tag{11}$$

where, δ indicates the parameter influencing the classification output, Ψ indicates the error rate of classification, f indicates the total count of features extracted in the proposed feature extraction process and F indicates the total count of features found in the dataset. The above fitness function is required to be minimized for appropriate selection of features.

Exploration The exploration process of the search agents involves the movement of the search agents from one place to the other based on the fitness function. The exploration process of MOSOA are subjected to three major conditions as follows:

(i). *Collision Avoidance:* While exploring the search space, there is a possibility of the occurrence of collision and hence a variable is employed for the computation of the position of the search agent. The formulation is as follows:

$$\overrightarrow{c}_s = A * \overrightarrow{p}_s(x) \tag{12}$$

where, $\overrightarrow{c}_s$ indicates the position of the search agent not involved in collision, $\overrightarrow{p}_s$ indicates the search agent's current position, x indicates the current iteration and the variable A indicates the movement of the search agent in the solution space. The formulation for the variable can be given as follows:

$$A = f_\iota - \left(x * \left(\frac{f_\iota}{Itr_{\max}} \right) \right); x = 0, 1, \ldots .Itr_{\max} \tag{13}$$

where, f is used to control the frequency of the variable A.

(ii). *Movement to best neighbour's position:* The search agents that avoided the collision then move towards their best neighbour and the formulation can be as follows:

$$\overrightarrow{m}_s = B * \left(\overrightarrow{p}_{bs}(x) - \overrightarrow{p}_s(x) \right) \tag{14}$$

where, $\overrightarrow{p}_s$ indicates the search agent, $\overrightarrow{p}_{bs}$ indicates the position of the best search agent and $\overrightarrow{m}_s$ indicates the movement of $\overrightarrow{p}_s$ *towards* $\overrightarrow{p}_{bs}$. The random value B is responsible for maintaining a balance between the exploration and exploitation strategies. The formulation of B is as follows:

$$B = 2 * A^2 * rnd \tag{15}$$

where, *rnd* is the random number and it ranges between 0 and 1.

(iii). *Position Update:* Finally, the search agent updates its position based on the position of the best search agent in the group. The position update formulation is as follows:

$$\overrightarrow{d}_s = |\overrightarrow{c}_s + \overrightarrow{m}_s| \tag{16}$$

where, $\overrightarrow{d}_s$ indicates the distance seen between the search agent and the best one in the group.

The MOSOA algorithm evaluates the fitness functions of the search agents and the best solutions are updated as archive. When the archive is found to overflow, the grid method is called to omit the crowded solutions from the available solutions in the archive. Then the new solution is updated to the archive and then the boundary of the search agents are evaluated and adjusted. Finally, the positions of the search agents in the archive are evaluated with the fitness function and the best search agent is updated with the new position.

Exploitation The exploitation process is imitated from the attacking behaviour of the search agents based on the history and experience in the exploitation. The search agents move spirally in the air in 3-dimensional axis and the movement can be described in three planes as follows:

$$x^{'} = \alpha * \cos(l) \tag{17}$$

$$y^{'} = \alpha * \sin(l) \tag{18}$$

$$z^{'} = \alpha * l \tag{19}$$

$$l = u * e^{lv} \tag{20}$$

In the above set of exploitation equations, α determines the radius of every turn in spiral movement, l is a random number chosen between the range 0 and 2π and u ***and*** v are the constants denoting the spiral movement. The final updated position of the search agent can be given as follows:

$$\overrightarrow{p}_s(x) = \left(\overrightarrow{d}_s * x^{'} * y^{'} * z^{'}\right) + \overrightarrow{p}_{bs}(x) \tag{21}$$

The MOSOA algorithm follows a strategy of comparing the best pareto optimal solution with the existing solutions. Thus, the algorithm chooses the leader for the group to achieve it. The least crowded space in the archive is filled using the roulette wheel selection method, where the best solutions in the optimal boundary is considered for the process. The formulation is as follows:

$$U_l = \frac{h}{N_l} \tag{22}$$

where, h is a constant value greater than 1 and N_l is the number of pareto optimal solutions for the segment.

Based on the exploration and exploitation process, the best search agents or the optimal features are stored in the archive and the remaining irrelevant features are omitted. The pseudocode for optimal feature selection using MOSOA is presented in algorithm 1.

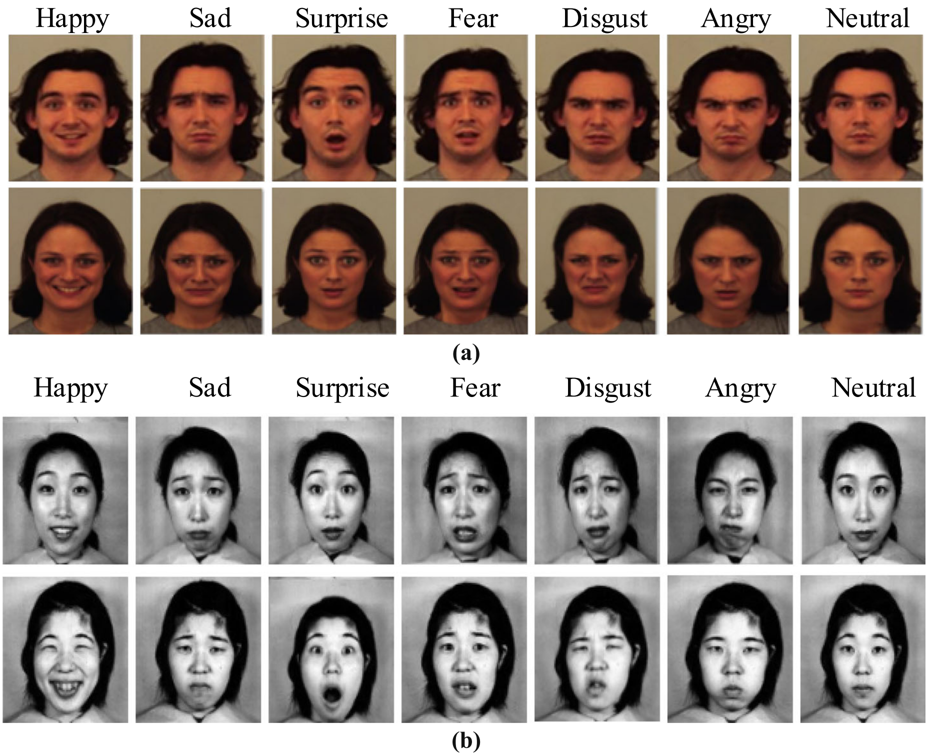
Algorithm 1: MOSOA for optimal feature selection

```
Input: Initial search agents (features) $\vec{p}_s$
Output: Archived optimal search agents (features) $\vec{p}_{bs}$
For every search agent → compute fitness using (11)
$Archive \leftarrow$ all non-dominated solutions
While $(x < Itr_{\max})$
    For every search agent
        Update position using (21)
    End for
    Evaluate fitness using (11)
    Find the optimal solutions from updated solutions
    $Append(solutions) \rightarrow Archive$
    $If\ overload(Archive)$
        Call grid function to omit the crowded solution
        Add $new\ solution \rightarrow Archive$
    End if
    Compute the search agents crossing the boundary limit
    Adjust the search agents within the boundary limit
    Evaluate the fitness using (11)
    $x \leftarrow x + 1$
End while
Return $Archive(optimal\ search\ agents)$
```

The selected features are the optimal features that can well describe the emotions of the individuals. Apart from selecting the required features, another major need of the feature selection phase is to reduce the dimensionality issues of the classifier. By reducing the dimensionality problem, the classifier attains the maximum accuracy rate by performing classification only on the needed features.

