

# Grey Wolf optimisation-based feature selection and classification for facial emotion recognition

Ninu Preetha Nirmala Sreedharan<sup>1</sup>✉, Brammya Ganesan<sup>1</sup>, Ramya Raveendran<sup>1</sup>, Praveena Sarala<sup>1</sup>, Binu Dennis<sup>1</sup>, Rajakumar Boothalingam R.<sup>1</sup>

<sup>1</sup>Resbee Info Technologies Private Limited, Thuckalay 629175, India

✉ E-mail: research@resbee.org

**Abstract:** The channels used to convey the human emotions consider actions, behaviours, poses, facial expressions, and speech. An immense research has been carried out to analyse the relationship between the facial emotions and these channels. The goal of this study is to develop a system for Facial Emotion Recognition (FER) that can analyse the elemental facial expressions of human, such as normal, smile, sad, surprise, anger, fear, and disgust. The recognition process of the proposed FER system is categorised into four processes, namely pre-processing, feature extraction, feature selection, and classification. After preprocessing, scale invariant feature transform -based feature extraction method is used to extract the features from the facial point. Further, a meta-heuristic algorithm called Grey Wolf optimisation (GWO) is used to select the optimal features. Subsequently, GWO-based neural network (NN) is used to classify the emotions from the selected features. Moreover, an effective performance analysis of the proposed as well as the conventional methods such as convolutional neural network, NN-Levenberg–Marquardt, NN-Gradient Descent, NN-Evolutionary Algorithm, NN-firefly, and NN-Particle Swarm Optimisation is provided by evaluating few performance measures and thereby, the effectiveness of the proposed strategy over the conventional methods is validated.

## 1 Introduction

Automatic identification of a person, termed as face recognition system, generally depends on diverse pattern recognition approaches as well as digital image processing techniques [1]. Basically, automatic face recognition (AFR) deals with recognising a person from a set of images or video frames using a computer [2, 3]. The discrimination of a peculiar personality from unfamiliar objects portrayed by facial images is a high-level vision task, known as AFR process. In daily life, contextualised conditions involve the expression regarding the emotion of faces; however, duration of emotional face-context integration remains unsettled [4]. Accordingly, the task of potent and valuable human communication rose as the fundamental skill, which accurately recognises and distinguishes the emotions from diverse facial expressions [5–7]. Thus, facial emotion recognition (FER) has become a challenging issue, as the appearance variation of the full face signals numerous classes of facial expressions, all with differing magnitudes [8, 9].

In general, the transformation of various feelings and emotions is promoted by the powerful channel called facial expressions [10–12]. The contraction of facial muscles brings out the facial expressions, which temporally contort the facial units such as mouth, lips, nose, eyebrows [13, 14]. On the academic basis, the appropriate movement and the location of the facial muscles beneath the skin convey significant facial expressions. Those muscle movements create patterns of non-verbal communication and transfer the emotional status to an observer [15, 16]. Subsequently, by relevant evidence, one's emotional status tendentious in the order of emotions, while interpreting an expression either through the face or the body [13]. Nevertheless, in the case of healthy social interactions, the recognition of accurate facial expression has become complex. However, the patients with neurodegenerative disease, especially with prefrontal or temporal atrophy, show dramatic socio-emotional impairment [17]. In addition, there is a shortage of observations in their own emotions for the person with head injury, who feels intricacy with expressions [18, 19]. Moreover, insufficient processing of social information or social cognition [20], constant social isolations, and

psychotic symptoms are faced by the patients with mental disorder. Even more, the loss of social cognition like FER is obvious in schizophrenia and the impact of functional prognosis differentially to neurocognitive measures such as memory and intelligence quotient [21, 22].

Recently, there were many techniques available to recognise the distinct facial expressions. Some of the common methods were deep belief network and convolutional neural network (CNN). These methods help to promote diverse tasks in the field of image processing and have obtained tremendous success [23]. Another technique called multi-layer neural network (NN) [24, 25] was also highly suited for the face recognition purpose. On considering the recognition process, the features associated with the face were generally comprised of geometric-based, appearance-based, and Action Unit-based features [26]. Furthermore, intense and vigorous results were obtained by sparse representation models, against the visual divergence of occlusion, brightness, and corruption of facial images [3]. Similarly, scale invariant feature transform (SIFT) [27] was also adopted for the purpose of object recognition.

This paper develops a novel FER system through the processes, such as pre-processing, feature extraction, feature selection, and classification. The main contribution of the paper is to use the logic of Grey Wolf optimisation (GWO) algorithm in the phases like feature extraction and classification. Thus, it can select the relevant features from the feature set to be applicable for classification and reduce the error of weights applied for training during the classification process. Here, the optimal feature extraction and classification uses the bio-inspired algorithm called GWO. The organisation of this paper is as follows: Section 2 describes the literature review on FER and its respective problem statement. Section 3 portrays the illustration behind the proposed method for FER and Section 4 presents the optimised feature extraction and classification using GWO algorithm. Section 5 provides the valuable results and discussions portraying the performance analysis, comparative analysis, and training performance. Section 6 concludes the paper with proper suggestions.

## 2 Literature review

### 2.1 Related works

In 2017, Zwick [1] have proposed the method of social cognition in two forms such as FER and reasoning. In general, they have performed the method to desire the remitted and accurately depressed patients (rMDD and aMDD) with naturally effective materials. Moreover, they have analysed the diminished facial mimicry arbitrates, the connotation among the FER and depressive indications. The shortages that exist in the reasoning and FER were also analysed by them. The social cognition was evaluated by the stimuli from Amsterdam Dynamic Facial Expression group. Meanwhile, they have recorded the movement of the Corrugator supercilii and Zygomaticus Major. The simplicity of decoding was certified by comparing the stimulus material. The comparison has shown that the FER deficits of aMDD and rMDD patients are highly noticeable.

In 2016, Zhang *et al.* [21] have developed the modified evolutionary firefly algorithm for the purpose of FER. The bio-inspired algorithm like firefly algorithm was contributed to optimise the features. Initially, the preliminary discriminative face representation was determined by the improved local binary pattern descriptor. In fact, the curved search performance of moths and attractiveness pattern of the fireflies were exploited in the proposed algorithm to alleviate the premature convergence of the moth-flame optimisation (MFO) algorithm. The improvement of the local manipulation of the fireflies was made possible by the employment of logarithmic spiral search capability. The global exploration was increased by both the attractiveness function and the best solution representation of the firefly when compared to the flames in MFO. Moreover, the most favourable solution was obtained by linking the levy flights to the simulated annealing method. This method was experimented for the detection of seven expressions from CK+, JAFFE, and BU-3DFE databases using various ensemble and single classifiers. Finally, the performance of the proposed method was proven to be effective than the state-of-the-art feature optimisation methods.

In 2016, Theurel *et al.* [4] have suggested a method for extracting the visual context information in the case of FER. They have carried the experimentation for the candidates from the age of 5–15 years. Collectively, 190 participants were involved in the respective experimentation. In addition, five different facial expressions, such as happiness, disgust, anger, sadness, and fear, were considered. They have experimented in two conditions: in the presence and in the absence of visual context. It is clear from the experiment that the recognition of emotions in the presence of visual context information has shown superior performance to the emotions in the lack of visual context information. Moreover, the performance was independent of the age of the participant, whereas, it was dependent on the gender and emotions. Ultimately, it has proven that the children from the age of 5 years were able to coordinate the visual context information and the facial expressions that have enhanced the FER performance.

