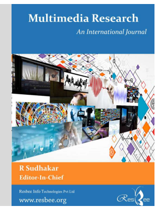
Vol.7 No.2 April 2024

Multimedia Research

Received 12 December, Revised 23 February, Accepted 28 March



Pose and Occlusion Invariant Face Emotion Recognition by Optimal Feature Extraction and Classification

Aravind Reddy Bobbala
Lamar University
Beaumont, Texas, United States.
aravind.reddybobbala@gmail.com

Abstract: Facial expressions play a major role since it is capable of sharing a lot of data, which could not be expressed by voice throughout interpersonal communication. As a result, Facial Emotion Recognition (FER) is turning out to be a rising topic in current years. In this paper, a novel FER model is implemented, which consists of four stages: (i) Face Detection, (ii) Feature extraction, (iii) Optimal feature selection, and (iv) Classification. In this context, face detection is done using the Viola-Jones (VJ) method. It is the initial technique to offer better object detection in real-time. Further, the features namely Scale-Invariant Feature Transform (SIFT), Discrete Cosine Transform (DCT), Hough Transform (HT), and Radon Transform are extracted during the feature extraction process. As the length of features seems to be high, the features are optimally selected using the improved Jaya Algorithm (JA). Subsequently, the optimally selected features are subjected to a classification process, which is done using Deep Belief Network (DBN). Here, in this paper, the activation function of DBN is optimized using the same improved JA. Since the improvement of JA is done with the modification of the probability function, the proposed algorithm is termed a Modified Probability-based Jaya Algorithm (MP-JA). Thus, the various emotions on the face (normal, smile sad, surprise, anger, fear, and disgust of face) are determined efficiently by the proposed framework.

Keywords: Facial Expression; Emotion Detection; Optimal Feature Selection; Modified Jaya Algorithm; Deep Belief Network

Nomenclature

Abbreviations	Descriptions
FER	Facial Emotion Recognition
DCT	Discrete Cosine Transform
JA	Jaya Algorithm
DBN	Deep Belief Network
SVM	Support Vector Machine
LDTP	Local Directional Ternary Pattern
CNN-RNN	Convolution-Recurrent Neural Network
RF	Random Forest
ESL	Extreme Sparse Learning
KE	Key Emotion
mGA	micro Genetic Algorithm
LBP	Local Bianry Pattern
VJ	Viola Jones
DoG	Differences of Gaussians
ROI	Region Of Interest
HT	Hough Transform
SHT	Standard Hough Transform
FT	Fourier Transform
2D	2Dimensional
FBP	Filtered Back Projection
PSO	Particle Swarm Optimization
FF	FireFly
GWO	Grey Wolf Optimization
WOA	Whale Optimization Algorithm
CD	Convergent Divergence
FNR	False Negative Rate
FPR	False Positive Rate
MCC	Mathews Correlation Coefficient
NPV	Negative Predictive Value
FDR	False Discovery Rate
MLFFTL-CNN	Multi-Level Feature Fusion and Transfer Learning Convolutional Neural Network
AFORET	Affinity-based overlap reduction technique

https://doi.org/10.46253/j.mr.v7i2.a4

1. Introduction

Face is considered the most generally used biological feature for the identification of a person. FER [1] [3] exists as a dynamic area of research in pattern recognition and computerization. Besides being natural, the most important advantage of FER [4] is that it can be captured even at long distances. The FER technique can be used for all facial analysis algorithms. Facial analysis algorithms include several techniques such as facial modeling, facial alignment, facial authentication, facial expression tracking, and facial relighting methods. Thus, several face recognition approaches were initiated and proposed. Image processing techniques in the face recognition method are used to improve the raw images obtained from cameras for various applications [9].

In FER technology [10] [11], there are two major steps, such as optimal feature extraction and classification. Hence, to enhance the performance in FER [12], it is necessary to find an efficient feature extractor and a classifier. To raise the efficiency of the categorization process, the dimension of extracted feature space should be reduced. For the reduction in the dimension of feature space, several methods can be used to enhance the accuracy and robustness of FER [13] [14] [15]. The FER approach is of two categories, such as the holistic feature-based approach and the local feature-based approach [16] [17]. In a holistic approach, it includes a single vector in which the complete facial features can be considered, whereas, in a local feature-based approach, individual components of the face can be considered.

FER [18] [19] has attained much consideration in the last few years and its exploration was rapidly grown by engineers and neuroscientists, as it has numerous prospective applications in automatic control and computer vision systems [20]. Even though its applications are very helpful in several fields, it is much more complex to employ, as the human face may vary depending on the varying conditions. Apart from these attempts, identifying facial emotions [21] [22] with high precision remains to be complicated, and it still requires an extended research topic. Most often, the researchers introduce novel approaches for improving the accuracy of FER [23] using several well-known FER databases such as the JAFFE database and the MMI Database. However, these techniques do not succeed in exactly clarifying the types of testing and training protocols, and it does not perform cross-database computations. These problems make it complex to evaluate diverse systems with one another obstructs the development of this field.

The major contribution of this paper is depicted below.

1. To present a novel FER model that includes four stages: (i) Face Detection, (ii) Feature extraction, (iii) Optimal feature selection, and (iv) Classification.
2. Here, face detection is done using the VJ method, and the features, namely, SIFT, DCT, HT, and radon transform are extracted during the feature extraction process.
3. Since the length of features seems to be high, the features are optimally selected using MP-JA. Further, the optimally selected features are classified using DBN.
4. Along with the optimal feature selection, the next contribution focuses on the classification stage, where the activation function of DBN is optimized using the MP-JA algorithm. Thus, the various emotions of the face are determined efficiently by the proposed framework.
5. Finally, the betterment of the implemented FER model is evaluated by comparing it to the conventional models, and its enhancement is confirmed.

The organization of this paper is offered as follows: Section 2 analyzes the review on this topic. Section 3 describes the face detection and feature extraction techniques utilized for the proposed FER. Section 4 portrays the optimal feature selection and classification process. Section 5 illustrates the outcomes, and Section 6 concludes the paper.

2. Literature Review

2.1 Related Works

In 2022, Susrutha Babu Sukhavasi *et al.* [1] designed driver assistance systems (ADAS) for vehicular intelligence. Here, there were two approaches, namely deep neural networks and machine learning methods. These ADAS helped to supervise various expressions in different illuminations, occlusions, and pose variations, which ensured safety and also tracked any possibility of road accidents. This method obtained better results on various datasets such as FER 2013, CK+, KMU-FED, KDEF, etc.

In 2022, Eaby Kollonoor Babu *et al.* [2] created a visual descriptor called SIM-LBP. Initially, a different LBP was used to extract facial expressions in different lighting conditions. To test the model's efficiency, it was applied to the JAFFE dataset, which converted all images into their corresponding transformed variants using SIM-LBP. Then, a convolution neural network-based deep learning model for recognizing facial expressions was trained using these converted images (FER). Finally, the estimation metrics were compared with traditional LBP descriptors and other existing counterparts. The output SIM-LBBP transformation outperformed baseline methods in all four matrices tested against them and also showed better results than state-of-the-art models when compared side-by-side.