3.4 Classification

After the selection of optimal features, the classification phase is carried out with the most important and relevant features. The major aim of this work is to classify the images of the individuals into seven different classes based on their facial expressions. To this application, the proposed work uses the hybrid DBRO technique with higher classification accuracy. The DBRO is used in labelling the images according to the input features provided. The sample images with different facial expressions of individuals acquired from the KDEF and JAFFE datasets are depicted in Fig. 3.



(a)

(b)

Fig. 3 Sample images of facial expressions taken from (**a**) KDEF and (**b**) JAFFE datasets

The proposed DBRO classification technique uses the DBN for training and labelling of images. To label the images based on the expressions, the DBRO classifier is trained with certain sample images of the individuals with varied expressions. The DBN [12] network model is a composition of unsupervised networks of Restricted Boltzmann Machines (RBMs) with three main types of layers such as the input layer, hidden layers and the output layer. The layers in the RBM are connected and the hidden layers are responsible for detecting the selected features provided as input. The training and testing process carried out in the DBRO is depicted in Fig. 4.

The features that are selected in the last phase are vectored and provided to the classifier for training. The hidden neurons in the deep model extract the features for classification. Based on the discriminative input features, the classifier trains itself to provide appropriate class labels to the features. The proposed DBRO classification technique holds the RBMs for training, where the mutual independence between the units of layers are ensured for fast training. The model is probabilistic with random weight and bias parameters. The input layer of the RBM can be called as visible denoted as $\nu$and the hidden layers are denoted as h. The random weight parameter of the classification model is fine-tuned with the ROA [21] algorithm for every classes in the training period. Based on this fine-tuned model, the testing samples are labelled during the testing period. The input features are provided to the visible layer for feature detection and the output layer restores the features provided as input so that the restored

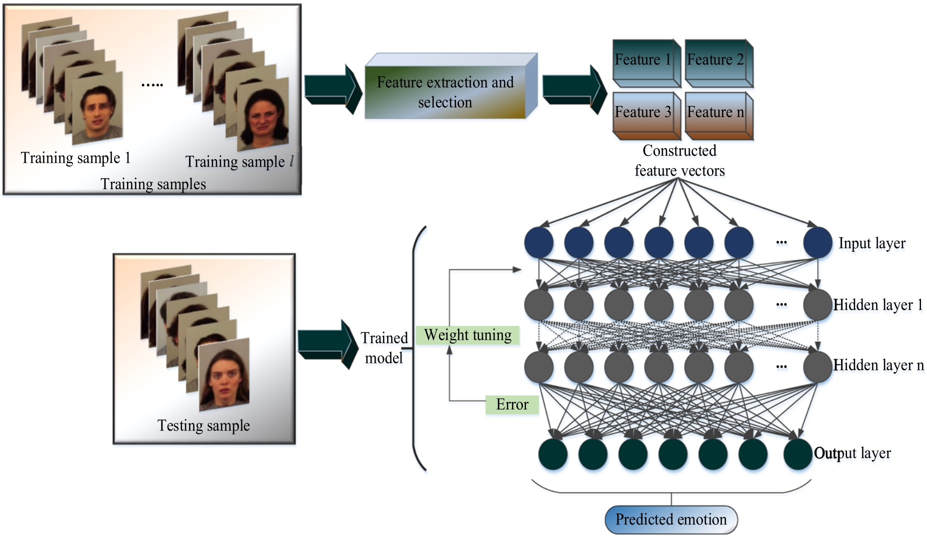


Fig. 4 Training and Testing Process of DBRO

features are similar to the submitted input. The joint probability of the visible and the hidden layers can be represented as follows:

$$P(\nu, h) = \frac{e^{-E(\nu,h)}}{\int\int\limits_{\nu,h} e^{-E(\nu,h)}} \tag{23}$$

where, $e^{-E(\nu,\ h)}$ indicates the energy function. The Gaussian energy function is seen to be better in place of binary energy function and it can be represented as follows:

$$E(\nu, h) = \sum_j \frac{(\upsilon_j - a_j)}{2\sigma_j^2} + \sum_k \frac{(h_k - b_k)^2}{2\sigma_k^2} - \sum_{j,k} \frac{\upsilon_j\, h_k}{\sigma_j\, \sigma_k}\omega_{jk} \tag{24}$$

where, υ_j ***and*** h_k indicates the activation states of visible and hidden layer units j ***and*** k, b_k indicates the bias values, σ_j ***and*** σ_k are the standard deviations for the input and hidden units with a value around 1 and ω_{jk} indicates the weight parameter connectingυ_j ***and*** h_k.

The conditional probabilities of RBM can be obtained through the Bayesian inference as follows:

$$P(h/\nu) = \frac{P(\nu, h)}{P(\nu)} = \frac{e^{-E(\nu,h)}}{\int\limits_h e^{-E(\nu,h)}} \tag{25}$$

$$P(\nu/h) = \frac{P(\nu, h)}{P(h)} = \frac{e^{-E(\nu,h)}}{\int\limits_\nu e^{-E(\nu,h)}} \tag{26}$$

The hidden units in the RBM can be updated using the given visible unit using the normal distribution as follows:

$$P(h_k|\nu) \approx N(\lambda_k, \sigma_k); \lambda_k = b_k + \sigma_k \sum \frac{\upsilon_j}{\sigma_j}\omega_{jk} \tag{27}$$

$$P\left(v_j|h\right)\approx N\left(\lambda_j, h_j\right); \lambda_j = a_j + \sigma_j \sum \frac{h_k}{\sigma_k}\omega_{jk} \tag{28}$$

The DBN model consists of RBMs stacked into it with an output layer in the end. The first hidden layer act as the visible layer for the next hidden layer and the training is proceeded one by one until the final RBM gets trained. In this way, the model gets trained with the input features and the images get labelled based on the provided input features appropriately. The overall model can be denoted with the parameters as follows:

$$\delta_{DBRO} = \{\delta_\omega, \delta_b\} \tag{29}$$

where, δ_ω denotes the overall weight values added with the input and δ_b indicates the overall bias values added. The random choice of weight parameter shows influence in the overall output of the training process and hence the tuning of the parameter is required to improve the accuracy of the model. The loss or the error function seen in the RBM is denoted as follows:

$$L_s(\omega, a, b) = -\sum_{j=1}^{l} \ln\left[P\left(\nu^{(j)}\right)\right] \tag{30}$$

where, l indicates the total count of image samples used for training.