In 2016, Rigon *et al.* [19] have found that the deterioration of facial affect recognition was familiar ensuing severe-moderate traumatic brain injury (TBI). In TBI community, there were various changes in large-scale practical brain networks in TBI population. The abilities of facial affect recognition were measured for the healthy comparison participants, and 26 participants with TBI and fMRI were used for resting-state functional connectivity. The network-based statistics did the examination of the recognition ability. Moreover, the relationship between the inter-individual differences in emotion recognition skill was determined within the facial affect processing network. After the analysis, it was noted that the participants with TBI had shown lesser resting-state functional connectivity in a component forming homotopic and within-hemisphere, anterior-posterior connections within the facial affect processing network. Furthermore, higher-emotion labelling skills were formed in stronger rs-FC, which should be within a network composing inter- and intra-hemispheric bilateral connections. The findings of the experimental results have certified the ability recognition of the facial-affect after TBI.

In 2017, Gola *et al.* [17] have performed the experiment associated with face-to-face socio-emotional testing, and the informants defined the characteristic socio-emotional behaviour of the participants. Here, voxel-based morphometry was used to connect the performance of the patients on emotions across all the utilised samples. They have compared the fundamental socio-emotional deficits with determined emotional imitation scores and have found that the patients with known socio-emotional deficits performed not as good as intentional emotion imitation control. The rightward cortical atrophy pattern was similar to the left lateralised speech production network and was affiliated with intentional emotional imitation deficits. Finally, they have found the complex socio-emotional communication was convenient to the neural mechanisms for the patients with neurodegenerative disease.

In 2017, Fan *et al.* [15] have utilised the spatiotemporal features for performing the emotion recognition that was based on the local Zernike moment. Collectively, the recognition scheme was accomplished using motion change frequency. They have utilised the dynamic features that consist of motion history and entropy of image. Further, the dynamic data of the facial expression was obtained by weighting strategy and was classified by the use of traditional support vector machine (SVM). The respective experiment has carried out in CK+ and MMI datasets. Moreover, the experimental outcome has revealed that the integrated framework has attained better performance than the individual descriptor and the state-of-the-art techniques.

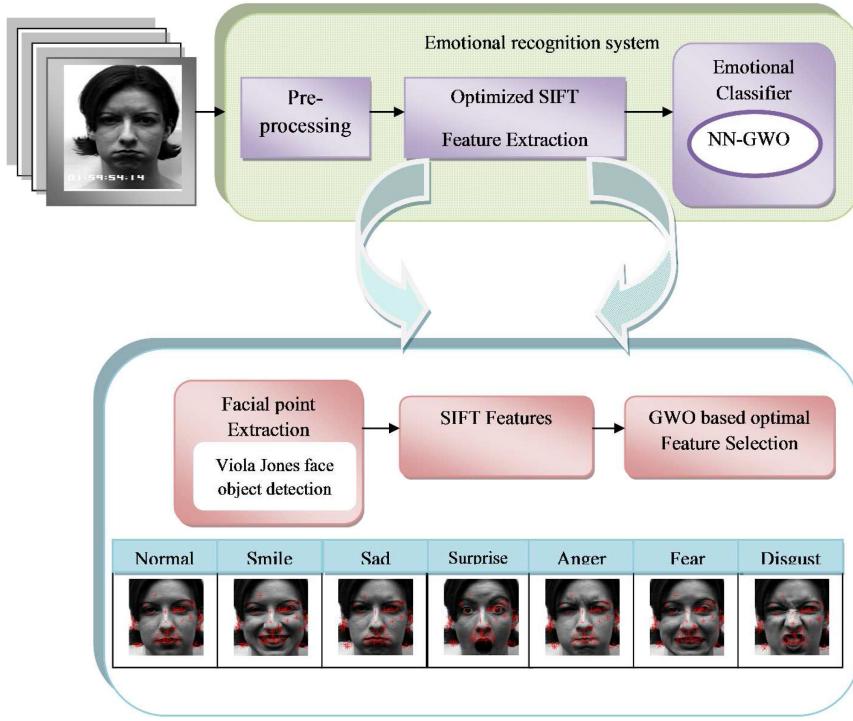
### 2.2 Research gaps in FER

Different methods used for the FER, as reported in the literature, include moth-firefly optimisation [21], visual context integration method [4], network-based statistics algorithm [19], voxel-based morphometry [17], and weighting strategy-based integration method [15]. Among those methods, moth-firefly optimisation provides increased convergence speed with sufficient diversity and capability of fine-tuning the solutions. However, there may be inefficiency in the transverse orientation, which leads to produce artificial lights and is needed to move in a straight way, when the light source is very far. Moreover, high accuracy and confirmed facial expressions are produced by visual context integration method [4], which in turn become complex when the criterion changes across trials, subjects, and conditions. The network-based statistics algorithm in [19] exhibits successful recognition capability and provides high emotional labelling skills, yet it is limited in certain conditions. In addition, the edges involved should be eliminated for entire subjects, if the vertex is absent in a single subject. Subsequently, voxel-based morphometry [17] contributes better knowledge regarding the mechanistic relationships and careful delineation of empathy-related processes, but there is a possibility of producing altered diagnostic groups, which failed in recognising the spontaneous form of expressions. The classification accuracy of weighting strategy-based integration method [15] is high, but requires increased computational speed in real-time performance and often offers worst recognition performance due to the degradation of information. These limitations have motivated a lot to promote an effective algorithm for developing the automatic FER system.

## 3 Facial emotion recognition system

### 3.1 Proposed methodology

The block diagram of the proposed FER is shown in Fig. 1. Here, the set of faces with various emotions such as normal, smile, sad, surprise, anger, fear, and disgust are applied to the pre-processing stage includes illumination and normalisation. Further, the optimised SIFT features are extracted from the pre-processed face images. The optimised SIFT feature extraction is categorised into three stages include facial point extraction using Viola face object detection, SIFT feature extraction and GWO-based optimal feature selection. To the next, the appropriate emotions of the face are recognised by applying the optimally selected features to the GWO-NN-based classification. The description regarding each step of the proposed FER is described in the subsequent sections.



**Fig. 1** Block diagram of the proposed facial emotion recognition

### 3.2 Pre-processing

The input image with different emotions is considered as  $E(x, y)$ , where  $(x, y)$  indicates the spatial coordinate. In the beginning, the input face image is given for pre-processing in order to control the noise and redundant information. As a result, the pre-processing can improve the features of the image, which thus, become adaptable for the further processing. Accordingly, this paper adopts illumination normalisation to pre-process the face image. In the particular method, the range of intensity values of the pixels are tuned to control the over lighting of the image. Thus the obtained pre-processed face image is denoted as  $V(x, y)$ .

### 3.3 Facial point extraction

Since all the human faces share analogous properties, it is necessary to differentiate the features among the different face images. As the experiment belongs to the emotion recognition of face images, it is needed to consider the objects which take responsibility to provide the suitable expressions. On considering the face images, the main objects which express the emotions are eye, nose, mouth, and ears. Hence, the facial point extraction extracts those objects from the face image based on Viola-Jones object detection algorithm, which uses the Haar features to locate the specific objects. Subsequently,  $O(x, y)$  is considered as the face image obtained after extracting the eyes, ears, mouth, and nose.

### 3.4 SIFT feature extraction

The SIFT-based feature extraction can extract the unique and precise informative features of the face image. Consider the two-dimensional face image  $O(x, y)$ , after the facial point extraction. The SIFT feature extraction basically depends on five steps, which are depicted as follows.

