



An efficient facial emotion recognition using convolutional neural network with local sorting binary pattern and whale optimization algorithm

Fereshteh Aghabeigi¹ · Sara Nazari¹ · Nafiseh Osati Eraghi¹

Received: 20 October 2023 / Accepted: 8 May 2024

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract

Facial emotion recognition is one of the fields of machine learning and pattern recognition. Facial expression recognition is used in a variety of applications. For robust automatic facial emotion recognition, feature extraction from input image data is challenging. To address this issue, we propose an emotion recognition system based on a new feature extraction method, whale optimization algorithm, and convolutional neural network. In the feature extraction phase, a new efficient human face descriptor is expressed using a local sorting binary pattern and a convolutional neural network. Also, the hyperparameters of the convolutional neural network are optimized using the whale optimization algorithm. Then, the convolutional neural network is applied for classification. The performance of the proposed method is evaluated using three well-known face databases CK+ (extended Cohn-Kanade) with facial expressions (happiness, sadness, fear, anger, disgust, contempt, and surprise), JAFFE (Japanese female facial expression) with (happiness, sadness, anger, fear, neutral, disgust, and surprise), and MMI (MMI facial expression database) with facial emotions (happiness, sadness, anger, fear, disgust, and surprise). The accuracies with CK+, JAFFE, and MMI are 100%, 99.93%, and 99.83%, respectively. Experimental results demonstrate that the proposed model can provide better performance compared to alternative methods.

Keywords Feature extraction · Convolutional neural network · Local sorting binary pattern · Facial emotion recognition · Deep learning · Whale optimization algorithm

1 Introduction

Emotion recognition is a way to determine or verify an individual's emotion using his/her face, body language, and speech. Due to the wide range of faces in various applications, the face is used in the emotion recognition system more than other biometrics. In the image processing field, the processing process includes three phases: preprocessing, feature extraction, and classification. Preprocessing is the preliminary processing of raw data to prepare for the emotion recognition system. After preprocessing phase, extracting distinctive features is a vital step in the emotion recognition system. Face feature extraction is divided into appearance and geometric methods [1, 19, 26, 31, 37]. In the geometric

feature extraction approach, the position and shape of facial components, such as eyes, mouth, nose, and eyebrows, are used to identify distinctive features. While in the appearance feature extraction approaches, the histogram of oriented gradients (HOGs) [14] discrete wavelet transform (DWT) [32], local binary pattern (LBP) [55], local directional pattern (LDP) [27], and local ternary pattern (LTP) [61]; Gabor [22] are employed. Furthermore, convolutional neural network (CNN) [3, 5, 13, 15, 21, 25, 33] is used as a type of deep learning algorithm to extract features from images and learn to recognize patterns.

To design a high-performance model, tuning hyperparameters can be effective. Both automatic and manual methods can be used to set hyperparameters in architecture. In manual approaches, adjusting the hyperparameters requires a lot of resources and time. At the same time, meta-heuristic algorithms can tune hyperparameters automatically. Meta-heuristic algorithms inspired by natural phenomena are used to search/optimize problems. There are many meta-heuristic

✉ Sara Nazari
sara.nazari@iau.ac.ir

¹ Department of Computer Engineering and Information Technology, Arak Branch, Islamic Azad University, Arak, Iran

algorithms for adjusting hyperparameters. The whale optimization algorithm (WOA) [41] can be used as one of the swarm intelligence algorithms to achieve optimal solutions.

In the following, several related works in the field of facial emotion recognition are expressed. Boughida et al. [11] to detect facial emotions have used from three steps: First, the face is detected. Then, it uses a Gabor filter to extract features from regions of interest. Finally, the genetic algorithm for feature selection and support vector machine (SVM) hyperparameters tuning is presented. In [26], a descriptor based on the gradient of the neighbors of the target pixel for pattern generation called neighborhood-aware edge direction pattern (NEDP) is proposed by Iqbal et al. Farkhod and Chae [19] introduced the local prominent directional pattern (LPDP) approach for feature extraction based on local edge-based descriptors for recognizing emotions from faces. Jeen Retna Kumar et al. [28] proposed a system for facial emotion recognition, which includes: face recognition, feature extraction by the new subband selective multilevel stationary biorthogonal wavelet transform (SM-SBWT) method, and SVM for classification. Kola and Samayamantula [35] use four neighbors and diagonal neighbors in LBP for feature extraction. In addition, adaptive windows and averaging in radial directions are introduced to complete the proposed method. Bendjillali et al. [8] described a method to recognize facial expressions. The steps of this work include the Viola–Jones face detection algorithm, the contrast limited adaptive histogram equalization (CLAHE) algorithm for improving facial image, the discrete wavelet transform (DWT) for feature extraction, and deep CNN for classification. Nigam et al. [44] proposed a structure to recognize facial emotions. The first is face detection by the Viola–Jones method. Next, the discrete wavelet transformation transforms the spatial domain into the frequency domain. Then, the features are extracted by the histogram of oriented gradients. Finally, SVM is used for classification. Boughanem et al. [10] suggested a multi-channel CNN based on three models, VGG19, GoogleNet, and ResNet101, to form a more robust feature vector. SVM classifier is used for emotion recognition. Mukhopadhyay et al. [43] presented a new facial expression recognition method by exploiting texture image features such as LBP, LTP, and completed local binary pattern (CLBP). The CNN model is used to recognize facial expressions. Using a histogram of oriented gradient (HOG) for feature extraction and graph signal processing (GSP) method to reduce the feature vector is a system suggested by Meena et al. [40]. Alphonse and Dharma [2] reported maximum response-based directional texture pattern (MRDTP) and maximum response-based directional number pattern (MRDNP) approaches for feature extraction. Besides, an effective generalized supervised dimension reduction system (GSDRS) has been introduced. It used an extreme learning

machine with radial basis function (ELM-RBF) for classification. Barra et al. [6] proposed an approach for facial emotion recognition based on geometry, a technique that analyzes salient points through a virtual spider web on the face. Emotion is classified using the K-nearest neighbor. Kola and Samayamantula [36] elaborated an approach for feature extraction for facial emotion recognition using a combination of the local gradient coding based on horizontal and diagonal (LGC-HD) with wavelet transform and singular value decomposition to calculate the facial image's singular values. Arora et al. [4] proposed a method for facial emotion recognition by gradient filter and PCA for feature extraction. Also, a random forest algorithm is used for classification. A novel graph-based texture transform for feature extraction to automatic facial expression detection is presented by Tuncer et al. [60]. Linear discriminant analysis (LDA) and SVM classifier are used for classification. Kar et al. [3] were used ripple transform type II (ripple-II) for the feature extraction. The principal component analysis (PCA) and linear discriminant analysis (LDA) approaches are expressed for the compress and discrimination features. Besides, the least squares variant of the support vector machine (LS-SVM) is considered for the classification. Baygin et al. [7] suggested an approach recognition of emotional expressions about individuals' social behavior. This model has four prime phases: facial areas are segmented, features are extracted with AlexNet and MobileNetV2, the most valuable 1000 features are selected by neighborhood component analysis (NCA), and these 1000 features are selected on an SVM. Bentoumi et al. [9] proposed method is based on CNN model (VGG16, ResNet50) for feature extraction with a multilayer perceptron (MLP) classifier.

The drawbacks of the mentioned methods above compared to the proposed method are described below. In [2, 31, 40, 60], the computational complexity is more than the proposed model. The results show a high accuracy in recognizing facial expressions, despite the similar dataset, for the proposed system compared to the proposed methods in [11, 19, 26, 60]. Also, the method of this study has been able to extract distinct features by combining the well-known feature extraction methods of HOG [14], DWT [32], LBP [55], LDP [27], and Gabor [22] with the automatic feature extraction method of CNN, so that using the advantages of both methods, it has a better performance in recognizing facial emotions for real samples. In addition, compared to the reviewed methods, the proposed work has used data augmentation techniques in order to extract more features and increase the generalization ability of the model and high accuracy in emotion recognition due to the variety of data. Besides, a meta-heuristic algorithm has been employed to fine-tune CNN hyperparameters to improve network efficiency. Therefore, this paper introduced a new technique called local sorting binary pattern (LSBP). Our model uses a combination of LSBP to