Weight update using ROA To improve the performance of the DBN, the weight parameter is optimized using ROA. The optimization strategy helps in finding the exact weight parameter for the neural network. The objective of ROA in the proposed technique is to minimize the loss function of the DBN and the fitness function formulation can be represented as follows:

$$F = \min(L_s) \tag{31}$$

The fitness value is evaluated for all the search agents in the solution space. The weights used in the DBN are placed randomly in the solution space. The radius of the weight is considered as the major property in the proposed technique to optimize the weight parameter. Higher radius indicates a higher weight value and a lower radius indicates a lower weight value. The radius of the weight increases with respect to the time. The limit of the weight parameter is checked to choose the optimal weight among all the weight values. The radius of the weight is identified based on the following conditions:

(i) Based on the neighbouring weight values, the position is updated by the search agent as below:

$$\Re = \left(\xi_1^n + \xi_2^n\right)^{\frac{1}{n}} \tag{32}$$

where, ξ_1 *and* ξ_2 indicates the radii of two weight values and n indicates the variables belonging to the search agent.

(ii) Based on the training error and the fitness value, the weight value is minimized using the following equation:

$$\Re = \left(\beta\xi_1^n\right)^{\frac{1}{n}} \tag{33}$$

where, β indicates the rate of minimization of the parameter value and it controls the exploration and exploitation between the search agents. The algorithm works iteratively until the optimal weight value is chosen for the training model with minimized error. The proposed deep learning model correlates well with the problem in hand as the downstream task in the model is to label the input features accurately. The strong inter-connection between the layers in the deep learning model helped to accurately predict the labels for the given facial features.

In the testing process, the features for all the classes are reconstructed separately. The reconstruction losses are computed, and after optimization of the reconstruction errors, the features are labelled by the classifier into different emotions.

4 Simulation analysis

The simulation analysis and the performance comparison of the proposed approach against the existing models are presented in this section. The models chosen for comparison are the ResNet [14], RetinaNet [36], YOLO [7] and AlexNet [10]. The scenario chosen, performance metrics considered and the comparative analysis for the proposed model is presented in the upcoming sections.

4.1 Simulation scenario

The efficiency of the proposed approach is proved by simulating it in the Matlab platform. The facial emotions of the individuals are identified using the proposed model. The facial features for different emotions varies in almost all aspects. This helps the classifier to understand the differences between the emotions and therefore the features are labelled accordingly. For analysis, the KDEF dataset (https://www.kdef.se/download-2/register.html) and the Japanese Female Facial Expression (JAFFE) dataset (https://www.kaggle.com/shawon10/jaffe-facial-expression-detection) have been chosen with facial images of different emotions. The KDEF dataset comprises 4900 images of individuals with different facial expressions of both males and females. For the proposed work, 970 images are chosen for training and 194 images are taken for testing purposes.

The JAFFE dataset is a facial emotion recognition dataset that comprises labels for seven different facial emotions. This dataset consists of 213 image samples that are collected from 10 Japanese female participants. The pixel resolution of each image in this dataset is 256 × 256 and it consists of training, testing and validation samples. The seven different labels in this dataset are happy, sad, fear, surprise, disgust, neutral and angry. From this dataset, 80% of the images are taken for training and the remaining 20% is considered for testing. The hyper-parameter settings for the proposed framework is provided in Table 1.

4.2 Performance metrics

The major performance metrics optimized in the proposed approach are precision, recall, f-measure, accuracy, specificity and Mean Square Error (MSE). The classification accuracy is determined to understand the best model that can efficiently identify the differences between

Table 1 Hyper-parameter settings of the proposed work

Sl. No	Hyper-parameters	DBRO
1.	Tuning algorithm	ROA
2.	Initial learning rate	0.1%
3.	Max epochs	100
4.	Mini batch size	32
5.	No of hidden units	10
6.	No of neurons in input layer	250
7.	Initial population	100
8.	No of raindrops	100
9.	Dimension	5
10.	β	10

the facial emotions of the individuals. The computations of the considered performance indices are as follows:

(i)Precision: This measure is used to identify the true values of the classification technique and has an impact with the overall accuracy of the model. Better precision values indicate the performance efficiency of the model in identifying the emotions more accurately. The computation for precision measure is as follows:

$$P = \frac{tp}{tp + fp} \tag{34}$$

where, tp indicates true positive values i.e. the capability of the model to identify the exact truth value or the positive value an fp indicates false positive value i.e. the capability of the model to accurately identify the false or the negative values.

(ii) Recall: Recall or the sensitivity measure indicates the truth values of the classification with an impact in the accuracy of the system. The computation for recall measure is as follows:

$$R = \frac{tp}{tp + fn} \tag{35}$$

where, fn indicates the false negative values i.e. the value that is actually false but classified as true.

(iii)F1-Score: F1-score is the harmonic mean of precision and recall depicting the measure of accuracy of the test. The value for this measure is expected to be 1 to gain better classification performance. The computation is as follows:

$$F1 = \frac{2^*P^*R}{P + R} \tag{36}$$

(iv)Accuracy: Accuracy measure helps in identifying the performance of classification in classifying the facial emotions. Higher accuracy value indicates that the system is more effective. The computation is based on the truth values of the system and is computed as follows:

$$A = \frac{tp + tn}{tp + tn + fp + fn} \tag{37}$$

where, *tn* is the true negative value i.e. the ability of the model in accurately predicting the negative class.

(v) Specificity: This measure determines the true negatives from the classifier output. The mathematic formulation is provided by determining the true negatives of the problem given as follows:

$$S = \frac{tn}{tn + fp} \tag{38}$$

(vi) MSE: This measure determines the overall error in classifying the images. The formulation for MSE is given as follows:

$$MSE = \sum_{i=1}^{n} \frac{\left(\widehat{c_i} - c_i\right)^2}{n} \tag{39}$$

where, n is the total number of images subjected to classification, $\widehat{c_i}$ denotes the number of images classified and c_i denotes the number of observed images.

