Check for updates

Optimized face-emotion learning using convolutional neural network and binary whale optimization

T. Muthamilselvan[1] · K. Brindha[1] · Sudha Senthilkumar[2] · Saransh[3] ·
Jyotir Moy Chatterjee[4] ⬤ · Yu-Chen Hu[5]

Abstract

Human emotion detection using facial expressions might be easy for humans, but computing technology to accomplish the same task is more challenging. We can recognize emotions from images using the latest computer vision and machine learning (ML) advancements. This research proposes a novel optimized face emotion learning method with binary whale optimization (OFELBW). The OFELBW is implemented in three phases, the first phase with a convolutional neural network (CNN) in which from the image the background noise is removed in the initial phase, and the facial feature extraction is performed in the second phase. Finally, the binary whale optimization algorithm is used for the feature selection to obtain the most relevant feature subset. The proposed OFELBW method was examined with more than 750 K images using SFEW, CK+, JAFFE, and FERG datasets. We have compared our proposed OFELBW model with other existing techniques to examine the accuracy of our models with the above-mentioned datasets and received an accuracy of 98.35% with the CK+ dataset, 99.42% with the FERG dataset, 96.6% with the JAFFE dataset and 64.98% with the SFEW with 80% training, 10% testing, and 10% validation set. This technique will be useful in various applications such as human social/physiological interaction systems, mental disease diagnosis and military environment, etc.

1 Introduction

Emotions accompany all interpersonal communication. Various forms of emotional expressions may be observable to the naked eye, but some may not be visible to the naked eye. Therefore, any indications before or after them can be detected and recognized with the right tools [24, 34]. Over the past few years, there has been increased interest in understanding a person's emotions.

✉ Jyotir Moy Chatterjee
  jyotirchatterjee@gmail.com

Extended author information available on the last page of the article

Many fields, including human-computer interfaces, animation, medicine, and security, have been interested in human emotion recognition [45, 50]. It has been proposed to use facial expressions as a feature for emotion recognition for various reasons; As a result, they are obvious, contain many useful features, and are easier to collect as a database of faces [2, 4].

An adequate facial expression recognition (FER) model can be obtained using deep learning (DL) and especially CNN [30]. In the case of facial expressions, however, a great deal of the traces comes from only some parts of the face, especially the eyes and mouth, whereas other parts are less influential, such as the ears and hair. Therefore, it makes sense for ML frameworks to focus only on the primary parts of the face and not pay much attention to other parts of the face [20, 28, 35, 51].

Fast and reliable performance in the wild is one of the major challenges in FER [33]. Humans can express an enormous variety of emotions besides the six basic ones of anger, sad, happy, disgusted, and surprised, which have received the most attention from researchers. Variations in head posture, partial occlusion, lighting conditions, identifying subjects, and distortions in the camera lens create major obstacles. Haar Cascade classifier is used to identify people's faces in images or videos. Then the convolutional neural network along with the decision tree used for classification and with the binary whale optimization produces better accuracy for predicting facial emotion in a fast and reliable manner [22].

The proposed work in this article involves novel approaches using the binary whale optimization method for recognizing facial expressions and facial landmark (FL) detection with the help of the dlib library. We proved that the convolutional neural network and the decision tree used for classification and binary whale optimization produce better accuracy than the existing approaches. The main contribution of the work is as follows:

- The primary objective of this work is to use a DL technique and a random forest algorithm. For better accuracy, binary whale optimization techniques are used.
- By utilizing CNN, we propose the OFELBW approach, which will pinpoint the features of a face with maximum accuracy.
- The proposed OFELBW method will visualize the critical regions in facial expression classification.

The remaining portion of the article is structured as follows: Section 2 explains the various works done in this area. Section 3 illuminates the various materials and methods used in this work like the Haar cascade classifier, CNN, whale optimizations algorithm, etc. Section 4 explains the proposed work in detail. Section 5 presents the research results with detailed discussions. Section 6 concludes the paper with possible future work.

2 Related work

The authors in [24] have shown DL alignments, which are vigorous face-calibration methods that are based on CNNs. According to their research, a deep alignment network (DAN) does face alignments mostly based on whole-face images as opposed to the face alignments performed by recent face alignment techniques. Because of this, it is very accurate to tremendous fluctuations in initializations and forehead poses. The use of heatmaps that have landmarks and that transmit the locations of landmarks among phases of the DAN, allowed them to use face images rather than locally available marks that are extracted around these

landmarks. Two challenging tasks were conducted with extensive performance evaluation resulting in a relative failure rate improvement of more than 70%.

Based on [34], the authors established an "Affective Computing" system that is aimed at developing systems, devices, and mechanisms that are recognizable, interpretable, and that imitate a person's effects through various attributes such as how he/she looks, the depth and modulation of his/her voice, and any biological signals he/she may possess. To shed light on emotional facial expressions, they have described several network architecture-driven models: 1) A direct measure is where the emotion gets picked up from a category related to emotions, such as FER datasets, which contain six basic human emotions. 2) Based on a simultaneous scale of valence and arousal in images, a numerical value is obtained from the extent of facial expression. In [50], the authors have shown the facial expression recognition system, which is a real-world application and solves the phases that occurred post changes are made. The authors have generated several new tests over FER datasets on these phases, and this model a new "Region Attention Network (RAN)" which itself depicts the essence of the facial landmarks. They further showed the implementation of a "Region Biased loss (RB-Loss)" function that is used to strengthen the high attention weight for regions that are the most salient. Additionally, the authors evaluated their method for collecting their data and conducted studies on FER-Plus and Affect-Net.

In [45], the authors have made their outlook on an effort-in-progress technique for facial expression recognition which enables the system to get much from the facial landmarks. The findings are figured on the JAFFE dataset, which suggested some signs of a place for development and more precision. The authors have made their overview, saying that the proposed method has strong potential to outperform the currently published methods. In [17], the authors propose a CNN technique which is a 3-Dimensional for FER in frames of videos. This model develops 3D Inception-ResNet layers followed by a unit called long short-term memory (LSTM) that simultaneously grasps the relations of space within images of faces and the temporal instances among different frames of the video. Facial curve dots are also used as samples for their network design which focus on the instances of facial landmarks rather than some noted facial patches that won't be beneficial and may not be able to generate facial expressions significantly.

In [2], there is research conducted by the author to categorize facial emotions over static facial pictures with the aid of DL techniques. The achieved results were non-futuristic and slightly better than other methods, including the characteristics engineering. Eventually, DL systems will be able to remove this problem, given an ample amount of labeled tuples. Characteristics engineering is not that essential, image pre-processing reduces the inconsistencies of the classification. That's why it increases the visibility and the quality of the input image. Today facial emotion detection software includes the use of characteristics engineering [37, 53, 57]. A finding that depends on the characteristic learning that does not seem near yet because of the major restraint shows the absence of a wide-ranging dataset of reactions. With the presence of a bigger dataset, systems with a larger ability that is used to learn structures could be applied [7–9, 26]. Thus, emotion classification could be attained with the help of DL approaches. With the help of the ML approach, authors in [11, 27] have tested the recognition on a set of 39 different Hindi hollow character classes and where some characters are distorted as well as multi-scaled and they have received good performance for the recognition of hollow characters that have different rotations and scales.

In [4, 12–14], the authors have proposed an architecture where CNN is trained to classify facial emotions/expressions. The authors have used Japanese Female Facial Expression (JAFFE) datasets of facial emotion images for training CNN to achieve good accuracy during the training phase. The concept of hybrid vehicles employing CNN has been used for detecting

drowsiness or alertness of the drivers in real-time [48]. In [30], the author has proposed a system of programmed facial expression recognition to detect and locate face landmarks in a muddled scene, a set of facial movement extraction, and facial emotions classification [18, 23]. This model is developed using CNN, which is dependent based on a network design called "Le-Net", Kaggle facial expression (FER2013) dataset with seven facial expression class labels which include happy, sad, surprise, disgust, fear, anger & neutral [21, 25, 47].