**3.4.1 Detection of image scale-space extreme:** In a two-dimensional image, the scale-space representation is based on the convolution of the input image with the Gaussian function as per (1), where  $G(x, y, \sigma)$  specifies the variable scale Gaussian function and  $O(x, y)$  indicates the input image after facial point detection and  $\sigma$  indicates the scaled coordinate

$$S(x, y, \sigma) = G(x, y, \sigma) * O(x, y) \quad (1)$$

**3.4.2 Demonstration of DOG scale space:** The extraction of invariant features is obtained by the difference of Gaussian (DOG). When the image is added with noise, the constancy associated with the extracted features is not convenient for the geometric transformation. Under such circumstances, DOG plays a significant role in SIFT feature extraction, as it helps to take out more scale invariant features. In fact, the ideal matching assists in extracting more scalable feature points. Therefore, using the scale space function of the image, DOG scale space is represented as

$$H(x, y, \sigma) = S(x, y, k\sigma) - S(x, y, \sigma) \quad (2)$$

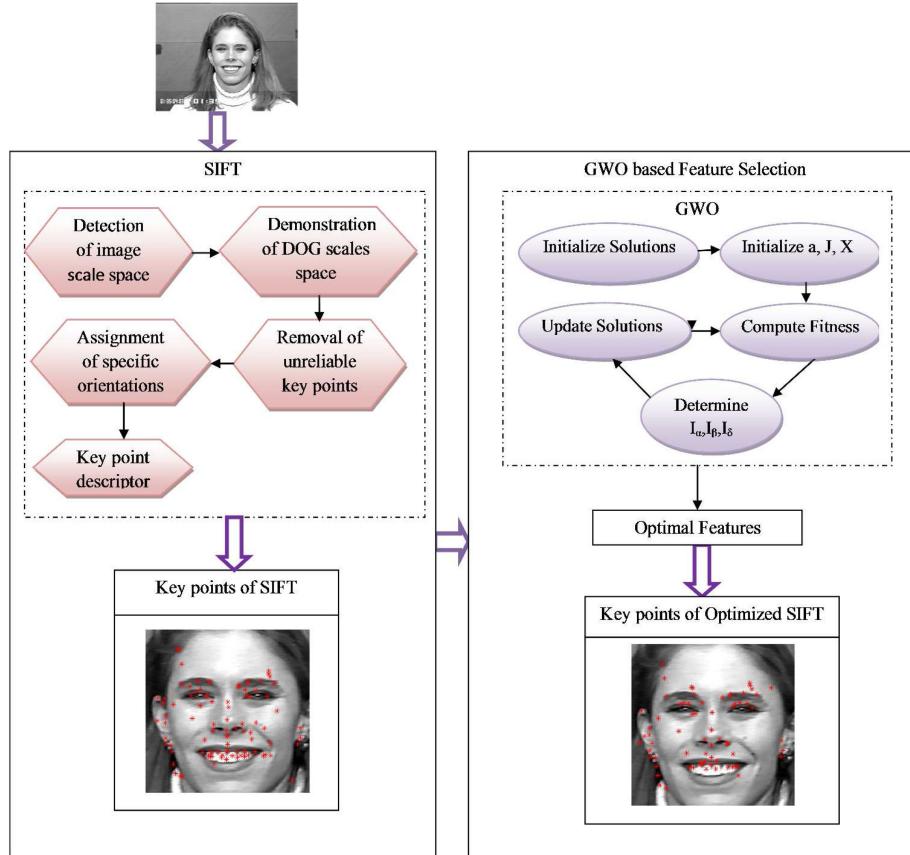
$$H(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * O(x, y) \quad (3)$$

In (3), the Gaussian function is represented as  $G(x, y, \sigma) = (1/2\sigma^2)e^{-(x^2+y^2)/2\sigma^2}$ . The extreme points with the 2D image space are determined by correlating each sample point of the image with 26 neighbours in the DOG scale space. Furthermore, it is needed to compare the target point with eight neighbours in the current image and also with 18 neighbours in the scale above and below.

**3.4.3 Removal of unreliable key points:** It estimates modulus of previous outcome at each candidate key point. Subsequently, it assigns the low contrast structure for the case of values below the threshold value. As a result, the particular structure will be delicate to noise, so that it eliminates certain key point. Further, it evaluates the ratio associated with the principle curvatures of each candidate key point. In fact, it is for the poorly determined peaks in the scale-normalised operators of Laplacian-of-Gaussian. The key point is kept if the ratio is below the threshold value.

**3.4.4 Assignment of specific orientations:** It is essential to assign a concerned orientation to each key point. The number of orientation should be one or more for each key point by the directions related to the local image gradient.

**3.4.5 Key point descriptor:** Around the location of the respective key point, it samples the magnitude and orientation of the image gradient. It is accomplished using the scale of the key point, which in turn chooses the range of Gaussian blur for the image. Further, it calculates the feature descriptor, as a group of orientation histogram on the neighbourhood pixels set as  $16 \times 16$  around the



**Fig. 2** Generation of SIFT features and optimised SIFT features

key point. Moreover, eight bins are assigned to each histogram, and  $4 \times 4$  arrays of histograms are assigned to each descriptor around the key point. Thus a vector of dimension,  $4 \times 4 \times 8 = 124$ , is obtained as the feature vector for each key point.

Finally, the extracted SIFT features of the facial image represented as  $I(x, y)$  are applied to the GWO-based optimisation for the optimal selection from the extracted features.

## 4 Optimised feature selection and classification

### 4.1 Optimal feature selection by GWO

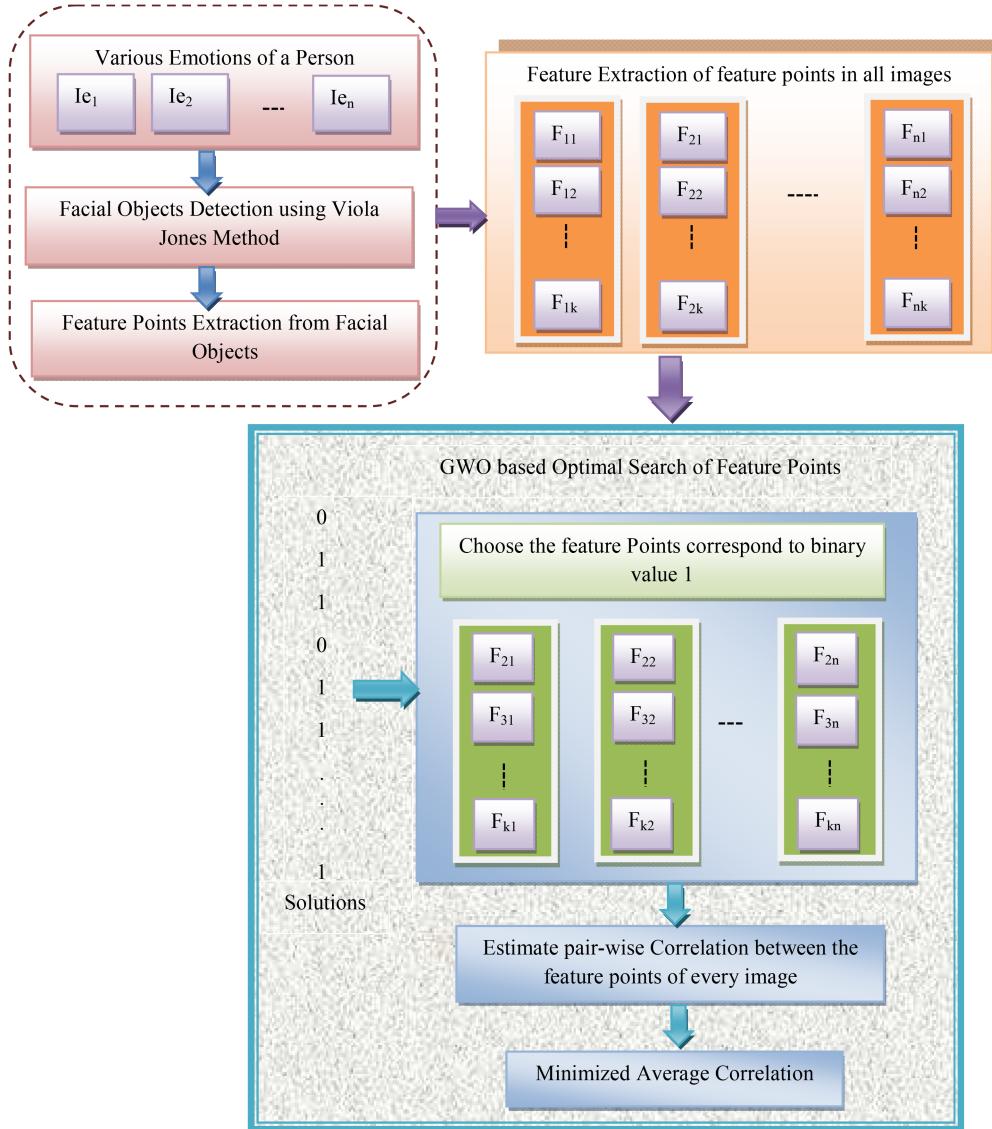
The proposed FER model exploits a novel feature selection technique using GWO algorithm from extracted SIFT features. Since there are numerous key points get extracted from SIFT technique, it is required to select the few key points optimally. Hence, GWO algorithm is used for optimally selecting the key points in such a way that the correlation between selected key points should be minimum. Moreover, it is certified that the same key points are selected for all emotions. If the correlation between key points is minimum, it is possible to distinguish various emotions accurately. The architecture used for the generation of SIFT and optimised SIFT features is shown in Fig. 2. Here, the steps employed for SIFT and optimised SIFT based on GWO are mentioned. The optimised SIFT generate the key points based on GWO principle, i.e. the position of wolves  $I_\alpha, I_\beta, I_\delta$  is found out and finally, the best keypoint is selected which is the best position of wolf. The objective function of the proposed method is the minimisation of correlation between the SIFT features that get optimised as shown in (4). The correlation between the two key points  $a$  and  $b$  is expressed in (5), where  $n$  indicates the number of point pairs