(vii) Kappa coefficient: The mathematical formulation for kappa coefficient can be given as follows:

$$K = \frac{2*(tp*tn)-(fp*fn)}{(tp + fp)(tn + fp) + (tp + fn)(tn + fn)} \tag{40}$$

(viii)Mathews correlation coefficient (MC): This metric is computed based on the following formulation:

$$M_C = \frac{(tp*tn)-(fp*fn)}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}} \tag{41}$$

4.3 Performance evaluation

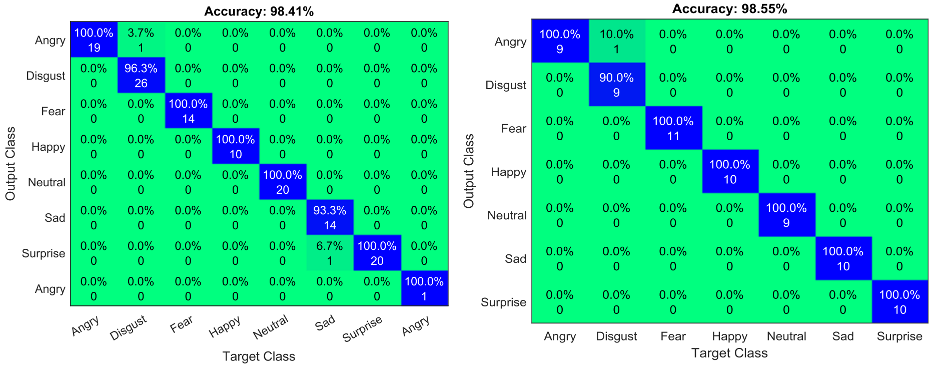
The performance is evaluated with the images collected from the dataset and with the extracted facial features. The facial images identified from the dataset comprises seven different emotions such as happy, sad, fear, disgust, angry, surprise and neutral. Among these emotions, the features such as the geometric facial points are extracted by the GF to understand the differences between the features. The categories of emotions of individuals vary in most of the aspects and hence the layers of DBN extract the differences identified in the features to improve the learning process. The layers learn with the help of the input features or the geometric points and then the same facial features are reconstructed for every iteration. The reconstructed features are evaluated to identify the accuracy of the classifier in learning the input features.

The proposed DBRO classifier classified the input images into seven different emotions based on the submitted input facial features. To put it clear, the framework worked on analysing the different input facial features and used it for differentiating the facial emotions of the individuals. From the observations, the facial emotions of the images are differentiated based on texture and intensity changes. The geometrical feature points described the shape of the face and the appearance features considered the skin texture, and extracted the intensity values from the facial points. The dimensions of the feature space are reduced through optimization. Finally, the features are labelled with seven different types of emotions based on the classifier training.

The confusion matrices for the proposed model for the KDEF and JAFFE datasets are provided in Fig. 5a and 5b. The proposed model achieved higher values of accuracy in classifying the emotions from the images. In the KDEF dataset, from the considered 19 angry images, all the 19 images are accurately classified under the class angry. Out of the 27 disgust images, 26 images are correctly classified as disgust and 1 image is misclassified as angry. From the 15 sad images, 14 are correctly classified as sad and 1 image is misclassified as surprise. The other images belonging to the classes such as fear, happy, neutral, surprise and angry images are correctly classified with 100% accuracy. The overall classification accuracy attained by the proposed model for the KDEF dataset is 98.41%.

Similarly, for the JAFFE dataset, except the disgust class, all the other 6 classes achieved 100% accuracy. Among the 10 disgust images, 9 are correctly classified as disgust and 1 image is misclassified as angry. The proposed model achieved higher accuracy for the images from the JAFFE dataset than the other considered dataset. In the JAFFE dataset, 9 images are classified as angry, 11 images are classified as fear, 10 images are classified as happy, 9 images are classified as neutral, 10 images are classified as sad and 10 images are classified as surprise emotion. The overall classification accuracy attained by the proposed model in the JAFFE dataset is 98.55%.

The ROC curve for the proposed and existing classification schemes for the KDEF and JAFFE dataset is provided in Fig. 6a and 6b. The ROC are plotted with false positive rate (FPR) in the x-axis vs. true positive rate (TPR) in the y-axis. For the KDEF dataset in Fig. 6a, the proposed model provided higher values of area under the curve (AUC) than the compared models. Higher values of AUC indicates that the model is capable of providing higher TPR. The AUC value of the proposed DBRO classification model for the KDEF dataset is 0.987. Among the compared models, AlexNet achieved higher values than the other models. For the JAFFE dataset in Fig. 6b, the proposed model provided better values of AUC compared to other deep models. The overall AUC value provided by the DBRO model for the JAFFE dataset is 0.989. Among the other compared models, AlexNet provided almost similar values of AUC to the proposed model. For both the datasets, the ResNet model provided minimum values of AUC for emotion recognition.



Accuracy: 98.41%

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise	Angry
Angry	100.0% 19	3.7% 1	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0
Disgust	0.0% 0	96.3% 26	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0
Fear	0.0% 0	0.0% 0	100.0% 14	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0
Happy	0.0% 0	0.0% 0	0.0% 0	100.0% 10	0.0% 0	0.0% 0	0.0% 0	0.0% 0
Neutral	0.0% 0	0.0% 0	0.0% 0	0.0% 0	100.0% 20	0.0% 0	0.0% 0	0.0% 0
Sad	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	93.3% 14	0.0% 0	0.0% 0
Surprise	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	6.7% 1	100.0% 20	0.0% 0
Angry	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	100.0% 1

(a)

Accuracy: 98.55%

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	100.0% 9	10.0% 1	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0
Disgust	0.0% 0	90.0% 9	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0
Fear	0.0% 0	0.0% 0	100.0% 11	0.0% 0	0.0% 0	0.0% 0	0.0% 0
Happy	0.0% 0	0.0% 0	0.0% 0	100.0% 10	0.0% 0	0.0% 0	0.0% 0
Neutral	0.0% 0	0.0% 0	0.0% 0	0.0% 0	100.0% 9	0.0% 0	0.0% 0
Sad	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	100.0% 10	0.0% 0
Surprise	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	0.0% 0	100.0% 10

(b)

Fig. 5 Confusion matrix of the proposed approach (**a**) for KDEF dataset (**b**) for JAFFE dataset

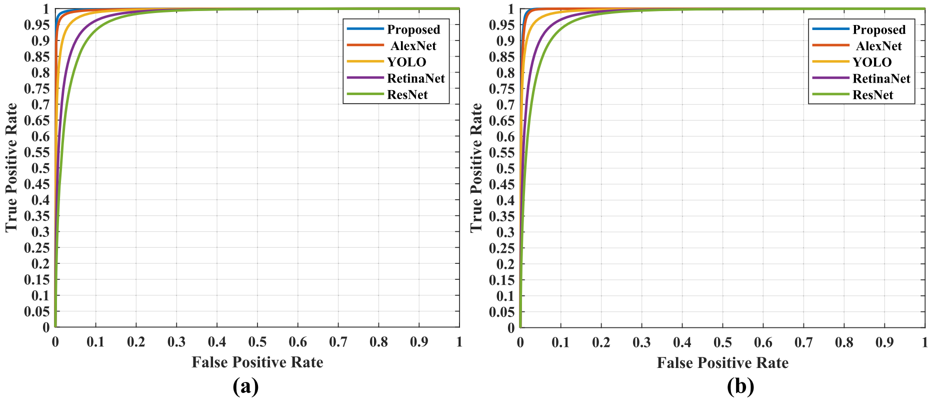


Fig. 6 ROC curves of the proposed and existing classification methods (**a**) for KDEF dataset (**b**) for JAFFE dataset