In [3, 31], the authors have developed a new design unit called the "Squeeze and Excitation (SE)" block, which is designed to manipulate channels channel-wise so that channel-wise features can be set. This paper has shown that chunks of patches can be loaded together to form SE-Net architecture to generalize extremely effectively across different datasets. "Squeeze and Excitation" Networks have included the basics of ILSRVC assortment submission. In [42–44, 46], the authors have provided a complete survey on a design that is deep "Facial Expression Recognition" which includes databases and algorithms that feature a selection of data acceptance and evolution designs for these sets of data. The authors have reviewed some already constructed deep neural network (DNN) models and related training modules designed for "Facial Expression Recognition 2013" based on sequential images, which are static and dynamic as well [19, 46, 49]. In [39, 55, 56], another Image super-resolution utilizing the Face Emotion Recognition approach has been presented.

Hence, in this work, we have used CNN with binary whale optimization using diverse datasets and the Haar Cascade classifier to overcome the limitations of existing methods and to find any facial emotion much faster in the given image or video. To remove irrelevant features and select the most appropriate features from the entire feature set and visualize the important region in the facial expression whale optimization approach is used. Thus, the proposed OFELBW method recognizes the expression of humans in real-time more efficiently and faster. Thus, the proposed OFELBW method recognizes the expression of humans in real-time more efficiently with higher accuracy by using CNN, Binary Whale Optimization, and Haar cascade classifier.

The limitation of the proposed system is the lack of identifying the expression of infants. Emotions are expressed differently by infants and children than by adults. In their facial expressions, they convey more than they can express verbally. Emotions are also not restrained in children. Table 1 depicts the comparison of various existing techniques.

3 Materials and methods

The following subsections describe the related methods used in the proposed optimized face emotion learning with the binary whale optimization (OFELBW) technique.

3.1 Haar Cascade classifier

It is a face detection approach used to identify people's faces in images or videos. It works with Haar features such as eyes, full body, upper and lower body, and frontal faces. It is calculated by summing the pixel intensities over many image regions and then computing the difference between these sums. The resulting feature map can be used to identify patterns in images by down-sampling them. The algorithm is proposed by Paul Viola and Michael Jones [41]. An ML technique is applied to this classifier for discovering items in additional photos using a cascade operation. Face detection and facial expressions can also be detected in an image. A

Table 1 Comparisons of the existing techniques

References	Method	Metrics	Limitations
Kowalski et al. [24]	DLN uses CNN for robust face alignment	Two challenging tasks were conducted with extensive performance evaluation resulting in a relative failure rate improvement of more than 70%.	Only the facial alignment method is discussed based on convolutional neural networks. Human facial expression reorganization is not performed and compared.
Mollahosseini et al. [34]	Facial emotion, reactivity, and activation computing in the wild database called AffectNet.	The attributes such as the person's appearance, the modulation and depth of the person's voice, and biological signals the person holds are used for designing affective computing.	The accuracy is predicted based on only static images further can be improved with video sequences.
Wang et al. [50]	Facial expression recognition is based on region attention networks.	The authors have generated several new tests over FER datasets on these phases and the model "RAN" depicts the importance of the facial landmarks.	The execution time is not a linear number of regions.
Tautkute et al. [45]	The model uses emotion recognition with facial landmarks	The authors have made their outlook on an effort-in-progress technique for identifying facial emotion, which enables the system to get much from the facial landmarks.	This work can be enhanced by the attention mechanism on facial landmarks.
Hasani et al. [17]	Deep 3D technique used for facial emotion recognition	This model develops 3D Inception-ResNet layers followed by a unit called LSTM that simultaneously grasps the relations of space within images of faces and the temporal instances among different frames of the video.	This work does not show improvement in the MMI database.
Spiers et al. [2]	Facial emotion detection using DL	The achieved results were non-futuristic and slightly better than other methods, including the characteristics engineering. Eventually, DL systems will be able to remove this problem, given an ample amount of labeled tuples.	This work yields the lowest classification accuracy towards the series combination of deep learning architectures.
Sivo et al. [4]	Facial Expression Detection using Different CNN Architecture Hybrid Vehicle Driving	The authors used the JAFFE dataset of facial emotion images for training CNN to achieve good accuracy during the training phase. The concept of Hybrid Vehicle Employing CNN has been used for detecting drowsiness or alertness of the drivers in real time.	Trained the model only with the JAFFE dataset. Using diverse datasets helps to get higher accuracy.
Lopes et al. [30]	Convolutional networks are used to recognize facial expressions.	This model is developed using CNN dependent on a network design called "Le-Net",	Could be implemented with more CNN and filter layers

Table 1 (continued)

References	Method	Metrics	Limitations
		Kaggle facial expression (FER2013) dataset with seven facial expression class labels, including happy, sad, disgust, disgust, fear, anger & neutral.	to get more accuracy with different datasets.
Hu et al. [3]	Squeeze-and-excitation networks	This paper has shown that chunks of patches can be loaded together to form SE-Net architecture to generalize extremely effectively across different datasets. "ILSRVC classification submissions are based on "Squeeze and Excitation" networks.	This model would have tested with the variance of datasets to prove its geniality.
Li et al. [46]	Deep facial expression recognition	A DLP – CNN called the deep locality-preserving CNN (DLP-CNN) model proposed to improve the discriminative deep features by protecting the locality closeness.	It is tested with only two archive datasets with static images; further study can be elaborated with more datasets including dynamic images.

positive and a negative picture are presented to the classifier at the end of the exercise. From the picture, characteristics can be drawn out. The individual characteristic values are acquired by subtracting the sum of pixels in white rectangles from the sum of pixels in black rectangles. The program detects faces in different environments of different individuals. Haar pixel value is calculated as follows using Eq. 1.

$$p\_v = (\text{sum of } d\_p \text{ value}/nd\_p) - (sum\ of\ l\_p\ value/nl\_p) \qquad (1)$$

in which p_v denotes pixel value, d_p denotes the dark pixels, nd_p denotes the total number of dark pixels, l_p denotes the light pixels and nl_p denotes the total number of light pixels.

3.2 Convolutional neural network (CNN)

CNN/Conv-Net is an algorithm of deep learning [28, 37]. An input image is fed for the algorithm to assign learnable weights and biases and try to find the importance of various characteristics in the picture. These networks help to differentiate each character from one another. The important feature of CNN is that pre-processing needed in this is much lower than in other algorithms (classification). The network neurons architecture in CNN is somewhat like patterns that human brain cell has while connecting [21, 47]. The receptive field is the visual field of the restricted region where single neurons respond to stimuli. The whole area (visual) is covered with a collection of such fields which overlap. Figure 1 shows how an input image of a facial emotion fed to the Conv-Nets goes through pooling layers [18, 40, 48]. An input layer contains only one feature map, which is used to feed the normalized face image to a CNN model. The C1 layer includes six feature maps, each of which is convoluted with a 5 × 5 random kernel. First, layer S1 calculates six feature maps from the output of layer C1. The mean convolution kernel connects a feature map to its corresponding feature map in layer C1. Therefore, a feature map and a feature map in the C1 layer will not overlap each other. In C2
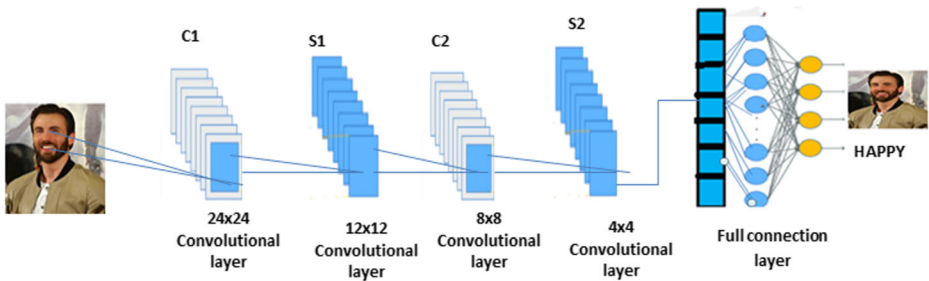
Fig. 1 CNN architecture

and S2, the second pooling and convolutional layer, the same feature maps are used, followed by the same calculations. In addition, the output layer is connected to the S2 layer with a fully connected perceptron. The final product is a 40-dimensional vector representing the classification of 40 individuals using sigmoid functions. In this module, Keras is used for pre-processing, and TensorFlow is used to enhance the model using the DNN algorithm. It goes through several iterations or epochs to train the model and test it. [1, 6, 15].