$$I(x, y) = \min (\text{correlation}) \quad (4)$$

$$\text{correlation} = \frac{n \sum ab - \sum a \sum b}{\sqrt{(n \sum a^2 - (\sum a)^2) - (n \sum b^2 - (\sum b)^2)}} \quad (5)$$

Fig. 3 shows the block diagram of the proposed correlation-based feature point selection process. Let  $I_e = \{I_{en}\}$  be the image set representing various facial emotions of a person, where  $n$  is the number of emotions to be recognised. As we consider seven facial emotions, the value  $n = 7$ . As shown in the figure, the facial objects from the  $n$  images are detected using the effective object detection method, Viola-Jones [28]. Then, the SIFT features are extracted, as described in Section 3.4, from the facial objects detected. The feature points extracted from each image is represented as,  $F_{nk}$ , where  $k$  is the dimension of the feature vector. Once the feature points are extracted, optimal features are selected using the GWO algorithm. The algorithm is initialised with a random set of solutions that takes the binary equivalent values. From each image, the feature points that correspond to the binary value 1 are selected by the algorithm. After the selection of optimal feature points, the correlation is estimated between each pair of feature points in every image. Here, the objective is to minimise the correlation, as it suggests that the similarity between the features of two images considered is less. In other words, the minimum correlation between the images indicates that the features considered are different and hence, more precise classification is possible. Thus, the feature points, whose average correlation is minimised, are selected finally.

GWO [29–32] is the recently introduced meta-heuristic algorithm which is operated based on the hunting behaviour of wolves. Here in this algorithm, the wolves are generally divided into three categories such as  $\alpha$ ,  $\beta$ , and  $\delta$ . Among these set of wolves,  $\alpha$  is assigned as the leader (best wolf) of the group that makes the decision related to hunting, sleeping time, sleeping place, waking time and so on. On the contrary,  $\beta$  and  $\delta$  are the second and the third best wolves, which help the best wolf to take decisions. The hunting behaviour of wolves is usually gone through three main phases that include (i) pursuing, tracking, and approaching the prey; (ii) following, enclosing, and harassing the prey till its completion of movement; and (iii) bout in the direction of prey. Accordingly, (6) and (7) depict the mathematical model associated with the encircling behaviour of the wolves towards the prey. In the subsequent equations,  $J$  and  $X$  indicate the coefficient



**Fig. 3** Correlation computation pattern for feature selection

vectors as shown in (8) and (9),  $\mathbf{I}$  indicates the position vector of the grey wolf,  $\mathbf{I}_p$  indicates the position vector of the prey and  $t$  refers to the current iteration

$$\mathbf{A} = |\mathbf{X} \cdot \mathbf{I}_p(t) - \mathbf{I}(t)| \quad (6)$$

$$\mathbf{I}(t+1) = \mathbf{I}_p(t) - \mathbf{J} \cdot \mathbf{A} \quad (7)$$

$$\mathbf{J} = 2\mathbf{a} \cdot \mathbf{r}_1 - \mathbf{a} \quad (8)$$

$$\mathbf{X} = 2\mathbf{r}_2 \quad (9)$$

Accordingly, the mathematical model for the hunting behaviour of wolves is expressed in (10)–(15). Finally, the updated position of the wolf that provides optimised feature selection is given in (16)

$$A_\alpha = |\mathbf{X}_1 \cdot \mathbf{I}_\alpha - \mathbf{I}| \quad (10)$$

$$A_\beta = |\mathbf{X}_2 \cdot \mathbf{I}_\beta - \mathbf{I}| \quad (11)$$

$$A_\delta = |\mathbf{X}_3 \cdot \mathbf{I}_\delta - \mathbf{I}| \quad (12)$$

$$I_1 = \mathbf{I}_\alpha - J_1 \cdot (A_\alpha) \quad (13)$$

$$I_2 = \mathbf{I}_\beta - J_2 \cdot (A_\beta) \quad (14)$$

$$I_3 = \mathbf{I}_\delta - J_3 \cdot (A_\delta) \quad (15)$$

$$\mathbf{I}(t+1) = \frac{\mathbf{I}_1 + \mathbf{I}_2 + \mathbf{I}_3}{3} \quad (16)$$

The pseudo code of the GWO-based feature extraction and classification for FER is given in Algorithm 1 (see Fig. 4).

The description of Algorithm 1 (Fig. 4) is illustrated as follows:

1. In the beginning, it needs to initialise the population of the grey wolves as  $I_i$  where  $i = 1, 2, \dots, k$ .
2. Then the components  $a$ ,  $\mathbf{J}$ , and  $\mathbf{X}$  are to be initialised.
3. Further, the fitness of each search agent is calculated.
4. The best search agent is assigned as  $I_\alpha$ , the second best search agent as  $I_\beta$ , and the third one as  $I_\delta$ .
5. For the particular iteration, the position of the current search agents is updated using (14).
6. Then the components  $a$ ,  $\mathbf{J}$ , and  $\mathbf{X}$  are to be updated.
7. To the next, it is essential to update the entire search agents  $I_\alpha$ ,  $I_\beta$ , and  $I_\delta$ .
8. The process is repeated until the generation of best position of the wolf.