In 2022, P. M. Ashok Kumar *et al.* [3] applied a MLFFTL-CNN. This method involved constructing a CNN classifier based on emotion databases. It was then followed by fusion at intermediate layers during the practice phase using E-CNN. The MLFFTL-CNN model was then fine-tuned with neutral or limited datasets for facial recognition purposes. The project was carried out on the Cohn-Kanade dataset. They compared results against other existing techniques to achieve better performance.

In 2022, Aayushi Chaudhari *et al.* [4] executed FER using the transformers and ResNet-18 model. It examined Vision Transformer's performance in tasks compared with other cutting-edge models using hybrid datasets. Cropping, face detection, and feature extraction via deep learning models as well as fine-tuned transformer technology were introduced. The outcome of the system was capable of emotion recognition successfully and used in practical settings.

In 2022, Rami Reddy Devaram *et al.* [5] introduced residual convolutional neural networks to recognize facial expressions using a combination of dilated and conventional convolutional layers. Capturing pixel displacement while using fewer parameters than other methods, makes it easier to distinguish between various facial expressions. Additionally, to lessen the dying RELU issue and increase model stability, the first layers of the model used an exponential linear unit activation function. Through one and fivefold cross-validation on numerous publicly accessible datasets as well as mixed datasets made from all of these sources, the performance was assessed. The number of parameters used was significantly reduced while still achieving comparable results.

In 2024, Chaoji Liu *et al.* [6] focused on pose-invariant facial expression recognition using a deep global multiple-scale and local patches attention dual-branch network. The network contained a feature extraction module, global multiple-scale module, local patches attention module, and model-level fusion model. It aimed to mitigate the impact of pose variation and self-occlusion on recognition accuracy. The structure of the network included global average pooling and fully connected operations. Local Patches Attention Network refined features by dividing mid-level feature maps into several local patches with an attention mechanism.

In 2022, Sankhadeep Chatterjee *et al.* [7] executed an emotion recognition system using facial images through an (AFORET), which can reduce class overlaps and improve the accuracy of predicting human emotions from facial images by training various well-known classifiers with performance indicators as metrics for comparison against existing methods such as NBU, OSM, and v-SVM techniques. The classifier presentation was greatly improved by the AFORET algorithm's output, which used the RVA model to predict human emotion from facial images.

In 2022, Antonio Greco *et al.* [8] focused on emotion recognition from face images to overcome corruption and perturbation issues. Initially, two data sets were created based on an existing test set called RAF-DB, which contains modified images in 18 types of corruptions, and 10 types of perturbations. Then the existing networks (VGG, Dense Net, SE Net, and Xception) trained using the original RAF-DB images were compared against ARM using the current state-of-the-art method used for testing the RAF-DB dataset. The output showed a 200% increase achieved by the ARM model while classifying the real RAf_Db Test Set images. This process increased robustness, reduced errors during the classification process, and outperformed Arm model results significantly.

2.2 Review

Table 1 shows the reviews of the FER models.

Table 1. *Features and challenges of fer models using various techniques*

Author [citation]	Adopted methodology	Features	Challenges
P. M. Ashok Kumar *et al.* [1]	MLFFTL-CNN	Limited training dataset and network training complexity in face recognition. Better accuracy and performance	Performed only on a single dataset.
Rami Reddy Devaram *et al.* [2]	LEMON	Compact and robust Less computation cost Improves stability of the model reducing the dying ReLU problem.	Only evaluated on facial expression recognition and no other aspects of social intelligence such as speech or gesture recognition.
Aayushi Chaudhari *et al.* [3]	FER	Better performance	ResNet model was adopted as a baseline architecture for many computer vision tasks.
Chaudhari, A. *et al.* [4]	RF	Better accuracy. Enhanced recognition of emotions.	Needs more consideration on hybridizations of feature selection techniques.
Sankhadeep Chatterjee *et al.* [5]	AFORET	Improves classifier performance significantly compared with other methods. Achieving higher classification accuracies.	The study had only been conducted on one dataset. Computational complexity involved in training deep neural networks.
Chaoji Liu *et al.* [6]	GMS-LPA	Provided more stable recognition results than individual GMS and LPA networks, demonstrating robustness in controlled and real-world scenarios.	The network structure may become complex in practical applications due to the requirement of several pre-trained models from diverse regions of interest for multi-branch feature learning.
Antonio Greco *et al.* [7]	LBP and SVM	Increased robustness to out-of-distribution variations, substantially reducing error rates on corrupted inputs and outperforming ARM algorithm.	Evaluated only two datasets. Constructed based on RAF-DB test set only. Not generalize well to other emotion recognition tasks or different image domains.
Susrutha Babu Sukhavasi *et al.* [8]	Viola-Jones algorithm and KLT	High accuracy Robustness Real-time processing capability.	Data sets are not sufficient to generalize the results for all possible scenarios. Focused only on six basic facial expressions of human emotions.

3. Face Detection And Feature Extraction Techniques Utilized for Proposed Face Emotion Recognition

3.1 Viola Jones Model

The image I_{im} is subjected to the VJ model [24], which is the first model for object detection,and it offers efficient detection in real-time. It comprises 3 phases for the detection of facial regions:

1. A feature extraction model, named Haar features is a rectangular model that is attained by an integral image.
2. 2. A machine-learning approach, known as Adaboost is exploited for recognizing the face. Here, the word 'boosted' indicates the classifiers, which are found complicated at every level that is built usinga boosting model.
3. In addition, a classifierin cascade arrangement is deployed to merge several features effectively. The word 'cascade' describes the numerous filters found on a resultant classifier.

On cascading the strong classifiers, the non-facial and facialare as could be isolated. This object detector more over recognizes the upper body, nose, pupil, eyes, and lips efficiently.Thus the face detected image obtained using the VJ model is indicated as I_{VJ}, which is the face image attained after extracting the eyes, mouth, nose, and ears. From the detected face image, four relevant features SIFT, DCT, Hough transform, and Radon transform are extracted.

3.2 SIFT Extraction

SIFT [25] is a technique for identifying and extracting local feature descriptors, which were invariant to image contrast, scaling, and rotation. The SIFT model is first established by Lowe [31]. In recent times, it was enhanced to offer a superior image extraction.

Scale-space extrema detection:, Initially,the forgery image $I(a,b)$is convolved with Gaussian filters at varied scales as given by Eq. (1), which, $A(a,b,\sigma)$ specifies the convolution of image $I(a,b)$along with the Gaussian filter $M(a,b,\sigma)$ at scale σ . Changes among both Gaussian images at scale σ and $o\sigma$ are given by Eq. (1) and Eq. (2).

$$A(a,b,\sigma)=M(a,b,\sigma)*I(a,b) \tag{1}$$

$$C(a,b,\sigma)=A(a,b,o\sigma)-A(a,b,\sigma) \tag{2}$$

The variations in both scales are known as DoG.