3.3 Whale optimization algorithm (WOA)

In WOA, a predation simulator is based on swarm intelligence optimization (SIO). The algorithm emulates the bubble net foraging style of whales [36]. Whales use bubble nets to catch their prey along spiral paths. Along the spiral path, they create bubble nets and move upstream to catch prey. There are 3 stages in this approach: surrounding prey, hunting prey, and attacking with bubble nets.

3.3.1 Surrounding prey

Although whales identify the position of their quarry first, they do not know where it is in advance. The current optimal position must be used as the target prey for the others to move to the optimal location. Mathematically, the enclosed stage may be expressed as follows:

$$X(j+1) = X^*(j) - B.E \tag{2}$$

In which $E = \mid T. X^*(j) - X(j) \mid$, j is the present iterations number, $X^*(j)$ is the prey location vector (current optimal solution), $X(j)$ is the prey location vector, and $B.E$ is the size of the surrounding steps.

$$B = 2v. \, rand - v$$
$$T = 2. \, rand \tag{3}$$

With increasing iteration number, v diminishes linearly from 2 to 0, and rand denotes the random number [0,1]. The final expression is as follows:

$$V = 2 - \frac{2j}{J_{max}} \tag{4}$$

In which the number of iterations is mentioned using the J_{max} variable.

3.3.2 Bubble net attack

During bubble net foraging, humpback whales move around their prey in a constricted encirclement on a spiral path. Whale predator behavior is described by two methods in WOA: shrinking and surrounding mechanisms, as well as spiral updates. In Eqs. (2) and (3), the convergence factor is reduced. The whale's position in a spiral will be updated by first measuring its distance from the optimal position, and then simulating its capture. The mathematical Eq. 5 is as follows:

$$X(j+1) = \acute{E}. \, e^{bl}.\cos(2\pi l) + X^*(j)$$
$$\acute{E} = \left| X^*(j) - X(j) \right| \tag{5}$$

One can choose l as the random number among $(-1, 1)$. In the logarithmic spiral form, the constant coefficient $\acute{E}$ is the distance between the j-th whale and its current optimal position.

3.3.3 Hunting prey

We can evolve the algorithm to replace the recent optimal solution randomly while searching for a better prey to replace the current whale reference; we can also enhance the global exploration capability by replacing the whale reference with a better prey. Mathematically, the model denotes as below:

$$X(j+1) = X_{rand} - B.E$$
$$E = |T.X_{rand} - X(j)| \tag{6}$$

Randomly selecting the location of the whale is indicated by X_{rand}.

3.4 Binary whale optimization (BWO)

The classical WOA determines that whales move in a continuous search area to reconfigure their positions, known as a continuous space. The only way to resolve the selection of features is with {0,1} values. Solving feature selection problems requires converting continuous (free position) solutions into their binary counterparts. The binary whale optimization algorithm with the S-shaped (BWOA-S) variant of BWOA is used to search over the adaptive feature space to select the best features. The features with a value of 1 are considered appropriate, while those with a value of 0 are deemed inappropriate. Using this algorithm, the higher accuracy with a smaller number of chosen features. Equation 7 denotes the fitness function F utilized in BWOA-S to assess the single positions of the whale.

$$F = \alpha \gamma_Z(D) + \beta \frac{|N-Z|}{|N|} \tag{7}$$

In which chosen feature subset length is denoted as Z, the complete number of features is mentioned as N. $\gamma_Z(D)$ is the condition attribute set R relative to decision D of classification accuracy. 2 arguments which are symmetric to the subset length and its accuracy classification denoted by α and β in which α indicates as $\alpha \in [0; 1]$ and $\beta = 1 - \alpha$.

Further, Eq. 7 is converted to a minimization problem according to the classification error rate and its features selected. The final acquired minimization problem is well-defined, as shown in Eq. 8.

$$F = \alpha E_Z(D) + \beta \frac{Z}{N} \tag{8}$$

Here, F is the fitness function and $E_Z(D)$ is the error rate of classification, Z is indicated as the selected feature subset length, and the total number of features is mentioned as N.

3.5 Dataset description

The following are the related datasets that are used for detecting diverse facial expressions. Table 2 depicts the characteristics of the datasets.

4 The proposed system

The architecture of the OFELBW system is depicted in Fig. 2. It consists of phases such as a pre-trained model based on CNN, feature extraction, feature selection, Haar Cascade classification, and performance evaluation. The input image captured through video was converted into a greyscale image in the pre-processing phase. A transformation is done by transforming one image ("I") into another ("J") using the transformation function $Tr()$. Here, a and b denote the pixels in images I and J, respectively.

$$a = Tr(b) \tag{9}$$

$Tr()$ corresponds to a transformation of pixel values a and b. The results of this transformation are mapped to the greyscale range since we are only concerned with greyscale images here. The normalization functions are applied to produce the normalization of the greyscale input image. Further, the scale of the image is represented using the range of 0 to 255. The linear normalization of the digital image is accomplished using the formula shown in Eq. 10.

$$I_N = (I - Min) \frac{newMax - newMin}{Max - Min} + newMin \tag{10}$$

where I is the grayscale input images, intensity values are (Min, Max) and new intensity values are denoted as ($newMin$, $newMax$).

Testing data sets are used to evaluate our method, 75% of training datasets and 25% of testing datasets are used to represent the accuracy. The following data sets CK+, JAFFE, FERG, and SFEW are used in the simulations. The performance of the face emotion learning of our proposed method is tested and compared with the various dataset discussed above.

In the pre-trained model based on the CNN phase, convolutional filters, and Max Pooling layers process the greyscale images. It goes through several iterations. The minimum iteration needed is 255 for CNN filtration and max pooling. In this module, Keras is used for pre-processing and TensorFlow is used to enhance the model based on deep neural networks. The model enters numerous iterations or epochs to be trained and tested in real time. In the extracted features phase, we extract the features of the essential part of the faces. To recognize a face, some parts of the face give more importance, such as the eyes and mouth. Other parts are considered less important. So, we are giving more importance to the primary parts of the face, such as the eyes and mouth. We give less importance to other parts of the face. The position and shape of the eyes and mouth are collected as features for face recognition.