#### 4.2 Grey Wolf optimisation-NN

The selected features by the GWO algorithm are given to the GWO-based NN. Here the weight is tuned both by the GWO and Levenberg –Marquardt (LM) method and check the error performance of each method. On comparing error from both LM

and GWO, GWO-NN tunes the weight with less training error and it is applied to the NN model, which is considered as the proposed classification model. The architecture of LM and GWO-NN model is shown in Fig. 5.

Thus, it adapts itself to respond accordingly for the appropriate classification. The LM algorithm [33] trains NN. Equation (17) shows the mathematical model of the GW-NN classification. In the respective equation,  $Z_m^{(k)}$  indicates the output from the  $m$ th node of the  $k$ th layer,  $M_I^{(k)}$  refers to the number of  $k$ th layer inputs,  $F_k(\cdot)$  defines the non-linear function of  $k$ th layer and  $b_m$  indicates the bias input applied to the  $k$ th layer

$$Z_m^{(k)} = F_k \left[ u_m^{(k)} b_m + \sum_{n=1}^{M_I^{(k)}} I_n^{(k)} v_{nm}^{(k)} \right], \quad 0 \leq k \leq M - 1 \quad (17)$$

Initialize the grey wolf population  $I_i$  where  $i = 1, 2, \dots, k$

Initialize  $a$ ,  $J$  and  $X$

Compute the fitness of each search agent

Set  $I_\alpha$  as the best search agent

Set  $I_\beta$  as the second best search agent

Set  $I_\delta$  as the third best search agent

While ( $t <$  total iterations)

For  $I_i$

    Update the position of the present search agent  
    using Eq. (16).

End for

    Update  $a$ ,  $J$  and  $X$

    Compute the fitness function for the entire search agents

    Update  $I_\alpha$ ,  $I_\beta$  and  $I_\delta$

$t = t + 1$

End while

Return  $I_\alpha$

**Fig. 4** Algorithm 1: GWO-based feature selection and classification for FER

The weight of each layer is given as  $w = [u; v]$ , where  $u$  and  $v$  represent the weight coefficients. The applied weight in (18), defined by assigning  $\hat{z} = Z_n^{(k)}$ ,  $n = 1; k = M - 1$ , is to be optimised by GWO algorithm. Equation (19) delivers the cardinality (length) of  $w$ . The minimisation of the objective function provides the required estimation process

$$w^* = \arg \min_{w=[u;v]} |z - \hat{z}| \quad (18)$$

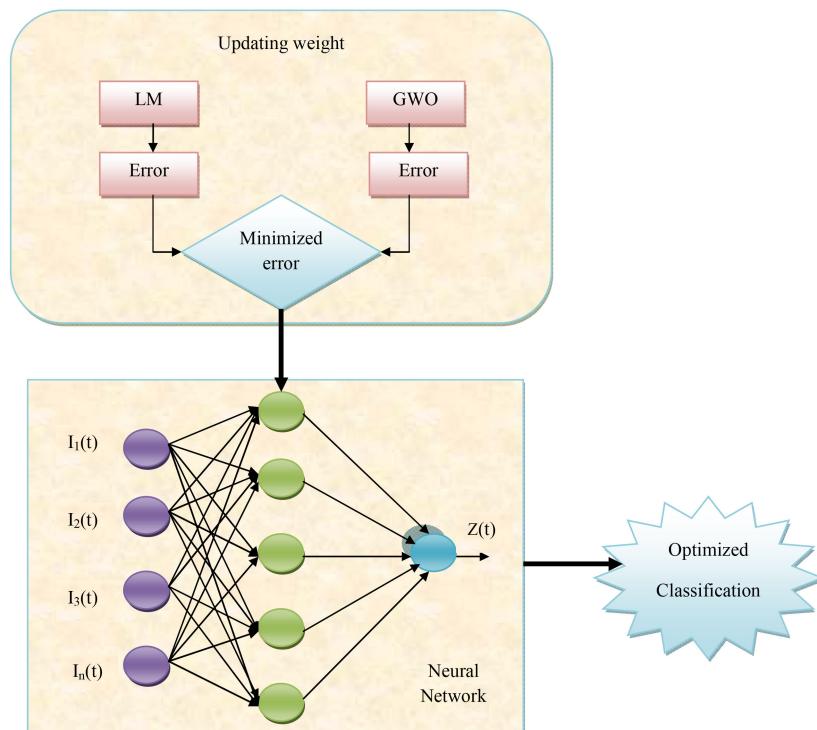
$$|w| = \sum_{k=1}^{M-1} (M_I^{(k-1)} \times M_I^{(k)}) + M_I^{(k)} \quad (19)$$

In (16), the function  $|\cdot|$  indicates the absolute error function. Thus, the optimal classification of face emotions is provided by the LM algorithm [33] with GWO algorithm [31].

## 5 Results and discussions

### 5.1 Experimental setup

The proposed method for FER is simulated in MATLAB. The experiment is carried out using two types of databases, such as JAFFE database and the Cohn–Kanade. Generally, JAFFE database and the Cohn–Kanade database are the popular databases for FER. In JAFFE database, 213 images are present with seven facial expressions and each image size is of  $256 \times 256$  pixels. On the contrary, the size of each image in Cohn–Kanade database used for FER is  $640 \times 490$  pixels. It comprises 486 image sequences with 97 poses. For the current experimentation, the proposed and the existing methods are analysed for with and without optimisation. Here, the performance measures are inspected based on the parameters, accuracy, sensitivity, specificity, precision, false positive rate (FPR), false negative rate (FNR), negative prediction value (NPV), false discovery rate (FDR), F1\_score, and Matthews correlation coefficient (MCC). Through the performance analysis, it compares the effectiveness of the proposed GWO-NN method with CNN [34], NN-LM [35], NN-gradient descent (GD) [36], NN-evolutionary algorithm (EA) [28], NN-firefly (FF) [37], NN-particle swarm optimization (PSO) [38], and NN-GWO [39]. The analysis of the experimental results is discussed in the subsequent sections. The sample images and its corresponding optimised SIFT



**Fig. 5** Architecture of LM and GWO-NN model

features of the Cohn–Kanade database and JAFFE database are shown in Figs. 6 and 7, respectively.

To calculate the accuracy, sensitivity, specificity, precision, FPR, FNR, NPV, FDR, F1-score, and MCC are presented in (20)–(29). Here, tp represents the true positive, tn represents the true negative, fp represents the false positive, and fn represents the false negative. The accuracy, specificity, sensitivity, precision, NPV, F1-score, and MCC should be increased, whereas the FPR, FNR, and FDR should be decreased

$$\text{accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}} \quad (20)$$

$$\text{sensitivity} = \frac{\text{tp}}{\text{tp} + \text{fn}} \quad (21)$$

$$\text{specificity} = \frac{\text{tn}}{\text{tn} + \text{fp}} \quad (22)$$

$$\text{precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \quad (23)$$

$$\text{NPV} = \frac{\text{tn}}{\text{tn} + \text{fn}} \quad (24)$$

$$\text{FNR} = \frac{\text{fn}}{\text{fn} + \text{tp}} \quad (25)$$

$$\text{FPR} = \frac{\text{fp}}{\text{fp} + \text{tn}} \quad (26)$$

$$\text{FDR} = \frac{\text{fp}}{\text{fp} + \text{tn}} \quad (27)$$

$$\text{F1-score} = \frac{2\text{tp}}{2\text{tp} + \text{fp} + \text{fn}} \quad (28)$$

$$\text{MCC} = \frac{\text{tp} \times \text{tn} - \text{fp} \times \text{fn}}{\sqrt{(\text{tp} + \text{fp})(\text{tp} + \text{fn})(\text{tn} + \text{fp})(\text{tn} + \text{fn})}} \quad (29)$$

## 5.2 Impacts of optimisation

The performance analysis of the proposed and the conventional methods for the condition with and without optimisation for the databases Cohn–Kanade is shown in Tables 1 and 2, respectively. From Table 1, the accuracy of the proposed method is 4.92% better than the NN-LM, 10.36% better than the NN-GD, 8.22% better than the NN-EA, 8.7% better than the NN-FF, and 3.47% better than the NN-PSO method. Similarly, the sensitivity and the specificity of the proposed method are 35 and 2.3% better than the NN-LM. The precision of the proposed method is 17% better than the NN-PSO method. The FPR, FNR, and FDR of the proposed method are 44, 46, and 50% better than the NN-FF method, respectively. Likewise, the NPV, F1-score, and MCC methods are 2.3, 31, and 42% better than the NN-LM method, respectively. Hence, Table 1 proves the recognition efficiency of the facial emotions by the proposed method than conventional methods.

