2-D canonical correlation analysis based image super-resolution scheme for facial emotion recognition

Zia ullah[1] · Lin Qi[1] · D. Binu[2] · B. R. Rajakumar[2] · B. Mohammed Ismail[3]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

In this research work, a new Image super-resolution-based Face Emotion Recognition Model has been introduced. The proposed work includes two major phases: (a) Facial image super-resolution and (b) Facial emotion recognition. Initially, the collected facial image is subjected to the facial image super-resolution phase, where the Higher Resolution (HR) facial images are subjected to two-dimensional canonical correlation analysis (2D CCA). The acquired HR facial images are considered as the input for facial emotion recognition. From the acquired HR facial images, the face region alone (lips, eyes, and cheeks) is detected by Viola-Jones facial detection model. Subsequently, from the acquired facial regions, the most relevant features like proposed "Geometric Mean based Weighted Local Binary Pattern (GM-WLBP), Gray Level Co-occurrence Matrix (*GLCM*)" and generalized low-rank model (*GLRM*) features are extracted. Then, Principal Component Analysis (PCA) technique is applied to solve the curse of dimensionality. Finally, the reduced dimensional features are given to the emotion classification phase to classify the emotions as sad, happy, fear, rage, disgust, and surprise. The proposed hybrid classifier framework includes the renowned Long-Short Term Memory Network (LSTM) and Convolutional Neural Network (CNN) models. These deep learning models is separately trained using the dimensionally reduced features, and the outcomes are combined. Then, the mean value is computed on the final combined outcome (output of LSTM+ output of CNN), which results in the type of emotions. To enhance the classification accuracy, the weight of CNN is fine-tuned by a new Improved Tunicate swarm Optimization Model (ITSA), which is the conceptual improvement of standard Tunicate swarm Optimization (TSA). The performance of the proposed work is evaluated over the existing model to show the supremacy of the proposed work.

Keywords Face image · Image super-resolution · Facial expression · 2D CCA · GM-WLBP · CNN · LSTM · ITSA

✉ Zia ullah
  ziaullahkhanhaji@gmail.com

Extended author information available on the last page of the article

1 Introduction

The face emotion-based interaction system among humans and computers has received great attention and concern [5, 7, 22, 25, 50]. The human face plays a significant role in the process of expressing emotions and identity in social situations. In general, humans tend to change their emotions in response to their settings and surroundings (either physical or mental). Despite the wide range of emotions in people, psychology has identified a set of six universal facial expressions, including joy, surprise, sadness, fear, rage, and disgust. These become the foundation for all other expressions [23, 49]. The main facial features are the eyes, eyebrows, nose, and mouth, used to express such emotions in human faces by facial muscle movements. Although humans are skilled at recognizing the facial emotions of known people, they get perplexed when confronted with unfamiliar faces. Besides, the variations in visual information get varied concerning age, the climate, and other natural factors (eyeglasses, beards, hairstyles, whiskers, and so forth). This leads to detection errors. It is observed that face emotions are the crucial component throughout many research areas [17, 41]. Thereby, different strategies have been created: "Feature-based methods and appearance-based methods" [6, 10, 14]. Feature-based techniques are one of the hottest topics [20, 39]. The features relating to human face emotions are derived from geometrical correlations of face attributes such as the chin, eyes, nose, and mouth. However, this approach gets affected due to practical application dependability [8, 38, 47]. The illumination of light in the image pixels can degrade the performance of the face-recognition systems. Faces identified in real-world applications, such as video surveillance, make recognition a challenging one. Face image Super-Resolution (SR), often known as face hallucination, is a suitable solution to resolve this issue [3, 12]. Though certain work handles SR and recognition without generating the HR images, the outcomes acquired weren't satisfactory [26, 27, 48, 52]. So, this would still be necessary to generate a "super-resolved face image from low-resolution inputs". The neural network has laid the cornerstone for the development of deep learning methods because of its outperformance on different applications [11, 40, 43].

Using a camera with acceptable sensor ranges is becoming the simplest way to create the HR image. However, it suffers from a variety of restrictions when it comes to increasing of the percentage of sensors [9, 29, 51]. An alternative option of constructing HR imaging is to post-process the collected photos with signal processing techniques to bring back the omitted features. Image SR is a promising solution to reconstruct the HR pictures from Low-Resolution (LR) pictures that have been down-sampled or blurred. It has a wide range of applications, including video surveillance, medical imaging, and object recognition, and satellite imaging [13, 28]. Different SR technologies have been implemented to overcome this concern, and these techniques could be roughly categorized into interpolation-based and learning-based [4, 16, 18, 33, 34, 44]. The bilinear, bicubic, and spline methods are being the sub-set of the Interpolation-based approaches. In this approach, a single kernel is required for approximating the unknown pixels in HR images based on the characteristics of nearby pixels. Further, the approach is more beneficial concerning computational simplicity, but they are highly prone to blurriness. To model a map amongst LR and HR patches, learning-based approaches such as random forest [30], sparse coding [5, 31, 32, 42], and neighbor embedding [22, 35, 36] are often utilized. However, most of the studies just have learned one-step mapping function from LR to HR images, and hence they aren't able to recover the high-frequency information existing on the edges.

While image sensing technology has progressed considerably in the last several decades, in certain circumstances, such as video surveillance and computer interaction, it's still very difficult to capture high-resolution face pictures. One of its issues that impact the effectiveness of the present facial expression recognition algorithms seems to be the poor resolution of facial images [19]. On the other hand, the SR method maximizes the efficiency of various vision-related activities with the enhanced picture quality. The article, therefore, examines further the prospect of employing the SR technology to enhance the identification of face emotion.

The major contribution of this research work are:

Introduces a new Geometric Mean based Weighted Local Binary Pattern (GM-WLBP) to extract the most relevant features from the detected face regions.

A new hybrid classifier model is modeled for the precise recognition of facial emotions from the input image. The hybrid classifier is constructed by amalgamating the two deep learning classifiers like CNN and LSTM.

Introduces a new Improved Tunicate Swarm Optimization model (ITSA) for fine-tuning the weights of CNN.

The rest of this paper is organized as: section 2 portrays the literature works undergone in this section. Section 3 tells about the proposed super-resolution facial emotion recognition model: an overview. Phase I: image super-resolution and phase II: a proposed model of face emotion recognition using a hybrid classifier is addressed in Section 4 and Section 5, respectively. The results acquired with the proposed work are discussed comprehensively in Section 6. Finally, this paper is concluded in Section 7.

2 Literature review

2.1 Related works

2021, Hu et al. [15] have used the face-specific priors to solve the face super-resolution issues in the photograph industry. The authors have used convolutional neural networks to improve the performance of face super-resolution. These new facial data were subsequently incorporated into the convolutional neural network via training. The findings have demonstrated that the proposed strategy was suitable for a variety of networks.

In 2021, Wang et al. [45] have suggested a unique "CNN-based Dual Closed-Loop Network (DCLNet)" to reduce the potential mapping space. Initially, they have limited the mapping space by including several dual learning networks into a "face super-resolution model". Moreover, the progressive facial prior estimation framework and progressive facial prior estimation framework were determined to guide the facial photograph's super-resolution. Thus, the proposed work had surpassed the existing works in terms of both quantitative and qualitative measurements.

In2019, Amiri et al. [1] have proposed a new approach for enhancing the super-resolution of facial videos. The proposed work attained less computational complexity and could handle the videos with movement patterns. These two benefits had added extra value for security. At first, from the collected video frames, the key points were

extracted from the facial inputs. Then, with the triangular patches extracted, the corresponding pixel in the frame was computed. Finally, the authors have obtained the SR facial images by using the available information in a "small window" located around the LR frames.

In 2021, Liu et al. [24] have proposed a novel Hierarchical Convolutional Neural Network (HCNN) for image SR. The proposed work was made up of a three-step hierarchical process: "edge extraction branch, an edge reinforcement branch, and an extraction branch an edge reinforcement branch, and an image reconstruction branch". In the "edge extraction and reinforcement branch", the authors have extracted the image's Informative edges. The proposed work was tested with several public datasets, and the outcome from them had exhibited the supremacy of the proposed work in terms of high-frequency edge information restoration.

In 2021, Yuan et al. [50] have developed the Orthonormalized Partial Least Squares (OPLS)-SR as a novel FSR learning approach, based on the OPLS. Initially, the authors have used recursive optimization to learn about the latent coherent feature space of low-dimensional HR and LR face embeddings. The LR face images were super-resolved using the facial detail compensation as well as global face reconstruction. The effectiveness of the proposed work was demonstrated in terms of quantitative and qualitative evaluations.

In 2020, Yan et al. [46] have proposed a new low-resolution facial expression recognition from the filter learning perspective. More particularly, the authors have introduced a new Image Filter Based Subspace Learning (IFSL) method to extract facial image representation. The proposed work included three major phases. The image filter learning was embedded into the Linear Discriminant Analysis (LDA's) optimization process. The authors have learned different facial expressions by optimizing the LDA's cost function. Secondly, learned DIF's filtered images were incorporated together for generating the combined images. Eventually, the authors have leveraged the "regression learning technique for subspace learning". Experimental results have shown the superior performance of the proposed IFSL method in terms of "low-resolution facial expression recognition".