The feature extraction consists of four convolutional layers, two merging layers, and an activation function that employs rectified linear units (ReLUs). Following these two layers are two fully connected layers and a dropout layer. The spatial transformer consists of two convolution layers and two completely coupled layers. Once the transformation parameters

Table 2 Characteristics of datasets

| JAFFE [50]	
Description	The JAFFE dataset is a small dataset that contains less than 1000 samples. To address the data scarcity problem in FER, image augmentation was applied in JAFFE, which significantly improved the model accuracy. 10 diverse female Japanese subjects
	exhibited 213 different facial expressions. There are seven facial expressions for every subject, and the images were interpreted based on average semantic ratings for each facial expression presented.
CK+ [24]	
Description	The extended Cohn-Kanade dataset (CK+) contains an initial small dataset as well as a laboratory-controlled benchmark dataset which represents the elementary emotions of anger, contempt, happiness, disgust, fear, sadness, and surprise, as well as a neutral expression. 123 different topics are represented in 593 video sequences.
FERG [32]	
Description	The database shows stylized characters that were modeled in MAYA. In total, 55,767 images represent six different style character types. Images are grouped into seven expression categories.
SFEW [47]	
Description	Static facial expressions in the wild (SFEW) provides a dataset for detecting facial expressions in the wild. By computing keyframes based on facial point clustering, static frames were selected from the AFEW database. In EmotiW 2015, the SREco sub-challenge benchmark was based on SFEW 2.0, which is the most frequently used version. The training comprises 958 samples, Val (436 samples), and Test (372 samples), and there are 7 expression categories – neutral, fear, anger, disgust, happy, sadness, and surprise.
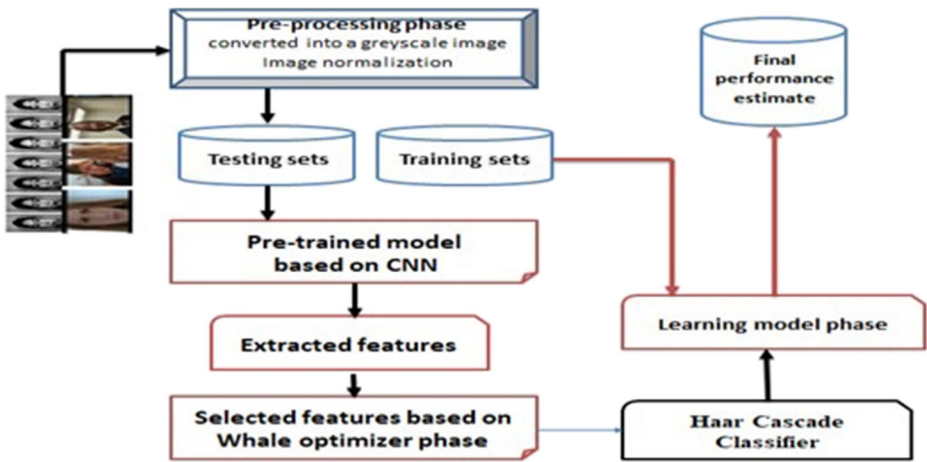
Fig. 2 The system architecture of the OFELBW method

are regressed, the output is transformed into the sampling grid T( ). The sampling grid in the spatial transformer module aims to focus on the most relevant portion of the image.

Selected features based on the whale optimizer phase; this feature selection phase involves selecting the appropriate features from the entire feature set using a binary whale optimization algorithm. As a result, WO is investigated for parameter selection, and the resultant classifier WO-OFEL is studied for its efficiency and effectiveness against different datasets. An optimization algorithm based on nature-motivated metaheuristics, called the WO algorithm, is used to fine-tune the parameters of the OFEL method. We successfully apply the WO-OFEL potential model to the OFEL method. It is intended that the WO-OFEL classifier will perform better than other competent schemes. For optimization, WO mimics the spiral bubble-net feeding mechanism of humpback whales. A Boolean search domain is used during this phase. In this method, features with a value of 1 are considered appropriate, while those with a value of 0 are deemed inappropriate. This method generates random locations of N whales (WOAk, K = 1,2, 3..., I) that provide solutions to the given problem as shown in Eq. 11. A random value (ε) is a small positive value used in the following equation to convert each solution into a binary solution.

$$WOA_K = \begin{cases} 1 & \text{if } WOA_K > \in \\ 0 & \text{if } WOA_k < \in \end{cases} \tag{11}$$

The facial expressions are detected in an image in the Haar Cascade classifier phase. A positive and a negative picture are presented to the classifier. The picture characteristics are drawn out at the end of the phase [5, 16]. In the learning model phase, a training data set is supplied to learn face recognition features. The proposed algorithm is compared with the previous relevant works in this field with different data sets in the final estimation phase [54].

5 Experimental results and discussions

We propose a method for visualizing the important region in facial expression classification. Based on a previously trained model on the occluded images, we zero out a square of size M x

M in the top left corner of an image. The saliency region in the neutral emotion image, for example, is vast enough to cover the entire face, which means that both regions are required to determine the neutral emotion of the image. In the feelings of happiness and surprise, the areas around the mouth become more important than other regions, as in the feelings of anger and sadness, where the areas around the eyebrow and eye become more significant. The salient regions for facial expression are shown in Fig. 3 where the proposed approach's reaction is compared with [29].

As shown in Fig. 4, the images are also captured in real time from the video captured by OpenCV. The images are then converted into grayscale images 48 × 48. Using a Haar family classifier, this module detects the faces of the users before it can create the grayscale images. Initially, the face is detected in real-time through the webcam using the Haar Cascade classifier. The captured face is then converted into grayscale in a 48 × 48 pixels frame. Later, the corner of the mouth, the end of the eyebrow, and the ala of the nose as facial landmarks [52, 54]. For facial marking, we used the dlib library in python. The library implements the face alignment with an ensemble of regression trees algorithm [29, 52]. These FLs are located near facial components


Fig. 3 Salient regions for different facial expressions

(a) Face detection (b) Conversion into grayscale (c) Facial landmark (d) Facial expression recognition
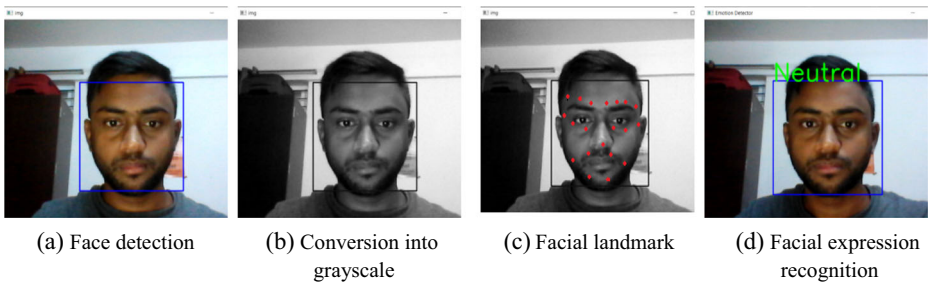
Fig. 4 Various phases involved in facial expression recognition (**a**) Face detection (**b**) Conversion into grayscale (**c**) Facial landmark (**d**) Facial expression recognition

and contours, capturing deformations caused by head movement. To establish a feature vector, facial landmarks can be compared. Finally, the model then predicts the expressions as presented in Fig. 5. The model is saved in a JSON file and then it will be imported into the OpenCV module to recognize the real-time images in the videos taken by the user.

Figure 5 presents the various facial expressions detected by the proposed system. The number of images selected in various datasets for different expressions is depicted in Fig. 6.