Key point localization: After the initial process, Interest points (also known as key points) are identified as local minima or maxima across scales for the DoG images. The total pixels in the DoG images are evaluated with its related eight neighbors at the same scale as shown by Eq. (3), in which $\hat{x}$ is computed by fixing the derivative $C(a,b,\sigma)$to 0. Furthermore, there is a need to attain the exact localization of key points by disregarding the points by a predetermined value.

$$C(\hat{a})=C+\frac{1}{2}\frac{\partial C^T}{\partial a}\hat{a} \tag{3}$$

Orientation allocation: To attain invariance to orientation, the gradient magnitude $l(a,b)$and orientation $\theta(a,b)$are computed as shown in Eq. (4) and Eq. (5).

$$l(a,b)=\sqrt{A(a+1,b)-A(a-1,b)^2+A(a,b+1)-A(a,b-1)^2} \tag{4}$$

$$\theta(a,b)=arc\tan\left(\frac{A(a,b+1)-A(a,b-1)}{A(a+1,b)-A(a-1,b)}\right) \tag{5}$$

Key point descriptor production: After choosing the key point orientation, the feature descriptor is computed on 4×4 pixel neighbor hoods as a group of orientation histograms.
Finally, the extracted features using SIFT are represented as $Feat_{SIFT}$.

3.3 Discrete Cosine Transform

In general, DCT [26] is a major concept in the image processing domain, and it has the better capability of information centralizing for image sources. Thus, it is renowned for its advanced performances of simple execution and energy compaction. Eq. (6) describes the 2D DCT of a block of image $N\times N$.

$$F(q,v)=\frac{2}{N}d(q)d(v)\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}f(i,j)\cos\left(\frac{\pi(2i+1)q}{2N}\right) \\ \cos\left(\frac{\pi(2j+1)q}{2N}\right),q,v=0,1,2...N-1 \tag{6}$$

$$f(q,v)=\frac{2}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}d(q)d(v)F(i,j)\cos\left(\frac{\pi(2i+1)q}{2N}\right) \\ \cos\left(\frac{\pi(2j+1)q}{2N}\right),q,v=0,1,2...N-1 \tag{7}$$

In Eq. (7), $f(q,v)$ and $F(q,v)$indicates $(q,v)^{th}$ image pixel value and DCT coefficient present in spatial and frequency domain respectively; $d(k)$ indicates multiplication parameter as shown in Eq. (8).

$$d(k)=\begin{cases}\frac{1}{\sqrt{2}}, & if\ \ k=0\\ 1, & if\ \ k>0\end{cases} \tag{8}$$

Hence, $Feat_{DCT}$ is considered as the feature vector extracted by DCT.

3.4 Hough Transform

HT [28] is carried out using a voting model that attains the most excellent peak values in the parameter space. Here SHT is considered for determining the straight lines as portrayed by Eq. (9), which u and m denotes the y-intercept and slope of the line.

$$p = mq + u \tag{9}$$

For a robust evaluation procedure, a diverse parameter space was portrayed ρ and θ was defined for improved recognition of lines using HT15. The formulation of the line is given in Eq. (10), which ρ denotes the vertical distance between the line and the origin and θ indicates the angle extended by the line in terms of one of its axis.

$$p = \left(-\frac{\cos\theta}{\sin\theta}\right)q + \left(\frac{\rho}{\sin\theta}\right) \tag{10}$$

Therefore, Eq. (10) is rewritten as given by Eq. (11).

$$\rho = x\cos\theta + y\sin\theta \tag{11}$$

The fundamental concept of HT is known by regarding a point in the spatial domain specified by (q, p), which is attained from edge detection procedure. Consequently, every point votes for a sequence of (ρ, θ) values for entire lines crossing over it. A line at edge detected image includes numerous points preferred for that specific line segment. The line that attains a big contribution of votes is considered a line of interest, and it portrays the facial feature regions of the image.

Therefore, the Hough transform features extracted are represented as $Feat_{hough}$.

3.5 Radon Transform

Radon transform [30] is a method that is deployed to identify the image features. It is dependent on the parameterization of straight lines, and also it depends on the computation of the image integrals along the straight lines. Assume the image $f(S,Y)$ specified in the unit circle which is interchangeably referred to $f(ra,\theta)$ as polar coordinates or as $f(s,z)$ for rotated $s-z$ coordinate system. "The Radon transform [30] $g(s,\phi)$ of the image is defined in terms of line integrals through the image along a line parallel to the z axis at position s" as shown by Eq. (12).

$$g(s,\varphi) = \int f(s,z)dz \tag{12}$$

The term $g(s,\phi)$ is indicated by a sinogram. The sinogram gratifies the term $g(-s,\phi+\pi) = g(s,\phi)$. To discover the "Inverse Radon transform", assume the FT, $G(w,\phi) = \aleph_s[g(s,\phi)]$ of the sinogram in terms of position s.

$$G(w,\varphi) = \int_{-\infty}^{\infty} g(s,\varphi)e^{-2\pi iws}ds \tag{13}$$

On considering Eq. (12) and on deploying the identity $s = S\cos\phi + Y\sin\phi$, Eq. (14) can be attained. In Eq. (14), $v = w\sin\phi$ and $u = w\cos\phi$.

$$G(w,\varphi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(S,Y)e^{-2\pi i(uS+vY)}dSdY \tag{14}$$

This term is the 2D-FT $F(u,v) = \aleph_2[f(S,Y)]$ of the image for the particular values of v and u portrayed by ϕ and w. Therefore, $F(u,v)$ is determined completely by considering angles in the level $0 \le \phi < \pi$, as shown in Eq. (15).

$$f(S,Y) = \aleph_2^{-1}[F(u,v)] = \aleph_2^{-1}[\aleph_s[g(s,\phi)]] \tag{15}$$

This association between the FT of the image and sinogram is much more significant in medical images and is termed FBP or Central Slice. FBP, the much generally deployed technique of "tomographic image reconstruction", is obtained from Eq. (15) on considering the integral of $u-v$ to ϕ and w using Eq. (13).In Eq. (16), a filter specified by $|w| = \aleph_s[\lambda(s)]$ is denoted by $\lambda(s)$ and convolution is indicated by $*$.

$$f(S,Y) = \int_0^{\pi} g(s,\varphi) * \lambda(s)d\varphi \tag{16}$$

$|w|$ is keptat zero for frequencies beyond Nyquist frequency for sampled computations and a window is deployed for minimizing the noise at increased frequencies.

Thence, the Radon transform features extracted from the detected face are denoted as $Feat_{radon}$.

After extracting the individual feature, the four feature vectors SIFT, DCT, Hough transform, and Radon transform are combined, which is based on Eq. (17). Hence, the combined feature vector is expressed as Eq. (18), which $i = 1,2,\cdots,N_{feat}$, and N_{feat} indicates the number of extracted features.

$$Feat_{(i)} = Feat_{SIFT} + Feat_{DCT} + Feat_{hough} + Feat_{radon} \tag{17}$$

$$Feat_{(i)} = Feat_{(1)}, Feat_{(2)}, \cdots, Feat_{(N_{feat})} \tag{18}$$

From the overall feature vectors, some relevant feature has to be selected through the optimal feature selection process by the proposed MP-JA algorithm.

