Journal of Engineering and Applied Sciences
MEDWELL PUBLICATIONS

Taylor-Swarm Ganet: Learning Illumination Invariant Feature Descriptor For Facial Expression Recognition using Deep Generative Adversarial Network

[1]Priyanka A. Gavade, [2]Vandana S. Bhat and [3]Jagadeesh Pujari
[1]*Department of ISE, SDM College of Engineering, Dharwad-02, Karnataka*
[2]*Department of Information Science and Engineering, SDM College of Engineering, Dharwad-02, Karnataka*
[3]*SDM College of Engineering and Technology, Dharwad-02, Karnataka*

Key words: Facial Expression Recognition (FER), Deep Generative Adversarial Network (GAN), video March 23, compression, Taylor series, Viola Jones algorithm

Corresponding Author:
Priyanka A. Gavade
Department of ISE, SDM College of Engineering, Dharwad-02, Karnataka

Page No.: 208-221
Volume: 16, Issue 6, 2021
ISSN: 1816-949x
Journal of Engineering and Applied Sciences

Abstract: Facial expression is the nonverbal way to express the human intensions and emotions. Facial Expression Recognition (FER) intends to understand and analyze the facial behavior of humans such that it has become an active research area in the field of pattern recognition, artificial intelligence and computer vision. Various FER methods are developed for classifying the facial expression in video sequences but to extract the discriminative video features from the facial expression images results a key challenging issue in FER system. Hence, an effective FER method is designed using proposed Taylor-Chicken Swarm Optimization-based Deep Generative Adversarial Network (Taylor-CSO based Deep GAN) for the recognition of facial emotions. However, the proposed method named Taylor-CSO is derived by the integration of Taylor series with Chicken Swarm Optimization (CSO), respectively. The process of Illuminant Invariant Local Binary Pattern (IILBP) is made by employing the LBP descriptor to the facial object. Based on the feature matrix, the process of FER is accomplished using Deep GAN. However, the proposed approach achieved the accuracy, precision and recall of 0.8846, 0.8996 and 0.8952 with respect to training data.

INTRODUCTION

Most of traffic in the telecommunication network is mainly connected to video in recent decades and this traffic factor can be steadily increases in future. However, it forces the requirement for efficient compression of video for minimizing the cost of data transmission[1]. Video coding is an active research topic as new standard is evolved for video compression in the market for each year. For multiple encoding and decoding the material is significantly essential for post production and the workflow of video editing. There exists a huge demand for the continuous delivering of video with high quality. However, the video content has reaches a significant portion of network traffic in worldwide and is still enhancing[2]. As we move forward from one generation to other, number of technologies is enduring us in accordance to our requirement. Hence, these technologies are acts as the baseline for human computer interaction. One among the recent technology is Facial Expression

Recognition (FER). Face plays a key factor in the social communication such that facial expression is a very important thing. The face biometric is widely used in different appliances such as forensic, security and other commercial appliances. Facial expression is the fastest way of data communication that conveys any categories of information[3]. The facial expression not only explains the feelings or sensitivity of the person but also it can be considered to judge the mental views.

Human reveals various categories of facial expression based on the status of mind. An accurate analysis of the human emotions helps the machine to provide the accurate response. The different applications, like car driving, software considers the facial emotion information of driver to take the correct action. Facial expression is the expressive and natural non-verbal channel for the humans to transfer their emotions[4, 5]. The system modeled to automatically analyze the facial emotions using the human computer interaction is terms as Automatic FER system (AFERS). Hence, FER plays an active and important area in the wide variety of appliances such as health care, biometrics, human computer interaction, robot systems and digital entertainment. Ekman and Friesen suggested university of Neural (Ne) and the six categories of human facial actions, like Surprise (Su), Sadness (Sa), Happiness (Ha), Disgust (Di), Anger (An) and Fear (Fe)[6]. In general, the facial emotions are classified into two categories, such as micro expression and macro expression. However, the formal expression can lasts among three quarters of second to two seconds whereas movements of muscle are simultaneously happened at various parts of face. Hence, macro expressions are professed by the humans at real time discussions[7, 8]. The FER considers different disciplines, like behavioral science, psychology, emotional computing, computer science, and artificial theory. However, it gains a practical value and significant in the fields of distance education, safe driving, human computer interaction system and character animation[5].

In the pattern recognition field, classifying the facial emotions of humans results a major issue for the researchers[9, 10]. The FER techniques are classified into two types, namely static or image based techniques and dynamic or sequence based techniques[6]. Based on the representation of features, the FER system is categorized into two different types, namely dynamic and static image FER. In the static based technique, the representation of features is encoded by the spatial data from single image, while the dynamic based model[14, 4, 8] focus on temporal relation between the contiguous frames of facial emotion sequence. By considering vision based techniques, some other modalities, like physiological and audio channels are widely utilized in the multimodal systems[5] for assisting the recognition of facial expression. Different machine learning methods are widely used for classifying the facial emotions. The Support Vector Machine (SVM)[12, 10] is the commonly used model for the binary and multiclass classification[13]. With respect to local minima solution of the Artificial Neural Network (ANN), the SVM provide global minima solution for the optimization problem. Most of the research words focused on the extraction of features from the facial images by considering the classification model to automatically detect the facial expression. The existing researchers considered Principal Component Analysis (PCA) to perform face recognition and facial expression[14, 10]. The methods such as wavelets[15] and Linear Discriminant Analysis (LDA)[16, 10] are also considered for FER. The constructive feed forward neural network model is designed in[11, 10] for the recognition of facial motions. The dynamic classifier considers the temporal patterns for showing facial expressions. Most of the recent words used Hidden Markov Model (HMM) for classifying the facial expression[5].

This research is modeled using proposed Taylor-CSO based Deep GAN for the recognition of facial expression. The proposed approach involved various phases such as frame extraction, pre-processing, face detection, feature extraction and expression recognition. Initially, the video frames are acquired from input video sequences collected from dataset. The video frames are allowed to the pre-process phase where individual video frame is pre-processed using ROI based extraction. The pre-processed result is fed to the face detection phase where the process of facial object detection is carried out using Viola Jones algorithm. The face object is employed to the LBP descriptor by varying the illumination intensity and the resulted factor is multiplied with the weight value in order to generate the feature matrix of IILBP. The final phase of FER is the classification phase where Deep GAN is used to detect the facial expressions in such a way that the training process of Deep GAN is carried out using Taylor-CSO algorithm which is designed by the incorporation of Taylor series with CSO algorithm, respectively. The major contribution of the research is explained as follows:

Proposed Taylor-Swarm GANet: An effective recognition model is designed to classify the facial expressions using proposed Taylor-CSO based Deep GAN. The extraction of IILBP feature is made by employing the face object to the LBP descriptor. Based on the feature matrix, the Deep GAN performs the recognition process more effectively.