In 2020, Nagar et al. [37] have introduced a new position-patch-based face SR method known as RLENR to enhance the quality of the LR facial images that are degraded by Gaussian and impulse noises. Initially, they applied the matrix with pixel-wise noise details and the PCA-oriented mate face onto the LR face image to suppress the IN. Then, the LR training set was updated using the residual-learning therefore, the effect of Gaussian Noise (GN) was found to be minimized. Meanwhile, the respective LR face was hallucinated using the updated LR training set. Thus, the proposed model was observed to be achieved the sparsity and locality for facial information recovery.

In 2020, Lu et al. [30] have designed a novel Global-Local Fused Network (GLFSR) to refine the HF information and to recover the fine details of facial images. The residual HF information was enhanced to a global level from the local level in the networks. The proposed work consisted of 4 sub-modules: reconstruction network for synthetic HR image super-resolution from pixel level, local and global residual enhancement networks for generating the residual information during the learning mechanism, and fusion module for final HR image generation. The experimental results had exhibited the supremacy of the proposed work. The features and challenges of the existing works are depicted in Table 1.

Table 1 Features and Challenges of existing works

Author [Citation]	Adopted methodology	Features	Challenges
Hu et al. [15]	CNNs	➢ Achieves higher PSNR/SSIM. ➢ Even without the fusion of any facial alignment components, this achieves outstanding performance.	✓Requires longer time in the training phase ✓Inaccurate localization of facial components
Wang et al. [45]	DCLNet	➢ Achieves accurate structure and landmark points. ➢ Provides more accurate auxiliary information.	✓Higher MSE ✓Lower PSNR
Amiri et al. [1]	Dense Registration	➢ Provides less computational complexity.	✓Time complexity ✓Lower PSNR.
Liu et al. [24]	Hierarchical Convolutional Neural Network	➢ Can restore high-frequency edges information. ➢ Higher recognition performance.	✓Higher computational complexity. ✓Slow convergence.
Yuan et al. [50]	OPLS-SR	➢ Shows uncorrelated projections with minimum redundancy. ➢ Reduce the noise and computational burden.	✓Higher total computational complexity. ✓Higher computational time and computational cost.
Yan et al. [46]	Image Filter Based Subspace Learning (IFSL)	➢ Preserve useful information. ➢ Reduces the influence of facial identity differences.	✓Suffer from ghosting artifacts.
Nagar et al. [37]	RLENR	➢ Can well preserve the fine details of face images such as sharp edges, the contextual relationship among the neighbor pixels. ➢ Can retain the local spatial structure of the input face image.	✓Higher reconstruction error.
Lu et al. [30]	CNN	➢ Higher consistency between the input image and ground-truth.	✓Lower PSNR and SSIM. ✓Longer running time. ✓Higher time complexity.

3 Proposed super- resolution facial emotion recognition model: An overview

3.1 Architectural description

Image SR and face expression analysis are the two key parts of the proposed study. The architecture is depicted in Fig. 1. The input face image I^{inp}(low-resolution face and higher resolution face) is subjected to image super-resolution phase. Then, the image SR phase is molded using a 2D CCA approach. The suggested face super-resolution methodology incorporates two phases: the first one is facial reconstruction, and the second one is information compensation, which refines the face that was
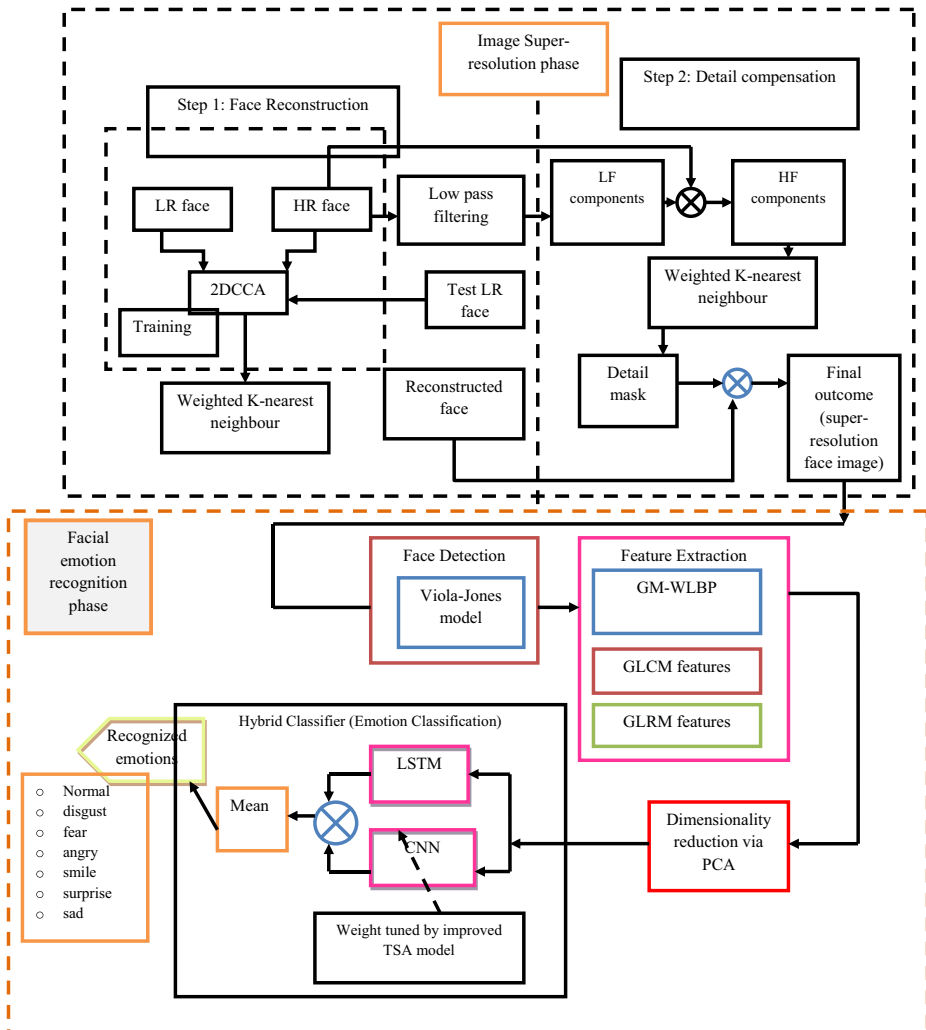
Fig. 1 The architecture of the proposed work

reconstructed in the first stage. At the end of the reconstruction phase, the HR images I^{HR} are formulated, from which the facial emotions can be recognized. The proposed Facial Emotion Recognition (FER) model includes four stages: (i) Face Detection: facial region like mouth, nose, chin, and eyes cheek identification via Viola Jones (VJ) model, (ii) Feature extraction from facial regions via proposed GM-WLBP, *GLCM*, and *GLRM* features, (iii) Feature dimensional reduction via PCA, and (iv) Emotion Classification via hybrid classifier (*CNN* and LSTM), respectively. To enhance the recognition accuracy, the weights of *CNN* are fine-tuned by the ITSA model, which is the conceptual improvement of standard TSA.

4 Phase I: Image super-resolution

4.1 Architectural overview

This is the initial phase, wherein the collected input images I^{inp} (have both LR face $I^{inp(LR)}$ and HR face images $I^{inp(HR)}$) are supplied. The image super-resolution is processed by a 2D CCA model. The suggested face super-resolution algorithm incorporates two phases: (i) facial reconstruction and (b) information compensation. The training set contains HR and LR face images and they are projected to a coherent subspace by 2D CCA in which the correlation between the HR face images and the LR face images is maximized. The HR face image is reconstructed using the same K nearest neighbors in the HR training set based on the structural similarities among the HR and LR facial images. This reconstruction of the input LR face takes place with minimal error. To obtain the super-resolved image, the same neighborhood is applied to the HR training set. Subsequently, following a detailed compensation phase, the high-frequency components are used to rebuild the high-frequency mask. The sum of the rebuilt facial picture and the detail mask is the final output. This is the output HR image I^{HR}.

4.2 Face reconstruction

Training and reconstruction are the two phases of facial reconstruction. The learning of the 2 D CCA model takes place in the training phase. Then, from the learned model, the HR faces are constructed from the LR input.

(i) **Training:** The projection matrices corresponding to the left as well as right are identified, and this helps to project the subspaces of both the LR and HR image. Here, the correlation among them is maximized. Let the HR training set be denoted as $H = h_i$; $i = 1,2,…N$ and LR training set $L = l_i$; $i = 1,2,…N$. The centered datasets $\hat{H}$ and $\hat{L}$ can be obtained by subtracting the mean faces μ_H and μ_L. In addition, the left transforms $left_{\hat{H}}$ and $left_{\hat{L}}$ as well as the right transforms $right_{\hat{H}}$ and $right_{\hat{L}}$ can be acquired by maximizing $\rho = \frac{Cov\left(left_{\hat{H}}^T.\hat{H}.right_{\hat{H}}, left_{\hat{L}}^T.\hat{L}.right_{\hat{L}}\right)}{\sqrt{Var\left(left_{\hat{H}}^T.\hat{H}.right_{\hat{H}}\right)}.\sqrt{Var\left(left_{\hat{L}}^T.\hat{L}.right_{\hat{L}}\right)}}$ Now, $\hat{H}$ and $\hat{L}$ are transformed to $P_H = left_{\hat{H}}^T.\hat{H}.right_{\hat{H}}$ and $P_L = left_{\hat{L}}^T.\hat{L}.right_{\hat{L}}$.