3.6 Classification Using DBN

The optimally selected feature vectors are applied to DBN for recognizing the specified emotions from the face. In general, the DBN [27] framework is a well-known intelligent approach that comprises numerous layers, visible neurons, and hidden neurons from the output layer. There exists a profound association between input and hidden neurons; however, association rules are not present between hidden neurons, and no associations are there in the visible neurons. The link between hidden and visible neurons is exclusive and symmetric. The output of neurons in the Boltzmann network is probabilistic. The output $\overline{PO}$ in Eq. (20) is based on probability function $\overline{P}_q(\zeta)$. Here, t^P denotes the pseudo-temperature. The DBN model is specified in Eq. (21).

$$\overline{P}_q(\zeta) = \frac{1}{1 + e^{\frac{-\zeta}{t^P}}} \tag{19}$$

$$\overline{PO} = \begin{cases} 1 & with\ 1 - \overline{P}_q(\zeta) \\ 0 & with\ \overline{P}_q(\zeta) \end{cases} \tag{20}$$

$$\lim_{t^P \to 0^+} \overline{P}_q(\zeta) = \lim_{t^P \to 0^+} \frac{1}{1 + e^{\frac{-\zeta}{t^P}}} = \begin{cases} 0 & for \quad \zeta < 0 \\ \frac{1}{2} & for \quad \zeta = 0 \\ 1 & for \quad \zeta > 0 \end{cases} \tag{21}$$

The mathematical design reveals theBoltzmann machine energy for the formation of a binary state bi or neuron, and that is portrayed in Eq. (22) and Eq. (23), which $L_{a,l}$ denotes the weights between neurons and θ_a refers to the biases.

$$EN(bi) = -\sum_{a<l} bi_a L_{a,l} - \sum_a \theta_a bi_a \tag{22}$$

$$\Delta EN(bi_a) = \sum_l bi_a L_{a,l} + \theta_a \tag{23}$$

The development of energy with respect to joint composition of hidden and visible neurons (x, y) is portrayed in Eq. (24), Eq. (25) and Eq. (26). Here, x_a and y_a denotes the binary state of a visible unit and hidden unit l . k_a and C_l points out the biases and $L_{a,l}$ specifies the weight among them.

$$EN(x, y) = -\sum_{(a,l)} L_{a,l} x_a y_l - \sum_a k_a x_a - \sum_l C_l y_a \tag{24}$$

$$\Delta EN(x_a, \overline{y}) = \sum_l L_{al} y_l + k_a \tag{25}$$

$$\Delta EN(\vec{x}, y_a) = \sum_l L_{al} x_a + C_l \tag{26}$$

Training of RBM achieves the distributed probabilities, and the resultant weight allocation is given in Eq. (27).

$$\hat{L}_{(\hat{Z})} = \max_{\hat{L}} \prod_{\vec{x} \in N} c(\vec{x}) \tag{27}$$

For the hidden and visible vectors pair $(\vec{x}, \overrightarrow{hi})$, the probability distributed RBM model is specified in Eq. (28), which PR^F denotes the partition function as given by Eq. (29).

$$c(\vec{x}, \overrightarrow{hi}) = \frac{1}{PR^F} e^{-EN(\vec{x}, \vec{y})} \tag{28}$$

$$PR^F = \sum_{\vec{x}, \vec{y}} e^{-EN(\vec{x}, \vec{y})} \tag{29}$$

DBN model exploits the CD learning model, whose steps are given below.

Step 1: Select the x training samples and brace them into visible neurons.

Step 2: Assess the probability of hidden neurons c_y by recognizing the product of the visible vector x and $\hat{L}$ weight matrix $c_y = \sigma(x.\hat{L})$ depending on Eq. (30), where, σ denotes the activation function.

$$c\left(\vec{y}_l \to 1 \mid \vec{x}\right) = \sigma\left(C_l + \sum_a x_a L_{a,l}\right) \quad (30)$$

Step 3: Scrutinize the y hidden states from c_y probabilities.

Step 4: Compute the x exterior product of vectors c_y which is evaluated as positive gradient $\phi^+ = x.c_y'^{\,t^P}$.

Step 5: Analyze the renovation of x' visible states from y hidden states as shown in Eq. (31). In addition, it is necessary to assess y' hidden states from the renovation x'.

$$c\left(\vec{x}_l \to 1 \mid \vec{y}\right) = \sigma\left(k_a + \sum_a x_l L_{a,l}\right) \quad (31)$$

Step 6: Compute the exterior product of x' and y', by its negative gradient $\phi^- = x'.y'^{\,t^P}$.

Step 7: Describe the weight as given by Eq. (32), which η denotes the learning rate.

$$\Delta\hat{L} = \eta\left(\varphi^+ - \varphi^-\right) \quad (32)$$

Step 8: Eq. (33) defines the weight update with novel values.

$$L'_{a,l} = \Delta L_{a,l} + L_{a,l} \quad (33)$$

Before the learning process of the MLP model, assume $\left(K^{\hat{Z}}, W^{\hat{Z}}\right)$ training patterns, which $\hat{Z}$ denotes the number of training patterns, $1 \le \hat{Z} \le \overline{P}$, $K^{\hat{Z}}$ and $W^{\hat{Z}}$ denote the output vector and desired output vector correspondingly. Eq. (34) demonstrates every neuron error in l of output layer.

$$e_l^{\hat{Z}} = K^{\hat{Z}} - W^{\hat{Z}} \quad (34)$$

Moreover, Eq. (34) describes the squared error of $\hat{M}$ the pattern after MSE, as in Eq. (35) and Eq. (36), which $\widetilde{P}$ denotes the number of training patterns.

$$SE_{\hat{Z}}^{mean} = \frac{1}{\tilde{o}_y} \sum_{l=1}^{\tilde{o}_y} \left(e_l^{\hat{Z}}\right)^2 = \frac{1}{\tilde{o}_y} \sum_{l=1}^{\tilde{o}_y} \left(K^{\hat{Z}} - W^{\hat{Z}}\right)^2 \quad (35)$$

$$SE_{avg} = \frac{1}{\tilde{P}} SE_{\hat{Z}}^{mean} \quad (36)$$

Moreover, as the main contribution, the optimal feature selection and activation function of DBN is optimized using the MP-JA algorithm to improve the accuracy of classification.

4. Optimal Feature Selection and Classification Process

4.1 Implemented Architecture

The diagrammatic representation of the proposed FER model is shown in Fig. 1. In this work, a novel FER model is implemented, which consists of four stages: (i) Face Detection, (ii) Feature extraction, (iii) Optimal feature selection, and (iv) Classification. In this context, face detection is done using the VJ method. Further, the features namely, SIFT, DCT, Hough Transform, and Radon Transform are extracted during the feature extraction process. As the length of features seems to be high, the features are optimally selected using the proposed MP-JA, which is the advanced version of conventional JA. Subsequently, the optimally selected features are subjected to a classification process, which is done using DBN. Here, in this proposal, the activation function of DBN is optimized using the same MP-JA algorithm, such that the accuracy of classification should be maximal. Thus, the various emotions, namely, normal, smile sad, surprise, angry, fear, and disgust of face, are determined efficiently by the proposed framework. Further, the efficiency of the FER classification is proved using experimentations, which are carried out using the JAFFE database.