Motivation: In this study, different FER techniques are explained along with their merits and drawbacks that motivate the researchers to design Taylor-CSO based GAN for the recognition of facial expression.

Literature reveiw: Various FER techniques are reviewed in this study. Yan[17] introduced a CDMML method for the recognition of facial expression using video. For each video frame, multiple feature vectors were calculated for describing the facial exterior as well as the motion data from different aspects. Here, the distance measures were learned using the features for revealing the discriminative and complementary information for the recognition process. However, this method was very effective, but failed to consider efficient feature learning schemes for increasing the performance of system. Hu *et al*.[18] introduced a LEMHI-CNN classifier for FER using video sequences. Here, a local and a global network were integrated based on the motion history image. The CNN-based Long Short Term Memory (LSTM) was used as the feature extractor as well as the classifier to recognize the facial emotions from video in global network. The prediction result was generated using random search weighted summation model. It achieved better accuracy but require more number of iterations. Cruz *et al*.[19] developed the TPOEM features to explore the temporal derivatives and the adjacent frames. Here, the features were calculated within the non overlapping image patches and based on the score of each patches, the final classification result was generated. This method showed better accuracy but failed to integrate the features with the deep learning model for learning spatial relationship among image patches. Makhmudkhujaev *et al*.[20] introduced an edge based descriptor termed as Local Prominent Directional Pattern (LPDP) that considered statistical data of pixel neighborhood for encoding reliable information to extract the features. The LPDP was used to monitor the neighborhood of pixel for retrieving the edges that correspond to local shape. It generated solid and clear pattern codes but failed to enhance the performance of system.

Fernandez *et al*.[21] developed an end-to-end architecture for recognition of facial expression. The image correction and the classification component utilized the encoder decoder network and the feature extractor to generate the feature map. The facial expression component was used to generate the embedded representation over representation space. The classification performance was increased by the attention module but this method failed to consider large sized database. Richhariya and Gupta[10] developed a Iterative Universum Twin Support Vector Machine (IUTWSVM) based on Newton model to perform multiclass classification. This method generated better performance with limited computation cost. The effectiveness of this method was analyzed with the real time datasets. It achieved better performance but required various smoothing methods and convergence methods for increasing the performance of system. Zhang *et al*.[2] developed a SVM for classifying the facial expression. The segment level spatial as well as the temporal features was acquired by CNN and the results were fused using Deep Belief Network (DBN) to learn the discriminative features. The average pooling was achieved based on segment level features using video sequences. This method increased the accuracy but failed to consider deep compression of deep models. Liu *et al*.[6] developed an Identity-Disentangled Facial Expression Recognition Machine (IDFERM) for FER. However, this method solved the threshold tuning issues and the anchor selection problem using deep metric learning scheme.

The gray scale image may lead to loss the information and the quality of image was not considered in this method. However, it reduced the computation time, but failed to apply visual quality assessment model.

Challenges: Some of the issues faced by the traditional FER methods are explained as follows:

To design the real time FER system by considering the hybrid deep learning model poses a major challenge and it is more interested to reveal the deep compression of the deep model for minimizing the factors of deep learning model in the large scale network[2].

A major issue faced by the recognition process is the orientation of face and the variation of size in input images. Due to camera angle, the pose of face may be different and it also shows some facial features[22].

An automatic FER system based on the multimodal sensor data does not analyze achievability and feasibility for the detection of emotions and also the effectiveness was not analyzed with real time experiments[23]. The accuracy measure achieved by LEMHI-CNN model was still unsatisfactory with AFEW dataset[18].

As, faces are partially occluded by some other objects, recognition of facial expression result a challenging task. To extract the features from facial expression is more complex while the face is occluded by other objects, like glasses, hair and mask[10].

Due to complexity and variety of facial expression, FER poses a challenging task as the facial movement varies with respect to time. The complexities of illumination variance make the FER process more complex. The natural acquisition of image faces the issue of improper or discriminative image collection such that it poses partial capturing, facial information loss, and pose variation.

MATERIALS AND METHODS

Proposed Taylor-chicken swarm optimization-based deep generative adversarial network for facial expression recognition: Facial expression is a natural nonverbal channel for the humans to converse the

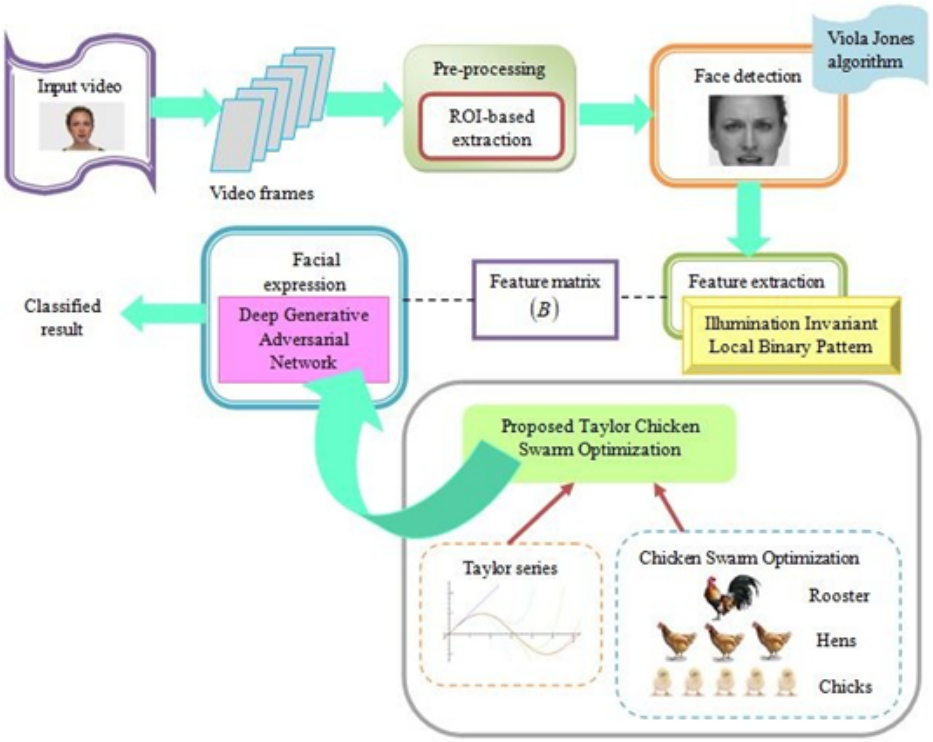

Fig. 1: Schematic view of proposed Taylor-CSO based Deep GAN for recognition of facial expression

emotions. This research focused to design the FER method using Taylor-CSO based Deep GAN. The steps involved in the proposed method are extraction of video frames, pre-processing, face detection, feature extraction, and FER phase. The input video sequences collected from the dataset are allowed to process of frame extraction step where the video frames are extracted using video sequences. The pre-processing stage is done by employing ROI based extraction. The process of face detection is accomplished using Viola Jones algorithm. The feature extraction phase is done by IILBP feature that is derived by modifying LBP descriptor. Based on the feature matrix, the process of FER is carried out using Deep GAN which is tuned by proposed Taylor-CSO algorithm. However, the developed Taylor-CSO is the integration of Taylor series[24] and CSO[25-27]. Figure 1 shows the schematic view of proposed developed for FER system.