The performance analysis of the proposed method and the conventional methods for the condition without optimisation for the JAFFE database is depicted in Table 2. Here, the accuracy of the proposed method is 5.2% better than the NN-LM method, which is proven by exhibiting higher values in the FPR, FNR, and FDR. Similarly, another three performance measures, such as sensitivity, specificity, and precision give 22%, 2 and 22% of higher values than the NN-PSO method. Hence, the proposed method outperforms all the existing methods by providing better results without optimisation.

Moreover, Table 3 depicts the accuracy of seven classes of facial emotions for JAFFE database. Here, the bold values specify the highest value of observation among various approaches. It is observed from Table 3 that the proposed NN-GWO method provides significantly higher performance in recognising the appropriate emotions when compared to the other existing

methods. It is observed from Table 3 that the proposed NN-GWO method provides significantly higher performance in recognising the appropriate emotions when compared to the other existing methods. However, in recognising some facial emotions such as angry and smile, the proposed methods attain lesser value than the existing methods but the overall performance of the proposed method is higher than the existing methods.

## 5.3 Comparative analysis

Table 4 demonstrates the performance of the proposed and the existing methods using the JAFFE database. From Table 5, it is clear that all the measures show better performance in the proposed NN-GWO method. Here, the accuracy of the proposed NN-GWO method is 13, 9.97, 18.08, 8.99, 12.62, 12.62% better than the CNN, NN-LM, NN-GD, NN-EA, NN-FF and NN-PSO, respectively, for the state with optimisation. Table 5 depicts the performance of the proposed method of seven classes of facial emotions for the JAFFE database. The proposed method yields promising results while using JAFFE database and even outperforms NN-LM by at least 20% for emotion (Angry). Besides, the proposed NN-GWO algorithm achieves nearly 5% higher recognition rate for the emotion Fear. The recognition rate of proposed method for emotion (normal and smile) is 20 and 8.7% better than the NN-GD method. In addition, Table 5 portrays the classification accuracy of the proposed GW-NN for JAFFE database and reveals that the accuracy of proposed method is higher than the existing methods.

## 5.4 Training performance

The examination of the training performance analysis at each period is given in Fig. 7. Here, the proposed GW-NN approach provides superior solution that is reliable over the conventional approaches.

## 5.5 Accuracy performance

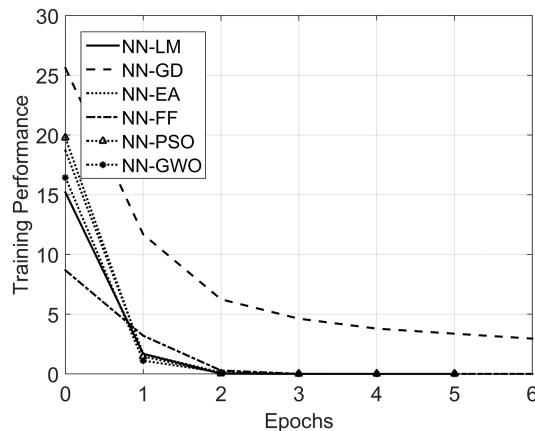
The performance analysis of the accuracy of proposed and the other conventional methods used in different papers are given in Table 6. The accuracy of proposed method of Cohn–Kanade database is 0.26% better than Sun and Akansu [40], 3.07% better than Yi *et al.* [41], 1.45% better than Sonmez and Albayrak [42] and 0.24% better than Ahmed [43]. The accuracy of the proposed method using JAFFE database is 89.1% better than Moeini *et al.* [44], 13% better than Wang *et al.* [45], 11% better than Yu *et al.* [46] and 9.4% better than Gaidhane and Singh [47]. By taking the average of accuracy obtained from specified emotions, the overall performance of the proposed method is justified.

## 6 Conclusion

The emotions of human have considered the actions, behaviours, poses, facial expressions, and speech to interact with other people. Since the facial emotions have been taken as the primary role for the communication purpose; several researchers have dealt their contributions on it. Accordingly, this paper was aimed to build up the FER system to differentiate the basic human expressions that include normal, smile, sad, surprise, anger, fear, and disgust. The steps of the FER system were composed of pre-processing, feature extraction, feature selection, and classification. SIFT-based feature extraction was used to capture the significant features from the facial point. Here, the optimal features were selected using the GWO algorithm. Further, NN-GWO was adopted to recognise the concerned emotions. Further, a beneficial performance analysis was carried out by comparing the proposed method with the existing methods, such as NN-LM, NN-GD, NN-EA, NN-FF, and NN-PSO. Collectively, the classification accuracy of the proposed GWO-NN is 91.22% for Cohn–Kanade database and 89.79% for JAFFE database. Thus, the proposed intelligent FER system has provided an effective performance over the conventional methods in terms of high accuracy.

Sample Images							
Extracted SIFT Features							
Emotions	Normal	Disgust	Fear	Angry	Smile	Surprise	Sad

**Fig. 6** Sample images and corresponding optimised SIFT features of Cohn–Kanade database



**Fig. 7** Training performance analysis on proposed and existing methods

**Table 1** Performance analysis of the proposed and existing methods with optimisation for Cohn–Kanade database

Methods	Accuracy	Sensitivity	Specificity	Precision	FPR	FNR	NPV	FDR	F1-score	MCC
CNN	0.89184	0.58571	0.94286	0.62381	0.057143	0.41429	0.94286	0.37619	0.6033	0.54175
NN-LM	0.8694	0.4857	0.9333	0.5433	0.0667	0.5143	0.9333	0.4567	0.5115	0.4384
NN-GD	0.8265	0.3571	0.9048	0.3714	0.09524	0.6429	0.9048	0.6286	0.3637	0.2639
NN-EA	0.8429	0.4143	0.9143	0.4452	0.0857	0.5857	0.9143	0.5548	0.4286	0.3384
NN-FF	0.8388	0.3571	0.9190	0.3905	0.08095	0.6429	0.9191	0.6096	0.3725	0.2836
NN-PSO	0.8816	0.5571	0.9357	0.5952	0.0643	0.4429	0.9357	0.4048	0.5747	0.5070
NN-GWO	0.9122	0.6571	0.9548	0.7000	0.0452	0.3429	0.9548	0.3000	0.6748	0.6267