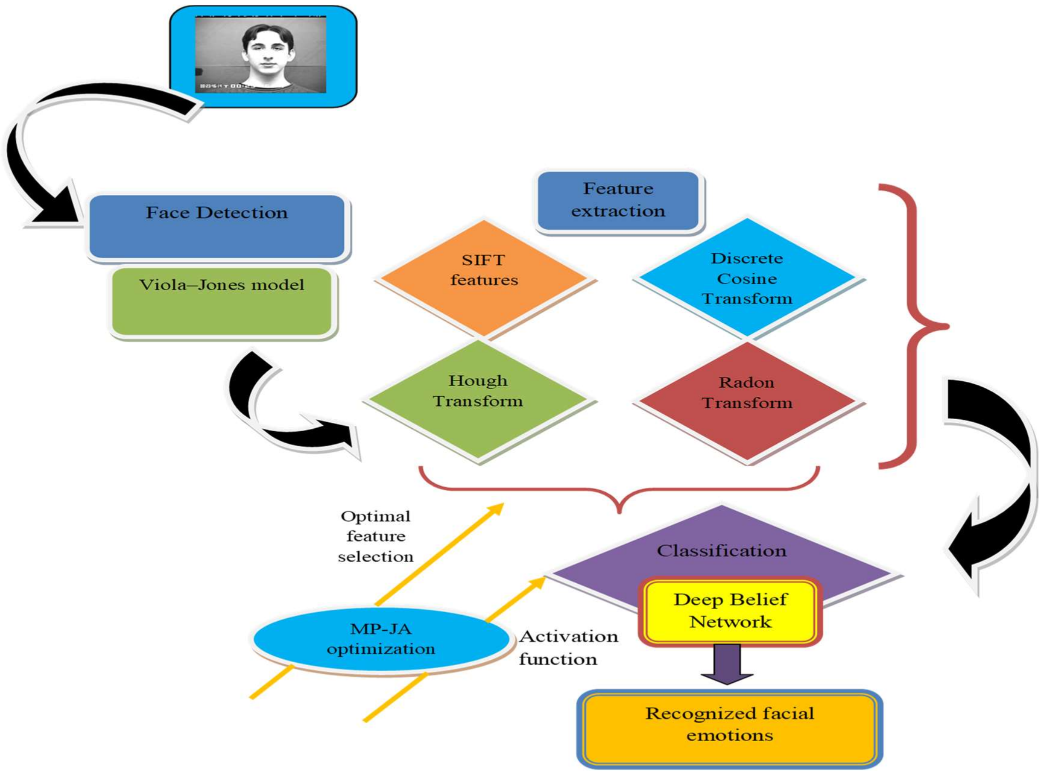


Fig. 1. *The overall architecture of the proposed Face Emotion Recognition model*

4.2 Objective Function with Solution Encoding

In the proposed framework, an improved optimization algorithm is utilized for two problems related to both feature selection and classification. In the optimal feature selection process, the extracted feature vector $Feat_{(i)}$ is given as input to the proposed MP-JA method and selects the optimal feature vector $Feat^*_{(i)}$ as shown in Eq. (37), which $N_{(opt)}$ specifies the number of optimally selected features.

$$Feat^*_{(i)} = Feat^*_{(1)}, Feat^*_{(2)}, \cdots, Feat^*_{(N_{opt})} \tag{37}$$

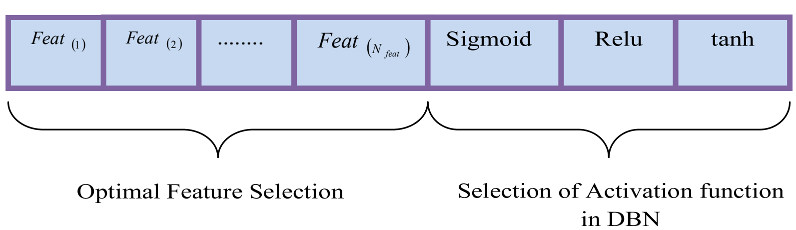


Fig. 2. *Solution encoding concerning feature selection and classification*

As DBN is used for the classification of facial emotions, the next contribution focuses on it. Here in DBN, the activation function is selected from three functions namely sigmoid, Relu, and tanh. The solution encoding for optimal feature selection, and classification is shown in Fig. 2. In this paper, the optimal feature selection, and classification is done in such a way that the recognition accuracy should be maximum. Accordingly, the maximized objective function is given in Eq. (38), where ACC denotes the accuracy, which is mathematically given in Eq. (39).

$$Obj = Max(ACC) \tag{38}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{39}$$

In Eq. (39), TP denotes true positive, TN denotes true negative, FP denotes false positive, and FN denotes false negative.

4.3 Jaya Algorithm

The JA [29] is dependent on the conception that the solution attained for a specified issue has to focus on the best solution and eliminate the worst solution. Consider that at i^{th} iteration, the count of modeling constraints is 'V'(i.e. $p = 1,2,...V$) and population size 'P'. If $X_{p,r,i}$ denotes the value of p^{th} factor for the r^{th} aspirant throughout the i^{th} iteration. Then, this value is customized as per Eq. (40), which $X_{p,best,i}$ denotes the value of the p^{th} constraint for the best solution and $X_{p,worst,i}$ refers to the value of p^{th} the constraint on considering the worst solution in the population.

$$X'_{p,r,i} = X_{p,r,i} + r_1\left(X_{p,best,i} - \left|X_{p,r,i}\right|\right) - r_2\left(X_{p,worst,i} - \left|X_{p,r,i}\right|\right) \tag{40}$$

In addition, $X'_{p,r,i}$ denotes the novel value of $X_{p,r,i}$, and r_1, and r_2 are arbitrary numbers with the limits of [0,1]. The term "$r_1\left(X_{p,best,i} - \left|X_{p,r,i}\right|\right)$" denotes that the solution attempts to move towards the best solution and the factor "- $r_2\left(X_{p,worst,i} - \left|X_{p,r,i}\right|\right)$" demonstrates the attempts of the solution to move away from the worst one. $X'_{p,r,i}$ is approved if the function value generated by it is improved. The pseudocode of the conventional JA model is given by Algorithm 1.

Algorithm 1: JA Algorithm [29]

Initialize the population
Recognize the best and worst solutions
While $i < Max\ Generation$
    Update the solution depending on best and worst solutions
    If the solution respective to $X'_{p,r,i}$ improved than $X_{p,r,i}$
        Accept and substitute the preceding solutions
    else
        Maintain the preceding solution
    If the termination condition satisfied
        Report the optimal solution
    else
        Repeat step 2
    end if
end while

4.4 Proposed MP-JA Model

Despite the interesting facts about the conventional JA model, it includes certain disadvantages such as premature convergence and worst exploitation capability. To overcome the shortcomings of the conventional JA approach, the presented model is modified based on a probability function based on Eq. (41), which, $\Pr$ indicates the probability function, and $f(i)$ denotes the fitness of i^{th} iteration. $mean(f(i))$ denotes the mean of all fitness. If the random variable $rand$ is greater than $\Pr$, the position is updated using Eq. (40), or else, the position is updated using Eq. (42), i.e. the worst solution is replaced by the mean of all solutions. The pseudo-code of the proposed MP-JA method is specified in Algorithm 2.