Acquisition of input video and video frames: The classification of facial expression from the consecutive frames in the video is more common as the video sequence offers the information for FER than the static facial images. Let us consider the dataset as D and Y be the training set of facial videos and is expressed as:

$$D = \{Y_1, Y_2, ..., Y_i, ..., Y_n\} \quad (1)$$

where, $Y_i \in R^{u \times v}$, i = {1,2, ..., n} n denotes total number of samples and [u×v] specifies the dimension of each sample. The facial video frame of each sample is considered as V and is represented as:

$$V = \{V_1, V_2, ..., V_j, ..., V_m\} \quad (2)$$

Here, j = {1,2, ..., m} and m specifies total number of video frames with the dimension of [u×v], respectively.

Pre-processing using ROI based extraction: The input video frame V_j is fed to the pre-processing phase, where the video frame is effectively pre-processed by employing ROI based extraction. The aim of ROI extraction is to extract a specific region of facial image from the video frame V_j in order to perform FER. By selecting the exact region, the computational complexity is effectively reduced. The outer region of the video frames is eliminated and in turn generates the object by removing the external artifacts at the pre-processing stage. Here, the process of face cropping is done to eliminate the useless region in the image. The pre-processed result of video frame V_j is represented as P_j which is passed to the faced detection phase for detecting the face region.

Face detection by Viola Jones algorithm: After the completion of pre-processing step, the face detection stage begins by employing Viola Jones algorithm[28] to detect the face region more accurately. The Viol Jones algorithm contains three schemes for detecting the face object from pre-processed result P_j.

It is the learning based model that is specifically used for the detection of objects. Here, the face detection process is carried out using Viola Jones algorithm. This algorithm uses Haar features and the classifier for identifying the objects. However, Haar features are calculated by integral image such that best features are selected by Adaptive Boost (Adaboost) algorithm.

Haar features: The video frame P_j is partitioned into rectangular region or small windows with the size of [M×M] The features are individually computed for each window. In general, three categories of features such as two-rectangle features, three-rectangle features and four-rectangle features are considered for face detection. However, the two-rectangular feature is computed based on the difference among sums of pixels inside two rectangular region and these rectangular regions have similar size and shape and are neighboring to each other vertically or horizontally. The three-rectangle feature computes sums of pixel of two outside rectangles and is subtracted from sum of pixels of center rectangle. The four- rectangle feature is computed based on the dissimilarity among diagonal pairs of rectangle.

Formation of integral image: The intermediate representation for image calculated using Haar feature is called as integral image. The mathematical model used to compute the integral image is expressed as:

$$I(U, V) = \sum_{\omega=1}^{U}\sum_{\rho=1}^{V} P(\omega, \rho); 1 \le U \le M; 1 \le V \le M \quad (3)$$

Where:
I = Integral image
M = Specifies the dimension of rectangular region
P = Pre-processed image

Adaboost algorithm: This algorithm is employed to minimize the redundancy of features computed using Haar. It is the learning classification function utilized to eliminate the redundant features and transform the large sized features into compact one. Finally, the detected face object is represented as A.

Feature extraction: The face object detected from the video frame is passed as input to feature extraction phase, where the features, like IILBP is effectively acquired to achieve the process of FER. The IILBP feature is extracted by applying Local Binary Pattern (LBP) descriptor to the face object. The process of extracting IILBP feature involves four steps, namely applying intensity variation, applying LBP descriptor, compute IILBP feature and computation of feature matrix. At first, the face object is selected by varying the intensity pixel value. In addition to the original face detected object, four objects are selected by varying the intensity value. By varying the intensity with the value of '-50' to the face object A is represented as a_1, the intensity value varied by '+50' to A is denoted as a_2, the intensity value varied by '+75' to A is denoted as a_3 and the intensity value varied by '-75' to A is specified as a_4, respectively. At the second step, the LBP descriptor is individually applied to A, a_1, a_2, a_3 and a_4 , respectively. After LBP is applied, the result obtained from each face object is specified as d_1, d_2, d_3, d_4 and d_5 in such a way that these results are multiplied with the weight factor at third step and is indicated as αd_1, $\alpha/2d_2$, $\alpha/3d_3$, $\alpha/4d_4$ and $\alpha/5d_5$, respectively.

Finally, the feature matrix of IILBP feature is generated at the fourth step and is expressed as:

$$B = \alpha d_1 + \frac{\alpha}{2} d_2 + \frac{\alpha}{3} d_3 + \frac{\alpha}{4} d_4 + \frac{\alpha}{5} d_5 \quad (4)$$

where, B denotes the feature matrix with the dimension of [280×256]. Figure 2 shows the steps of IILBP feature extraction phase.

Facial expression recognition using proposed Taylor-Swarm GANet: After the computation of feature extraction process using, the process of FER is accomplished by employing proposed Taylor-CSO based Deep GAN. The Deep GAN takes the feature matrix as input to perform the recognition process of facial expression. The features generated by the IILBP descriptor is termed as feature matrix such that it is extracted based on the LBP descriptor and weighted factor with the varying intensity pixel value.

Structure of deep GAN: Deep GAN[29] is the latent variable generative scheme that effectively generates the classification result from the feature matrix through the adversarial process. The major idea behind GAN is the simultaneous training of generator H and discriminator M. The discriminator is used to distinguish the real samples from the fake samples generated by H. The generator takes the random noise *b* as input and it effectively trained for generating false samples. The major benefit of Deep GAN is that it does not require any labeled data as the training is done by the unlabeled data. This feature effectively minimizes the training speed of the classifier. Figure 3 portrays the structure of Deep GAN.