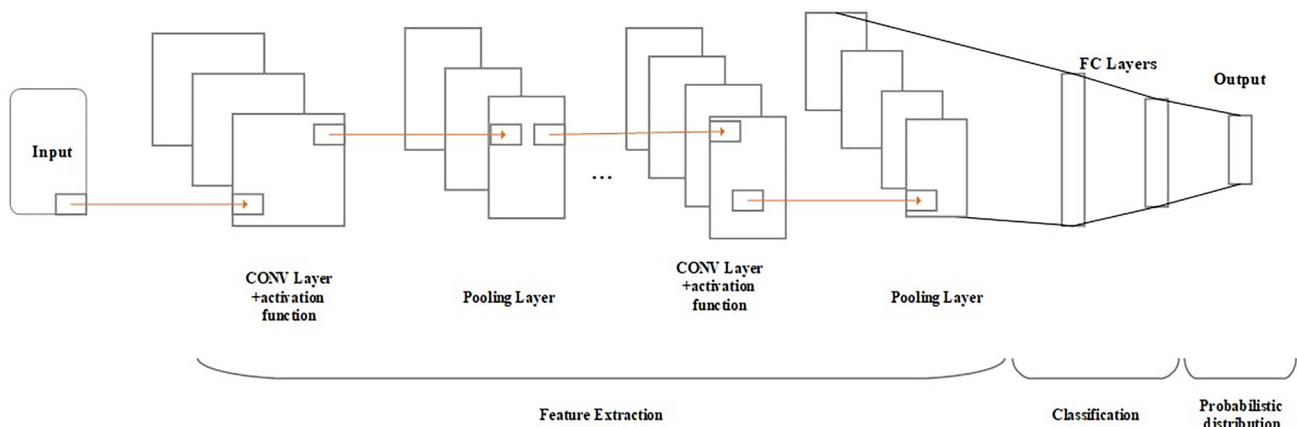


Fig. 1 CNN structure

consider spatial features and CNN optimized with WOA algorithm to consider features of hyperspectral image data for facial emotion recognition. In addition, CNN is used to classify facial expressions (happiness, sadness, anger, contempt/natural, disgust, fear, and surprise). Experiment of the proposed method is done with CK+ [38], JAFFE [39], and MMI [47] datasets. The contributions of our work are as follows:

- Preparing data and data augmentation in the preprocessing step for emotion recognition.
- Employing WOA to tune hyperparameters of CNN architecture.
- Obtaining high accuracy in emotion recognition by using a combination of LSBP and CNN for feature extraction.
- CNN classifier is employed in the learning and testing phases.
- Using different facial image datasets to evaluate the proposed model compared to related facial emotion recognition models.

The rest of the paper is organized as follows: Work-related concepts are explained in the Sect. 2. In the Sect. 3, the proposed method is presented. Experimental results are presented in the Sect. 4. Finally, Sect. 5 concludes the paper.

2 Background

This section discusses the CNN algorithm, Whale optimization algorithm, and LBP approach as three basic concepts for the proposed method.

2.1 Convolutional neural network

Artificial intelligence is simulated human intelligence. Machine learning is a subset of artificial intelligence.

Machine learning allows computers to learn. Learning is applied with the help of algorithms based on the features in the collected data.

Deep learning (DL) is a machine learning method. As a subset of machine learning, neural networks are the basis of deep learning algorithms. The neural network consists of input, hidden, and output layers that are connected similarly to neurons in the human brain. DL emphasizes representational learning [16]. Representation learning uses raw data to extract features automatically. As a representation learning technique, DL transforms the representation to more abstract levels by hierarchically using nonlinear modules in each layer [42]. Machine learning methods can be divided into two categories: supervised and unsupervised learning. In supervised learning, labeled datasets are used in algorithms to predict data classes. CNN is a well-known supervised DL algorithm for image data processing and classification. As an artificial neural network with multiple hidden layers, CNN can be considered for feature extraction. There are different architectures for CNN, such as LeNet, AlexNet, VGG, GoogleNet, ResNet, DenseNet, and SENet [31, 48]. The CNN network consists of the following components, as shown in Fig. 1:

- Convolutional layer: a layer containing one or more filters. The size of the filters is smaller than the size of the input image. Convolve a filter with an image produces a feature map. The outputs of the convolution process are passed through a nonlinear activation function. Selecting the appropriate activation function is effective in the performance of the network. Some activation functions include swish, ReLU [62], tanh, softmax, and sigmoid [3, 45, 51, 58]. Some activation functions such as ReLU are more popular for this layer [Eq. (1)].

$$\text{ReLU}(x) = \max(0, x) \quad (1)$$

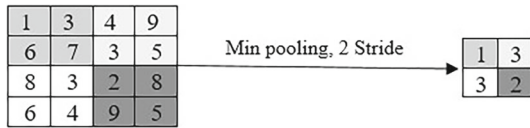


Fig. 2 Min pooling algorithm with stride 2

Where x is the input value.

- Pooling layer: The pooling layer follows the convolutional layer. It is used to reduce the dimensionality of the feature map with keeping the vital data and the computational cost of the network. The size of the feature map from the previous layer is reduced according to the determined pooling size and stride value. The most commonly used down-sampling algorithms include min-pooling, max-pooling, average-pooling, global-pooling, and global average-pooling. A sample of this process is shown in Fig. 2.
- Fully connected layer: It is mostly used at the end of the network. The input of this layer is the output of the last convolution or pooling layer. The obtained features are flattened (Converting 2D feature maps to 1D feature maps) with the complete connection. In other words, the fully connected (FC) layer is suitable for making decisions about data classification. Also, in the last layer of the CNN model, it is used an activation function such as softmax or sigmoid which outputs a probability distribution of multi-class classification.

2.2 Whale optimization algorithm

WOA [41] is a nature-inspired meta-heuristic optimization algorithm that mimics the social behavior of humpback whales. Humpback whales prefer to prey on groups of krill or small fish near the surface of the water by creating distinctive bubbles along a circle or ‘9’-shaped. This interesting behavior for hunting is called the bubble-net feeding method. The WOA algorithm consists of three steps: encircling prey, spiral bubble-net feeding maneuver (exploitation phase), and search for prey (exploration phase), which is explained as follows:

- Encircling prey: Whales can detect and encircle the prey’s location. The WOA algorithm assumes that the current best candidate solution is the target prey or is close to the optimal state. After defining the best search agent, other search agents update their position relative to the best agent. Its mathematical model is as follows:

$$\vec{D} = \left| \vec{C} \cdot \vec{X}^*(t) - \vec{X}(t) \right| \quad (2)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \quad (3)$$

$$\vec{A} = 2 \cdot \vec{a} \cdot \text{rnd}_1 - \vec{a} \quad (4)$$

$$\vec{C} = 2 \cdot \text{rnd}_2 \quad (5)$$

where t is the current iteration, \vec{X}^* is the location vector of the best solution obtained, and \vec{X} is the location vector, rnd_1 and rnd_2 are random vectors in $[0,1]$, \vec{a} decreases linearly from 2 to 0 during iterations, and \vec{A} and \vec{C} are the coefficient vectors.

- Exploitation phase: For mathematical model of the bubble-net behavior of whales, two techniques are expressed:

- Shrinking encircling approach: This behavior is achieved by decreasing the value of a in Eq. (4).
- Adjusting random values for A in $[-1,1]$, the new position of a search agent can be defined anywhere in between the original position of the agent and the position of the current best agent.
- Spiral update position: This approach first calculates the distance between the whale located at (X, Y) and the prey located at (X^*, Y^*) . The simulation of the spiral movement of whales according to the spiral equation between the position of the whale and the prey is as follows:

$$\vec{X}(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \quad (6)$$

Where D' represents the distance of the i th whale to the prey (the best solution obtained so far), b is a constant to determine the shape of the logarithmic spiral, and l is a random number in $[-1,1]$.

The whales swim simultaneously in a small circle along a spiral path around the prey. Therefore, assuming a probability of $\frac{1}{2}$ for choosing each method, the mathematical model is as follows:

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D} & P < 0.5 \\ D' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & P \geq 0.5 \end{cases} \quad (7)$$

where p is a random number in $[0,1]$.

Exploration phase: whales search randomly according to each other’s position. Therefore, random values of \vec{A} greater

than 1 or less than -1 are used to force the search agent to move away from the reference whale. The random selection of the agent is used to update the position of a search agent. This mechanism and $|\vec{A}| > 1$ emphasize exploration and allow the WOA algorithm to perform a global search. The equations are as follows:

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \tag{8}$$

$$\vec{X}(t + 1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \tag{9}$$

In this equation, \vec{X}_{rand} is a randomly selected position vector (random whale) from the current population. A random search agent is selected in the case $|\vec{A}| > 1$, while the best solution is chosen when the search agents position update is $|\vec{A}| < 1$.