(ii) **Reconstruction:** The LR images $I^{inp(LR)}$ are provided for super-resolution. Onto the subspace, the LR image is projected using Eq. (1).

$$P^{I^{inp(LR)}} = left_{\hat{H}}^T\left(I^{inp(LR)} - \mu\right).right_{\hat{L}} \tag{1}$$

In P_L, the reconstruction of $P^{I^{inp(LR)}}$ can be achieved by the linear combination of its K nearest neighbors in P_H. Then, by minimizing the reconstruction error, the w_j coefficients can be obtained as per Eq. (2). The LR sample set is denoted as $\sum_{j=1}^{K} w_j = 1$. P_{L_j} and the Frobenius norm is computed by $\|FMS\|$. By acquiring the reconstruction weights $\left(w_j\right)_{j=1}^{K}$, the desired HR image $I^{inp(HR)}$ can be projected into the 2D CCA space. The reconstruction is as per Eq. (3).

$$\left(w_j\right)_{j=1}^K \arg \ \min \left\| P^{I^{inp(LR)}} - \sum_{j=1}^K w_j.P_{L_j} \right\|_{FMS} \tag{2}$$

$$P^{I^{inp(HR)}} - \sum_{j=1}^K w_j.P_{H_j} \tag{3}$$

Similar to Eq. (2), $P^{I^{inp(GHR)}}$ relating to $I^{inp(HR)}$ is defined as per Eq. (4).

$$P^{I^{inp(GHR)}} = left_{\hat{H}}^T \left(I^{inp(HR)} - \mu_L \right).right_{\hat{H}} \tag{4}$$

Such that $I^{inp(HR)}$ can be derived as per Eq. (5).

$$I^{inp(HR)} = left_{\hat{H}}^{T\kappa} \left(P_i^{I^{inp(HR)}}.right_{\hat{H}}^\kappa + \mu_L \right) \tag{5}$$

Here, κ denotes the "Moore–Penrose pseudoinverse operation". The whole face reconstruction seems to be a bit challenging. Therefore it is split into three regions: "divide a face into three parts from top to bottom: eyes part, nose part, and mouthpart (for the aligned face images, each part is taken within a predefined region)". Then, the same super-resolution procedure is applied directly to each of the parts. Finally, the outcome can be generated by amalgamating the outcome acquired from these three regions.

4.3 Detail compensation step

The detail compensation step is added to alleviate the issues acquired in the reconstruction phase, such as the loss of information during the data projection and the averaging process of neighborhood reconstruction. At the end of the detail compensation step, the face images with high-frequency details can be obtained.

The Gaussian filter is applied to h_i in the training set of the HR images H. To generate the HF compensation mask HF_{comp}, the computation of the reconstruction weights takes place during the training stage. This HF_{comp} generation is mathematically given in Eq. (6). Here, K is the chosen nearest neighbors. The final outcome can be modeled as per Eq. (7).

$$HF_{comp} = \sum_{j=1}^K w_j.H_j \tag{6}$$

$$I^{HR} = HF_{comp} + I^{inp(HR)} \tag{7}$$

The facial emotions of the persons are identified from the generated I^{HR} image.

5 Phase II: Proposed model of face emotion recognition using hybrid classifier

5.1 The framework of the proposed emotion recognition model

The proposed model includes four stages: "(i) Face Detection, (ii) Feature extraction, (iii) Dimension reduction, and (iv) Emotion Classification". In this context, the face detection from

I^{HR} is done by the Viola Jones method (VJ). It is the first object recognition model to offer better recognition rates in real-time. The detected face regions like mouth, eyes, nose chin and cheek as well are denoted as I^{face}. Further, the features like proposed GM-WLBP $f^{GW-WLBP}$, GLCM f^{GLCM}, and GLRM f^{GLRM} features are extracted from the detected images I^{face}. The combined feature set is $F = f^{GLCM} + f^{GLRM} + f^{GW-WLBP}$. To solve the issue of curse of dimensionality, PCA model is used to obtain the reduced dimensional features. The acquired dimensionally reduced features F^{dim} are used to train the proposed hybrid classifier to classify the emotions. The emotion classification phase is modeled with hybrid deep learning models: CNN and LSTM. Each of these deep learning models is trained separately with F^{dim}. The outcome from CNN (Out^{CNN}) and the outcome from LSTM (Out^{LSTM}) are combined as $(Out) = (Out^{CNN}) + (Out^{LSTM})$. Then, the mean value is computed onto the combined outcome (Out) as $O = mean(Out)$. The outcome (O) portrays the type of emotion ("Normal, disgust, fear, angry, smile, surprise, sad") expressed by the individual in the input image. Moreover, we have fine-tuned the weight of CNN using a new ITSA to enhance the overall classification accuracy. This ITSA model is a conceptual improvement of standard TSA.

5.2 Face identification via Viola-Jones approach

Viola-Jones Model The identification of facial areas is the first stage in the autonomous face recognition model. The viola-jones model undergoes four major stages.

> "(a) Choosing Haar-like features.
> (b) Creating an integral image.
> (c) AdaBoost training.
> d) Establishing classifier cascades."

This method examines numerous tiny subregions of the input image and uncovers the face by searching certain characteristics within each subregion. Since a photograph includes numerous faces with varied sizes, it should verify the distinct locations and scales. To detect the faces, Viola and Jones have employed the Haar-like characteristics.

Haar-Like Features Most of the facial features have similar characteristics. While looking at the image of a person's face, we'll notice that the eye region is darker than the bridge of the nose. The cheeks seem to be lighter than that of the space all around the eyes. These features are used to determine whether the image contains a human face. Summing up and comparing the pixel values of both regions is a straightforward approach to determine whether the region is lighter or darker. The sum of the pixel values in the darker area is minimal than the total pixel values in the lighter area. This can be accomplished using the Haar-like Features, wherein a rectangular section of an image is taken and divided into several portions. They are frequently shown as the neighboring rectangles of black and white. The Viola-Jones algorithm, as previously stated, assesses more properties throughout the subregions of an input image. This rapidly increases the cost of computations. Moreover, employing a computer's limited resources requires more time. "Viola and Jones" used integral images to solve this problem.

Integral Image It's being used to calculate the sum of pixel values in the rectangular section of an image in an efficient manner. The value of each point in an integral image is the summation of all the pixels, which is above and left of the target pixels. The integral image

may be computed in a single pass over the original image. Following that, the relevant and irrelevant characteristics retrieved from the face are detected using Ada boost, a machine-learning methodology that weighs all characteristics. Finally, there is a powerful cascade classifier in the linear combination to blend the multiple facets effectively. The cascade of powerful classifiers promptly eliminates the non-facial and facial parts. Finally, the image's recognized face areas, such as the nose, cheek, eyes, and lips, are represented by the notation I^{face}. Then from the identified facial regions, I^{face} the most relevant features like the (a) improved LBP, (b) GLRM, and (c) GLCM features are extracted as the subsequent process.

5.3 Feature extraction

(i) **Proposed GM-WLBP:** The LBP is a simple and effective statistical texture descriptor with higher discriminative power. It extracts the discriminative feature which is encoded constantly via binomial concatenation among the given referenced pixel and neighbor-hoods. The LBP labeled with decimal values is compared with its neighboring pixels by subtracting the center pixel value during the labeling process. As a result, the decimal pixel values get converted into binary one (i.e. negative values are encoded as 0 and positive value as 1). Further, all of these binary codes are "concatenated clockwise" from the top-left to generate a binary number. These binary numbers are known as Local Binary Patterns or LBP codes. In the LBP approach, the texture descriptor creates numerous local descriptions of the face, which are then combined to make a global description. However, the existing LBP suffers from 3 major drawbacks: (1) They create fairly lengthy histograms that slow down the identification performance, particularly on vast facial images; (2) They overlook local structure in some circumstances because they don't account for the influence of the centre pixel; (3) The binary data they create is susceptible to noise. Therefore, we have introduced a new weighted geometric mean-based LBP model.

Weighted Geometric mean-based LBP It computes the Nth root of the result of the data values of I_i^{face}. This can be mathematically defined as per Eq. (8). We have introduced a new weight function wt to the geometric mean. The neighboring pixels are denoted as X_i and the centre pixel is pointed as X_C.

$$Geometric\ Mean \left(\prod_{i=1}^{N} wt.X_i \right)^{\frac{1}{N}} \tag{8}$$

$$wt = 2*\frac{1}{1+\exp(-E(X_i))} \tag{9}$$

$$E(X_i) = \sum_{t=1}^{\mu(X_i)} Q_t . \log Q_t \tag{10}$$

The operation logic of weighted geometric mean is expressed as per Eq. (11).

$$f^{GW-WLBP} = \sum_{i=1}^{8} B*2^{i-1} \tag{11}$$

Here,

$$\text{Here,} \quad B = \begin{cases} 1 & \text{if } WGM - LBP > X_c \quad WGM - LBP \le X_i \\ 0 & \text{if } WGM \ge X_c \quad\quad\quad WGM > X_i \\ 1 & \text{if } (WGM \\ 0 \end{cases} \tag{12}$$

The extracted GM-WLBP is denoted as $f^{GW-WLBP}$.