The GAN comprises two components, namely generator H and discriminator M that follow the practice of two player min-max game:

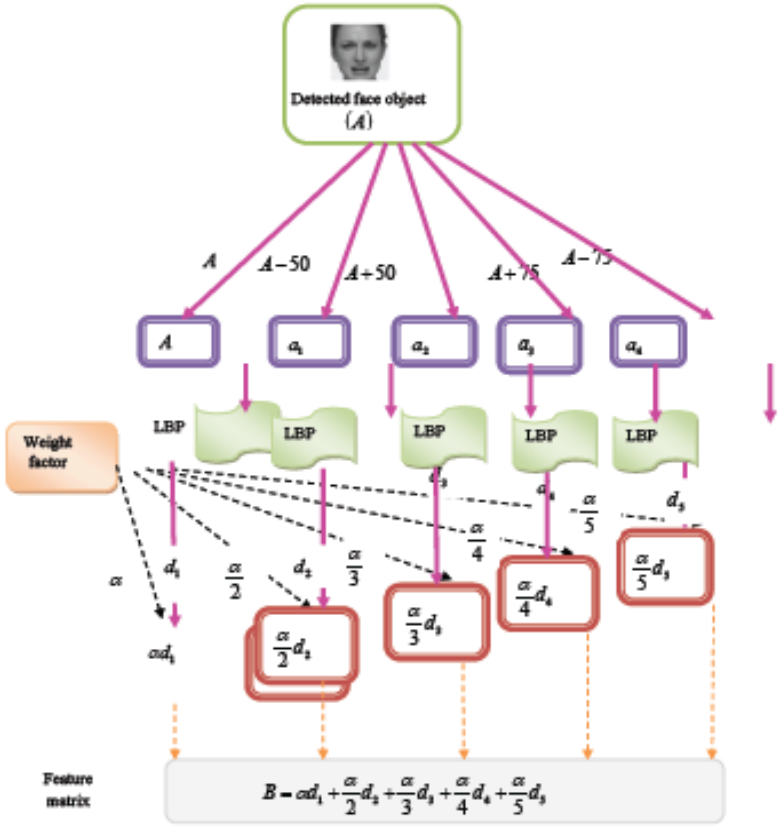

Fig. 2: Structure of IILBP feature extraction phase

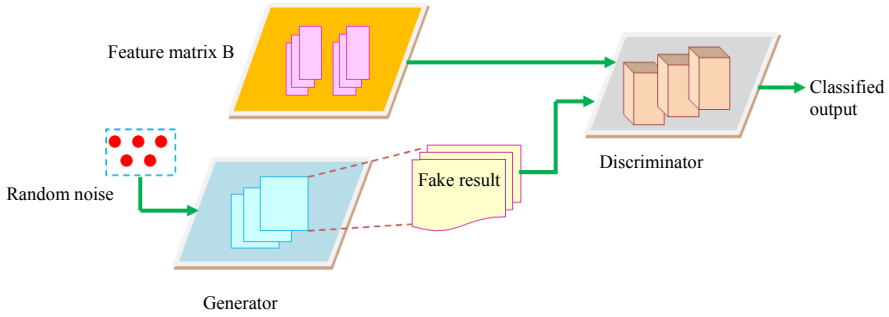

Fig. 3: Structure of deep GAN

$$\min_{H} \max_{M} pB[\log(M(B))] + p_b[\log(1 - M(H(b)))] \quad (5)$$

The generator is used to map the latent vector from same known prior p_b to the sample space. The task of discriminator is to differentiate among the samples generated by the generator H(b) and the real data samples M(B). The adversarial loss function of H is given as:

$$S_{adv} = \log(1 - M(H(B))) \quad (6)$$

In addition to adversarial loss, the data mismatch terms is employed for facilitating generator:

$$S_{data} = \| DM - H(B) \|_2 \quad (7)$$

Here, DM denotes data mismatch term. The training process of adversarial encourages the network for generating sharp and accurate images results. The data mismatch term forces the network to map the degraded results correctly to original ones. Hence, the final loss for H is the weighted sum of S_{data} and S_{adv}:

$$S = S_{data} + \mu S_{adv} \quad (8)$$

where, μ specifies hyper parameter used to control the weight of each loss term. Accordingly, H and M are iteratively trained using proposed Taylor-CSO.

Algorithmic procedure of proposed Taylor-CSO algorithm: The training process of Deep GAN is done using proposed Taylor-CSO algorithm which is the integration of Taylor series[24] and CSO[25], respectively. CSO is the bio-inspired optimization algorithm that mimics the foraging behavior and the hierarchical order or chicken swarm. It considers three categories of chickens, namely roosters, hen and chicks.

The swarm is partition into different groups and each group has a single rooster with number of chicks and hens. The competition between various chickens is exists under certain hierarchical order. Based on the fitness value, the identity of chickens is determined. The chickens having best fitness value is acted as rooster, whereas the chickens with worst fitness value is treated as chicks, and the remaining are called as hens. Here, the relationship between the chicks and hens are randomly specified. However, the chickens follow the group mate roosters for searching the food. By integrating the foraging behavior of chicken swarm with the expansion of Taylor series, the performance of FER is more accurate. The proposed method showed highly robustness and achieved global best by eliminating the local minima. The integration of Taylor series with the CSO shows the efficiency of developed scheme and effectively minimized the computational complexity. The algorithmic steps involved in proposed Taylor-CSO are explained as follows:

Initialization: Let us initialize the population with x number of chickens, rx number of roosters, hx number of hens, cx number of chicks and mx number of mother hens. The chickens with the best value are considered as roosters whereas the worst value of chickens is considered as chicks and remaining chickens are called as hens. All the x number of virtual chickens have the position as:

$$T^{k}_{g,s}(g \in [1, ..., x], s \in [1, ..., D]) \quad (9)$$

where, k denotes time step and C specifies dimensional space.

Fitness measure: The fitness function is used to compute the optimal solution by determining the best fitness value and the function used to compute fitness measure is expressed as:

$$F = \frac{1}{L}\sum_{c=1}^{L}[S_c - O_c]^2 \quad (10)$$

Where:
F = Fitness measure
L = Number of samples
O = Specifies target output

Update solution of rooster: The roster having higher fitness value get more priority to access the food rather than considering worst fitness value. The roosters search their food in wide area and this mechanism is represented as:

$$T^{k+1}_{g,s} = T^{k}_{g,s} * \left(1 + GD\left(0, \sigma^2\right)\right) \quad (11)$$

$$\sigma^2 = \begin{cases} 1;\ F_g \le F_r;\ r \in [1, x],\ r \ne g \\ \exp\left(\dfrac{(F_r - F_g)}{|F_g| + \tau}\right);\ \text{otherwise} \end{cases} \quad (12)$$

where, GD indicates Gaussian distribution with the mean value of '0' and the standard deviation of σ^2 τ denotes constant term that is used for eliminating zero division error, r specifies index of rooster that is chosen randomly from rooster's group.

Update solution of hen: The hen follow the group mater rooster's for accessing the food. The hen steals the food found by other virtual chickens randomly and hence they can be repressed by the virtual ones.