The simulation outcomes of the proposed and existing classifiers for emotion recognition using the KDEF dataset is provided in Table 2. The performance values of precision, recall, f-measure, accuracy, specificity, kappa, FPR, MC and error are considered for analysis. From the values, it is clear that the proposed model provided better values than the other classification models in emotion recognition. The overall accuracy of the proposed model in classifying the facial emotion of the KDEF dataset is 98.41% and the error value of the model is 1.59. Among the compared classifiers, AlexNet achieved better accuracy rate of 97.92%. The least value of accuracy is provided by the ResNet classifier with an accuracy rate of 91.87%. Also, the error value of ResNet in emotion recognition is 8.13 which is higher than the other models. In the proposed model, the influence of error has been minimized by the hybridization of ROA that optimally selected the weight value for the classifier for every iterations. For the other compared models, the weight value is taken randomly by the model itself that resulted in certain misclassifications. Apart from accuracy, the precision, recall, f-measure, specificity, kappa, FPR and MC values of the proposed model for the KDEF dataset are 98.20%, 98.78%, 98.46%, 98.96%, 92.74%, 0.24% and 98.49%. The proposed model is thus capable of accurately identifying the facial emotions without misclassifications.

The performance values of the proposed and existing classifiers in emotion recognition for the images from the JAFFE dataset is provided in Table 3. From the values in the table, it is clear that the proposed model provided better results for emotion recognition compared to the other models. The overall accuracy achieved by the proposed model in emotion classification for the JAFFE dataset is 98.55% with an error value of 1.45. Among the compared classifiers, AlexNet model provided better accuracy rate of 98.11% and the least accuracy rate is provided

Table 2 Performance values of the classification models for the KDEF dataset

Methods	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)	Specificity (%)	Kappa (%)	FPR (%)	MC (%)	MSE
ResNet	91.64	89.41	90.51	91.87	93.06	83.50	6.14	83.53	8.13
RetinaNet	91.64	92.54	92.09	93.38	93.22	86.52	5.98	86.52	6.62
YOLO	93.82	96	94.90	95.84	94.92	91.51	4.28	91.52	4.16
AlexNet	96.88	97.81	97.34	97.92	97.21	95.76	1.99	95.76	2.08
Proposed	98.20	98.78	98.46	98.41	98.96	92.74	0.24	98.49	1.59

Table 3 Performance values of the classification models for the JAFFE dataset

Methods	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)	Specificity (%)	Kappa (%)	FPR (%)	MC (%)	MSE
ResNet	91.64	89.79	90.70	92.06	93.08	83.88	6.12	83.90	7.94
RetinaNet	91.64	92.95	92.29	93.57	93.24	86.90	5.96	86.90	6.43
YOLO	93.82	96.43	95.11	96.03	94.94	91.89	4.26	91.91	3.97
AlexNet	96.88	98.24	97.55	98.11	97.21	96.15	1.99	96.15	1.89
Proposed	98.07	98.57	98.25	98.55	98.96	94.08	0.24	98.31	1.45

by the ResNet with an accuracy rate of 92.06%. Also, the error value of the ResNet model is 7.94 which is higher than the other compared models. Since the ROA algorithm diminished the error rate in classification, the proposed DBRO model achieved higher percentage of accuracy without any misclassifications. The random weight initialization of the compared models reduced the accuracy of the classifiers and also resulted in a particular range of misclassifications. The values of precision, recall, f-measure, specificity, kappa, FPR and MC of the proposed model for the JAFFE dataset are 98.07%, 98.57%, 98.25%, 98.96%, 94.08%, 0.24% and 98.31% respectively.

The overall accuracy comparison of the proposed and existing models in emotion recognition for the KDEF and JAFFE dataset is graphically depicted in Fig. 7. From the graph, it is clear that the proposed model provided better outcomes than the other models. Also, the features selected for classification are optimal and provided more context for the model to understand the difference between the emotions. Among the evaluations with two datasets, higher values are provided by the model for JAFFE dataset. Since the hybridization of ROA improves the training of the model, the model has learned the features to classify without error

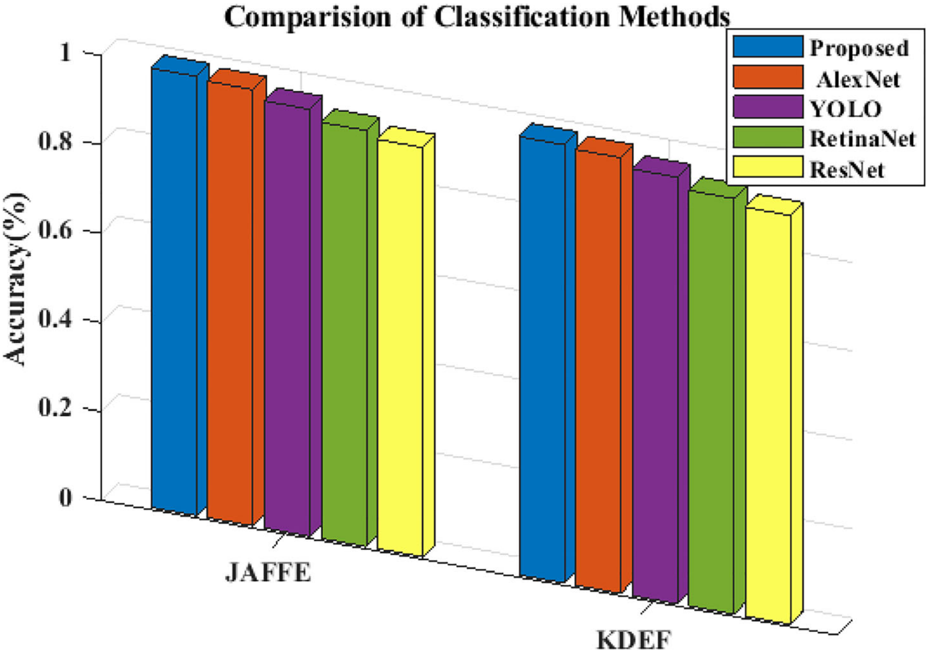


Fig. 7 Accuracy comparison of the proposed and existing classifiers for facial emotion recognition

rates. Among the compared models, AlexNet achieved higher values of accuracy than the other models. The overall accuracy of the proposed model for the KDEF dataset is 98.41% and for the JAFFE dataset is 98.55%. Least performance is provided by the ResNet model with high range of misclassifications.

The convergence of the proposed MOSOA algorithm for feature selection is graphically depicted in Fig. 8. The graph is plotted with number of iterations in the x-axis vs. fitness function in the y-axis. From the graph, it is seen that the proposed model converged faster when the iteration number is increased. The feature selection phase resulted in optimal features for training the classifier without dimensionality issues. Through the use of this algorithm, the model obtained optimal features that defined the individual emotions accurately. The fitness function of the model helped to identify the optimal features based on the objective. MOSOA evaluated the fitness for each iteration over the training and testing samples. When the iterations are between 0 and 40, there is slight increase in the fitness and when the iterations are increased, the MOSOA converged faster and provided optimal results.