GLRM [18] Typically, the texture analysis is performed on the matrix form where the "texture features" are extracted. The GLRM features from I^{face} is denoted as f^{GLRM}. The properties of GLRM are tabulated in Table 2.

GLCM [44] "The matrix that denotes the various grouping distribution over the pixel brightness values i.e. grey level arise within an image is referred as GLCM". The extracted GLCM feature from I^{face} is denoted as f^{GLCM}. The properties of GLCM are tabulated in Table 3.

The extracted features are denoted as $F = f^{GLCM} + f^{GLRM} + f^{ILBP}$. However, these features seem to large and hence are highly prone to "curse of dimensionality issue". Therefore, the dimensions of the extracted features F are reduced via the PCA approach.

5.4 Dimensionality reduction via PCA

The PCA is an unsupervised model that is utilized for reducing the dimensionality of the data (i.e. the data in the higher dimensions are reduced to its lowest dimension without any loss). In this research work, the dimensionality of the extracted features F is demined using PCA. In PCA, the eigenvectors of the correlation matrix are first identified, with the eigenvectors having higher eigenvalues being the most important. It's defined as the transformation of input vectors F in a a-dimensional vector $a = [a_1, a_2, \ldots a_I]^T$ with the same length D into a vector as stated in Eq. (13).

$$b = Y(a - me_a) \tag{13}$$

Table 2 GLRM features

Parameter	Mathematical formula
GLNU	$GL = \frac{1}{N} \sum_H \left(\sum_K p(H,K)^2 \right)$
	In which p denotes the probability of (H,K) in several distances of I^{face}.
HGRE	$HG = \frac{1}{N} \sum_{H,K} p(H,K).H^2$
LGRE	$LG = \frac{1}{N} \sum_{H,K^2} \frac{p(H,K)}{H^2}$
LRE	$LR = \frac{1}{N} \sum_{H,K} p(H,K).K^2$
RP	$R = \sum_{K^2} \frac{N_H}{p(H,K)K}$
Run length Non uniformity	$RL = \frac{1}{N} \sum_K \left(\sum_H p(H,K)^2 \right)$
SRE	$SR = \frac{1}{N} \sum_{H,K} \frac{p(H,K)}{K^2}$

Table 3 GLCM features

Parameter	Mathematical formula
Contrast	$Con = \sum_u \sum_v (u-v)^2 g_{uv}$
Correlation	$Corr = \frac{\sum_u \sum_v (uv)g_{uv} - \mu_u \mu_v}{\sigma_u \sigma_v}$,
	where $\sigma_u, \sigma_v \rightarrow$ std. deviations of g_u, g_v
	$\mu_u, \mu_v \rightarrow$ mean of g_u, g_v
Difference Entropy	$DE = \sum_{u=0}^{N_g-1} g_{u-v}(u) \log\{g_{u-v}(u)\}$
Difference Variance	$DV =$ variance of g_{u-v}
Energy	$Energy = \sum_u \sum_v g_{uv}$
	g_{uv} denotes the $(u,v)^{th}$ entry in I^{face}
Entropy	$Ent = -\sum_u \sum_v g_{uv} \log_2 g_{uv}$
Homogeneity	$Hog = \sum_u \sum_v \frac{1}{1+(u-v)^2} g_{uv}$
Information Measures of Correlation 1	$IMC1 = \frac{HXY - HXY1}{\max\{HX, HY\}}$
Information Measures of Correlation 2	$IMC2 = \sqrt{(1 - \exp[-2.0[HXY2 - HXY]])}$
	where $HXY = -\sum_u \sum_v g_{uv} \log_2 g_{uv}$
	$HXY1 = -\sum_u \sum_v g_{uv} \log_2 \{g_u(u)g_v(v)\}$
	$HXY2 = -\sum_u \sum_v g_u(u)g_v(v) \log_2 \{g_u(u)g_v(v)\}$
MCC (2nd higher Eigen value of Q)$^{0.5}$	$MCC = \sum_k \frac{g(u,k)g(v,k)}{g_u(u)g_v(v)}$
Sum Average	$SA = \sum_{u=2}^{2N_g} u \cdot g_{u+v}(u)$
	Where $N_g \rightarrow$ varied gray levels in image.
Sum Entropy	$SE = \sum_{i=2}^{2N_g} g_{i+j}(i) \log\{g_{i+j}(i)\}$
Sum Variance	$SV = \sum_{u=2}^{2N_g} (u - SA)^2 g_{u+v}(u)$
Variance	$Va = \sum_u \sum_v (u - \mu)^2 g_{uv}$
	Where, $\mu \rightarrow$ mean of g_{uv}

Each a row comprises of F values. The me_a vector indicates the mean vector, and it is computed as per Eq. (14).

$$me_a = Mat\{a\} = \frac{1}{F}\sum_h^D a_h \tag{14}$$

From the covariance matrix CM_a, the Mat matrix is portrayed as per Eq. (14). The determination of CM_a is given by Eq. (15).

$$CM_a = M\left\{(a - me_a)(a - me_a)^T\right\} = \frac{1}{D}\sum_{h=1}^D a_h a_h^T - me_a me_a^T \tag{15}$$

In between $a_l, a_{l'}$, the covariance is specified by Eq. (16).

$$CM_a\left(l, l'\right) = M\{(a_l - me_l)(a_{l'} - me_{l'})\} \tag{16}$$

The count of elements for retaining is specified in Eq. (17), in which VA denotes the Eigenvalues, D points out the count of retained principal components, NU denotes the total count of Eigenvalues.

$$\frac{\sum_{l=1}^{NU} VA_l}{\sum_{h=1}^{D} VA_h} = \alpha \quad where \ 0 \tag{17}$$

The dimensionally reduced features are denoted as F^{dim}. With this F^{dim}, the hybrid classifier in the emotion classification phase is trained.

5.5 Emotion classification via hybrid classifier

In order to recognize the emotions expressed in the collected input image, designed a new hybrid classifier. The proposed hybrid classifier includes CNN and LSTM, respectively. Both the CNN and LSTM are separately trained with the F^{dim}.

(a) CNN: CNN is a well-known deep learning network consisting of three unique layers: a "convolutional layer," a "pooling layer," and "fully-connected layers." In general, the convolution layer contains more convolution kernels that help to evaluate various feature maps. Each neuron in the feature map is specifically related to its neighbors in the preceding layer. In the feature map c, the location of the pixel (a, d) in the lth layer is denoted as per Eq. (18).

$$Z_{a,d,c}^{l} = W_{c}^{l^{T}} V_{a,d}^{l} + Bias_{c}^{l} \tag{18}$$

Here, W_{c}^{l}, $Bias_{c}^{l}$ and $V_{a,b}^{l}$ denotes the weights, biases, and the patched input of cth filter corresponding to l^{th} layer residing at the centred location (a,d). As per the proposed work, the weight function W is fine-tuned by the new ITSA model. Moreover, the activation function $\left(act_{a,d,c}^{l}\right)$ creates non-linearity, which identifies the non-linear properties in multi-layer networks. Equation (19) evaluates the activation value corresponding to the convolutional features $Z_{a,d,c}^{l}$. Pooling layer: "In the DCNN, pooling layers execute downsampling operations using the results obtained from the convolutional layers." The value of $P_{a,d,c}^{l}$ is assessed as shown in Eq. (20) for each pooling function that corresponds to $act_{m,s,c}^{l}$. The output layer is the final layer of CNN, where the categorization takes place. The loss of CNN is represented by $Loss$, and is calculated by Eq. (21). The $Loss$ value has to be minimized, and this is considered to be the fitness or objective while tuning the weights (as given in Eq. (22)).

$$act_{a,d,c}^{l} = act\left(Z_{a,d,c}^{l}\right) \tag{19}$$

$$P_{a,d,c}^{l} = pool\left(act_{m,s,c}^{l}\right), \forall (m, s) \in Y_{a,d} \tag{20}$$

$$Loss = \frac{1}{R} \sum_{r=1}^{R} l\left(\theta; P^{(r)}, O^{r}\right) \tag{21}$$

$$Obj = Min(Loss) \tag{22}$$

The overall parameter associated with W_{c}^{l} and B_{c}^{l} is signified by θ. Here exists R count of input-output relations $\{(V^{(r)}, P^{(r)}); r \in [1, \Lambda, R]\}$. The rth input feature, the corresponding target labels, and the output of CNN is indicated by $V^{(r)}$, $P^{(r)}$ and $O^{(r)}$, correspondingly. The detected emotions using the CNN model is represented by $Outh^{CNN}$. For the weight optimization, the input solutions (weight) are fed as the input to ITSA, which is shown in Fig. 2.