However, the dominant hens gain more benefit than the submissive ones in competition of food. This behavior is mathematically expressed as:

$$T_{k+1}(g,s) = T_k(g,s) + G_1 rand(T_k(l_1, s) - T_k(g,s)) + G_2 rand(T_k(l_2, s) - T_k(g,s)) \quad (13)$$

$$T_{k+1}(g,s) = T_k(g,s) + G_1 randT_k(l_1,s) - G_1 randT_k(g,s) + G_2 randT_k(l_2,s) - G_2 randT_k(g,s) \quad (14)$$

$$T_{k+1}(g,s) = T_k(g,s)[1 - G_1 rand - G_2 rand] + G_1 randT_k(l_1,s) + G_2 randT_k(l_2, s) \quad (15)$$

The standard equation of Taylor series is represented as:

$$T_{k+1}(g,s) = 0.5T_k(g,s) + 1.3591T_{k-1}(g,s) - 1.359T_{k-2}(g,s) + 0.6795T_{k-3}(g,s) - 0.2259T_{k-4}(g,s) + 0.0555T_{k-5}(g,s) - 0.0104T_{k-6}(g,s) + 1.3e^{-3}T_{k-7}(g,s) - 9.92e^{-5}T_{k-8}(g,s) \quad (16)$$

$$T_k(g,s)=2\begin{bmatrix}T_{k+1}(g,s)-1.3591T_{k-1}(g,s)+1.359T_{k-2}(g,s)-\\0.6795T_{k-3}(g,s)+0.2259T_{k-4}(g,s)-\\0.0555T_{k-5}(g,s)+0.0104T_{k-6}(g,s)-\\1.3e^{-3}T_{k-7}(g,s)+9.92e^{-5}T_{k-8}(g,s)\end{bmatrix} \quad (17)$$

By substituting the above Eq. (17) in Eq. (15) is expressed as:

$$T_{k+1}(g,s)=2\begin{bmatrix}T_{k+1}(g,s)-1.3591T_{k-1}(g,s)+1.359T_{k-2}(g,s)-\\0.6795T_{k-3}(g,s)+0.2259T_{k-4}(g,s)-\\0.0555T_{k-5}(g,s)+0.0104T_{k-6}(g,s)-\\1.3e^{-3}T_{k-7}(g,s)+9.92e^{-5}T_{k-8}(g,s)[1-G_1rand-\\G_2rand]+G_1randT_k(l_1,s)+G_2randT_k(l_2,s)\end{bmatrix} \quad (18)$$

$$T_{k+1}(g,s)-2Tk+1(g,s)[1-G_1rand-G_2rand]=2\begin{bmatrix}-1.3591T_{k-1}(g,s)+1.359T_{k-2}(g,s)-\\0.6795T_{k-3}(g,s)+0.2259T_{k-4}(g,s)-\\0.0555T_{k-5}(g,s)+0.0104T_{k-6}(g,s)-\\1.3e^{-3}T_{k-7}(g,s)+9.92e^{-5}T_{k-8}(g,s)[1-G_1rand-\\G_2rand]+G_1randT_k(l_1,s)+G_2randT_k(l_2,s)\end{bmatrix} \quad (19)$$

$$T_{k+1}(g,s)[-2(1-G_1rand-G_2rand)]=2\begin{bmatrix}-1.3591T_{k-1}(g,s)+1.359T_{k-2}(g,s)-\\0.6795T_{k-3}(g,s)+0.2259T_{k-4}(g,s)-\\0.0555T_{k-5}(g,s)+0.0104T_{k-6}(g,s)-\\1.3e^{-3}T_{k-7}(g,s)+9.92e^{-5}T_{k-8}(g,s)[1-G_1rand-\\G_2rand]+G_1randT_k(l_1,s)+G_2randT_k(l_2,s)\end{bmatrix} \quad (20)$$

$$T_{k+1}(g,s)=\frac{1}{1-2(1-G_1rand-G_2rand)}\left\{2\begin{bmatrix}-1.3591T_{k-1}(g,s)+1.359T_{k-2}(g,s)-\\0.6795T_{k-3}(g,s)+0.2259T_{k-4}(g,s)-\\0.0555T_{k-5}(g,s)+0.0104T_{k-6}(g,s)-\\1.3e^{-3}T_{k-7}(g,s)+9.92e^{-5}T_{k-8}(g,s)[1-G_1rand-\\G_2rand]+G_1randT_k(l_1,s)+G_2randT_k(l_2,s)\end{bmatrix}\right\} \quad (21)$$

Here, the term G_1 and G_2 is expressed as:

$$G_1=\exp\left(\frac{F_g-Fl_1}{abs(F_g-\tau)}\right) \quad (22)$$

$$G_2=\exp\left(Fl_2-F_g\right) \quad (23)$$

where, rand denotes random number that lies in the range of [0,1], l_1 denotes the index of rooster such that $l_1\in[1, ..., x]$, l_2 represents the index of chicken such that $l_2\in[1, ..., x]$ and this l_2 can be either hen or rooster that is selected randomly from swarm.

Update solution of chicks: The chicks go near to their mother for accessing the food and this behavior is represented as:

$$T_{g,s}^{k+1}=T_{g,s}^{k}+Q*\left(T_{y,s}^{k}-T_{g,s}^{k}\right) \quad (24)$$

where, $T_{y,s}^{k}$ indicates the position of gth chicks mother such that $y\in[1, ..., x]$ and Q denotes a parameter and it ranges between 0 and 2, respectively.

Evaluate feasibility: The feasibility of the solution is evaluated to find the solution best value. When the new solution has the best value than previous one, then the solution can be updated with the new value.

Termination: The above steps are repeated until the best solution is attained. Algorithm 1 portrays the pseudo code of developed Taylor-CSO based Deep GAN.

Algorithm 1; Pseudo code of proposed Taylor-CSO based Deep GAN

Pseudo code of proposed Taylor-CSO based Deep GAN
Input: $T_{g,s}$
Output: $T_{k+1}(g, s)$
Initialize the population with x virtual chickens
Compute fitness measure
while ($k<k_{max}$); k_{max}-maximum generation
if ($k\%\eta == 0$)
Rank the fitness value of chickens and form a hierarchical order in swarm
Partition the swarm to various groups and find the relation among mother hens and chicks in the group
end if
for (g = 1 to x)
if (g == rooster)
Update the solution using Eq. (11)
if (g == hen)
Update the solution using Eq. (21)
if (g == chick)
Update the solution using Eq. (24)
Evaluate feasible solution
Replace the existing solution with the new best solution
end for
end while

RESULTS AND DISCUSSION

This study explains the results and discussion of proposed Taylor-CSO based Deep GAN with respect to performance metrics.

Experimental setup: The implementation of the developed method is carried out in MATLAB tool by Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) dataset[30].