The algorithm of the WOA is shown in following:

Algorithm 1 WOA

```

Initialize: the population of whales

Calculate the fitness of initial population and
select the best search agent ( $X^*$ )

 $t = 0$ 
While  $t < Max\text{-iteration}$  do
  foreach all search agent do
    Update  $a, A, C, l$  and  $p$  for each agent
    if  $P < 0.5$  then
      if  $|\vec{A}| < 1$  then
        Update the position of the current
        agent based on Eq. (2)
      else
        Select a random search agent
        ( $X_{rand}$ )
        Update the position of the current
        agent based on Eq. (9)
      else if  $p \geq 0.5$  then
        Update the position of the current
        agent based on Eq. (6)

    Calculate the fitness of all agents and Update  $X^*$ 

   $t = t + 1$ 

Return  $X^*$ 

```

2.3 Local binary pattern

Local binary pattern (LBP) [17, 35, 37, 54, 56] is a local texture pattern presented by Ojala et al. [46]. Researchers have studied LBP in various areas of image processing, including face recognition and facial emotion recognition. LBP is used to extract features of grayscale images with robustness against monotonic illumination changes. This technique has efficiency and low complexity in calculations. In the LBP, a

block with a specific size, such as 3×3 , is used to scroll the input image (Fig. 3a). In each step, the difference between the value of the neighboring pixels and the central pixel is calculated. Values greater than zero are encoded with 1, and other values are encoded with 0 for adjacent pixels (Fig. 3b). By moving counterclockwise, binary values are written from left to right. The decimal value is calculated. Formally, the LBP technique can be represented according to Eq. (10):

$$LBP(g_i, g_c) = \sum_{i=0}^{p-1} 2^i f(g_i, g_c), i = 0, 1, \dots, p \tag{10}$$

$$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{11}$$

where g_i is the pixel value of the i th neighbor of the central pixel. g_c is the central pixel value. p is the total number of neighbors. The resulting value is placed in the center of the block, which is shown in Fig. 3c. This process continues until a 3×3 block scans the image.

Recently, there have been numerous proposed modified methods based on local binary patterns for extracting distinctive features used for classifying input data in various fields, including texture and facial recognition. In [24] introduced a method for texture-based image classification invaginating to changes in scale and rotation. Meanwhile, Ryu et al. [58] proposed a new approach called sorted consecutive local binary pattern (scLBP) for texture classification. The scLBP method can encode patterns with varying spatial transitions while still being rotation-invariant, achieved by sorting consecutive patterns. Additionally, the researchers utilized dictionary learning based on kd-tree to partition the data in space. Song et al. [57] introduced a histogram sorting method to preserve the distribution information of LBP codes and their complements, called first- and second-order sorted LBP (SLBP) which are robust to inverse grayscale changes and image rotation. Kalyoncu [29] employed Sorted Uniform LBP (SULBP), a rotation-invariant LBP variant for identifying leaf images.

The proposed method is similar to the proposed methods in terms of the input data format, which is an image. However, the context of the proposed method is the classification of emotions from faces. The proposed filter uses different steps and parameters than existing methods. In Sect. 3, the proposed method is described in detail.

3 Proposed methods

The proposed approach for classifying facial emotions is discussed in this section (Fig. 4). The combination of the new proposed LSBP and the CNN network is considered to extract appropriate features to recognize emotions with high

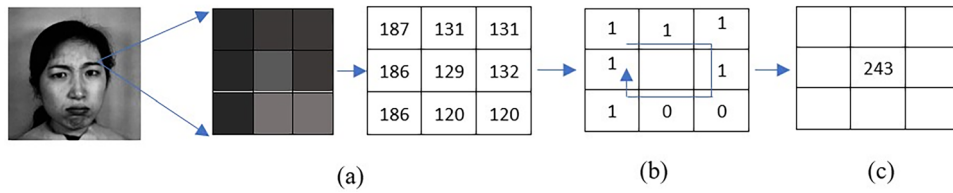


Fig. 3 A sample of LBP descriptor. **a** 3 × 3 block, **b** binary coding resulting from the difference between the neighboring pixels of the central pixel, and **c** converting the binary value to decimal and storing it in the central pixel

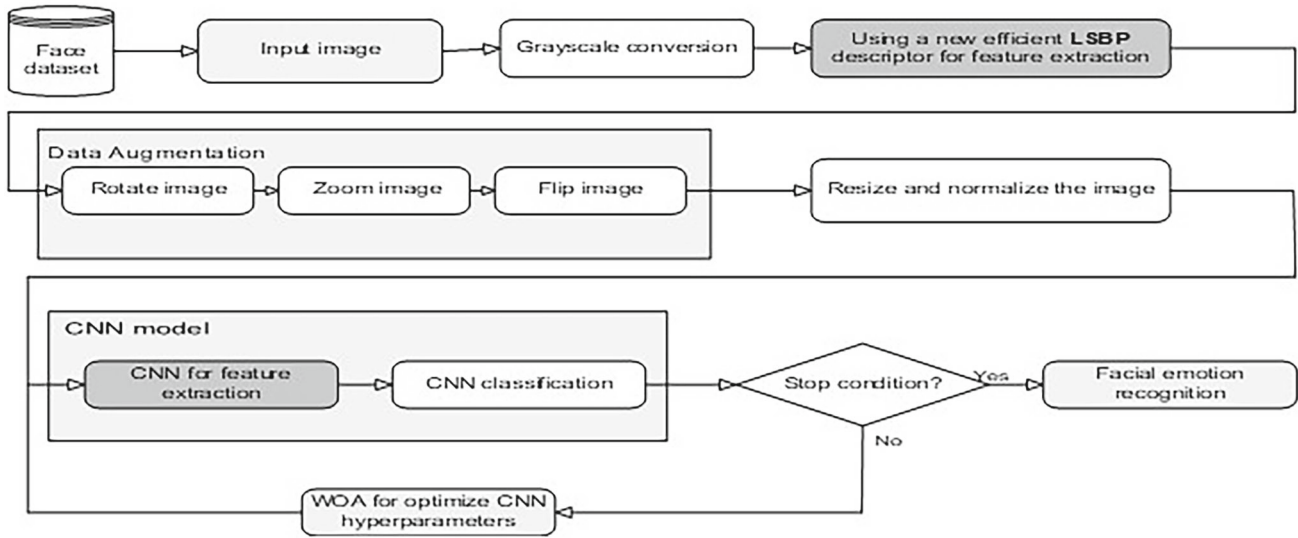


Fig. 4 Diagram of the proposed method

accuracy. In the preprocessing step, several operations are employed to the input data. LSBP is a hand-crafted approach for feature extraction from input images. The proposed pattern applies changes to the image pixels to extract the feature of the input image. Then, the output image is augmented (to create variety and increase the number of input data). The obtained data are normalized and resized. In the next step, the achieved data are considered as the input of the CNN network whose hyperparameters are optimized with the WOA algorithm to extract higher-level features and classification in order to recognize emotions from facial image data with high accuracy. Hence, data preprocessing is presented in Sect. 3.1. Section 3.2 introduces the new LSBP descriptor. The WOA algorithm for optimizing CNN hyperparameters and CNN structure is discussed in Sect. 3.3.

3.1 Preprocessing

Preprocessing is the process of preparing raw data for a deep learning model. The input images in the proposed scheme are considered in grayscale. Data augmentation is done by zooming, rotating, and flipping. Also, normalization and resizing are applied to the data.

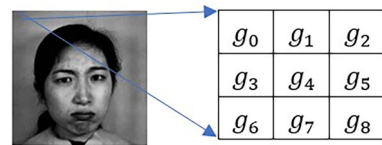


Fig. 5 3 × 3 segmentation of input grayscale image

3.2 The suggested local sorting binary pattern

In this paper, a simple method with low computational complexity and high emotion recognition accuracy is described. The LSBP technique, based on the LBP method, is proposed to extract the local features of input grayscale images, as shown in Fig. 5. In this pattern, image data pixels are divided into 3 × 3 blocks [Eq. (12)]:

$$S = \{g_0, g_1, g_2, g_3, g_4, g_5, g_6, g_7, g_8\} \tag{12}$$

where S is the set of 3 × 3 pixels selected. Within this set, the pixel values of a 3 × 3 block are sorted in ascending order. Indexes of sorted values are saved. The central element $c = g_4$ is used as the threshold value for the eight neighbors. Index values greater than the threshold value are replaced

with one and otherwise with 0, which is shown in Eq. (13):

$$f(\text{Index value}) = \begin{cases} 1 & \text{Index value} \geq c \\ 0 & \text{Index value} < c \end{cases} \quad (13)$$

The values 0 and 1 are written from left to right clockwise, starting from the pixel g_0 . The binary value is converted to decimal according to Eq. (14):

$$\text{result} = \sum_{i=0}^8 2^i f(g_i), \quad i = 0, 1, \dots, 8 \quad (14)$$

The resulting value is stored in the center pixel. This process is applied until the scan the whole image. The steps of the LSBP descriptor are described in Algorithm 2.