$$\Pr = \left(rand * f(i) / mean(f(i)) + \left(\frac{1}{it}\right)\right) \tag{41}$$

$$X'_{p,r,i} = X_{p,r,i} + r_1\left(X_{p,best,i} - \left|X_{p,r,i}\right|\right) - r_2\left(X_{p,mean-score,i} - \left|X_{p,r,i}\right|\right) \tag{42}$$

Algorithm 2: Proposed MP-JA Algorithm

Initialize the population
Recognize the best and worst solutions
While $i < Max\ Generation$
    Update the solution depending on best and worst solutions
    Determine $\Pr$ as per Eq. (41)
    If $rand > \Pr$
        Update the position using Eq. (40)
    Else
        Update the position using Eq. (42)
    If the termination condition satisfied
        Report the optimal solution
    else
        Repeat step 2
    end if
end while

5. Results and Discussions

5.1 Simulation Procedure

The implemented MP-JA-based feature selection and classification for FER was simulated in MATLAB, and the corresponding results were obtained. The analysis for the presented FER model was done using the JAFFE database. The implemented FER model was compared over other traditional methods namely, FF [32], PSO [31], GWO [33] WOA [34] and JA [29] in terms of different performance metrics like, Accuracy, Specificity, Sensitivity, Precision, NPV, F_1-Score and MCC, FNR, FPR and FDR, and the corresponding outcomes were obtained. Furthermore, the analysis was done based on six facial emotions of the JAFFE database such as normal, happy, sad, surprise, angry, fear, and disgust as shown in Fig. 3.

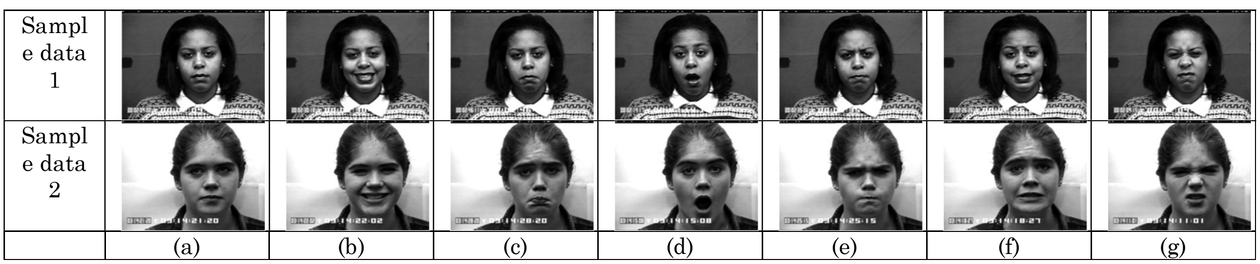


Fig. 3. *Sample images of the JAFFE database for (a) Normal (b) Smile (c) Sad (d) Surprise (e) Angry (f) Fear (g) Disgust.*

5.2 Performance Analysis on Occlusion

The performance analysis of the proposed FER system using MP-JA-based optimal feature selection and classification on considering occlusion is specified in Fig. 4. Here, the experiment is analyzed for 4 ways of occlusion left, right, top, and bottom. From Fig. 4(a), the adopted MP-JA scheme in terms of accuracy at right face occlusion is 5.06%, 3.79%, 5.06%, 6.33%, and 5.06% better than PSO, FF, GWO, WOA, and JA algorithms. From Fig. 4 (b), the sensitivity of the proposed method at right face occlusion is 37.04%, 29.63%, 40.74%, 55.56%, and 44.44% superior to PSO, FF, GWO, WOA, and JA algorithms. From Fig. 4(c), the adopted MP-JA model for specificity at right face occlusion is 3.45%, 2.29%, 3.45%, 4.59%, and 3.45% superior to PSO, FF, GWO, WOA, and JA models. In addition, at top face occlusion, the adopted model for specificity is 3.45%, 2.29%, 3.45%, 4.59%, and 3.45% superior to PSO, FF, GWO, WOA, and JA models. The precision from Fig. 4(d) at right face occlusion is 37.04%, 29.63%, 40.74%, 55.56%, and 44.44% superior to PSO, FF, GWO, WOA, and JA algorithms. From Fig. 4(e), the implemented model for FPR at top face occlusion is 7.69%, 5.38%, 15.38%, 11.54%, and 7.69% superior to PSO, FF, GWO, WOA and JA schemes. Further, from Fig. 4(f), the FNR of the proposed MP-JA model is 12.33%, 9.59%, 13.69%, 19.18% and 15.07% superior to PSO, FF, GWO, WOA and JA algorithms. In addition, the NPV of the adopted MP-JA model is observed from Fig. 4(g), where the presented scheme at right face occlusion is 3.45%, 2.29%, 3.45%, 4.59% and 3.45% superior to PSO, FF, GWO, WOA, and JA models. In addition, at top face occlusion, the adopted MP-JA model for NPV is 3.45%, 2.29%, 3.45%, 4.59%, and 3.45% superior to PSO, FF, GWO, WOA, and JA models. The FDR of the proposed model is attained from Fig. 4(h), which is 12.33%, 9.59%, 13.69%, 19.18%, and 15.07% superior to PSO, FF, GWO, WOA, and JA algorithms. Also, from Fig. 4(i), the F1-score of the presented MP-JA scheme at right face occlusion is

37.04%, 29.63%, 40.74%, 55.56%, and 44.44% superior to PSO, FF, GWO, WOA, and JA algorithms. Finally, from Fig. 4(j), the adopted MP-JA scheme of MCC at right face occlusion is 74.07%, 62.96%, 85.19%, 74.07%, and 99.26% superior to PSO, FF, GWO, WOA, and JA. Therefore, the improvements of the adopted MP-JA technique based on occlusion have been substantiated successfully.