Dataset description: This dataset contains 7356 files and each file is rated 10 times on the genuineness, intensity, and emotional validity. The ratings are offered by 247

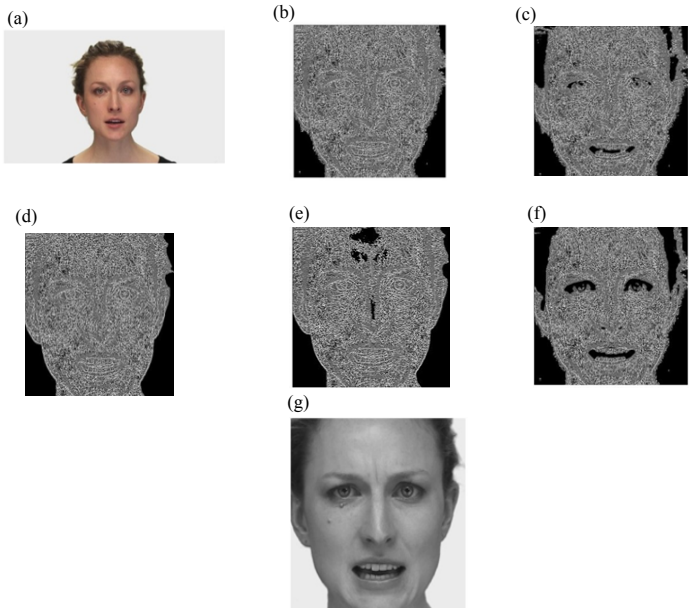

Fig. 4(a-g): Experimental results, (a) Input image, (b) Extracted LBP feature (d_1), (c) Extracted LBP feature (d_2), (d) Extracted LBP feature (d_3), (e) Extracted LBP feature (d_4), (f) Extracted LBP feature (d_5) and (g) Detected face object

individuals who are the features of untrained adult research participants from the North America. Here, 72 participants offered test-retest data. It consists of 24 professional actors with 12 female and 12 male actors. The speech includes happy, angry, fearful, sad, calm, disgust and fearful expressions and the songs contain fearful angry, happy, calm and sad emotions. Each of the expression is generated at two levels with the emotional intensity of strong and normal by integrating the neutral expression.

Evaluation metrics: The performance of developed method is evaluated by considering the metrics, such as accuracy, precision and recall.

Accuracy: It is the measure that shows the ratio of accurately classified observation to total number of observations and is represented as:

$$\beta = \frac{J_p + J_n}{J_p + J_n + K_p + K_n} \tag{25}$$

where, β denotes accuracy, J_p denotes true positive, J_n denotes true negative, K_p specifies false positive and K_n signifies false negative.

Precision: It is the ratio of accurately classified positive observations to the total number of positive observations and is specified as:

$$PR = \frac{J_p}{J_p + K_p} \tag{26}$$

Here, PR denotes precision.

Recall: It is the measure that defines the ratio of correctly classified positive observations to all observations and is represented as:

$$RC = \frac{J_p}{J_p + K_p} \tag{27}$$

Here, RC specifies recall.

Experimental results: Figure 4 portrays the experimental results of proposed Taylor-CSO based Deep GAN. Figure 4a represents the input image, Fig. 4b-f portrays the extracted LBP features d_1-d_5. Figure 4g portrays detected face object.

Comparative methods: The performance improvement of the developed scheme is analyzed using the traditional approaches, like Collaborative Discriminative multi-metric learning (CDMML)[19], Local Enhanced Motion History Image based Convolutional Neural Network (LEMHI-CNN)[17], Temporal Patterns of Oriented Edge Magnitudes (TPOEM)[19] and Generative Adversarial Network (GAN)[29].

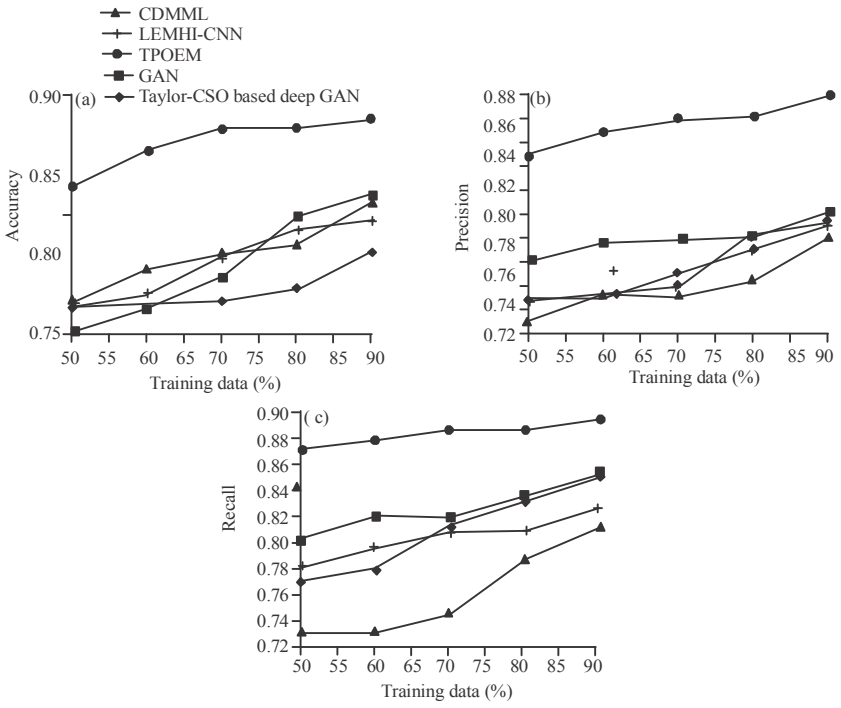

Fig. 5(a-c): Analysis with training data, (a) Accuracy, (b) Precision and (c) Recall

Comparative analysis: This study explains the comparative analysis made by the proposed scheme based on the feature size, training data and illumination pixel intensity.

Analysis based on training data: Figure 5 portrays the analysis with respect to training data. The performance evaluation carried out with respect to accuracy is shown in Fig. 5a. For training data of 60%, accuracy measured by the existing CDMML, LEMHI-CNN, TPOEM, GAN is 0.7650, 0.7680, 0.7738 and 0.7896 whereas the proposed Taylor-CSO based Deep GAN achieved the accuracy of 0.8645 that shows the percentage of enhancement while comparing the developed model with the traditional techniques, such as CDMML, LEMHI-CNN, TPOEM and GAN is 11.5, 11.2, 10.5 and 8.7%. When increasing the training data to 90%, accuracy measured by the approaches, like CDMML, LEMHI-CNN, TPOEM, GAN and proposed Taylor-CSO based Deep GAN is 0.8355, 0.8012, 0.8205, 0.8320 and 0.8846 that reports the performance improvement with CDMML is 5%, LEMHI-CNN is 9.4%, TPOEM is 7% and GAN is 5.9%.