Algorithm 2 LSBP

```

Initialize:
    Size of segmentation: 3×3
While process all images of a dataset do
    Get input image
    foreach all 3 × 3 segments an image do
        Ascending sorting of the pixels
        value of the selected segment and
        save indexes;
        Save the index value of the center
        of the segment (c);
    foreach i as neighbors of the central pixel
    do
        if Index value < c then
            gi = 0
        else if Index value ≥ c then
            gi = 1
        Convert binary value to decimal
        according to Eq. (14) and store result
        in center;
Return result
    
```

3.3 The proposed CNN architecture and tuning the hyperparameters using WOA

In this paper, WOA is used to optimize CNN. The time complexity of GA, PSO, and WOA optimization algorithms is presented in Table 1. According to the displayed results, WOA in the proposed network can achieve high accuracy faster than GA and PSO. Therefore, the hyperparameters of CNN, including the activation function, optimizer function, learning rate, number of epochs, and batch size, are adjusted by the WOA algorithm. Every hyperparameter vector (\vec{HP}) is known as an agent in WOA, where each parameter (p) has a predetermined range/set (i). The initial population (S_n) of the WOA is a random set of n agents:

$$\vec{HP}_i = \{P_{i0}, P_{i1}, P_{i2}, P_{i3}, P_{i4}, P_{i5}\} \quad (15)$$

Table 1 Time complexity of optimization algorithms

Algorithm	Time complexity	Description
GA	$O(gNk)$	Number of generation (g), population size (N), and size of the individuals
PSO	$O(MND + ND)$	Population size (N), parameter dimension (D), and number of iterations (M)
WOA	$O(ND + F)$	Population size (N), dimensions of the problem (D), and time taken to calculate the fitness value (F)

$$S_n = \{\vec{HP}_1, \vec{HP}_2, \vec{HP}_3, \dots, \vec{HP}_n\} \quad (16)$$

In this proposed method, WOA parameters are unchanged. And, the fitness function is the accuracy value obtained by each hyperparameter vector. Eventually, the best agent with the highest accuracy is considered the optimal solution to set the hyperparameters of CNN.

After tuning the hyperparameters, the optimized CNN network is considered for feature extraction from the obtained data using the LSBP method and classification of facial emotions. Therefore, the recognition of facial emotions by CNN is described in three parts: designing CNN architecture, training, and evaluation of results:

- Model design (CNN architecture): The architecture of CNN used is a combination of convolution and pooling layers based on inception modules [59] and residual blocks [23]. The inception module provides parallelization of several convolution and pooling operators. A residual block is a stack of network layers. Moreover, ResNet can merge the connection between the output value of a block and its input value. The layering of the neural network is shown in Fig. 6.
- Training model: The image data are passed to the designed model. First, the loss function is considered to calculate the error. There are different loss functions, such as mean square error (MSE), mean absolute error (MAE), and cross-entropy [49]. In this work, cross-entropy has been used. It is one of the most famous methods for classification problems. The value of the cross-entropy (CE) function increases according to the probability difference between the predicted value and the actual value, which is expressed in Eq. (17):

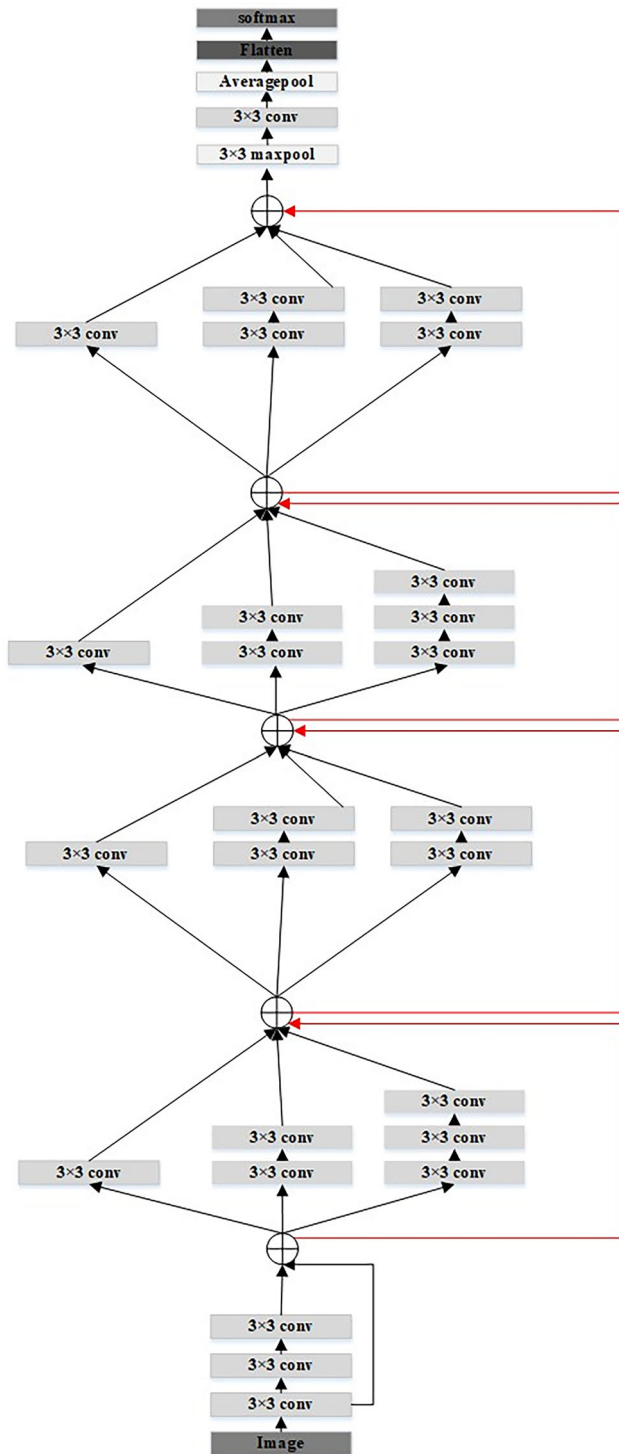


Fig. 6 The proposed CNN architecture

$$CE = -(y_i \log \hat{y}_i + (1 - \hat{y}_i) \log(1 - \hat{y}_i)) \quad (17)$$

where y_i expected output value and \hat{y}_i expected output value.

Then, the optimizer function is used to modify the weights and parameters of the neural network. The selection of the optimizer function has an influential role in the network results. Different functions, such as Adamax, SGD, Nadam, and Adagrad [52], are used. Finally, the weights and parameters of the neural network are updated, and the data are repassed to the CNN model. This cycle continues until the minimum expected error is reached.

- Result analysis: To evaluate the trained model, the specificity, sensitivity, F1-score, and accuracy of data classification are checked. The k -fold cross-validation method [34] and data split approach can express the accuracy of emotion recognition. The parameter in the k -fold cross-validation method is the number of data separation groups to evaluate. This method is suitable for preventing overflow. The data split validation technique describes the data partitioning into two or more parts. For example, the data are divided into training and testing parts in two-part splitting. Also, the confusion matrix is used to display the results. The confusion matrix is a two-dimensional table that expresses the performance of the proposed method for classification prediction.

4 Experimental results

This section, the proposed system's performance, is compared to that of other state-of-the-art methods. In evaluating facial emotion recognition, three popular datasets CK+, JAFFE, and MMI are used. The data augmentation techniques including rotation 5° , zoom [1,1.5], and horizontal flip are applied to improve the classification. Therefore, the number of data in each dataset has been tripled. Besides, input images are grayscale, normalized, and resized. The values employed to adjust the network hyperparameters with the WOA are given in Table 2. In addition, Table 3 shows the best value of hyperparameters in 100 iterations of the WOA algorithm.

The fivefold validation method and data split approach (85–15%) have been used to test the proposed method. For example, in the data split approach (85–15%), 85% of the data in experiments is set for the training phase, and 15% is employed for the testing phase. Furthermore, standard metrics of specificity, sensitivity, F1-score, and accuracy have been stated to evaluate facial emotion recognition. These metrics can be calculated from the confusion matrix.

Table 2 Values of hyperparameters

Hyperparameter	Values
Activation function	Swish, LeakeyReLU, ReLU, SELU
Optimizer function	Adamax, SGD, Nadam, Adagard
Learning rate	0.1, 0.01, 0.001, 0.0001
Number of epochs	20, 30, 40, 50
Batch size	8, 10, 12, 14

Table 3 Best values of hyperparameters

Hyperparameter	Best value
Activation function	Swish
Optimizer function	Adamax
Learning rate	0.01
Number of epochs	30
Batch size	12

The confusion matrix includes the labels of actual and predicted classes, which are employed to analyze the performance of the proposed algorithm. Equations (18–21) are used to measure the efficiency of the proposed model where TP, FP, TN, and FN are true positives, false positives, true negatives, and false negatives and refer to the results that correctly predict the positive class, incorrectly predicts the positive class, correctly predicts the negative class, and incorrectly predicts the negative class, respectively:

- Specificity (true negative rate) is the proportion of correctly classified positive cases out of all cases classified in a particular class, and the specificity formula is as follows:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (18)$$

- Sensitivity (true positive rate) is the proportion of actual positive cases which are correctly classified. The sensitivity formula is:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (19)$$

- F1-score is the harmonic mean of Specificity and Sensitivity values. The goal of this metric is when we have data with unbalanced distribution. The formula for F1-score is given by:

$$F1 - \text{score} = 2 \times \frac{\text{Specificity} \times \text{Sensitivity}}{\text{Specificity} + \text{Sensitivity}} \quad (20)$$

- Accuracy is an important evaluation metric. It is the proportion of true results among the total number of items checked for a particular class. The formula for accuracy is as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

The accuracy and loss of the proposed method are tracked to assess the classification network's performance. For this purpose, first, the datasets are introduced. Then, the results of our experiments are discussed. All experiments are conducted in the Google colaboratory environment with Python language in the Tensorflow framework.