ITSA model for weight optimization The TSA model is based on the jet propulsion of the tunicate swarms. The TSA is good in solving complex engineering problems with higher

precision. The three major conditions that are satisfied by the tunicates are: "avoid the conflicts between search agents, the movement towards the position of best search agent, and remains close to the best search agent". In order to further enhance the TSA model, we've introduced a new updated model in the Converge towards the best search agent phase. The proposed updating might increase the convergence of the solution. The mathematical model of the proposed ITSA model is depicted below:

(i) *Avoid the conflicts among search agents:* Among the search agents, the conflicts can be avoided by employing $\vec{G}$, which is being used to compute the new location of the search agent. This is mathematically shown in Eq. (23). Here, $\overrightarrow{Gravity}$ denotes the gravity force and $\overrightarrow{SoF}$ represents the social force between search agents. In addition, $\overrightarrow{flow}$ points to the flow of water towards the deep ocean.

$$\vec{G} = \frac{\overrightarrow{Gravity}}{\overrightarrow{SoF}} \tag{23}$$

Where,

$$\overrightarrow{SoF} = [D_{min} + e_1.D_{max} - D_{min}] \tag{24}$$

$$\overrightarrow{Gravity} = e_2 + e_3 - \overrightarrow{flow} \tag{25}$$

$$\overrightarrow{flow} = 2.e_1 \tag{26}$$

The variables e_1, e_2, and e_3 refers to the random number ranges among [0, 1]. In addition, D_{min} and D_{max} represents the initial and subordinate speeds to make the social interaction.

(ii) *Movement towards the best neighbor's direction:* Here, the search agent moves towards the best search agent to escape the conflict between them. This behavior is modeled as per Eq. (27).

$$\vec{D}ist = \left| \overrightarrow{food} - rand.\overrightarrow{Y(itr)} \right| \tag{27}$$

Here, *itr* specifies the current iteration, $\vec{D}ist$ denotes the distance among the food source and tunicate. In addition, $\overrightarrow{food}$ portrays the position of the food source and $\overrightarrow{Y(itr)}$ refers to the current tunicate position, and *rand* indicates the random number lies among [0, 1].

Fig. 2 Input Solution to ITSA

(iii)   *Proposed levy-flight-based Converge towards the best search agent:* Traditionally, the search agent could sustain its position towards the best in the convergence phase. However, even though the solution acquired with the existing models is higher in convergence, they still get trapped into the local optima. Therefore, we have introduced a new levy- flight *Levy(β)* based convergence model. The levy flight is a random walk exhibited by the search agent while searching for the best solutions. As a result, global solutions can be acquired. The newly formulated position update model is shown in Eq. (28).

$$\overrightarrow{Y(itr)} = \begin{cases} \overrightarrow{food} + \overrightarrow{G}.\left|\overrightarrow{food} - rand.P_p(u)\right| + \sigma \oplus Levy(\beta) & if \ \ rand \geq 0.5 \\ \overrightarrow{food} - \overrightarrow{G}.\left|\overrightarrow{food} - rand.P_p(u)\right| - \sigma \oplus Levy(\beta) & if \ \ rand \end{cases} \qquad (28)$$

Here, $Y(itr+1) \rightarrow$ indicates the updated position of tunicate in terms and *rand* value is generated using the chaotic map.

(iv)   *Swarm behavior:* The swarm behavior of the search agent is modeled by finding the first two optimal best solutions, and the rest is updated concerning the best search agents. Mathematically, the swarm behavior is modeled as per Eq. (29).

$$Y(itr+1) \rightarrow \ = \frac{Y(u+1) \rightarrow + Y(u) \rightarrow}{2 + e_1} \qquad (29)$$

(b)   **LSTM:** In the recurrent hidden layer, the LSTM has a special unit referred as memory blocks. The memory cells are available within the memory blocks, and these cells are embedded with self-connections for storing the network's temporal state. In addition, the gates being the special multiplicative units for managing the information flow. In the LSTM architecture, there are two gates: an input gate and an output gate. In the memory cell, the input activations-related flows are controlled by the input gate, while the output gate controls the output flow of the activities corresponding to the cells in the leftover parts of the network. In addition, there is a forget gate in the memory block. The forget gate adaptively forgets or resets the cell's memory by scaling the cell's internal state before adding it as input to the cell via the cell's self-recurrent link. Let the hidden state be A and C be the cell states. The input and output layers are shown by (x_t, C_{t-1}, A_{t-1}) and (A_t, C_t). At the time phase t, the output gate is denoted as $Ogate_t$, input gate as $Igate_t$ and forget gate as F_t. The formulation of F_t is given in Eq. (30).

$$F_t = \sigma\left(J_{Igate.F}x_t + S_{Igate.F} + J_{AF}A_{t-1} + S_{AF}\right) \qquad (30)$$

Here, (J_{AF}, S_{AF}) and $(J_{Igate.F}, S_{Igate.F})$ are the weight matrix and bias parameter that map the hidden layer and input layer to the forget gate, respectively. The activation function of the gate

is denoted as σ. Onto the LST cell, the appropriate information is provided by the input gate. This is mathematically shown in Eqs. (31), (32), and (33), respectively. In addition, the notation (J_{AG}, S_{AG}) represents the weight matrix and $(J_{Igate.G}, S_{Igate.G})$ represents the bias parameter that maps the hidden and input layers to the cell gate.

$$G_t = \tanh\left(J_{Igate.G}x_t + S_{Igate.G} + J_{AG}A_{t-1} + S_{AG}\right) \tag{31}$$

$$Igate_t = \sigma\left(J_{Igate.Igate}x_t + S_{Igate.Igate} + J_{A.Igate}M_{t-1} + S_{A.Igate}\right) \tag{32}$$

$$C_t = F_t C_{t-1} + Igate_t G_t \tag{33}$$

In the above equations, the weight and bias parameters map the hidden and input layers to $Igate_t$ are $(J_{A.Igate}B_{A.Igate})$ and $(J_{Igate.Igate}B_{Igate.Igate})$, respectively.

$$Ogate_t = \sigma\left(J_{I.Ogate}x_t + S_{I.Ogate} + J_{A.Ogate}A_{t-1} + S_{A.Ogate}\right) \tag{34}$$

$$M_t = Ogate_t \tanh(C_t) \tag{35}$$

Finally, the LSTM cell obtains the hidden output layer from the output gate, as shown in Eqs. (34) and (35), where $(J_{M.Ogate.}, B_{M.Ogate.})$ and $(J_{Igate.Ogate}, B_{Ogate.Igate.O})$ denote the weight matrix and bias parameter that map the hidden and input layers to O_t, respectively. The recognized emotional outcome acquired from LSTM is denoted as (Out^{LSTM}).

The outcome from CNN (Out^{CNN}) and the outcome from LSTM (Out^{LSTM}) are combined as $(Out) = (Out^{CNN}) + (Out^{LSTM})$. Then, the mean value is computed onto the combined outcome (Out) as $O = mean(Out)$. The outcome (O) portrays the type of emotion ("Normal, disgust, fear, angry, smile, surprise, sad") expressed by the individual in the input image.

6 Results and discussion

6.1 Simulation procedure

The proposed Face emotion recognition model with ITSA+Hybrid classifier model was developed in MATLAB, and the results of each study were recorded. The databases used for this evaluation were the JAFFE database and the CK+ dataset. "The JAFFE dataset contains 213 images of 10 distinct Japanese female participants with various facial expressions. Each subject was requested to make seven facial expressions (6 basic and one neutral), and the photographs were annotated by 60 annotators with average semantic evaluations for each facial emotion. The Extended Cohn-Kanade (CK+) dataset comprises 593 video sequences from 123 distinct people ranging in age from 18 to 50 years old, as well as gender and ethnicity. Each movie depicts a face transition from neutral to a selected peak emotion, captured at 30 frames per second (FPS) in 640x490 or 640x480 pixels resolution. Three hundred and twenty-seven of these videos have been assigned to one of seven expression categories: rage, contempt, disgust, fear, happiness, sorrow, and surprise. The CK+ database,

which is utilized in most facial expression classification methods, is largely recognized as the most frequently used laboratory-controlled facial expression classification database available".

The proposed ITSA+Hybrid classifier was evaluated over the traditional models like LA + Hybrid classifier [40], SLnO+Hybrid classifier [21], PSO + Hybrid classifier, FF + Hybrid classifier, WOA + Hybrid classifier, and TSA + Hybrid classifier in terms of "accuracy, sensitivity, specificity, precision, FPR, FNR, NPV, FDR, F1-score and MCC". The positive measures like "accuracy, specificity, sensitivity, and precision" should be sustained at a higher level for the most favorable results. The error or negative measures are FPR, FNR, and FDR, which must be as low as possible. The F1-score, MCC, and NPV are additional value-added indicators that exhibit the supremacy of the proposed work. The training of proposed data is undergone with 70%, and the rest 30% is used for testing. Figure 3 represents the sample images with seven emotions, namely, Normal, disgust, fear, angry, smile, surprise, sad for JAFFE database and CK+ database.

6.2 Performance analysis

The positive performance evaluation of the ITSA+Hybrid classifier is shown in Fig. 4. The accuracy of the ITSA+Hybrid classifier with the JAFFE database is 92% and the accuracy of the proposed work with the CK+ database is 93%. On observing the proposed work in the case of the CK+ database, it is 2.97%, 3.72%, 8.6%, 11.8%, 10.7%, and 6.5% better than the traditional models like LA + Hybrid classifier, SLnO+Hybrid classifier, PSO + Hybrid classifier, FF + Hybrid classifier, WOA + Hybrid classifier, and TSA + Hybrid classifier, respectively. The specificity of the proposed work with the JAFFE database is 98% and the CK+ database is 98.5%, which are the most favorable outcomes than the existing ones. The sensitivity of the proposed work is 8.54% better than the LA + Hybrid classifier, 3.56% better than the SLnO+Hybrid classifier, 5.6% better than the PSO + Hybrid classifier, 4.1% better than the FF + Hybrid classifier, 6.3% better than the WOA + Hybrid classifier, and 8% better than TSA + Hybrid classifier in the case of the CK+ database. In addition, the precision of the ITSA+Hybrid classifier model in the case of the JAFFE database is 30.41%, 29.23%, 2.7%, 2.17%, 1.8%, and 1.5% superior to traditional models like the LA + Hybrid classifier, SLnO+ Hybrid classifier, PSO + Hybrid classifier, FF + Hybrid classifier, WOA + Hybrid classifier, and TSA + Hybrid classifier, respectively. Thus from the evaluation, it is vivid that ITSA+Hybrid is much appropriate for Facial image super-resolution-based Face emotion recognition model.