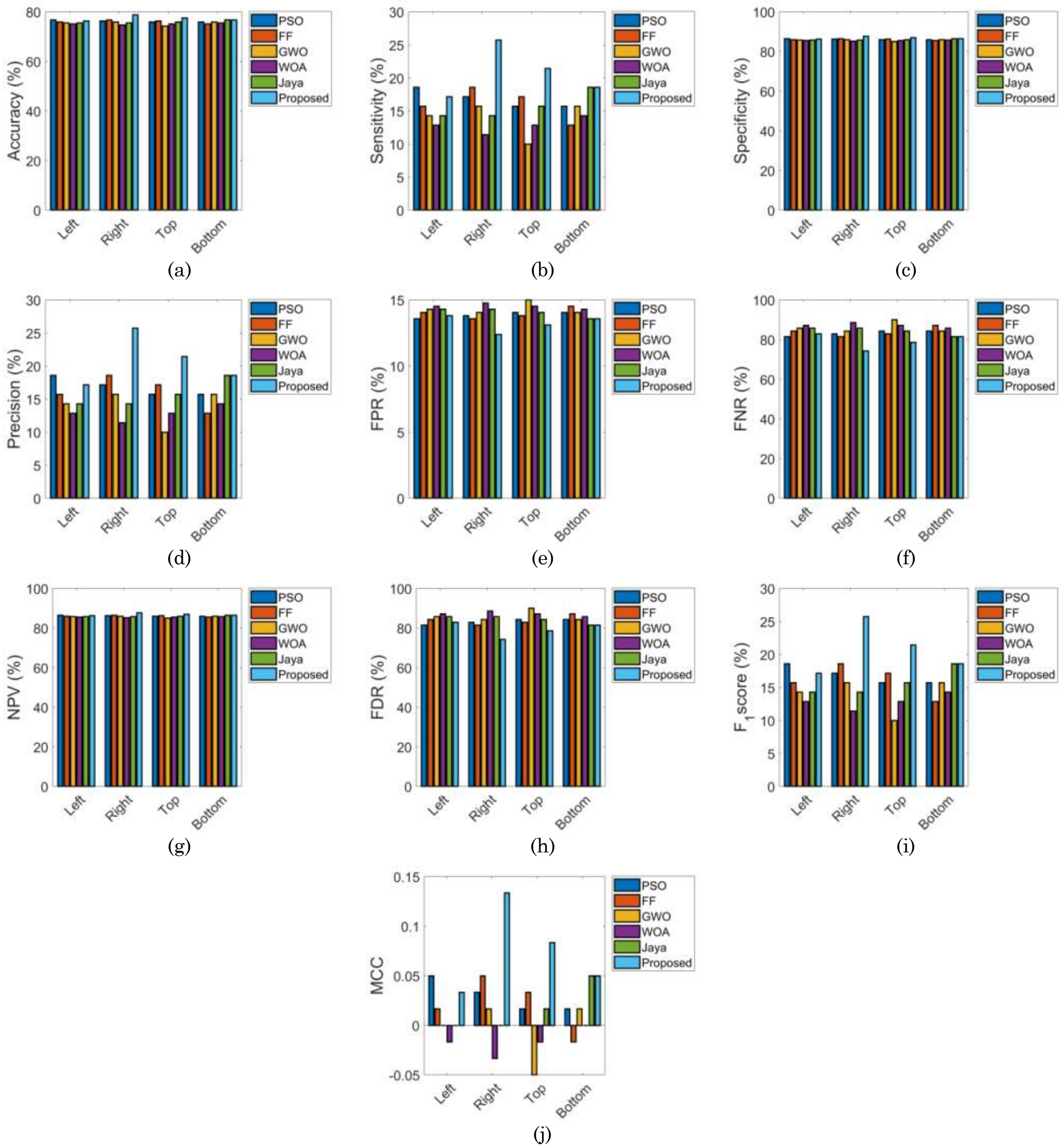


Fig. 4. *Performance analysis of proposed and traditional FER models for different types of occlusion (a) Accuracy (b) Sensitivity (c) Specificity (d) Precision (e) FPR (f) FNR (g) NPV (h) FDR (i) F1-score (j) MCC*

5.3 Overall Performance Analysis

The overall performance of the adopted MP-JA-based FER is given in Table II. From the analysis, the accuracy of the proposed scheme is 2.16%, 2.16%, 4.41%, 1.28%, and 5.33% superior to PSO, FF, GWO, WOA, and JA schemes. The sensitivity of the adopted MP-JA model is 8.77%, 8.77%, 19.23%, 5.08%, and 23.99% better than PSO, FF, GWO, WOA, and JA schemes. Also, the FNR of the adopted scheme is 38.46%, 38.46%, 55.55%, 27.27%, and 59.99% superior to PSO, FF, GWO, WOA, and CS algorithms. The F1-score of the presented MP-JA model is 8.77%, 8.77%, 19.23%, 5.08%, and 23.99% better than PSO, FF, GWO, WOA and JA algorithms. Thus, the superiority of the suggested MP-JA model has been verified effectively.

Table2. *Overall performance analysis of proposed and conventional FER models*

Measures	PSO[31]	FF [32]	GWO [33]	WOA [34]	JA [29]	MP-JA
Accuracy	0.94694	0.94694	0.92653	0.9551	0.91837	0.96735
Sensitivity	0.81429	0.81429	0.74286	0.84286	0.71429	0.88571
Specificity	0.96905	0.96905	0.95714	0.97381	0.95238	0.98095
Precision	0.81429	0.81429	0.74286	0.84286	0.71429	0.88571
FPR	0.030952	0.030952	0.042857	0.02619	0.047619	0.019048
FNR	0.18571	0.18571	0.25714	0.15714	0.28571	0.11429
NPV	0.96905	0.96905	0.95714	0.97381	0.95238	0.98095
FDR	0.18571	0.18571	0.25714	0.15714	0.28571	0.11429
F1-score	0.81429	0.81429	0.74286	0.84286	0.71429	0.88571
MCC	0.78333	0.78333	0.7	0.81667	0.66667	0.86667

6. Conclusion

This paper has presented an enhanced FER model that includes four phases, namely; face detection, feature extraction, optimal feature selection, and classification. In the presented work, face detection was done using the VJ technique, and in the feature extraction process, the features namely, SIFT, DCT, Hough Transform, and Radon Transform were extracted. Owing to the increased length of features, the features were optimally selected using the proposed MP-JA. Further, the optimally selected features were classified using DBN, where the activation function of DBN was optimized using the same MP-JA model. Thus, the various emotions on the face were efficiently determined by the proposed optimal model. From the analysis, the accuracy of the proposed scheme was 2.16%, 2.16%, 4.41%, 1.28%, and 5.33% superior to PSO, FF, GWO, WOA, and JA schemes. The sensitivity of the adopted model was 8.77%, 8.77%, 19.23%, 5.08%, and 23.99% better than PSO, FF, GWO, WOA, and JA schemes. On considering the occlusion analysis, the proposed scheme in terms of accuracy at right face occlusion was 5.06%, 3.79%, 5.06%, 6.33%, and 5.06% better than PSO, FF, GWO, WOA, and JA algorithms. Thus, it has been finalized that the adopted methodology performs better when compared to the other traditional models.

Compliance with Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows the ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

[1] Sukhavasi, S.B., Sukhavasi, S.B., Elleithy, K., El-Sayed, A. and Elleithy, A., “Deep Neural Network Approach for Pose, Illumination, and Occlusion Invariant Driver Emotion Detection”, International Journal of Environmental Research and Public Health, Vol. 19(4), pp.2352, 2022.
[2] Babu, E.K., Mistry, K., Anwar, M.N. and Zhang, L., “Facial Feature Extraction Using a Symmetric Inline Matrix-LBP Variant for Emotion Recognition”, Sensors, Vol. 22(22), p.8635, 2022.
[3] Kumar, P.A., Raj, L.A., Sagayam, K.M. and Ram, N.S., “Expression invariant face recognition based on multi-level feature fusion and transfer learning technique”, Multimedia Tools and Applications, Vol. 81(26), pp.37183-37201, 2022.
[4] Chaudhari, A., Bhatt, C., Krishna, A. and Mazzeo, P.L., “ViTFER: Facial Emotion Recognition with Vision Transformers”, Applied System Innovation, Vol. 5(4), pp.80, 2022.
[5] Devaram, R.R., Beraldo, G., De Benedictis, R., Mongiovì, M. and Cesta, A., “LEMON: a lightweight facial emotion recognition system for assistive robotics based on dilated residual convolutional neural networks”, Sensors, Vol. 22(9), pp.3366, 2022.
[6] Liu, C., Liu, X., Chen, C. and Zhou, K., “Deep Global Multiple-Scale and Local Patches Attention Dual-Branch Network for Pose-Invariant Facial Expression Recognition”, CMES-Computer Modeling in Engineering & Sciences, Vol. 139(1), 2024.
[7] Chatterjee, S., Das, A.K., Nayak, J. and Pelusi, D., “Improving Facial Emotion Recognition Using Residual Autoencoder Coupled Affinity Based Overlapping Reduction”. Mathematics, Vol. 10(3), pp.406, 2022.
[8] Greco, A., Strisciuglio, N., Vento, M. and Vigilante, V., “Benchmarking deep networks for facial emotion recognition in the wild”, Multimedia Tools and Applications, pp.1-32, 2022.
[9] Liliana Lo Presti, Marco La Cascia, “Boosting Hankel matrices for face emotion recognition and pain detection”, Computer Vision and Image Understanding, vol. 156, pp. 19-33, March 2017.