4.1 Datasets

The extended Cohn-Kanade (CK+) dataset contains 593 video sequences from 123 objects. CK+ includes seven facial emotions (happiness, sadness, anger, contempt, disgust, fear, and surprise) and a resolution of either 640×490 or 640×480 pixels, which we employed 327 image sequences labeled with seven basic facial emotions in this study.

The images of the Japanese Female Facial Expression (JAFFE) dataset contain 213 images of 10 Japanese female models demonstrating seven facial emotions (happiness, sadness, fear, anger, neutral, disgust, and surprise) and size of 256×256 pixels.

MMI facial expression dataset (MMI) includes 2900 samples of static and sequence images of faces in frontal and profile views. The high-resolution images consist of 75 subjects, in which 235 videos have emotional labels. In this work, we used 238 sequence images of faces in frontal (sessions 1767-2004). This dataset displays six basic emotions (happiness, sadness, anger, disgust, fear, and surprise). The sample of images from CK+, JAFFE, and MMI datasets is depicted in Fig. 7. In the proposed system, input data with size of 48×48 pixels are considered.

4.2 Discussion

This sub-section describes the experiments conducted with the proposed facial emotion recognition system on datasets. Table 4 shows the fivefold and split (85–15%) validation techniques on three popular datasets (CK+, JAFFE, and MMI) to evaluate the accuracy of the proposed model. The results

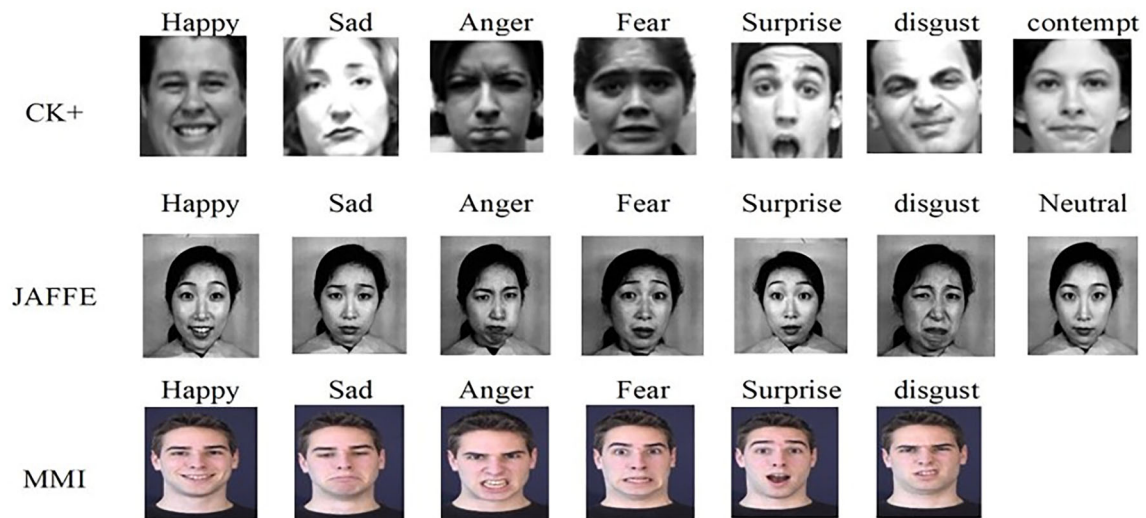


Fig. 7 The sample of images from the datasets (CK+, JAFFE, and MMI)

Table 4 Results of the proposed model on standard datasets (CK+, JAFFE, and MMI) using validation schemes [fivefold and split (85–15%)]

Dataset	Split (85–15%)	Fivefold (%)
CK+	100	100
JAFFE	99.57	99.93
MMI	99.84	99.83

Table 5 Confusion matrix for CK+ dataset

	Happiness	Sadness	Anger	Fear	Contempt	Disgust	Surprise
Happiness	100	0	0	0	0	0	0
Sadness	0	100	0	0	0	0	0
Anger	0	0	100	0	0	0	0
Fear	0	0	0	100	0	0	0
Contempt	0	0	0	0	100	0	0
Disgust	0	0	0	0	0	100	0
Surprise	0	0	0	0	0	0	100

Table 6 Confusion matrix of the proposed method for JAFFE dataset

	Happiness	Sadness	Anger	Fear	Neutral	Disgust	Surprise
Happiness	100	0	0	0	0	0	0
Sadness	0	99.59	0	0.4	0	0	0
Anger	0	0	100	0	0	0	0
Fear	0	0	0	100	0	0	0
Neutral	0	0	0	0	100	0	0
Disgust	0	0	0	0	0	100	0
Surprise	0	0	0	0	0	0	100

Table 7 Confusion matrix of the proposed method for MMI facial expression dataset

	Happiness	Sadness	Anger	Fear	Disgust	Surprise
Happiness	100	0	0	0	0	0
Sadness	0	100	0	0	0	0
Anger	0	0	100	0	0	0
Fear	0	0	0	100	0	0
Disgust	0	0	0	0	100	0
Surprise	0	0	0	0	1.30	98.68

Table 8 Statistical performance of the proposed model for facial emotion recognition on CK+, JAFFE, and MMI datasets

Labels	CK+			JAFFE			MMI		
	Specificity	Sensitivity	F1-score	Specificity	Sensitivity	F1-score	Specificity	Sensitivity	F1-score
Happiness	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sadness	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00
Anger	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Fear	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00
Neutral	–	–	–	1.00	1.00	1.00	–	–	–
Disgust	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.98
Surprise	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00	0.98
Contempt	1.00	1.00	1.00	–	–	–	–	–	–
Total	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99

demonstrate that our method is effective in recognizing emotions. Tables 5, 6, and 7 show the confusion matrix of the CK+, JAFFE, and MMI datasets, respectively.

In Table 8, the performance of the proposed method is tabulated based on the metrics of specificity, sensitivity, and F1-score on the datasets of CK+, JAFFE, and MMI to recognize basic facial emotions. Besides, Fig. 8 shows the diagram related to the accuracy of the datasets with the fivefold cross-validation method.

Friedman non-parametric hypothetical test [50] is used to illustrate the effectiveness of the proposed method on JAFFE dataset. This test consists of two hypotheses: H_0 (null hypothesis) samples uniform distribution between groups and H_1 (substituent hypothesis) represents the effect of the method on samples in groups. The Friedman test statistic consists of four components. First, the sample size (N) represents the total number of observations in each group. Second, the Chi-square distribution, similar to variance over the mean ranks, approximates the test statistic distribution. It is used to examine whether two categorical groups influence the test statistic independently. Third, the degrees of freedom (df) equal the number of groups in your data minus one. Fourth, the p -value/significance level (Asymp. Sig.) is the asymptotic probability and the first type's error probability ($\alpha = 0.05$) that is employed to recognize two hypotheses in the proposed method. The Friedman test between the proposed method and

Table 9 Friedman test on the JAFFE dataset between proposed method with model [28]

Subject	Value
N	5
Chi-square	5.000
df	1
Asymp. sig	0.025

P -value < 0.05

each of the models [28] and [60] ($df = 1$) is performed based on the accuracy of the fivefold approach ($N = 5$). According to Tables 9 and 10, the results show that the proposed method has a significant effect (rejection of the null hypothesis) on the accuracy of facial emotion recognition with the JAFFE dataset compared to the models of [28] and [60].

High prediction accuracy is not only sufficient to improve user confidence in the proposed CNN model and ensure their deployment for real-world applications. Therefore, the model's explainability and robustness features are examined to evaluate its performance.