On the other hand, the negative measures like FPR, FNR, and FDR are computed for the proposed work with the JAFFE database and CK+ database. The acquired outcomes are

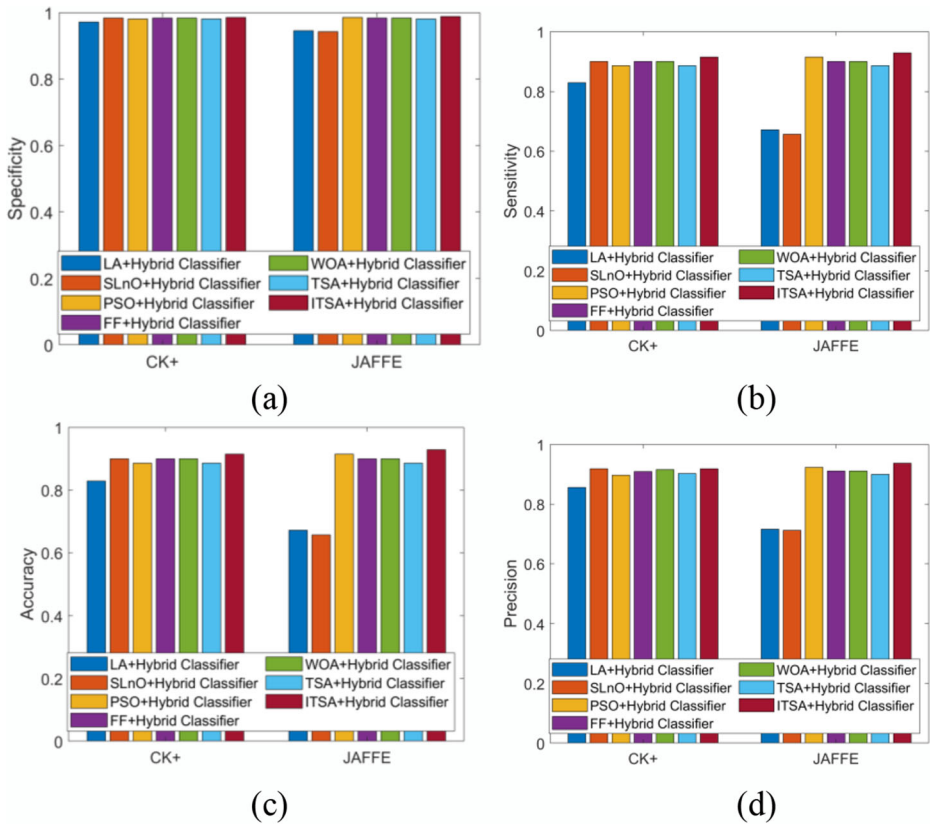| "CK+ database								
	Anger	Disgust	Feat	happy	Neutral	Sadness	Surprise	Super resolution
JAFFE database								
	Anger	Disgust	Feat	happy	Neutral	Sadness	Surprise	Super resolution"

Fig. 3 Sample Image

Fig. 4 Performance of adopted method over extant models for "(**a**) specificity (**b**) sensitivity (**c**) accuracy (**d**) precision"

shown in Fig. 5. The graphical outcomes acquired exhibit the supremacy of the proposed work, as it shows the least error values. Here, the FDR of the proposed work with the JAFFE database is 0.06, which is better than the existing models like LA + Hybrid classifier = 64.28, SLnO+Hybrid classifier = 28.57, PSO + Hybrid classifier = 0.08, FF + Hybrid classifier = 0.09, WOA + Hybrid classifier = 0.09, and TSA + Hybrid classifier = 0.1. On the other hand, the FPR of the proposed work is lower than the existing models in terms of the CK+ database also. In addition, the FNR of the proposed work for the JAFEE database is 0.07, which is 78.12%, 79.41%, 2.8%, 2.6%, 9.3%, and 13.3% better than the traditional models like PSO + Hybrid classifier, FF + Hybrid classifier, WOA + Hybrid classifier, and TSA + Hybrid classifier, respectively. The same improvement is faced by the proposed work when tested with the CK+ database. Thus, from the evaluation, a clear conclusion can be derived that the proposed work had achieved superior performance (fewer error values) in terms of error measures.

In addition, the other measures like NPV, MCC, and F_1-Score are computed for the proposed and existing works with both the JAFFE database and CK+ database. The outcomes acquired are manifested in Fig. 6. Here, the proposed work seems to exhibit better performance (higher values) than the existing works in the case of both the JAFFE database and CK+
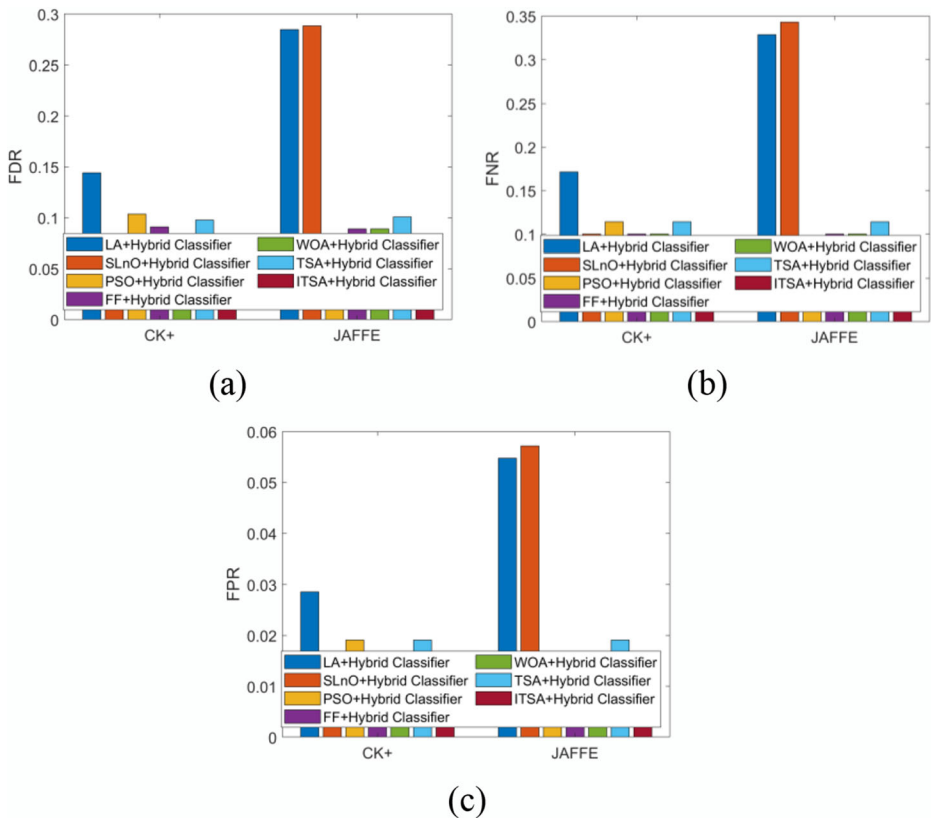
(a)

(b)

(c)

Fig. 5 Performance of proposed method over traditional models for "(**a**) FDR, (**b**) FNR and (**c**) FPR"

database. These results show the influence of super-resolution that makes the emotion classification more precise from the input image.