**Table 2** Performance analysis of the proposed and existing methods without optimisation for Cohn–Kanade database

Methods	Accuracy	Sensitivity	Specificity	Precision	FPR	FNR	NPV	FDR	F1-score	MCC
CNN	0.87347	0.55714	0.92619	0.55714	0.07381	0.44286	0.92619	0.44286	0.55714	0.48333
NN-LM	0.853061	0.471429	0.916667	0.485714	0.083333	0.528571	0.916667	0.514286	0.478022	0.392916
NN-GD	0.826531	0.385714	0.900000	0.385714	0.100000	0.614286	0.900000	0.614286	0.385714	0.28475
NN-EA	0.895918	0.614286	0.942857	0.645238	0.057143	0.385714	0.942857	0.354762	0.628571	0.568873
NN-FF	0.881633	0.571429	0.933333	0.580952	0.066667	0.428571	0.933333	0.419048	0.575824	0.507333
NN-PSO	0.867347	0.514286	0.92619	0.530952	0.07381	0.485714	0.92619	0.469048	0.521978	0.445458
NN-GWO	0.897959	0.628571	0.942857	0.651429	0.057143	0.371429	0.942857	0.348571	0.638095	0.579972

**Table 3** Classification accuracy of different emotions for Cohn–Kanade database

Methods	Normal	Smile	Sad	Surprise	Angry	Fear	Disgust
CNN	88.571	90	90	90	88.571	85.714	91.429
NN-LM	87.14286	85.71429	82.85714	87.14286	<b>91.42857</b>	85.71429	88.57143
NN-GD	82.85714	87.14286	74.28571	85.71429	78.57143	85.71429	84.28571
NN-EA	87.14286	81.42857	88.57143	80.0000	84.28571	84.28571	84.28571
NN-FF	88.57143	90.0000	81.42857	78.57143	84.28571	82.85714	81.42857
NN-PSO	88.57143	<b>95.71429</b>	84.28571	81.42857	90.0000	88.57143	88.57143
NN-GWO	<b>94.28571</b>	92.85714	<b>94.28571</b>	<b>90.0000</b>	84.28571	<b>90.0000</b>	<b>92.85714</b>

The bold values indicate the highest value of observation among various approaches.

**Table 4** Performance of the proposed and existing methods with optimisation on JAFFE database

Methods	Accuracy	Sensitivity	Specificity	Precision	FPR	FNR	NPV	FDR	F1-score	MCC
CNN	0.84082	0.4	0.91429	0.43333	0.085714	0.6	0.91429	0.56667	0.4141	0.32382
NN-LM	0.867347	0.528571	0.92381	0.528571	0.07619	0.471429	0.92381	0.471429	0.528571	0.451417
NN-GD	0.808163	0.242857	0.902381	0.278571	0.097619	0.757143	0.902381	0.721429	0.256264	0.150421
NN-EA	0.87551	0.5	0.938095	0.560714	0.061905	0.5	0.938095	0.439286	0.524875	0.457306
NN-FF	0.846939	0.414286	0.919048	0.442857	0.080952	0.585714	0.919048	0.557143	0.427473	0.340224
NN-PSO	0.846939	0.442857	0.914286	0.461905	0.085714	0.557143	0.914286	0.538095	0.451648	0.363249
NN-GWO	0.953878	0.631429	1.007619	0.68381	0.002381	0.378571	1.007619	0.32619	0.654322	0.595797

**Table 5** Accuracy rate of different emotion on JAFFE database

Methods	Angry	Disgust	Fear	Normal	Smile	Surprise	Sad
CNN	82.857	81.429	84.286	84.286	80	85.714	90
NN-LM	75.71429	81.42857	81.42857	82.85714	85.71429	85.71429	84.28571
NN-GD	81.42857	85.71429	80.00000	75.71429	81.42857	82.85714	82.85714
NN-EA	85.71429	72.85714	82.85714	80.00000	85.71429	90.00000	92.85714
NN-FF	82.85714	74.28571	81.42857	80.00000	78.57143	78.57143	88.57143
NN-PSO	88.57143	<b>90.0000</b>	84.28571	85.71429	84.28571	<b>85.71429</b>	<b>94.28571</b>
NN-GWO	<b>91.42857</b>	84.28571	<b>88.57143</b>	<b>91.42857</b>	<b>88.57143</b>	82.85714	88.57143

**Table 6** Performance comparison of accuracy of proposed with existing methods

Author	Accuracy, %
Cohn–Kande database	
Sun and Akansu [40]	90.98
Yi <i>et al.</i> [41]	88.5
Sonmez and Albayrak [42]	89.92
Ahmed [43]	91
NN-GWO	91.22
JAFFE database	
Moeini <i>et al.</i> [44]	89.1
Wang <i>et al.</i> [45]	84.3
Yu <i>et al.</i> [46]	85.7
Gaidhane and Singh [47]	94.5
NN-GWO	95.39

## 7 References

- [1] Zwick, J.C., Wolkenstein, L.: ‘Facial emotion recognition, theory of mind and the role of facial mimicry in depression’, *J. Affective Disorders*, 2017, **210**, pp. 90–99
- [2] Lenc, L., Kral, P.: ‘Automatic face recognition system based on the SIFT features’, *Comput. Electr. Eng.*, 2015, **46**, pp. 256–272
- [3] Fasel, B.: ‘Robust face analysis using convolutional neural networks’. Object Recognition Supported by User Interaction for Service Robots, Switzerland, 2002, vol. 2, pp. 40–43
- [4] Theurel, A., Witt, A., Malsert, J., *et al.*: ‘The integration of visual context information in facial emotion recognition in 5- to 15-year-olds’, *J. Exper. Child Psychol.*, 2016, **150**, pp. 252–271
- [5] Brak, L.B., Abby, L., Richman, D.M., *et al.*: ‘Facial emotion recognition among typically developing young children: a psychometric validation of a subset of NimStim stimuli’, *Psychiatry Res.*, 2017, **249**, pp. 109–114
- [6] Cruz, A.C., Bhanu, B., Thakoor, N.S.: ‘Vision and attention theory based sampling for continuous facial emotion recognition’, *IEEE Trans. Affective Comput.*, 2014, **5**, (4), pp. 418–431
- [7] Matamoros, A.H., Bonarini, A., Hernandez, E.E., *et al.*: ‘Facial expression recognition with automatic segmentation of face regions using a fuzzy based classification approach’, *Knowl.-Based Syst.*, 2016, **110**, pp. 1–14
- [8] Balas, B., Huynh, C., Saville, A., *et al.*: ‘Orientation biases for facial emotion recognition during childhood and adulthood’, *J. Exper. Child Psychol.*, 2015, **140**, pp. 171–183
- [9] Lee, S.Y., Bang, M., Kim, K.R., *et al.*: ‘Impaired facial emotion recognition in individuals at ultra-high risk for psychosis and with first-episode schizophrenia, and their associations with neurocognitive deficits and self-reported schizotypy’, *Schizophrenia Res.*, 2015, **165**, (1), pp. 60–65
- [10] Neoh, S.C., Zhang, L., Mistry, K., *et al.*: ‘Intelligent facial emotion recognition using a layered encoding cascade optimization model’, *Appl. Soft Comput.*, 2015, **34**, pp. 72–93
- [11] Prado, C.E., Treeby, M.S., Crowe, S.F.: ‘Examining relationships between facial emotion recognition, self-control, and psychopathic traits in a non-clinical sample’, *Personality Individual Differ.*, 2015, **80**, pp. 22–27
- [12] Moeini, A., Faez, K., Sadeghi, H., *et al.*: ‘2D facial expression recognition via 3D reconstruction and feature fusion’, *J. Vis. Commun. Image Represent.*, 2016, **35**, pp. 1–14
- [13] Meng, H., Berthouze, N.B., Deng, Y., *et al.*: ‘Time-delay neural network for continuous emotional dimension prediction from facial expression sequences’, *IEEE Trans. Cybern.*, 2016, **46**, (4), pp. 916–929
- [14] Russo, M., Mahon, K., Shanahan, M., *et al.*: ‘The association between childhood trauma and facial emotion recognition in adults with bipolar disorder’, *Psychiatry Res.*, 2015, **229**, (3), pp. 771–776
- [15] Fan, X., Tjahjadi, T.: ‘A dynamic framework based on local zernike moment and motion history image for facial expression recognition’, *Pattern Recognit.*, 2017, **64**, pp. 399–406
- [16] Zhou, Q., Shafiq, R., Zhou, Y., *et al.*: ‘Face recognition using dense SIFT feature alignment’, *Chin. J. Electron.*, 2016, **25**, (6), pp. 1034–1039
- [17] Gola, K.A., Shany-Ur, T., Pressman, P., *et al.*: ‘A neural network underlying intentional emotional facial expression in neurodegenerative disease’, *NeuroImage: Clin.*, 2017, **14**, pp. 672–678
- [18] Alqahtani, M.M.J.: ‘An investigation of emotional deficit and facial emotion recognition in traumatic brain injury: a neuropsychological study’, *Postepy Psychiatrii I Neurologii*, 2015, **24**, (4), pp. 217–224
- [19] Rigon, A., Voss, M.W., Turkstra, L.S., *et al.*: ‘Relationship between individual differences in functional connectivity and facial-emotion recognition abilities in adults with traumatic brain injury’, *NeuroImage: Clin.*, 2017, **13**, pp. 370–377
- [20] Tang, X.W., Yu, M., Duan, W.W., *et al.*: ‘Facial emotion recognition and alexithymia in Chinese male patients with deficit schizophrenia’, *Psychiatry Res.*, 2016, **246**, pp. 353–359
- [21] Zhang, L., Mistry, K., Neoh, S.C., *et al.*: ‘Intelligent facial emotion recognition using moth-firefly optimization’, *Knowl. Based Syst.*, 2016, **111**, pp. 248–267
- [22] Hargreaves, A., Mothersill, O., Anderson, M., *et al.*: ‘Detecting facial emotion recognition deficits in schizophrenia using dynamic stimuli of varying intensities’, *Neurosci. Lett.*, 2016, **633**, pp. 47–54
- [23] Zhang, T., Zheng, W., Cui, Z., *et al.*: ‘A deep neural network-driven feature learning method for multi-view facial expression recognition’, *IEEE Trans. Multimed.*, 2016, **18**, (12), pp. 2528–2536
- [24] Lopes, A.T., Aguiar, E., Souza, A.F.D., *et al.*: ‘Facial expression recognition with convolutional neural networks: coping with few data and the training sample order’, *Pattern Recognit.*, 2017, **61**, pp. 610–628
- [25] Vinay, A., Kathiresan, G., Mundroy, D.A., *et al.*: ‘Face recognition using filtered EOH-sift’, *Procedia Comput. Sci.*, 2016, **79**, pp. 543–552