Figure 5b shows the analysis of precision by altering training value. By considering 70% of training value, the precision measured by existing CDMML, LEMHI-CNN, TPOEM, and GAN is 0.8201, 0.8391, 0.8245 and 0.8152, whereas the proposed Taylor-CSO based Deep GAN has the precision of 0.8894 that outcomes the performance improvement with that of CDMML, LEMHI-CNN, TPOEM and GAN is 7.8, 5.7, 7.3 and 8.3%, respectively. When increasing the value of training to 80%, the precision computed by CDMML, LEMHI-CNN, TPOEM, GAN, proposed Taylor-CSO based Deep GAN is 0.8415, 0.8399, 0.8350, 0.8214 and 0.8912 which reports the percentage of improvement while comparing the proposed with CDMML, LEMHI-CNN, TPOEM and GAN is 5.6, 5.8, 6.3 and 7.8%.

The analysis of recall by altering the training data is illustrated in Fig. 5c. For 60% of training value, recall achieved by CDMML, LEMHI-CNN, TPOEM and GAN is 0.8201, 0.8301, 0.8456 and 0.7867 whereas proposed Taylor-CSO based Deep GAN measured the recall value of 0.8845 such that the developed model shows the performance improvement with traditional CDMML, LEMHI-CNN, TPOEM, and GAN is 7.3, 6.1, 4.4 and 11.1%, respectively. By considering the training value as 80%, the recall achieved by CDMML is 0.8546, LEMHI-CNN is 0.8391, TPOEM is 0.8569, GAN is 0.8245 and proposed Taylor-CSO based Deep GAN is 0.8899 that reports the performance improvement with CDMML is 4%, LEMHI-CNN is 5%, TPOEM is 3% and GAN is 7%. The proposed method achieved better performance by considering the feature descriptor named IILBP.

Analysis based on illumination pixel intensity: Figure 6 represents the comparative analysis made by

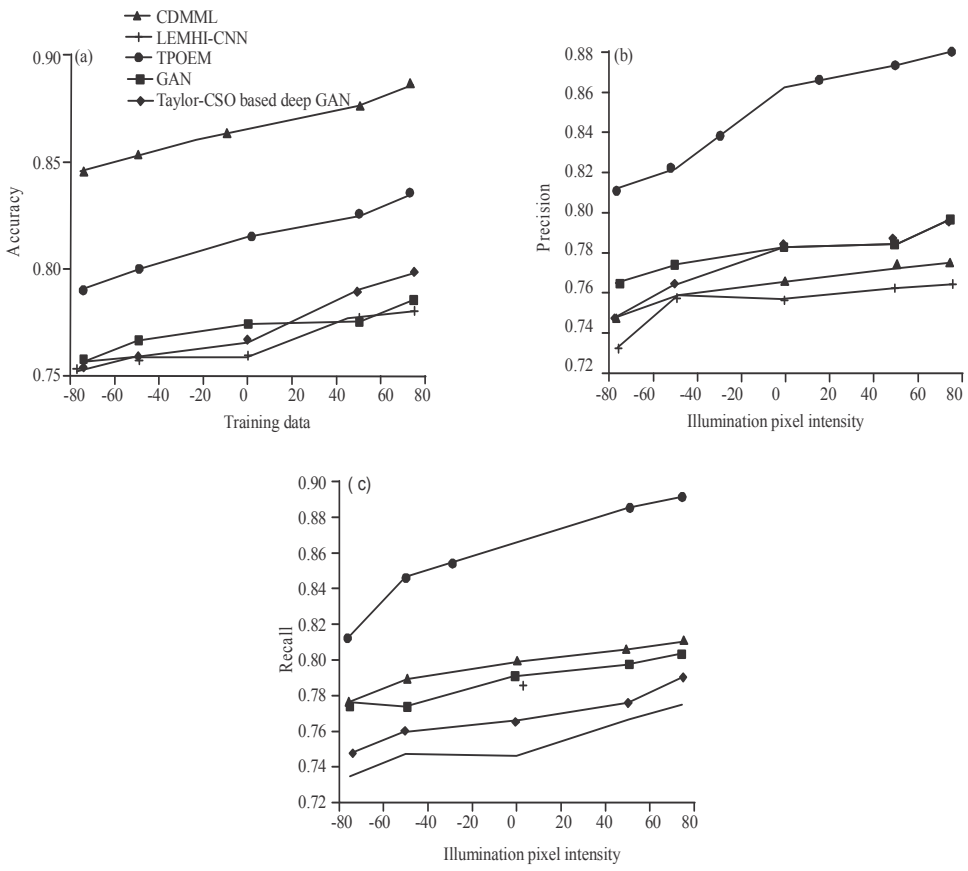

Fig. 6(a-c): Analysis based on illumination pixel intensity, (a) Accuracy, (b) Precision and (c) Recall

varying illumination pixel intensity. The analysis carried out in terms of accuracy is represented in Fig. 6a. When illumination pixel intensity is '-50', the accuracy obtained by CDMML, LEMHI-CNN, TPOEM, GAN, and proposed Taylor-CSO based Deep GAN is 0.7582, 0.7589, 0.7658, 0.80 and 0.8524 in such a way that the proposed model shows the performance enhancement by analyzing the developed scheme with CDMML is 11%, LEMHI-CNN is 11%, TPOEM is 10% and GAN is 6%. When considering the illumination pixel intensity as '75' accuracy of CDMML is 0.7789, LEMHI-CNN is 0.7972, TPOEM is 0.7849, GAN is 0.8345 and proposed Taylor- CSO based Deep GAN is 0.8856, respectively. However, the proposed method shows the performance improvement for the same illumination intensity by CDMML is 12%, LEMHI- CNN is 10%, TPOEM is 11%, and GAN is 5%.

Figure 6b depicts the analysis of precision. When the illumination pixel intensity is considered as '-50', the precision achieved by conventional CDMML, LEMHI-CNN, TPOEM and GAN is 0.7589, 0.7745, 0.7589 and 0.7654 whereas proposed Taylor-CSO based Deep GAN achieved the precision of 0.8235 that shows the percentage of improvement by considering the existing CDMML, LEMHI-CNN, TPOEM, and GAN is 7.8, 6%, 7 and 7%, respectively. For the illumination pixel intensity of '75', the precision computed by CDMML, LEMHI-CNN, TPOEM, GAN, and proposed Taylor-CSO based Deep GAN is 0.7758, 0.7965, 0.7658, 0.7956 and 0.88 such that it shows the performance enhancement with that of CDMML, LEMHI-CNN, TPOEM, GAN is 11.8, 9.5, 13 and 9.6%.