The performance of the face emotion learning of our proposed method is tested and compared with the dataset [5, 10, 29, 32]. To evaluate our method, we used 80% training, 10% validation, and 10% testing datasets to represent the accuracy. Each model is separately trained in our experiments and chosen the architecture and hyper-parameters same for all the models with 550 epochs with a varying batch sizes of 32 and 64. All the programs were executed in a system with Intel Core i5 8th generation CPU and 8 GB RAM with 512 GB HDD hard disk. The camera used here is of 5 megapixels (2592 × 1944) resolution and 25 frames per second. We have used the Gaussian distribution variable with zero mean and standard deviation as one. Adam optimizer with a learning rate of 0.001 along with weight decay used for optimization purposes. Our experiment uses data augmentation to populate the dataset with more relevant data. The execution time for training the data sets was
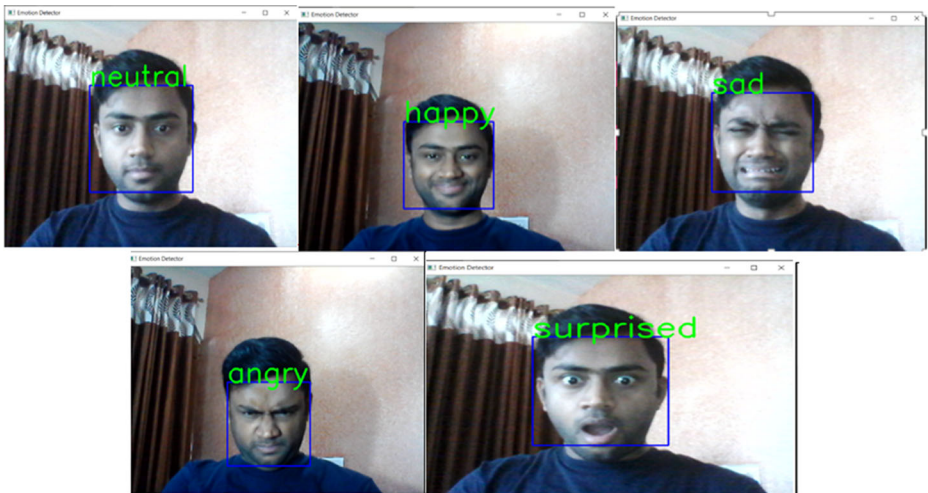
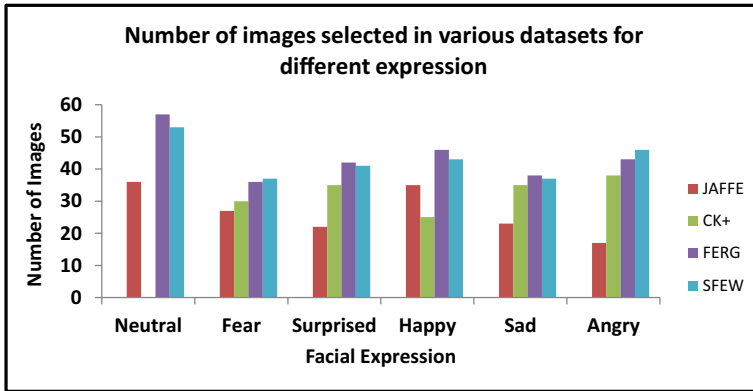**Fig. 5** Various facial expressions detected by the proposed system

Fig. 6 Number of images selected in various datasets

approximately 15.92 ms. The execution time change is considerable, as the execution time for the additional feature extraction for all the datasets is 32. 6% higher than the execution time for original features. The pixel size of the 48 * 48 image considered in our method takes 7.1 ms of
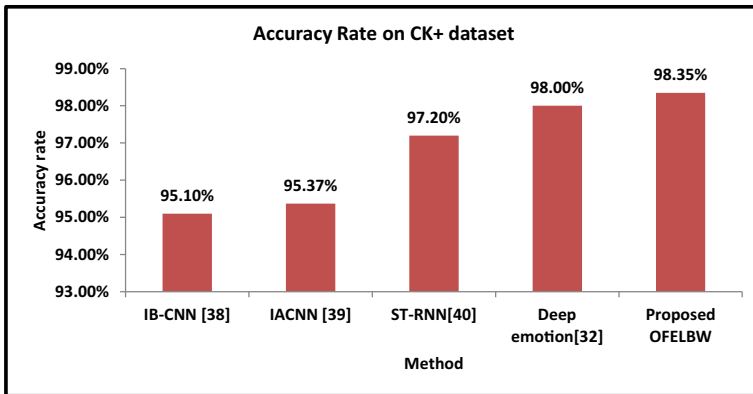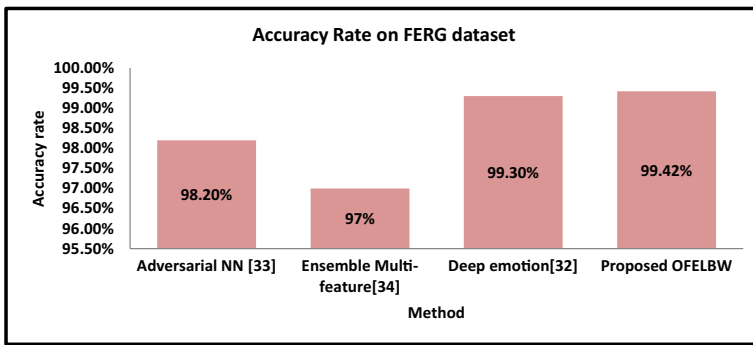

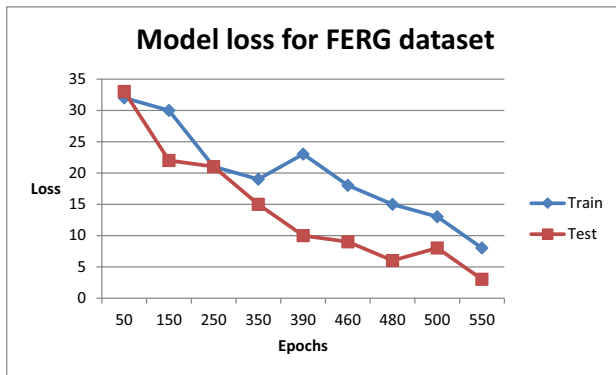**a** Accuracy rate on the CK+ dataset
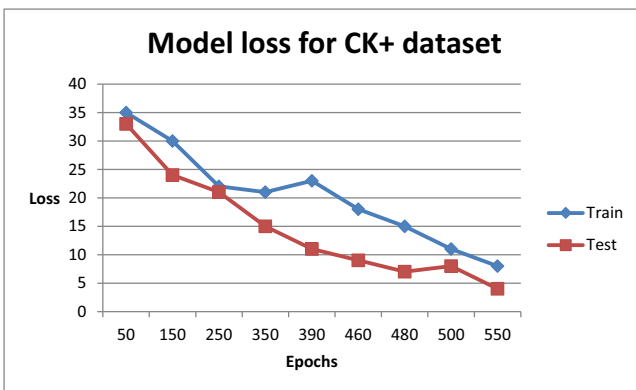
b Loss rate on the CK+ dataset

Fig. 7 **a** Accuracy rate on the CK+ dataset **b** Loss rate on the CK+ dataset

a Accuracy rate on the FERG dataset

b Loss rate on the FERG dataset

Fig. 8 **a** Accuracy rate on the FERG dataset **b** Loss rate on the FERG dataset

execution time which is much lower than the existing method mentioned in the paper which requires 14.5 ms [38].

The proposed algorithm is compared with the previous relevant works in this field on the CK+ dataset for accuracy and loss rate as shown in Fig. 7a and 7b. We have used the whole dataset of CK+, 80% used as the training, 10% for validation, and 10% as the testing set. As an experimental result, we were able to get an accuracy of 98.35%., which is relatively higher than the previous approaches.

The proposed OFELBW algorithm is compared with the previous relevant works in this field on the FERG dataset for accuracy and loss rate are shown in Fig. 8a and 8b. We have used the whole dataset of FERG, 80% used as the training, 10% for validation, and 10% as the testing set. As an experimental result, we were able to get an accuracy of 99.42%, which is relatively higher than the previous approaches.