Explainability [12, 20, 30] refers to techniques used to make the output produced by intelligent systems understandable to humans. Two methods for global-level and local-level

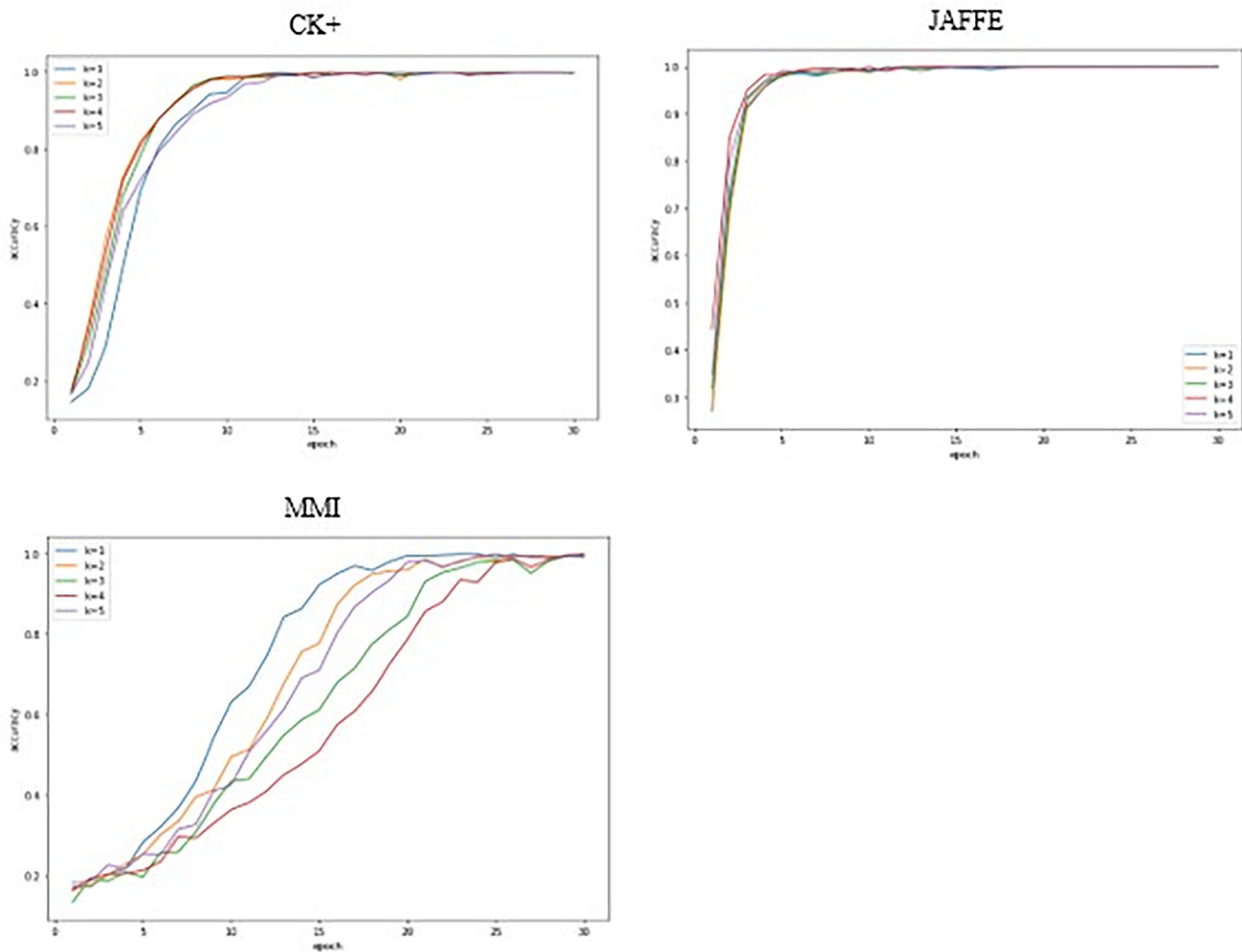


Fig. 8 Accuracy of CK+, JAFFE, and MMI datasets related to fivefold method

Table 10 Friedman test on the JAFFE dataset between proposed method with model [60]

Subject	Value
N	5
Chi-square	5.000
df	1
Asymp. Sig	0.025

P -value < .05

explanations are proposed. Global-level explanations are focused on the model behavior in general.

They can be explored in detail using partial dependence plots (PDPs). PDPs describe how the response of the model changes with a shift of a single feature's value. Local-level explanations are expressed as model behavior around a single model prediction. Many methods have been developed to determine the importance of features at the level of a single

prediction. The Shapley additive explanations (SHAP) use Shapley values from cooperative game theory to attribute the effects of individual additive features to individual model predictions. Figure 9 shows the appropriate explainability of the proposed CNN network in the classification of four images on the JAFFE dataset with the shap technique.

Robustness [18, 20] is to evaluate model performance on manipulated/modified input data and generalizability. In the proposed model, the preprocessing phase and especially data augmentation are employed. Data augmentation is an effective technique for improving the robustness of machine learning models by diversifying the input data. In addition, there are other methods to evaluate data robustness, such as applying noise to the input data [18]. Figure 10 illustrates the good performance of a test sample on the JAFFE dataset to recognize emotions despite the presence of noise (with Gaussian and Blur methods) as a type of data corruption.

In the following, the experimental results of our method and well-known approaches (LBP, HOG, LDP, Gabor filters, and DWT) are shown in Table 11. Also, the proposed model

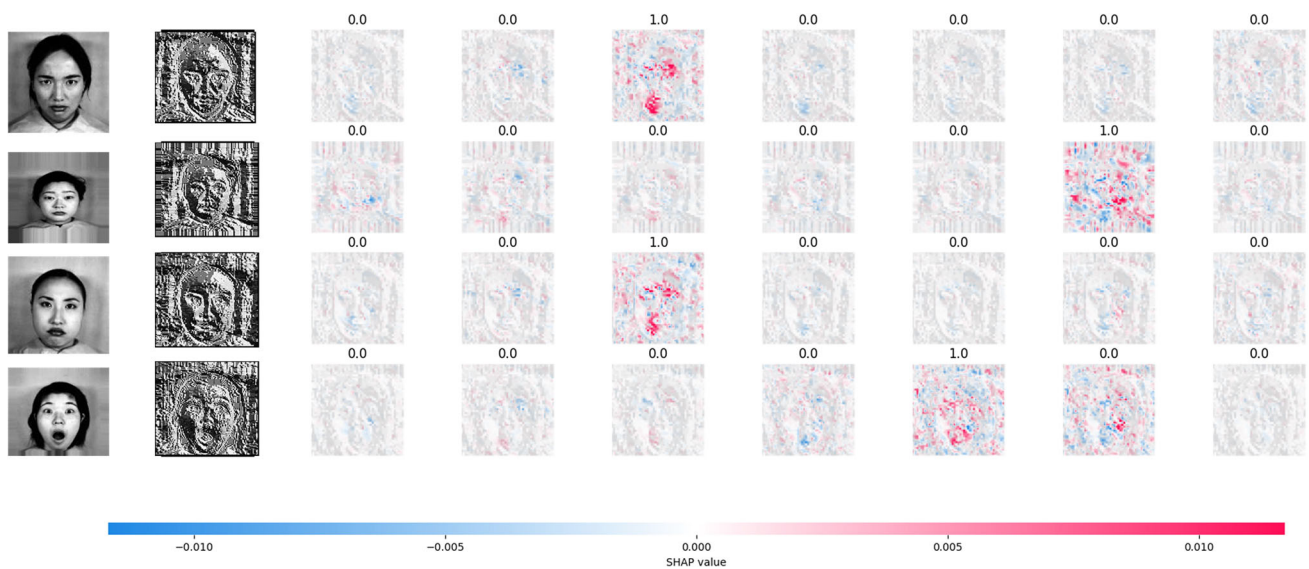


Fig. 9 Correct emotions recognition of the proposed method using the shap method on the JAFFE dataset

Fig. 10 The correct prediction of the emotion of happiness on one sample of JAFFE: **a** the original image, **b** the image filtered with LSBP, **c** the image with blur, and **d** the image with Gaussian noise

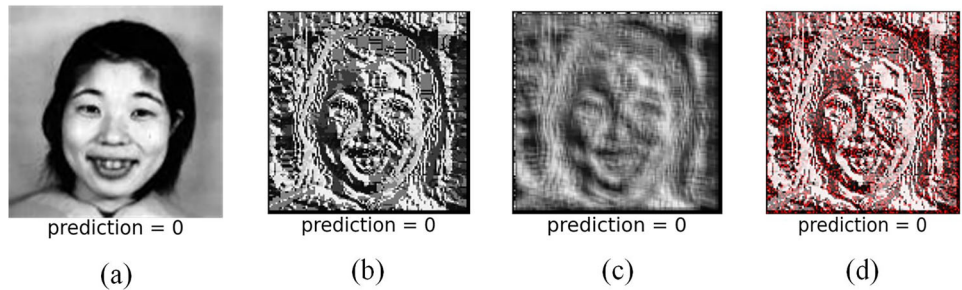


Table 11 Comparison of proposed system with existing approaches

References	Dataset/s	Accuracy rate (%)
LBP [55]	JAFFE	88.9
HOG [14]	CK + , JAFFE	88.7, 94.3
LDP [27]	JAFFE	85.4
Gabor filters [22]	JAFFE	92.4
DWT [32]	JAFFE	96
Proposed method	CK+, JAFFE, MMI	100, 99.93, 99.83

is compared with state-of-art methods in Table 12. The evaluations confirm that the proposed system has improved the accuracy of the method compared to previous approaches, as depicted in Fig. 11.