4.3.1 Ablation study

Various ablation experiments were performed to evaluate the performance of each phase of the proposed approach. The experiments have been conducted by partitioning the models into 4 modules such as Module 1, Module 2, Module 3 and Module 4. The overall accuracy offered by these modules are analyzed separately to identify the importance of all the four phases of

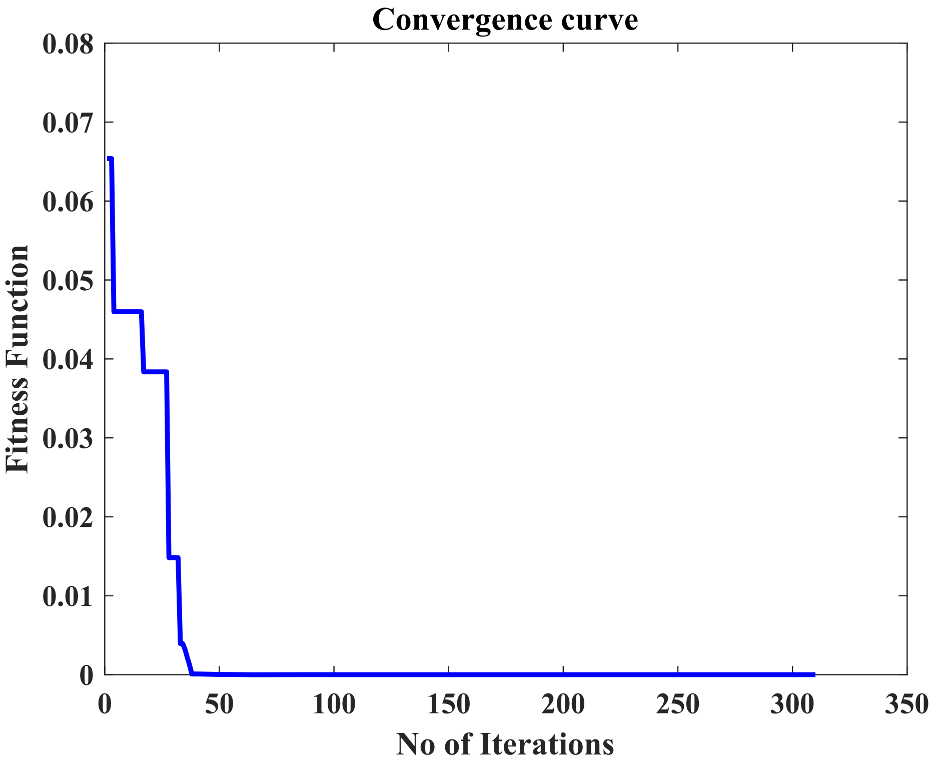


Fig. 8 Convergence curve for the proposed MOSOA algorithm for feature selection

the proposed work. Module 1 indicates the proposed work with all the four phases, Module 2 indicates all the four phases of the proposed model but the classification is carried out without parameter tuning, Module 3 includes the pre-processing, feature extraction and classification phases and finally Module 4 includes the feature extraction and classification phases.

The overall performance achieved by each model is graphically shown in Fig. 9. From the figure, it is clear that each phase of the proposed approach contributes equally in emotion classification. The least performance is provided by Module 4 where only the feature extraction and classification phases are considered. In this model, the features of the images are extracted without pre-processing. Therefore, the distortions in the images reduced the accuracy in classification. Similarly in Module 3, the feature selection phase is omitted that increased the dimensionality of the features. This model took more time for training and also provided minimum accuracy rate. Compared to Modules 3 and 4, Module 2 provided optimal results but this model had an influence of error rate that diminished the accuracy rate. Best performance is provided by Module A that combined all the four phases of the proposed approach.

This analysis show the importance of the each phase of the proposed model in emotion recognition. The noisy images cannot be directly subjected to feature extraction and classification. The features extracted from such noisy images consists of additional unwanted information that reduces the efficiency and effectiveness of classification. Also, the enormous amount of features extracted from each image increases the training time leading to time complexity issues. Thus Module 4 implies the importance of pre-processing and feature selection phases. The MOSOA algorithm played an effective role in selecting the optimal features from the feature space thereby reducing the dimensionality issues. The variations in the performance of Module 2 and Module 3 implies the importance of feature selection phase

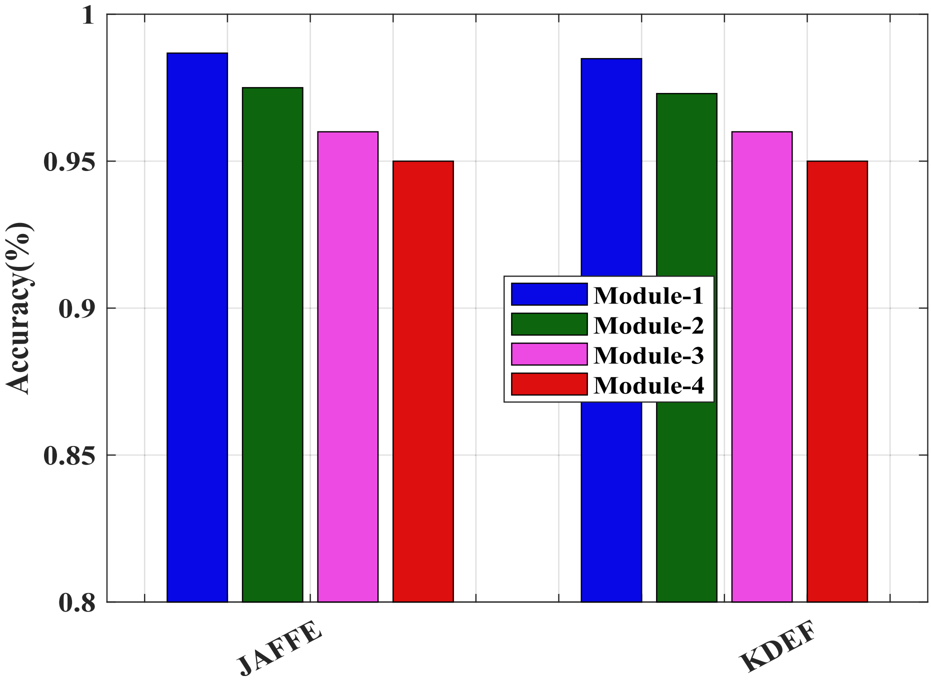


Fig. 9 Ablation study for different modules of the proposed approach

in the proposed model. Also, the Module 1 and Module 2 show minor variations in performance. This is due to the influence of error rates in Module 2 that diminished the performance and leaded to some misclassifications. The ROA algorithm helped to achieve optimal results by selecting suitable weight values for the classifier during the training time. For all the iterations, the algorithm is the optimal value for learning. Thus the overall accuracy emotion recognition in both the datasets are enhanced. From the ablation experiments, it can be concluded that all the four phases of the proposed approach have equal contribution in emotion recognition.