The proposed algorithm, compared with the previous relevant works in this field on the JAFFE dataset for accuracy and loss rate is shown in Fig. 9a and 9b. We have used the whole dataset of JAFFE 80% used as training, 10% for validation, and 10% as a testing set. As an experimental result, we were able to get an accuracy of 96.6%, which is higher than the previous approaches.

The proposed algorithm is compared with the previous relevant works in this field on the SFEW dataset for accuracy and loss rate are shown in Fig. 10. We have used from the whole
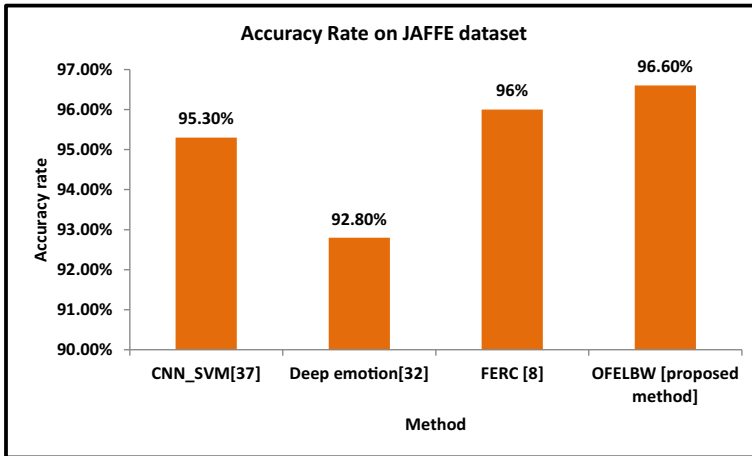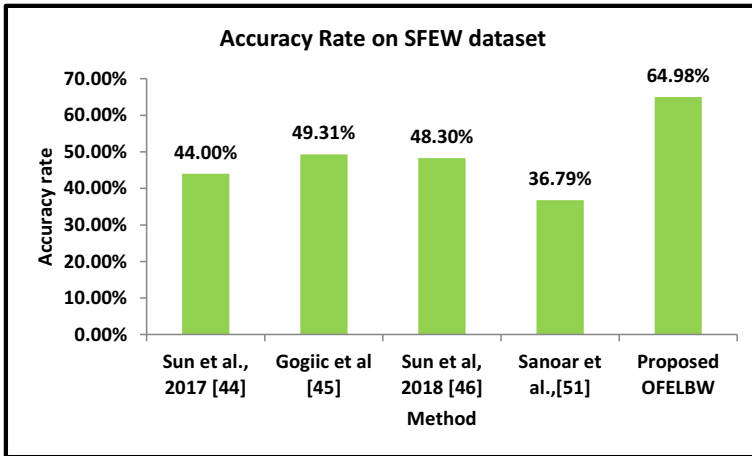
a Accuracy rate on the JAFFE dataset
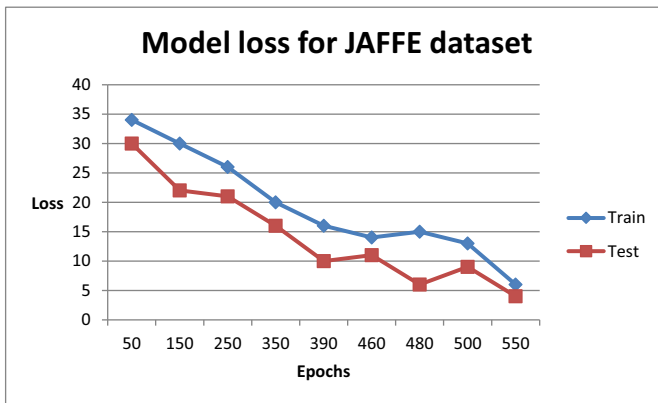
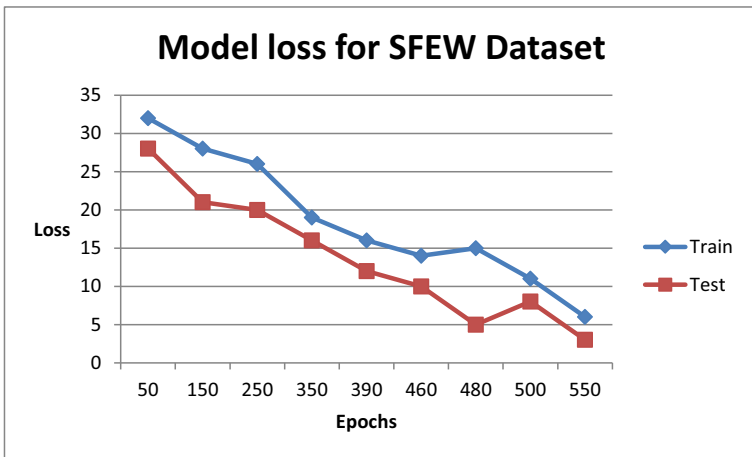**b Loss rate on the JAFFE dataset**

Fig. 9 **a** Accuracy rate on the JAFFE dataset **b** Loss rate on the JAFFE dataset

dataset of SFEW, 80% used as the training, 10% for validation, and 10% as the testing set. As an experimental result, we were able to get an accuracy of 64.98%, which is relatively higher than the earlier approaches. The accuracy of proposed method is less with respect to SFEW dataset because of more diverse features included in it. Since the SFEW dataset included images with both high and very low resolutions and more diverse features, the accuracy was significantly lower, despite the images being more representative of real-world conditions. Different head poses, a wide range of ages, different face resolutions, occlusions, varying focus, and almost real-world illumination were included, along with varying illumination levels [10].

Even though the proposed method results are slightly higher than the existing work mentioned in the article [32], we have obtained a higher accuracy rate concerning the SFEW dataset which is not focused on the article [32]. Also, the execution time of the existing is 14.5 ms, but our proposed method is much lower than the existing method as it requires 7.1 ms of execution time for the pixel size of 48 * 48 images considered in our method [38].

a Accuracy rate on the SFEW dataset

b Loss rate on the SFEW dataset

Fig. 10 **a** Accuracy rate on the SFEW dataset **b** Loss rate on the SFEW dataset

6 Conclusion and future work

This research work deals with the face emotion learning model based on CNN and a binary whale optimization algorithm to recognize facial expressions more accurately. It has various applications, such as receiving feedback from customers at restaurants and other profitable businesses. The extensive experimental analysis is presented in this article compared with other existing models and obtained the best accuracy score in our model. The model is trained and tested with 5 different database datasets globally used for human facial expression recognition. This work can be further studied and researched to find more accurate models using advanced deep learning techniques such as AlexNet, VGGNet, ResNet algorithms, and some advanced image processing techniques. With people coming more into this research

field, there are chances that a fully automated facial expression recognition system can be brought to the markets with 100% of accuracy. It can also feed into any microcontroller to make it a live project or an internet of things project.

Data availability The dataset used in this work can be accessed from [5, 10, 29, 32].

Declarations

Ethics approval No animals were involved in this study. All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

Conflict of interest The authors declare that they do not have any conflicts of interest that influence the work reported in this paper.