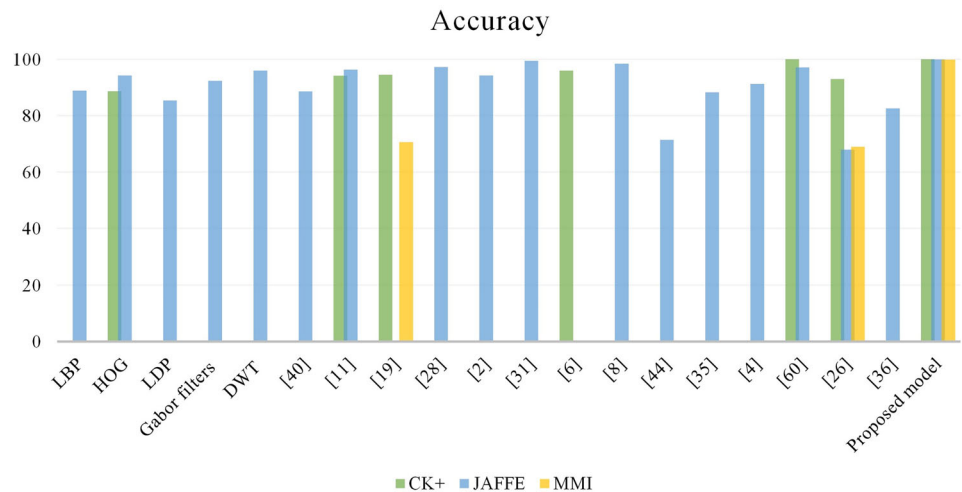
5 Conclusion

The robust feature extraction method plays a vital role in emotion recognition. This paper proposes a new face feature extraction method for a facial emotion recognition system. Combining the novel LSBP descriptor and CNN is proposed to extract the effective features. LSBP, based on LBP, is

defined to scroll image datasets on the sort value of pixels to extract features. Then, CNN is used for feature extraction from data obtained by the LSBP technique. In the proposed method, CNN hyperparameters are optimized using the WOA technique. In evaluation, we used the three well-known face datasets, namely CK+, JAFFE, and MMI. The results show that the proposed method guarantees high performance of emotion recognition. In the future, the scope of the proposed model can be expanded by providing new and combined methods of feature extraction, optimization of more hyperparameters with meta-heuristic algorithms, and different biometrics so that applications with an efficient

Table 12 Comparison of the proposed model with the other facial expression recognition models

References	Dataset/s	Accuracy rate (%)
Meena et al. [40]	JAFFE	88.57
Boughida et al. [11]	CK+, JAFFE	94.2, 96.3
Farkhod and Chae [19]	CK+, MMI	94.5, 70.63
Jeen Retna Kumar et al. [28]	JAFFE	97.3
Alphonse and Dharma [2]	JAFFE	94.3
Kar et al. [31]	JAFFE	99.46
Barra et al. [6]	CK+	96
Bendjillali et al. [8]	JAFFE	98.43
Nigam et al. [44]	JAFFE	71.43
Kola and Samayamantula [35]	JAFFE	88.3
Arora et al. [4]	JAFFE	91.3
Tuncer et al. [60]	CK+, JAFFE	100, 97.09
Iqbal et al. [26]	CK+, JAFFE, MMI	92.97, 67.97, 69.01
Kola and Samayamantula [36]	JAFFE	82.6
Proposed method	CK+, JAFFE, MMI	100, 99.93, 99.83

Fig. 11 Comparison between proposed method and the state-of-the-art model

emotional recognition system can be used in commercial, entertainment, and industrially fields.

Author contributions All authors discussed the results and commented on the writing of the manuscript.

Funding The authors have not received any financial support from any person, institution, or organization for this research work.

Data availability The data used to support the findings of this study are available from the corresponding author upon request.

Declarations

Conflict of interest The authors reported no potential conflicts of interest with respect to the research, writing, or publication of this article.

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

gotten through simulation with available datasets. All authors declare that they have no conflict of interest.

Consent to participate All authors declare that they have the consent to participate.

Consent for publication All authors declare that they have consent for publication.

References

- Ahmed, F., Hossain, E.: Automated facial expression recognition using gradient-based ternary texture patterns. *Chin. J. Eng.* (2013). <https://doi.org/10.1155/2013/831747>
- Alphonse, A.S., Dharma, D.: Novel directional patterns and a generalized supervised dimension reduction system (GSDRS) for facial emotion recognition. *Multim. Tools Appl.* **77**(8), 9455–9488 (2021). <https://doi.org/10.1007/s11042-017-5141-8>

3. Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L.: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* (2021). <https://doi.org/10.1186/s40537-021-00444-8>
4. Arora, M., Kumar, M., Garg, N.K.: Facial emotion recognition system based on PCA and gradient features. *National Academy Science Letters* **41**(6), 365–368 (2018). <https://doi.org/10.1007/s40009-018-0694-2>
5. Banharsakun, A.: Towards improving the convolutional neural networks for deep learning using the distributed artificial bee colony method. *Int. J. Mach. Learn. Cybern.* **10**(6), 1301–1311 (2018). <https://doi.org/10.1007/s13042-018-0811-z>
6. Barra, P., Maio, L.D., Barra, S.: Emotion recognition by web-shaped model. *Multim. Tools Appl.* (2022). <https://doi.org/10.1007/s11042-022-13361-6>
7. Bayğın, M., Tuncer, I., Doğan, S., Barua, P.D., Tuncer, T., Cheong, K.H., Achary, U.R.: Automated facial expression recognition using exemplar hybrid deep feature generation technique. *Soft. Comput.* **27**(13), 8721–8737 (2023). <https://doi.org/10.1007/s00500-023-08230-9>
8. Bendjillali, R., Beladgham, M., Merit, K., Taleb-Ahmed, A.: Improved facial expression recognition based on DWT feature for deep CNN. *Electronics* **8**(3), 324 (2019). <https://doi.org/10.3390/electronics8030324>
9. Bentoumi, M., Daoud, M., Benaouali, M., Taleb Ahmed, A.: Improvement of emotion recognition from facial images using deep learning and early stopping cross validation. *Multim. Tools Appl.* (2022). <https://doi.org/10.1007/s11042-022-12058-0>
10. Boughanem, H., Ghazouani, H., Barhoumi, W.: Multichannel convolutional neural network for human emotion recognition from in-the-wild facial expressions. *Vis. Comput.* (2022). <https://doi.org/10.1007/s00371-022-02690-0>
11. Boughida, A., Kouahla, M.N., Lafifi, Y.: A novel approach for facial expression recognition based on Gabor filters and genetic algorithm. *Evol. Syst.* (2021). <https://doi.org/10.1007/s12530-021-09393-2>
12. Bücken, M., Szepannek, G., Gosiewska, A., Biecek, P.: Transparency, auditability, and explainability of machine learning models in credit scoring. *J. Op. Res. Soc.* (2021). <https://doi.org/10.1080/01605682.2021.1922098>
13. Canal, F.Z., Müller, T.R., Matias, J.C., Scotton, G.G., de Sa Junior, A.R., Pozzebon, E., Sobieranski, A.C.: A survey on facial emotion recognition techniques: a state-of-the-art literature review. *Inf. Sci.* **582**, 593–617 (2022). <https://doi.org/10.1016/j.ins.2021.10.005>
14. Chen, J., Chen, Z., Chi, Z., Fu, H.: Facial expression recognition based on facial components detection and HOG features. In: International workshops on electrical and computer engineering, Turkey (2014)
15. Cîrmeanu, A.-L., Popescu, D., Iordache, D.: New trends in emotion recognition using image analysis by neural networks, a systematic review. *Sensors* **23**(16), 7092 (2023). <https://doi.org/10.3390/s23167092>
16. Darwish, A., Hassanien, A.E., Das, S.: A survey of swarm and evolutionary computing approaches for deep learning. *Artif. Intell. Rev.* **53**(3), 1767–1812 (2019). <https://doi.org/10.1007/s10462-019-09719-2>
17. Dornaika, F., Bosaghzadeh, A., Salmanem, H., Ruichek, Y.: Chapter 9—object categorization using adaptive graph-based semi-supervised learning. In: Samui, P., Sekhar, S., Balas, V. E. (eds.). ScienceDirect; Academic Press. (2017) <https://www.sciencedirect.com/science/article/abs/pii/B9780128113189000090>
18. Drenkow, N., Sani, N., Ilya Shpitsner, & Mathias Unberath.: A systematic review of robustness in deep learning for computer vision: Mind the gap?, ArXiv.org (2021) <https://doi.org/10.48550/arxiv.2112.00639>
19. Farkhod, M., Chae, O.: Local prominent directional pattern for gender recognition of facial photographs and sketches. *J. Info. Secur.* **19**(2), 91–104 (2019). <https://doi.org/10.33778/kcsa.2019.19.2.091>
20. Giudici, P., Centurelli, M., Turchetta, S.: Artificial intelligence risk measurement. *Expert Syst. Appl.* **235**, 121220–121220 (2024). <https://doi.org/10.1016/j.eswa.2023.121220>
21. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. The Mit Press (2016)
22. Guo, G., Dyer, C.R.: Simultaneous feature selection and classifier training via linear programming: a case study for face expression recognition. In: 2003 IEEE computer society conference on computer vision and pattern recognition (2003). <https://doi.org/10.1109/cvpr.2003.1211374>
23. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. ArXiv.org. (2015). <https://arxiv.org/abs/1512.03385>
24. Hegenbart, S., Uhl, A.: A scale-and orientation-adaptive extension of local binary patterns for texture classification. *Pattern Recogn.* **48**(8), 2633–2644 (2015). <https://doi.org/10.1016/j.patcog.2015.02.024>
25. Hung, J.C., Lin, K.C., Lai, N.X.: Recognizing learning emotion based on convolutional neural networks and transfer learning. *Appl. Soft Comput.* **84**, 105724 (2019). <https://doi.org/10.1016/j.asoc.2019.105724>
26. Iqbal, M.T.B., Abdullah-Al-Wadud, M., Ryu, B., Makhmudkhujaev, F., Chae, O.: Facial expression recognition with neighborhood-aware edge directional pattern (NEDP). *IEEE Trans. Affect. Comput.* **11**(1), 125–137 (2020). <https://doi.org/10.1109/taffc.2018.2829707>
27. Jabid, T.: Robust facial expression recognition based on local directional pattern. *ETRI J.* **32**(5), 784–794 (2010). <https://doi.org/10.4218/etrij.10.1510.0132>
28. Kumar, J.R., Sundaram, M., Arumugam, N., Kavitha, V.: Face feature extraction for emotion recognition using statistical parameters from subband selective multilevel stationary biorthogonal wavelet transform. *Soft. Comput.* **25**(7), 5483–5501 (2021). <https://doi.org/10.1007/s00500-020-05550-y>
29. Kalyoncu, C.: Sorted uniform local binary patterns. Springer eBooks, pp. 733–739. (2022) https://doi.org/10.1007/978-981-16-8129-5_112
30. Kamakshi, V., Krishnan, N.C.: Explainable image classification: the journey so far and the road ahead. *AI* **4**(3), 620–651 (2023). <https://doi.org/10.3390/ai4030033>
31. Kar, N.B., Babu, K.S., Sangaiah, A.K., Bakshi, S.: Face expression recognition system based on ripplelet transform type II and least square SVM. *Multim. Tools Appl.* **78**(4), 4789–4812 (2017). <https://doi.org/10.1007/s11042-017-5485-0>
32. Kazmi, S.B., Qurat-ul-Ain, A., Jaffar, M.: Wavelets-based facial expression recognition using a bank of support vector machines. *Soft. Comput.* **16**(3), 369–379 (2011). <https://doi.org/10.1007/s00500-011-0721-4>
33. Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S.: A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* (2020). <https://doi.org/10.1007/s10462-020-09825-6>
34. Kohavi, R.: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model selection, in appears in the international joint conference on artificial intelligence (IJCAI). (1995)
35. Kola, D.G.R., Samayamantula, S.K.: A novel approach for facial expression recognition using local binary pattern with adaptive window. *Multim. Tools Appl.* **80**(2), 2243–2262 (2020). <https://doi.org/10.1007/s11042-020-09663-2>
36. Kola, D.G.R., Samayamantula, S.K.: Facial expression recognition using singular values and wavelet-based LGC-HD operator. *IET Biometrics.* **10**(2), 207–218 (2021). <https://doi.org/10.1049/bme2.12012>