4.3.2 Analysis of inference time

The inference time is analysed to identify the performance and efficiency of the proposed model in emotion classification. The major role in reducing the inference time of the model is played by the MOSOA algorithm as it provided only the optimal features to the classifier for training. Since the training has been enhanced, the proposed model took minimum time to yield an inference result. The analysis of the inference time for the proposed and existing models are detailed below:

The analysis of inference time in classifying one image by the proposed and existing classifiers for the KDEF and JAFFE datasets are provided in Table 4. To classify a single image of the KDEF dataset, the proposed model approximately takes 0.08 s and for the image of JAFFE dataset, the inference time of the model is 0.05 s. Compared to the other models, the proposed model provided optimal results with reduced time complexities. Among the compared models, the ResNet took more time to yield an inference result for a single image, with an average inference time of 0.25 s for the KDEF dataset and 0.22 s for the JAFFE dataset. The AlexNet model provided optimal results with an average inference of 0.11 s for a single image from KDEF dataset and 0.09 s for a single image from JAFFE dataset. The proposed model took only minimum time to process an image as it involved feature selection phase and parameter tuning methods. Therefore, it can be suggested that the proposed model is suitable for the emotion classification in order to attain higher classification accuracy.

4.4 Discussion

The proposed DBRO based framework for facial emotion recognition provided best results than the other compared models for all performance metrics considered. The different phases of the proposed model have been evaluated and the results are demonstrated. From the evaluations, it is also clear that each phase is important to attain performance improvement. The pre-processing phase used the JBF to pre-process the image that enriched the image

Table 4 Inference time analysis of the proposed and existing classifiers for 1 image for the KDEF and JAFFE datasets

Methods	KDEF (s)	JAFFE (s)
ResNet	0.25	0.22
RetinaNet	0.19	0.17
YOLO	0.14	0.13
AlexNet	0.11	0.09
Proposed	0.08	0.05

quality and also preserved the edge information. Moreover, the features are smoothened that supported the feature extraction phase to achieve higher efficiency. The HOG and GF features extracted in the feature extraction phase are capable of providing more information about the emotions of individuals. Also, the feature selection phase reduced the dimensionality issues occurring in the classification level to attain better accuracy rate. The combination of all these phases helped the proposed model to achieve the desired performance improvement compared to the baseline methods. The hybridization helped in increasing the values of the performance measures to a drastic level. It optimized the weight values chosen for the iterations in the training phase with optimal values selected by the ROA algorithm. For every iteration, the ROA algorithm is made to run within the layers of DBN to provide tuned-weights to the model. This helped the model to achieve higher values of accuracies than the other models. The proposed MOSOA based feature selection helped to reduce the number of features in the training phase that improved the overall training accuracy. Moreover, it helped to reduce the inference time of the model during testing. Upon evaluations, it is also seen that the model have never shown any over-fitting or under-fitting issues for both the datasets in the testing phase. This indicates that the model is suitable to classify the facial emotions even for new unseen samples.

The conventional model of DBN include flaws related to the robustness in classification. Moreover, the conventional DBN is evidenced with misclassifications of images with poor accuracy. This is proved through the ablation experiments carried out in section 4.3.3. Therefore, the proposed model improved the conventional DBN with optimized weight values. The model effectively optimized the classification process much better than the compared models. The facial points are well-recognized by the proposed DBRO technique as the model consisted of stacked RBMs to learn the input features. Compared with the existing models, the training of the proposed model is observed to be more efficient because of the stacked layers. The model converged within a limited number of iterations and hence the computational complexity of the model is found to be low. The hybrid model effectively achieved an overall accuracy value of 97% incomparable to the other existing models available in the field of facial emotion recognition. The analysis also concluded that the chosen ROA algorithm for fine-tuning is a much efficient bio-inspired algorithm to fine-tune the weight value of the DBN for facial emotion recognition process.

In comparison with some of the existing state-of-the-art schemes such as [19, 27], the proposed model is able to identify the emotions more accurately. The model named FERC discussed in [19] utilized the CNN model to classify the facial emotions. But the CNN model required larger datasets for training and included time complexities. Compared to this model, our proposed model provided optimal results due to the feature selection and weight optimization methods. The Bi-LSTM classification is carried out in [27] and it involved an attention layer to better classify the emotions. An extended approach of emotion classification was introduced in [13] that introduced several layers in a deep model for classification. But these models involved computational complexity as the deep model has to be trained for longer time to yield better results. Also, the structure of the models are complex and are difficult to be interpreted. An optimized training model was implemented in [33] to improve the classification accuracy. But the model provided inappropriate results when the training samples are increased. The overall comparison suggests that our model is effective and efficient in emotion recognition and requires only minimum time for training. Moreover, the proposed model is well-generalized in learning the features of newer samples. Thus, it can be concluded that the proposed model is suitable for classifying the facial emotions accurately.

5 Conclusion

In this article, a novel approach for automatic classification of the facial emotions based on the dataset images has been proposed. The proposed article presented a novel hybrid DBRO technique for classification with higher accuracy. The proposed model is formulated by initially pre-processing the images with the aid of improving the quality of the images through de-noising. The major features such as the geometric and appearance based features are extracted from the images using HOG and GF. After extraction, the dimensionality of the features are reduced through feature selection where, the optimal features are selected using the MOSOA algorithm. Finally, the classification is performed for the selected features where, the randomly initialized weight parameter is optimized based on the optimization algorithm. Experimental results demonstrated that the proposed technique performed well in accurately classifying the images into seven different emotions based on the input features. The overall performance of the proposed approach provided 97% of accuracy value incomparable to the other classification models. This model can also be extended for emotion classification in ASD affected individuals in future with the intake of several data and information collected in real-time. Apart from this, we aim to develop more optimized models that can improve the global stability of training even when trained for longer iterations with different data samples.

Funding No funding is provided for the preparation of manuscript.

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest Authors Fakir Mashuque Alamgir, Md. Shafiul Alam declares that they have no conflict of interest.

References

1. Alreshidi A, Ullah M (2020) Facial emotion recognition using hybrid features. In: Informatics Multidisciplinary Digital Publishing Institute 7(1):1–13
2. Anguraju K, Kumar NS, Kumar SJ, Anandhan K, Preethi P (2020) Adaptive feature selection based learning model for emotion recognition. J Critic Rev
3. Arora M, Kumar M (2021) AutoFER: PCA and PSO based automatic facial emotion recognition. Multimed Tools Appl 80(2):3039–3049
4. Chen Y, He F, Li H, Zhang D, Wu Y (2020) A full migration BBO algorithm with enhanced population quality bounds for multimodal biomedical image registration. Appl Soft Comput 93:106335
5. Choudhary D, Shukla J (2020) Feature extraction and feature selection for emotion recognition using facial expression. In: 2020 IEEE sixth international conference on multimedia big data (BigMM), pp 125–133
6. Dhiman G, Singh KK, Soni M, Nagar A, Dehghani M, Slowik A, Kaur A, Sharma A, Houssein EH, Cengiz K (2021 Apr 1) MOSOA: a new multi-objective seagull optimization algorithm. Expert Syst Appl 167: 114150
7. Garg D, Goel P, Pandya S, Ganatra A, Kotecha K (2018) A deep learning approach for face detection using YOLO. In: 2018 IEEE Punecon, pp 1–4