References

1. Abadi, M (2016) TensorFlow: learning functions at scale. In proceedings of the 21st ACM SIGPLAN international conference on functional programming (pp. 1-1). https://doi.org/10.1145/3022670.2976746
2. Ali, MF, Khatun, M, Turzo, NA (2020) Facial emotion detection using neural network. Int J Sci Eng Res
3. Ambert-Dahan E, Giraud AL, Mecheri H, Sterkers O, Mosnier I, Samson S (2017) Emotional recognition of dynamic facial expressions before and after cochlear implantation in adults with progressive deafness. Hear Res 354:64–72
4. Bairaju, SPR, Ari, S, Garimella, RM (2019) Emotion detection using visual information with deep auto-encoders. In 2019 IEEE 5th international conference for convergence in technology (I2CT) (pp. 1-5). IEEE. https://doi.org/10.1109/i2ct45611.2019.9033902
5. Burns EJ, Martin J, Chan AH, Xu H (2017) Impaired processing of facial happiness, with or without awareness, in developmental prosopagnosia. Neuropsychologia 102:217–228. https://doi.org/10.1016/j.neuropsychologia.2017.06.020
6. Ch S (2021) An efficient facial emotion recognition system using novel deep learning neural network-regression activation classifier. Multimed Tools Appl 80(12):17543–17568. https://doi.org/10.1007/s11042-021-10547-2
7. Dantas, AC, Do Nascimento, MZ (2022) Recognition of emotions for people with autism: An approach to improve skills. Int J Comput Games Technol, 2022. https://doi.org/10.1155/2022/6738068
8. Dantas, AC, do Nascimento, MZ (2022) Face emotions: improving emotional skills in individuals with autism. Multimed Tools Appl, 1–23
9. Demochkina, P, Savchenko, AV (2021) Neural network model for video-based facial expression recognition in-the-wild on mobile devices. In 2021 international conference on information technology and nanotechnology (ITNT) (pp. 1-5). https://doi.org/10.1109/itnt52450.2021.9649076
10. Dhall, A, Goecke, R, Lucey, S, Gedeon, T (2011) Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark. In 2011 IEEE international conference on computer vision workshops (ICCV workshops) (pp. 2106-2112). IEEE
11. Feutry, C, Piantanida, P, Bengio, Y, Duhamel, P (2018) Learning anonymized representations with adversarial neural networks. arXiv preprint arXiv:1802.09386. (pp. 1-20)
12. Georgescu MI, Ionescu RT, Popescu M (2019) Local learning with deep and handcrafted features for facial expression recognition. IEEE Access 7:64827–64836
13. Giannopoulos, P, Perikos, I, Hatzilygeroudis, I (2018) Deep learning approaches for facial emotion recognition: a case study on FER-2013. In advances in hybridization of intelligent methods (pp. 1–16). Springer, Cham
14. Gogić I, Manhart M, Pandžić IS, Ahlberg J (2020) Fast facial expression recognition using local binary features and shallow neural networks. Vis Comput 36(1):97–112. https://doi.org/10.1007/s00371-018-1585-8
15. Greco, A, Strisciuglio, N, Vento, M, Vigilante, V (2022) Benchmarking deep networks for facial emotion recognition in the wild. Multimed Tools Appl, 1–32

16. Han, S, Meng, Z, Khan, AS, Tong, Y (2016) Incremental boosting convolutional neural network for facial action unit recognition. Adv Neural Inf Proces Syst, 29

17. Hasani, B, Mahoor, MH (2017) Facial expression recognition using enhanced deep 3D convolutional neural networks. In proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 30-40)

18. Hong SW, Yoon KL (2018) Intensity dependence in high-level facial expression adaptation aftereffect. Psychon Bull Rev 25(3):1035–1042

19. Hossain S, Umer S, Asari V, Rout RK (2021) A unified framework of deep learning-based facial expression recognition system for diversified applications. Appl Sci 11(19):9174. https://doi.org/10.3390/app11199174

20. Hu, J, Shen, L, Sun, G (2018) Squeeze-and-excitation networks. In proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141)

21. Ismail, HA, Hashim, IA, Abd, BH (2019) A survey on linguistic interpretation of facial expressions and technologies. In 2019 2nd international conference on engineering technology and its applications (IICETA) (pp. 161-166). IEEE

22. Izen SC, Ciaramitaro VM (2020) A crowd of emotional voices influences the perception of emotional faces: using adaptation, stimulus salience, and attention to probe audio-visual interactions for emotional stimuli. Attention Percept Psycho 82(8):3973–3992

23. Kazemi, V, Sullivan, J (2014) One millisecond face alignment with an ensemble of regression trees. In proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1867-1874). https://doi.org/10.1109/cvpr.2014.241

24. Kowalski, M, Naruniec, J, Trzcinski, T (2017) Deep alignment network: a convolutional neural network for robust face alignment. In proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 88-97)

25. Kumar A, Jaiswal A, Garg S, Verma S, Kumar S (2019) Sentiment analysis using cuckoo search for optimized feature selection on Kaggle tweets. Int J Inf Retriev Res (IJIRR) 9(1):1–15

26. Kumar M, Jindal MK, Kumar M (2021) A novel attack on monochrome and greyscale Devanagari CAPTCHAs. Trans Asian Low-Resource Language Inf Process 20(4):1–30

27. Kumar M, Jindal MK, Kumar M (2022) Distortion, rotation and scale invariant recognition of hollow Hindi characters. Sādhanā 47(2):1–6. https://doi.org/10.1145/3439798

28. Li S, Deng W (2019) Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. IEEE transactions on image processing, image processing, IEEE transactions on. IEEE Trans Image Process 28(1):356–370

29. Lyons, MJ, Kamachi, M, Gyoba, J (2014) Japanese female facial expressions (JAFFE), Database Digit Images 2007

30. Mehendale N (2020) Facial emotion recognition using convolutional neural networks (FERC). SN Appl Sci 2(3):1–8. https://doi.org/10.1007/s42452-020-2234-1

31. Meng, Z, Liu, P, Cai, J, Han, S, Tong, Y (2017) Identity-aware convolutional neural network for facial expression recognition. In 2017 12th IEEE international conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 558-565)

32. Minaee S, Minaei M, Abdolrashidi A (2021) Deep-emotion: facial expression recognition using attentional convolutional network. Sensors 21(9) 3046:1–16

33. Mohammed AR, Kosonogov V, Lyusin D (2021) Expressive suppression versus cognitive reappraisal: effects on self-report and peripheral psychophysiology. Int J Psychophysiol 167:30–37. https://doi.org/10.1016/j.ijpsycho.2021.06.007

34. Mollahosseini A, Hasani B, Mahoor MH (2017) Affectnet: a database for facial expression, valence, and arousal computing in the wild. IEEE Trans Affect Comput 10(1):18–31. https://doi.org/10.1109/taffc.2017.2740923

35. Müller T, Schäfer R, Hahn S, Franz M (2019) Adults' facial reaction to affective facial expressions of children and adults. Int J Psychophysiol 139:33–39. https://doi.org/10.1016/j.ijpsycho.2019.01.001

36. Ning, GY, Cao, DQ (2021) Improved Whale Optimization Algorithm for Solving Constrained Optimization Problems Discrete Dyn Nature Soc, 2021. https://doi.org/10.1155/2021/8832251

37. Otberdout N, Kacem A, Daoudi M, Ballihi L, Berretti S (2019) Automatic analysis of facial expressions based on deep covariance trajectories. IEEE Trans Neural Netw Learn Syst 31(10):3892–3905. https://doi.org/10.1109/tnnls.2019.2947244

38. Owusu E, Zhan Y, Mao QR (2014) A neural-AdaBoost based facial expression recognition system. Expert Syst Appl 41(7):3383–3390. https://doi.org/10.1016/j.eswa.2013.11.041

39. Qi L, Binu D, Rajakumar BR, Mohammed Ismail B (2022) 2-D canonical correlation analysis-based image super-resolution scheme for facial emotion recognition. Multimed Tools Appl 81(10):13911–13934. https://doi.org/10.1007/s11042-022-11922-3