37. Lakshmi, D., Ponnusamy, R.: Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders. *Microprocess. Microsyst.* **82**, 103834 (2021). <https://doi.org/10.1016/j.micpro.2021.103834>
38. Lucey, P., Cohn, J. F., Kanade, T., Saraçih, J., Ambadar, Z., Matthews, I.: The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression, In: 2010 IEEE computer society conference on computer vision and pattern recognition—workshops. (2010)
39. Lyons, M.J., Kamachi, M., Gyoba, J.: Japanese female facial expressions (JAFFE), database of digital images. (1997)
40. Meena, H.K., Sharma, K.K., Joshi, S.D.: Improved facial expression recognition using graph signal processing. *Electron. Lett.* **53**(11), 718–720 (2017). <https://doi.org/10.1049/el.2017.0420>
41. Mirjalili, S., Lewis, A.: The whale optimization algorithm. *Adv. Eng. Softw.* **95**, 51–67 (2016)
42. Moradi, R., Berangi, R., Minaei, B.: A survey of regularization strategies for deep models. *Artif. Intell. Rev.* **53**(6), 3947–3986 (2019). <https://doi.org/10.1007/s10462-019-09784-7>
43. Mukhopadhyay, M., Dey, A., Kahali, S.: A deep-learning-based facial expression recognition method using textural features. *Neural Comput. Appl.* (2022). <https://doi.org/10.1007/s00521-022-08005-7>
44. Nigam, S., Singh, R., Misra, A.K.: Efficient facial expression recognition using histogram of oriented gradients in wavelet domain. *Multim. Tools Appl.* **77**(21), 28725–28747 (2018). <https://doi.org/10.1007/s11042-018-6040-3>
45. Nwankp, C., Ijomah, W., Gachagan, A., Marshall, S.: Activation functions: comparison of trends in practice and research for deep learning. *ArXiv.org.* (2018). <https://arxiv.org/abs/1811.03378>
46. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987 (2002). <https://doi.org/10.1109/tpami.2002.1017623>
47. Pantic, M., Valstar, M., Rademaker, R., Maat, L.: Web-based database for facial expression analysis, In: 2005 IEEE international conference on multimedia and exp. (2005)
48. Park, S.: A 2021 Guide to improving CNNs-network architectures: historical network architectures. *Geek Culture.* <https://medium.com/geekculture/a-2021-guide-to-improving-cnns-network-architectures-historical-network-architectures-d23f32afb1bd>. (2021). Accessed 15 July 2023
49. Parmar, R.: Common loss functions in machine learning. *Medium.* <https://towardsdatascience.com/common-loss-functions-in-machine-learning-46af0ffc4d23>. (2018). Accessed 15 July 2023
50. Ramachandran, K. M., Tsokos, P. C.: Chapter 12—Nonparametric Statistics. In: *Mathematical statistics with applications in R* (Third Edition) pp. 491–530. Elsevier (2021)
51. Ramachandran, P., Zoph, B., Le, Q. V.: Searching for activation functions. *ArXiv.org.* (2017). <https://arxiv.org/abs/1710.05941>
52. Ruder, S.: An overview of gradient descent optimization algorithms. *ArXiv.org.* (2016). <https://arxiv.org/abs/1609.04747>
53. Ryu, J., Hong, S., Yang, H.S.: Sorted consecutive local binary pattern for texture classification. *IEEE Trans. Image Process.* **24**(7), 2254–2265 (2015). <https://doi.org/10.1109/tip.2015.2419081>
54. Sairamya, N.J., Susmitha, L., Thomas, George. S., Subathra, M.S.P.: Chapter 12—hybrid approach for classification of electroencephalographic signals using time–frequency images with wavelets and texture features. In: Hemanth, D. J., Gupta, D., Emilia Balas, V. (eds.). *ScienceDirect; Academic Press.* (2019) <https://www.sciencedirect.com/science/article/abs/pii/B9780128155530000136>
55. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. *Image Vis. Comput.* **27**(6), 803–816 (2009). <https://doi.org/10.1016/j.imavis.2008.08.005>
56. Sharifnejad, M., Shahbahrami, A., Akoushideh, A., Hassannpour, R.Z.: Facial expression recognition using a combination of enhanced local binary pattern and pyramid histogram of oriented gradients features extraction. *IET Image Proc.* **15**(2), 468–478 (2020). <https://doi.org/10.1049/ipr2.12037>
57. Song, T., Han, Y., Feng, J., Wang, Y., Gao, C.: First- and second-order sorted local binary pattern features for grayscale-inversion and rotation invariant texture classification. (n.d.). *Ieeexplore.ieee.org.* (2020) <https://ieeexplore.ieee.org/document/9412246>
58. Suk, H.I.: Chapter 1—an introduction to neural networks and deep learning. In: Zhou, S. K., Greenspan, H., Shen, D. (eds.) *ScienceDirect; Academic Press.* (2017) <https://www.sciencedirect.com/science/article/abs/pii/B978012810408800002X>
59. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. *ArXiv.org.* (2014). <https://arxiv.org/abs/1409.4842>
60. Tuncer, T., Dogan, S., Abdulhamit, Subasi.: Automated facial expression recognition using novel textural transformation. (2023) <https://doi.org/10.1007/s12652-023-04612-x>
61. Vasudha, Kakkar, D.: Facial expression recognition with LDPP & LTP using deep belief network. (2018). <https://doi.org/10.1109/spin.2018.8474035>
62. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. *ArXiv.org.* (2015). <https://arxiv.org/abs/1505.00853>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.