8. Graumann L, Duesenberg M, Metz S, Schulze L, Wolf OT, Roepke S, Otte C, Wingenfeld K (2021 Jan 1) Facial emotion recognition in borderline patients is unaffected by acute psychosocial stress. J Psychiatr Res 132:131–135
9. Hassan AK, Mohammed SN (2020 Oct 1) A novel facial emotion recognition scheme based on graph mining. Defence Technology 16(5):1062–1072
10. Hosny KM, Kassem MA, Fouad MM (2020) Classification of skin lesions into seven classes using transfer learning with AlexNet. J Digit Imaging 33(5):1325–1334
11. Hou N, He F, Zhou Y, Chen Y (2020) An efficient GPU-based parallel tabu search algorithm for hardware/software co-design. Front Comput Sci 14(5):1–18
12. Jahanjoo A, Naderan M, Rashti MJ (2020) Detection and multi-class classification of falling in elderly people by deep belief network algorithms. J Ambient Intell Humaniz Comput 11(10):4145–4165
13. Jain DK, Shamsolmoali P, Sehdev P (2019) Extended deep neural network for facial emotion recognition. Pattern Recogn Lett 120:69–74
14. Li B, Lima D (2021) Facial expression recognition via ResNet-50. Int J Cogn Comput Eng 2:57–64
15. Li H, He F, Chen Y, Luo J (2020) Multi-objective self-organizing optimization for constrained sparse array synthesis. Swarm Evol Comput 58:100743
16. Liang Y, He F, Zeng X (2020) 3D mesh simplification with feature preservation based on whale optimization algorithm and differential evolution. Integrated computer-aided engineering preprint, pp 1–19
17. Luo J, He F, Yong J (2020) An efficient and robust bat algorithm with fusion of opposition-based learning and whale optimization algorithm. Intell Data Anal 24(3):581–606
18. Ma T, Benon K, Arnold B, Yu K, Yang Y, Hua Q, Wen Z, Paul AK (2020) Bottleneck feature extraction-based deep neural network model for facial emotion recognition. In: International Conference on Mobile Networks and Management, Springer, Cham, pp 30–46
19. Mehendale N (2020 Mar) Facial emotion recognition using convolutional neural networks (FERC). SN Appl Sci 2(3):1–8
20. Mistry K, Rizvi B, Rook C, Iqbal S, Zhang L, Joy CP (2020) A multi-population FA for automatic facial emotion recognition. In: 2020 international joint conference on neural networks (IJCNN), IEEE, pp 1–8
21. Moazzeni AR, Khamehchi E (2020 Dec 1) Rain optimization algorithm (ROA): a new metaheuristic method for drilling optimization solutions. J Pet Sci Eng 195:107512
22. Nawaz R, Cheah KH, Nisar H, Yap VV (2020 Jul 1) Comparison of different feature extraction methods for EEG-based emotion recognition. Biocybern Biomed Eng 40(3):910–926
23. Nguyen TD (n.d.) Multimodal emotion recognition using deep learning techniques (Doctoral dissertation, Queensland University of Technology), pp 1–138
24. Patwari M, Gutjahr R, Raupach R, Maier A (2020) Low dose CT Denoising via joint bilateral filtering and intelligent parameter optimization, pp 1–4. arXiv preprint arXiv:2007.04768
25. Rahul M, Shukla R, Goyal PK, Siddiqui ZA, Yadav V (2021) Gabor filter and ICA-based facial expression recognition using two-layered hidden Markov model. In: Advances in computational intelligence and communication technology. Springer, Singapore, pp 511–518
26. Saha S, Ghosh M, Ghosh S, Sen S, Singh PK, Geem ZW, Sarkar R (2020 Jan) Feature selection for facial emotion recognition using cosine similarity-based harmony search algorithm. Appl Sci 10(8):2816
27. Sepas-Moghaddam A, Etemad A, Pereira F, Correia PL (2020) Facial emotion recognition using light field images with deep attention-based bidirectional LSTM. In: ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 3367–3371
28. Siddiqui MF, Javaid AY (2020 Sep) A multimodal facial emotion recognition framework through the fusion of speech with visible and infrared images. Multimodal Technol Interact 4(3):46
29. Simcock G, McLoughlin LT, De Regt T, Broadhouse KM, Beaudequin D, Lagopoulos J, Hermens DF (2020 Jan) Associations between facial emotion recognition and mental health in early adolescence. Int J Environ Res Public Health 17(1):330
30. Slimani K, Kas M, El Merabet Y, Ruichek Y, Messoussi R (2020 Aug 1) Local feature extraction based facial emotion recognition: a survey. Int J Electr Comput Eng 10(4):4080
31. Staff AI, Luman M, Van der Oord S, Bergwerff CE, van den Hoofdakker BJ, Oosterlaan J (2021 Jan 7) Facial emotion recognition impairment predicts social and emotional problems in children with (subthreshold) ADHD. Eur Child Adolesc Psychiatry 31:1–3
32. Ulusoy SI, Gülseren ŞA, Özkan N, Bilen C (2020 Jul 1) Facial emotion recognition deficits in patients with bipolar disorder and their healthy parents. Gen Hosp Psychiatry 65:9–14
33. Wang S-H, Phillips P, Dong Z-C, Zhang Y-D (2018) Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm. Neurocomputing 272:668–676
34. Wang K, Su G, Liu L, Wang S (2020 Jul 20) Wavelet packet analysis for speaker-independent emotion recognition. Neurocomputing 398:257–264

35. Wieckowski AT, White SW (2020 Jan) Attention modification to attenuate facial emotion recognition deficits in children with autism: a pilot study. J Autism Dev Disord 50(1):30–41
36. Yang M, Xiao X, Liu Z, Sun L, Guo W, Cui L, Sun D, Zhang P, Yang G (2020) Deep RetinaNet for dynamic left ventricle detection in multiview echocardiography classification. Sci Program 2020:1–6
37. Yarasca FA, Henríquez SD (2020) Intelligent system based on wavelets for automatic facial emotion recognition. In: 2020 IEEE engineering international research conference (EIRCON), pp 1–4
38. Yildirim S, Kaya Y, Kılıç F (2021 Feb) A modified feature selection method based on metaheuristic algorithms for speech emotion recognition. Appl Acoust 173:107721
39. Yin Z, Liu L, Chen J, Zhao B, Wang Y (2020) Locally robust EEG feature selection for individual-independent emotion recognition. Exp Sys Appl 162:113768
40. Zhou W, Gao S, Zhang L, Lou X (2020 Mar 13) Histogram of oriented gradients feature extraction from raw Bayer pattern images. IEEE Trans Circuits Syst II Express Briefs 67(5):946–950

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.