- [26] Pu, X., Fan, K., Chen, X., *et al.*: ‘Facial expression recognition from image sequences using twofold random forest classifier’, *Neurocomputing*, 2015, **168**, pp. 1173–1180
- [27] Azeem, A., Sharif, M., Shah, J.H., *et al.*: ‘Hexagonal scale invariant feature transform (H-SIFT) for facial feature extraction’, *J. Appl. Res. Technol.*, 2015, **13**, (3), pp. 402–408
- [28] Boutorh, A., Guessoum, A.: ‘Complex diseases SNP selection and classification by hybrid association rule mining and artificial neural network – based evolutionary algorithms’, *Eng. Appl. Artif. Intell.*, 2016, **51**, pp. 58–70
- [29] Mirjalili, S., Saremi, S., Mirjalili, S.M., *et al.*: ‘Multi-objective grey wolf optimizer: novel algorithm for multi-criterion optimization’, *Expert Syst. Appl.*, 2016, **47**, pp. 106–119
- [30] Song, X., Tang, L., Zhao, S., *et al.*: ‘Grey Wolf optimizer for parameter estimation in surface waves’, *Soil Dyn. Earthq. Eng.*, 2015, **75**, pp. 147–157
- [31] Mirjalili, S., Mirjalili, S.M., Lewis, A.: ‘Grey Wolf optimizer’, *Adv. Eng. Softw.*, 2014, **69**, pp. 46–61
- [32] Komaki, G.M., Kayvanfar, V.: ‘Grey Wolf optimizer algorithm for the two-stage assembly flow shop scheduling problem with release time’, *J. Comput. Sci.*, 2015, **8**, pp. 109–120
- [33] Celika, O., Tekkeb, A., Yildirima, H.B.: ‘The optimized artificial neural network model with Levenberg–Marquardt algorithm for global solar radiation estimation in Eastern Mediterranean region of Turkey’, *J. Clean Prod.*, 2016, **116**, pp. 1–12
- [34] Yang, M., Liu, Y., You, Z.: ‘The Euclidean embedding learning based on convolutional neural network for stereo matching’, *Neurocomputing*, 2017, **267**, pp. 195–200
- [35] Jian, J., Jun, L., Hua, Z.X., *et al.*: ‘Inversion of neural network Rayleigh wave dispersion based on LM algorithm’, *Procedia Eng.*, 2011, **15**, pp. 5126–5132
- [36] Kobayashi, M.: ‘Gradient descent learning for quaternionic Hopfield neural networks’, *Neurocomputing*, 2017, **260**, pp. 174–179
- [37] Sánchez, D., Melin, P., Castillo, O.: ‘Optimization of modular granular neural networks using a firefly algorithm for human recognition’, *Eng. Appl. Artif. Intell.*, 2017, **64**, pp. 172–186
- [38] Lin, C.H.: ‘Novel application of continuously variable transmission system using composite recurrent Laguerre orthogonal polynomials modified PSO NN control system’, *ISA Trans.*, 2016, **64**, pp. 405–417
- [39] Ozturk, S., Akdemir, B.: ‘Automatic leaf segmentation using grey wolf optimizer based neural network’. *Electronics. Palanga*, 2017, pp. 1–6
- [40] Sun, Y., Akansu, A.N.: ‘Facial expression recognition with regional hidden Markov models’, *Electron. Lett.*, 2014, **50**, (9), pp. 671–673
- [41] Yi, J., Mao, X., Chen, L., *et al.*: ‘Facial expression recognition considering individual differences in facial structure and texture’, *IET Comput. Vis.*, 2014, **8**, (5), pp. 429–440
- [42] Sonmez, E.B., Albayrak, S.: ‘Critical parameters of the sparse representation-based classifier’, *IET Comput. Vis.*, 2013, **7**, (6), pp. 500–507
- [43] Ahmed, F.: ‘Gradient directional pattern: a robust feature descriptor for facial expression recognition’, *Electron. Lett.*, 2012, **48**, (19), pp. 1203–1204
- [44] Moeini, A., Faez, K., Moeini, H., *et al.*: ‘Facial expression recognition using dual dictionary learning’, *J. Vis. Communun. Image Represent.*, 2017, **45**, pp. 20–33
- [45] Wang, S., Liu, Z., Wang, J., *et al.*: ‘Exploiting multi-expression dependences for implicit multi-emotion video tagging’, *Image Vis. Comput.*, 2014, **32**, (10), pp. 682–691
- [46] Yu, K., Wang, Z., Zhuo, L., *et al.*: ‘Learning realistic facial expressions from web images’, *Pattern Recognit.*, 2013, **46**, (8), pp. 2144–2155
- [47] Gaidhane, V.H., Singh, Y.V.H.V.: ‘Emotion recognition using eigenvalues and Levenberg–Marquardt algorithm-based classifier’, *Sadhana*, 2016, **41**, (4), pp. 415–423