6.3 Performance analysis on classifiers

The performance of the hybrid classifier in the case of the CK+ dataset and JAFFE database is shown in Tables 4 and 5, respectively. On observing the CK+ database, the accuracy of the hybrid classifier is 0.914, which is better than the hybrid classifier = 0.85, CNN = 0.9, LSTM = 0.9, 2D CCA [2], and NN = 0.857. In addition, the specificity of the hybrid classifier is 0.9142, which is 6.25%, 1.56%, 1.56%, and 0.96% better than the existing models like CNN, LSTM, 2D CCA [2], and NN, respectively. Moreover, the FNR of the hybrid classifier is 0.085714, which is 40%, 14.25%, 14.2%, and 40% better than the existing models like a hybrid classifier, CNN, LSTM, 2D CCA [2], and NN, respectively. Moreover, the FDR of the hybrid classifier for the CK+ database is 0.082107, which is 31.3%, 7.54%, 0.53%, and 31.3% better than the existing models like CNN, LSTM, 2D CCA [2], and NN, respectively. On the other hand, the negative measures like FNR, FDR, and FPR of the hybrid classifier seems to be lower. The FNR of the hybrid classifier is 0.08571, which is the least value while compared to existing models like CNN = 0.14286, LSTM = 0.1, 2D CCA [2]=0.1, and NN = 0.142.
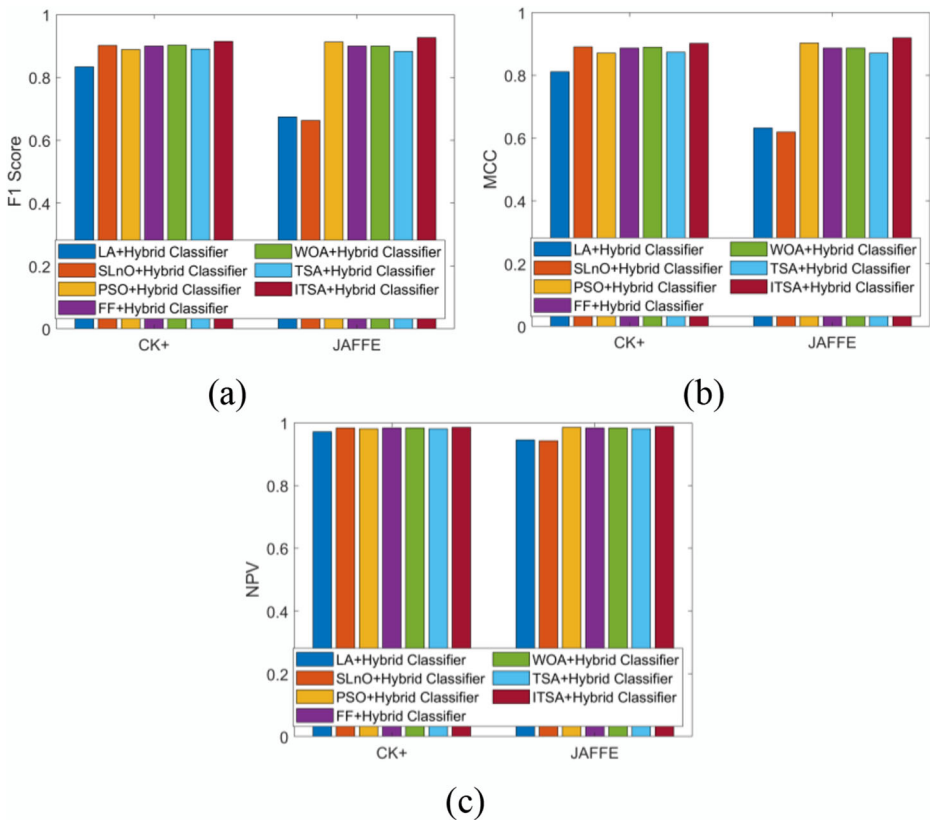
Fig. 6 Performance of the proposed method over the traditional models for "(**a**) F₁-measure (**b**) MCC (**c**) NPV"

Thus, from the overall evaluation, it is clear that the hybrid classifier is the best fitting for the facial image super-resolution based on the Face emotion recognition model.

In addition, the performance of the proposed hybrid classifier, as well as existing classifiers for the JAFFE database, is shown in Table 5. The accuracy of the hybrid classifier is 0.92857, which is 9.23%, 6.15%, 1.5%, and 9.23% better than the existing models like CNN, LSTM, 2D CCA [2], and NN, respectively. In addition, the Specificity of the propose work is 0.9881, better than the existing models like CNN = 0.97381, LSTM = 0.97857, 2D CCA [2]= 0.98571 and NN = 0.97381, respectively. In addition, the precision, as well as sensitivity, shows better performance with the proposed classifier. On the other hand, the FPR, FNR, and FDR of the hybrid classifier is lower than the existing models. The FPR of the hybrid classifier is 0.011905, which is the least value than CNN = 0.02619, LSTM = 0.021429, 2D CCA [2]= 0.014286 and NN = 0.02619. Therefore, from the evaluation, it is clear that the hybrid classifier is much suitable for Face emotion recognition.

7 Conclusion

In this research work, a new Image super-resolution-based Face Emotion Recognition Model was introduced. The proposed work includes two major phases: (a) facial image SR and (b)

Table 4 Performance of the proposed and the existing classifiers for CK+ dataset

Measures	Hybrid Classifier	CNN	LSTM	2D CCA [2]	NN
Specificity	0.98571	0.97619	0.98333	0.98333	0.97619
Sensitivity	0.91429	0.85714	0.9	0.9	0.85714
Precision	0.91789	0.88034	0.91119	0.91833	0.88034
NPV	0.98571	0.97619	0.98333	0.98333	0.97619
MCC	0.90139	0.84213	0.88742	0.89015	0.84213
FPR	0.014286	0.02381	0.016667	0.016667	0.02381
FNR	0.085714	0.14286	0.1	0.1	0.14286
FDR	0.082107	0.11966	0.088811	0.081668	0.11966
F1	0.91482	0.86005	0.9008	0.90181	0.86005
Accuracy	0.91429	0.85714	0.9	0.9	0.85714

facial emotion recognition. Initially, the collected facial image is subjected to the facial image super-resolution phase, where the HR facial images can be obtained by using the 2D CCA. The acquired HR facial images are passed into the facial emotion recognition phase, which is constructed by following four major phases: "(i) Face Detection, (ii) Feature extraction, (iii) Dimension reduction, and (iv) Emotion Classification". Initially, from the acquired HR facial images, the face region alone (lips, eyes, and cheeks) is detected by applying the Viola-Jones facial detection model. Then, from the acquired facial regions, we've extracted the most relevant features like proposed GM-WLBP, *GLCM*, and *GLRM* features. Since the dimensionality of these combined features is huge is size, we've applied the PCA technique to lessen the higher dimensionality features into the lower one. Then, we've constructed a hybrid classifier framework in the Emotion Classification phase to classify the emotions as: sad, happy, fear, rage, disgust and surprise. The proposed hybrid classifier framework includes the renowned LSTM and CNN models. Each of these deep learning models (LSTM and CNN) are trained separately using the dimensionally reduced features, and the outcome from them are combined together. Then, the mean value is computed on the final combined outcome (output of LSTM+ output of CNN). This mean computed result displays the type of emotion. Moreover, to enhance the classification accuracy of the Emotion Classification phase, we've fine-tuned the weight of CNN by introducing ITSA, which is a conceptual improvement of standard TSA. Ultimately, the acquired mean computed results from the proposed work are evaluated over the existing model to show the supremacy of the proposed work. The accuracy of the hybrid classifier is 0.92857, which is 9.23%, 6.15%, 1.5% and 9.23% better than the existing models like CNN, LSTM, 2D CCA [2] and NN, respectively. In addition, the

Table 5 Performance of the proposed and the existing classifiers for JAFFE dataset

Measures	Hybrid Classifier	CNN	LSTM	2D CCA [2]	NN
Specificity	0.9881	0.97381	0.97857	0.98571	0.97381
Sensitivity	0.92857	0.84286	0.87143	0.91429	0.84286
Precision	0.93723	0.85287	0.87847	0.92677	0.85287
NPV	0.9881	0.97381	0.97857	0.98571	0.97381
MCC	0.9191	0.82048	0.85172	0.90426	0.82048
FPR	0.011905	0.02619	0.021429	0.014286	0.02619
FNR	0.071429	0.15714	0.12857	0.085714	0.15714
FDR	0.062771	0.14713	0.12153	0.073227	0.14713
F1	0.92707	0.84405	0.86894	0.91421	0.84405
Accuracy	0.92857	0.84286	0.87143	0.91429	0.84286

Specificity of the propose work is 0.9881, better than the existing models like CNN = 0.97381, LSTM = 0.97857, 2D CCA [2]= 0.98571 and NN = 0.97381, respectively. Therefore, from the evaluation, it is clear that the hybrid classifier is much suitable for Face emotion recognition from SR images.