[10] Guillermo Lahera, Sara Herrera, Cristina Fernández, Marta Bardón, Alberto Fernández-Liria, “Familiarity and face emotion recognition in patients with schizophrenia”, Comprehensive Psychiatry, vol. 55, no. 1, pp. 199-205, January 2014.

[11] Deborah J. Wallis, Nathan Ridout, Emma Sharpe, “The influence of non-clinical eating-related psychopathology on the recognition of emotion from static faces and realistic social interactions”, Eating Behaviors, vol. 29, pp. 19-24, April 2018.

[12] Alison Heck, Alyson Chroust, Hannah White, Rachel Jubran, Ramesh S. Bhatt, “Development of body emotion perception in infancy: From discrimination to recognition”, Infant Behavior and Development, vol. 50, pp. 42-51, February 2018.

[13] Carla Liedtke, Waldemar Kohl, Mariska Esther Kret, Katja Koelkebeck, “Emotion recognition from faces with in- and out-group features in patients with depression”, Journal of Affective Disorders, vol. 227, pp. 817-823, February 2018.

[14] Vladimir Kurbalija, Mirjana Ivanović, Miloš Radovanović, Zoltan Geler, Weidong Zhao, “Emotion perception and recognition: An exploration of cultural differences and similarities”, Cognitive Systems Research, vol. 52, pp. 103-116, December 2018.

[15] Umesh Thonse, Rishikesh V. Behere, Samir Kumar Praharaj, Podila Sathya Venkata Narasimha Sharma, “Facial emotion recognition, socio-occupational functioning and expressed emotions in schizophrenia versus bipolar disorder”, Psychiatry Research, vol. 264, pp. 354-360, June 2018.

[16] Yue Ming, “Robust regional bounding spherical descriptor for 3D face recognition and emotion analysis”, Image and Vision Computing, vol. 35, pp. 14-22, March 2015.

[17] Nicolò Zarotti, Jane Simpson, Ian Fletcher, Ferdinando Squitieri, Simone Migliore, “Exploring emotion regulation and emotion recognition in people with presymptomatic Huntington's disease: The role of emotional awareness”, Neuropsychologia, vol. 112, pp. 1-9, April 2018.

[18] Jac N. Airdrie, Kate Langley, Anita Thapar, Stephanie H. M. van Goozen, “Facial Emotion Recognition and Eye Gaze in Attention-Deficit/Hyperactivity Disorder With and Without Comorbid Conduct Disorder”, Journal of the American Academy of Child & Adolescent Psychiatry, vol. 57, no. 8, pp. 561-570, August 2018.

[19] Kevin B. Meehan, Chiara De Panfilis, Nicole M. Cain, Camilla Antonucci, Fabio Sambataro, “Facial emotion recognition and borderline personality pathology”, Psychiatry Research, vol. 255, pp. 347-354, September 2017.

[20] Jingwei Yan, Wenming Zheng, Zhen Cui, Chuangao Tang, Yuan Zong, “Multi-cue fusion for emotion recognition in the wild”, Neurocomputing, vol. 309, pp. 27-35, 2 October 2018.

[21] Marion Timmermann, Haang Jeung, Ruth Schmitt, Sabrina Boll, Sabine C. Herpertz, “Oxytocin improves facial emotion recognition in young adults with antisocial personality disorder”, Psychoneuroendocrinology, vol. 85, pp. 158-164, November 2017.

[22] Carolien Rieffe, Carin H. Wiefferink, “Happy faces, sad faces: Emotion understanding in toddlers and preschoolers with language”, impairments Research in Developmental Disabilities, vol. 62, pp. 40-49, March 2017.

[23] Alla Yankouskaya, Moritz Stolte, Zargol Moradi, Pia Rotshtein, Glyn Humphreys, “Integration of identity and emotion information in faces: fMRI evidence”, Brain and Cognition, vol. 116, pp. 29-39, August 2017.

[24] Amr El Maghraby, Mahmoud Abdalla, Othman Enany, Mohamed Y. El, "Detect and Analyze Face Parts Information using Viola- Jones and Geometric Approaches", International Journal of Computer Applications, vol.101, no.3, pp. 0975 – 8887, September 2014

[25] Nilkanta Sahu, Arijit Sur, “ SIFT based video watermarking resistant to temporal scaling”, Journal of Visual Communication and Image Representation, vol. 45, pp. 77-86, May 2017.

[26] Xin Jin, Qian Jiang, Shaowen Yao, Dongming Zhou, Kangjian He, “:Infrared and visual image fusion method based on discrete cosine transform and local spatial frequency in discrete stationary wavelet transform domain”, Infrared Physics & Technology, vol. 88, pp. 1-12, January 2018.

[27] H.Z. Wang, G.B. Wang, G.Q. Li, J.C. Peng, and Y.T. Liu, " Deep belief network based deterministic and probabilistic wind speed forecasting approach", Applied Energy, vol. 182, pp. 80–93, 2016.

[28] R. Varun, Yadunandan Vivekanand Kini, K. Manikantan, S. Ramachandran, “Face Recognition Using Hough Transform Based Feature Extraction”, Procedia Computer Science, vol. 46, pp. 1491-1500, 2015.

[29] H. M. Pandey, "Jaya a novel optimization algorithm: What, how and why?," 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence), Noida, pp. 728-730, 2016.

[30] Ned C. Rouze, Victor C. Soon, Gary D. Hutchins, “On the connection between the Zernike moments and Radon transform of an image”, Pattern Recognition Letters, vol. 27, no. 6, pp. 636-642, 15 April 2006.

[31] Junhao Zhang, Pinqi Xia, “An improved PSO algorithm for parameter identification of nonlinear dynamic hysteretic models”, Journal of Sound and Vibration, vol. 389, pp. 153-167, 17 February 2017.

[32] Hui Wang, Wenjun Wang, Xinyu Zhou, Hui Sun, Zhihua Cui, “ Firefly algorithm with neighborhood attraction”, Information Sciences, vol. 382–383, pp. 374-387, March 2017.

[33] Seyedali Mirjalili a, Seyed Mohammad Mirjalili, Andrew Lewis, "Grey Wolf Optimizer", Advances in Engineering Software, vol. 69, pp.46–61, 2014.

[34] Seyedali Mirjalili, Andrew Lewis, “The Whale Optimization Algorithm”, Advances in Engineering Software, vol. 95, pp. 51-67, May 2016.