40. Said Y, Barr M (2021) Human emotion recognition based on facial expressions via deep learning on high-resolution images. Multimed Tools Appl 80(16):25241–25253
41. Shetty AB, Rebeiro J (2021) Facial recognition using Haar cascade and LBP classifiers. Global Trans Proceed 2(2):330–335
42. Shima, Y, Omori, Y (2018) Image augmentation for classifying facial expression images by using deep neural network pre-trained with object image database. In proceedings of the 3rd international conference on robotics, control and automation (pp. 140-146). https://doi.org/10.1145/3265639.3265664
43. Sun W, Zhao H, Jin Z (2017) An efficient unconstrained facial expression recognition algorithm based on stack binarized auto-encoders and binarized neural networks. Neurocomputing 267:385–395
44. Sun W, Zhao H, Jin Z (2018) A visual attention based ROI detection method for facial expression recognition. Neurocomputing 296:12–22
45. Tautkute, I, Trzcinski, T, Bielski, A (2018) I know how you feel: emotion recognition with facial landmarks. In proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 1878-1880)
46. Tejada, J, Freitag, RMK, Pinheiro, BFM, Cardoso, PB, Souza, VRA, Silva, LS (2021) Building and validation of a set of facial expression images to detect emotions: a transcultural study. Psychol Res, 1–11
47. Teufel C, Westlake MF, Fletcher PC, von dem Hagen E (2019) A hierarchical model of social perception: psychophysical evidence suggests late rather than early integration of visual information from facial expression and body posture. Cognition 185:131–143. https://doi.org/10.1016/j.cognition.2018.12.012
48. Varcin KJ, Nangle MR, Henry JD, Bailey PE, Richmond JL (2019) Intact spontaneous emotional expressivity to non-facial but not facial stimuli in schizophrenia: an electromyographic study. Schizophr Res 206:37–42. https://doi.org/10.1016/j.schres.2018.12.019
49. Verma B, Choudhary A (2021) Affective state recognition from hand gestures and facial expressions using Grassmann manifolds. Multimed Tools Appl 80(9):14019–14040
50. Wang K, Peng X, Yang J, Meng D, Qiao Y (2020) Region attention networks for pose and occlusion robust facial expression recognition. IEEE Trans Image Process 29:4057–4069. https://doi.org/10.1109/tip.2019.2956143
51. Wong SF, Trespalacios F, Ellenbogen MA (2020) Poor inhibition of personally-relevant facial expressions of sadness and anger predicts an elevated cortisol response following awakening six months later. Int J Psychophysiol 150:73–82
52. Wu Y, Ji Q (2019) Facial landmark detection: a literature survey. Int J Comput Vis 127(2):115–142
53. Yu, J, Yu, L (2018) Synthesizing photo-realistic 3D talking head: learning lip synchronicity and emotion from audio and video. In 2018 25th IEEE international conference on image processing (ICIP) (pp. 1448-1452)
54. Zhang, Z, Luo, P, Loy, CC, Tang, X (2014) Facial landmark detection by deep multi-task learning. In European conference on computer vision (pp. 94-108). Springer, Cham. https://doi.org/10.1007/978-3-319-10599-4_7
55. Zhang T, Zheng W, Cui Z, Zong Y, Li Y (2018) Spatial temporal recurrent neural network for emotion recognition. IEEE Trans Cybern 99:1–9
56. Zhao, H, Liu, Q, Yang, Y (2018) Transfer learning with ensemble of multiple feature representations. In 2018 IEEE 16th international conference on software engineering research, management and applications (SERA) (pp. 54-61)
57. Zhao Y, Oveneke MC, Jiang D, Sahli H (2019) A video prediction approach for animating single face image. Multimed Tools Appl 78(12):16389–16410

T. Muthamilselvan is an Associate Professor in the School of Information Technology and Engineering, at VIT University, India. He received his Ph.D. from VIT, and M. Tech in Computer Science and Engineering from SASTRA University, Thanjavur. He published around 20 international journal papers. He received the active researcher award from VIT university. His current research interest includes Machine learning deep learning, cloud computing, knowledge discovery, and Blockchain technology. He is a member of the Computer Society of India and the Indian Science Congress.

K. Brindha is an Associate Professor in the School of Information Technology and Engineering, at VIT University, India. She received her Ph.D. from VIT, ME in Computer Science and Engineering from Sathyabama University. She published around 40 international journal papers, many international conference papers, and 10 books. She received the active researcher award from VIT for the past twelve years. She is one of the reviewers in the JVLC Elsevier Journal. She is a life member of the Computer Society of India. Her research interest includes Deep learning, Machine learning, Cryptography, Network Security, Cloud computing, and Block chaining.
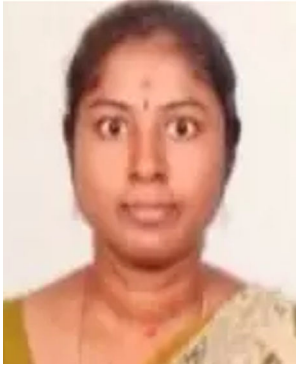
Sudha Senthilkumar is currently working as an Associate Professor in the School of Computer Science and Engineering at Vellore Institute of Technology, Vellore. She received her Ph.D. degree in the School of Information Technology and Engineering from VIT University, Vellore. She completed her M. Tech in Information Technology and Engineering from VIT University, Vellore. Her current research interest includes C*ryptography*, *Network security*, Bigdata, Blockchain Technologies, Machine learning deep learning, and *remote user authentication using smart cards*. She has authored over 55 research articles at reputed international conferences. She is a life member of the Computer Society of India and the Indian Science Congress.

Saransh is an Assistant System Engineer at Tata Consultancy Services, Pune, India. He received his MCA from VIT University, Vellore. His work includes developing to supporting Sales and Finance applications for World's 2nd largest Agri Firm based in Canada. His current research interest mainly focuses on Machine Learning and how to best utilize it in inventory applications for any Company.

Jyotir Moy Chatterjee is currently working as an Assistant Professor in the Department of Information Technology at Lord Buddha Education Foundation, Kathmandu, Nepal. Before this, he worked as an Assistant Professor in the Department of Computer Science Engineering at G. D. Rungta College of Engineering & Technology (Chhattisgarh Swami Vivekananda Technical University), Bhilai, India. He received M. Tech in Computer Science & Engineering from Kalinga Institute of Industrial Technology (KIIT), Bhubaneswar, Odisha in 2016, and a B. Tech in Computer Science & Engineering from Dr. MGR Educational & Research Institute, Maduravoyal, Chennai in 2013. His research interests include Machine Learning & Deep Learning.

Yu-Chen Hu received his Ph.D. degree in computer science and information engineering from the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan. Currently, Dr. Hu is a professor in the Department of Computer Science and Information Management, at Providence University, Sha-Lu, Taiwan. He joins the editorial boards of Advances in Multimedia (Hindawi), IET Image Processing, Electronics (MDPI), Mathematical Problems in Engineering (Hindawi), etc. His research interests include image and signal processing, data compression, information hiding, information security, computer network, and machine learning.

Affiliations

T. Muthamilselvan[1] **· K. Brindha**[1] **· Sudha Senthilkumar**[2] **· Saransh**[3] **· Jyotir Moy Chatterjee**[4] **· Yu-Chen Hu**[5]

T. Muthamilselvan
tmuthamilselvan@vit.ac.in

K. Brindha
brindha.k@vit.ac.in

Sudha Senthilkumar
sudha.s@vit.ac.in

Saransh
saranshsrivastava4916@gmail.com

Yu-Chen Hu
ychu@pu.edu.tw

[1] School of Information Technology and Engineering, VIT University, Vellore, India

[2] School of Computer Science and Engineering in Vellore Institute of Technology, Vellore, India

[3] Tata Consultancy Services (TCS), Pune, India

[4] Department of IT, Lord Buddha Education Foundation, Kathmandu, Nepal

[5] Department of Computer Science and Information Management, Providence University, Taipei, Taiwan, China