References

1. Amiri M, Ahmadyfard A, Abolghasemi V (2019) A fast video super resolution for facial image. Signal Process Image Commun
2. An L, Bhanu B (2014) Face image super-resolution using 2D CCA. Signal Process 103
3. Anita JS, Abinaya JS (2019) Impact of supervised classifier on speech emotion recognition. Multimed Res 2(1):9–16
4. Cai J, Han H, Shan S, Chen X (2020) FCSR-GAN: joint face completion and super-resolution via multi-task learning. IEEE Trans Biom Behav Identity Sci 2(2):109–121. https://doi.org/10.1109/TBIOM.2019.2951063
5. Cao L, Liu J, Du K, Guo Y, Wang T (2020) Guided cascaded super-resolution network for face image. IEEE Access 8:173387–173400. https://doi.org/10.1109/ACCESS.2020.3025972
6. Chen L, Pan J, Li Q (2019) Robust face image super-resolution via joint learning of subdivided contextual model. IEEE Trans Image Process 28(12):5897–5909. https://doi.org/10.1109/TIP.2019.2920510
7. Chen J, Chen J, Wang Z, Liang C, Lin C-W (2020) Identity-aware face super-resolution for low-resolution face recognition. IEEE Signal Process Lett 27:645–649. https://doi.org/10.1109/LSP.2020.2986942
8. Chen L, Pan J, Jiang J, Zhang J, Wu Y (2020) Robust face super-resolution via position relation model based on global face context. IEEE Trans Image Process 29:9002–9016. https://doi.org/10.1109/TIP.2020.3023580
9. Chen L, Pan J, Hu R, Han Z, Liang C, Wu Y (Dec. 2020) Modeling and optimizing of the multi-layer nearest neighbor network for face image super-resolution. IEEE Trans Circuits Syst Video Technol 30(12):4513–4525. https://doi.org/10.1109/TCSVT.2019.2917511
10. Chen C, Gong D, Wang H, Li Z, Wong K-YK (2021) Learning spatial attention for face super-resolution. IEEE Trans Image Process 30:1219–1231. https://doi.org/10.1109/TIP.2020.3043093
11. Chen K, Franko K, Sang R (2021) Structured model pruning of convolutional networks on tensor processing units, arXiv preprint arXiv:2107.04191
12. Darekar RV, Dhande AP (2019) Emotion recognition from speech signals using DCNN with hybrid GA-GWO algorithm. Multimed Res 2(4):12–22
13. Farrugia RA, Guillemot C (Sept. 2017) Face hallucination using linear models of coupled sparse support. IEEE Trans Image Process 26(9):4562–4577. https://doi.org/10.1109/TIP.2017.2717181
14. Grm K, Scheirer WJ, Štruc V (2020) Face hallucination using cascaded super-resolution and identity priors. IEEE Trans Image Process 29:2150–2165. https://doi.org/10.1109/TIP.2019.2945835
15. Hu X, Fan Z, Xuan Z (2021) Towards effective learning for face super-resolution with shape and pose perturbations. Knowl-Based Syst
16. Ismail M, Anjum G Reddy TB (2020) Variable block size hybrid fractal technique for image compression. Proceedings IEEE 6th International Conference on Advanced Computing & Communication Systems, pp 510–515
17. Jiang J, Ma J, Chen C, Jiang X, Wang Z (2017) Noise robust face image super-resolution through smooth sparse representation. IEEE Trans Cybern 47(11):3991–4002. https://doi.org/10.1109/TCYB.2016.2594184
18. Jiang J, Chen C, Ma J, Wang Z, Wang Z, Hu R (2017) SRLSP: a face image super-resolution algorithm using smooth regression with local structure prior. IEEE Trans Multimed 19(1):27–40. https://doi.org/10.1109/TMM.2016.2601020
19. Kaura S, Awasthia LK, Sangala AL, Dhimanb G (2020) Tunicate Swarm Algorithm: a new bio-inspired based metaheuristic paradigm for global optimization. Eng Appl Artif Intell 90
20. Koulierakis I, Siolas G, Efthimiou E, Fotinea E, Stafylopatis A-G (2020) Recognition of static features in sign language using key-points. In Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives, pp 123–126
21. Kumar Gola K, Chaurasia N, Gupta B, Singh Niranjan D (2021) Sea lion optimization algorithm based node deployment strategy in underwater acoustic sensor network. Int J Commun Syst 34(5):e4723

22. Li J, Zhou Y, Ding J, Chen C, Yang X (2020) ID preserving face super-resolution generative adversarial networks. IEEE Access 8:138373–138381. https://doi.org/10.1109/ACCESS.2020.3011699
23. Li M, Zhang Z, Yu J, Chen CW (2021) Learning face image super-resolution through facial semantic attribute transformation and self-attentive structure enhancement. IEEE Trans Multimed 23:468–483. https://doi.org/10.1109/TMM.2020.2984092
24. Liu B, Ait-Boudaoud D (2019) Effective image super resolution via hierarchical convolutional neural network. Neurocomputing
25. Liu Z-S, Siu W-C, Chan Y-L (2019) Reference based face super-resolution. IEEE Access 7:129112–129126. https://doi.org/10.1109/ACCESS.2019.2934078
26. Liu L, Wang S, Wan L (2019) Component semantic prior guided generative adversarial network for face super-resolution. IEEE Access 7:77027–77036. https://doi.org/10.1109/ACCESS.2019.2921859
27. Liu Q, Jia R, Zhao C, Liu X, Sun H, Zhang X (2020) Face super-resolution reconstruction based on self-attention residual network. IEEE Access 8:4110–4121. https://doi.org/10.1109/ACCESS.2019.2962790
28. Lu T, Xiong Z, Zhang Y, Wang B, Lu T (2017) Robust face super-resolution via locality-constrained low-rank representation. IEEE Access 5:13103–13117. https://doi.org/10.1109/ACCESS.2017.2717963
29. Lu T, Chen X, Zhang Y, Chen C, Xiong Z (2018) SLR: semi-coupled locality constrained representation for very low resolution face recognition and super resolution. IEEE Access 6:56269–56281. https://doi.org/10.1109/ACCESS.2018.2872761
30. Lu T, Wang J, Zhang Y (2020) Global-local fusion network for face super-resolution. Neurocomputing 387:309–320
31. Mohammad I, Harsha Vardhan V, Aditya Mounika V, Padmini KS (2019) An effective heart disease prediction method using artificial neural network. Int J Innov Technol Explor Eng 8(8):1529–1532
32. Mohammed Ismail B, Shaik MB, Reddy BE (2012) High rate compression based on luminance & chrominance of the image using binary plane technique. J Theor Appl Inf Technol 42(2):191–195
33. Mohammed Ismail B, Shaik MB, Reddy BE (2015) Improved fractal image compression using range block size. Proceedings of IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS), pp 284–289
34. Mohammed Ismail B, Reddy TB, Reddy BE (2016) Spiral architecture based hybrid fractal image compression. IEEE 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)
35. Mohammed Ismail B, Rajesh P, Alam M (2020) A machine learning based improved logisticregression method for prostate cancer diagnosis. Int J Emerg Trends Eng Res 8(9):5693–5698
36. Mohammed Ismail B, Alam M, Tahernezhadi M, Vege HK, Rajesh P (2020) A machine learning classification technique for predicting prostate cancer. 2020 IEEE International Conference on Electro Information Technology (EIT) July 2020, pp 228–232
37. Nagar S, Jain A, Kumar A (2020) Mixed-noise robust face super-resolution through residual-learning based error suppressed nearest neighbor representation. Inf Sci 546:121–145
38. Pei X, Dong T, Guan Y (2019) Super-resolution of face images using weighted elastic net constrained sparse representation. IEEE Access 7:55180–55190. https://doi.org/10.1109/ACCESS.2019.2913008
39. Potamias R-A, Siolas G, Stafylopatis A (2019) A robust deep ensemble classifier for figurative language detection. In: International conference on engineering applications of neural networks. Springer, Cham, pp 164–175
40. Rajakumar B R (2018) Optimization using lion algorithm: a biological inspiration from lion's social behavior. Evolutionary Intelligence, Special Issue on Nature inspired algorithms for high performance computing in computer vision, Vol. 11, No. 1–2, pages 31–52, DOI: https://doi.org/10.1007/s12065-018-0168-y
41. Sarkar A (2020) Optimization assisted convolutional neural network for facial emotion recognition. Multimed Res 3(2)
42. Shahne R, Ismail M, Prabhu CSR (2019) Survey on deep learning techniques for prognosis and diagnosis ofCancer from microarray gene expression data. J Comput Theor Nanosci 16(12):5078–5088
43. Sharma S Self supervised methods towards human activity recognition. IOSR Journal of Computer Engineering (IOSR-JCE) 22(6):51–56
44. Shi J, Liu X, Zong Y, Qi C, Zhao G (June 2018) Hallucinating face image by regularization models in high-resolution feature space. IEEE Trans Image Process 27(6):2980–2995. https://doi.org/10.1109/TIP.2018.2813163
45. Wang H, Hu Q, Wu H (2021) DCLNet: dual closed-loop networks for face super-resolution. Knowl-Based Syst 222:106987
46. Yan Y, Zhang Z, Wang H (2020) Low-resolution facial expression recognition: a filter learning perspective. Signal Process 169:107370

47. Yang S, Liu J, Fang Y, Guo Z (2018) Joint-feature guided depth map super-resolution with face priors. IEEE Trans Cybern 48(1):399–411. https://doi.org/10.1109/TCYB.2016.2638856
48. Yu X, Porikli F (2018) Imagining the unimaginable faces by Deconvolutional networks. IEEE Trans Image Process 27(6):2747–2761. https://doi.org/10.1109/TIP.2018.2808840
49. Yu X, Fernando B, Hartley R, Porikli F (2020) Semantic face hallucination: super-resolving very low-resolution face images with supplementary attributes. IEEE Trans Pattern Anal Mach Intell 42(11):2926–2943. https://doi.org/10.1109/TPAMI.2019.2916881
50. Yuan Y-H, Li J (2021) Furong Peng," OPLS-SR: a novel face super-resolution learning method using orthonormalized coherent features". Inf Sci 561:52–69
51. Yun JU, Jo B, Park IK (2020) Joint face super-resolution and Deblurring using generative adversarial network. IEEE Access 8:159661–159671. https://doi.org/10.1109/ACCESS.2020.3020729
52. Zhang Y, Tsang IW, Li J, Liu P, Lu X, Yu X (2021) Face hallucination with finishing touches. IEEE Trans Image Process 30:1728–1743. https://doi.org/10.1109/TIP.2020.3046918

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

Affiliations

Zia ullah[1] · Lin Qi[1] · D. Binu[2] · B. R. Rajakumar[2] · B. Mohammed Ismail[3]

[1] Zhengzhou University, Zhengzhou, China

[2] Resbee Info technologies Private Limited, Thuckalay, Tamil Nadu, India

[3] Department of Information Technology, Kannur University campus, Mangattuparamba, Kerala, India