The analysis of recall by varying illumination pixel intensity is shown in Fig. 6c. When it is considered the illumination pixel intensity as '-50', the recall measured by CDMML, LEMHI-CNN, TPOEM, GAN and proposed Taylor-CSO based Deep GAN is 0.7589, 0.7745, 0.7456, 0.789 and 0.8456 and it reported the performance improvement of developed scheme with that of CDMML, LEMHI-CNN, TPOEM, GAN is 10.3, 8.4, 11.8 and 6.7%. When considering the illumination pixel intensity as '75', the recall of CDMML, LEMHI-CNN, TPOEM, and GAN is 0.7896, 0.8025, 0.7745 and 0.81 whereas the proposed Taylor-CSO based Deep GAN measured the recall of 0.89 that shows the percentage of improvement with the existing techniques such as CDMML, LEMHI-CNN,

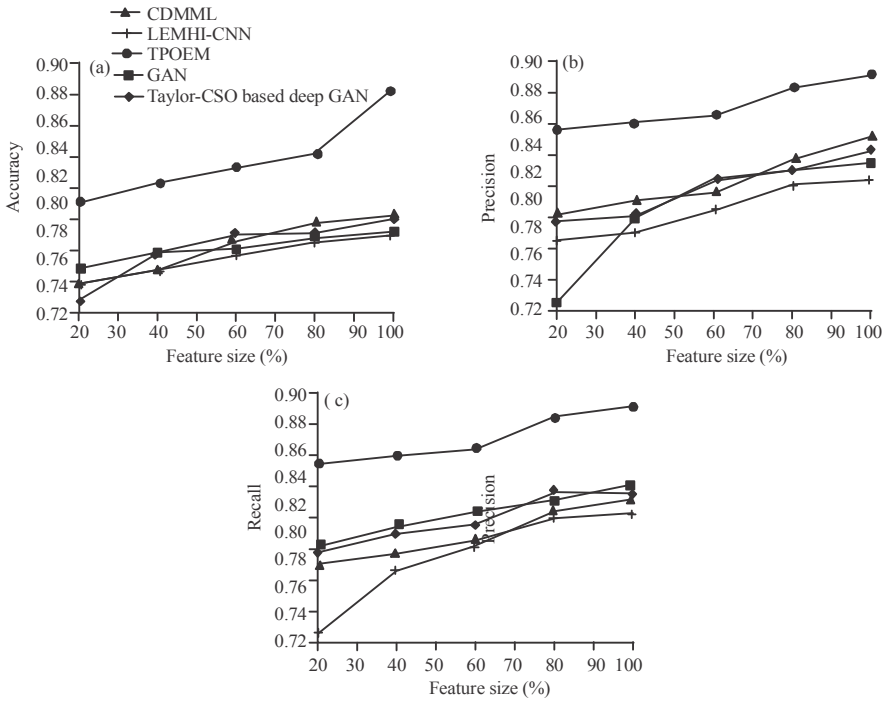

Fig. 7(a-c): Analysis based on feature size, (a) Accuracy, (b) Precision and (c) Recall

TPOEM and GAN is 11.3, 9.8, 13 and 9%, respectively. The proposed method is highly robust and showed higher efficiency in the recognition performance facial expression by employing the deep learning classifier.

Analysis based on feature size: The analysis made by varying the feature size in terms of the performance metrics is represented in Fig. 7. The analysis carried out in terms of accuracy is depicted in Fig. 7a. When the feature size is considered as 40%, the accuracy metric computed by CDMML, LEMHI-CNN, TPOEM, and GAN is 0.7658, 0.7767, 0.7663 and 0.7767 whereas the accuracy of proposed Taylor-CSO based Deep GAN is 0.8214 such that it shows the performance improvement with the traditional CDMML is 6.8%, LEMHI-CNN is 5.4%, TPOEM is 6.7% and GAN is 5.4%. By considering the feature size as 80%, the accuracy computed by CDMML, LEMHI-CNN, TPOEM, GAN, proposed Taylor-CSO based Deep GAN is 0.7845, 0.7896, 0.7956, 0.7852 and 0.8415 9001such that it reports the performance improvement with that of CDMML, LEMHI-CNN, TPOEM and GAN is 6.8, 6.2, 5.5 and 6.7%, respectively.

Figure 7b depicts the analysis of precision by varying the feature size. For 40% of feature size, the precision computed by the CDMML, LEMHI-CNN, TPOEM and GAN is 0.8101, 0.8015, 0.7903 and 0.8022 whereas the proposed Taylor-CSO based Deep GAN achieved the precision of 0.8606 that reports the performance improvement while verifying the developed approach with that of existing CDMML, LEMHI-CNN, TPOEM, and GAN is 5.9, 6.9, 8.2 and 6.8%, respectively. When the feature size is increased to 80%, the precision measured by CDMML, LEMHI-CNN, TPOEM, GAN, proposed Taylor-CSO based Deep GAN is 0.8361, 0.8311, 0.8205, 0.8312 and 0.8843 and it shows the performance enhancement with CDMML is 5.4%, LEMHI-CNN is 6%, TPOEM is 7% and GAN is 6%.

The analysis of recall measured in terms of feature size is shown in Fig. 7c. When it is considered the feature size as 60%, the recall achieved by CDMML, LEMHI-CNN, TPOEM and GAN is 0.8161, 0.8015, 0.8052 and 0.8237 while the developed Taylor-CSO based Deep GAN achieved the recall of 0.8658 that shows the performance improvement by considering the developed scheme with the traditional CDMML, LEMHI-CNN, TPOEM, GAN is 5.7, 7.4, 7 and 4.9%. When increasing the feature size to 80%, the recall obtained by the CDMML is 0.8356, LEMHI-CNN is 0.8245, TPOEM is 0.8205, GAN is 0.8312 and proposed Taylor-CSO based Deep GAN is 0.8843. However, the proposed method shows the performance improvement for 80% of feature size by analyzing the proposed approach with CDMML is 5.5%, LEMHI-CNN is 6.8%, TPOEM is 7.2% and GAN is 6%, respectively. The training process optimally tunes the classifier to generate optimal best solution by eliminating the local minima solution.

Table 1: Comparative discussion

Metrics/Methods	Variables	CDMML	LEMHI- CNN	TPOEM	GAN	Proposed Taylor-CSO based Deep GAN
Training data = 90%	Accuracy	0.8355	0.8012	0.8205	0.8320	0.8846
	Precision	0.8456	0.8502	0.8457	0.8404	0.8996
	Recall	0.8659	0.8502	0.8668	0.8404	0.8952
Illumination pixel intensity = 75	Accuracy	0.7789	0.7972	0.7849	0.8345	0.8856
	Precision	0.7758	0.7965	0.7658	0.7956	0.8800
	Recall	0.7896	0.8025	0.7745	0.8100	0.8900
Feature size = 100%	Accuracy	0.7905	0.7986	0.8000	0.7901	0.8812
	Precision	0.8511	0.8356	0.8236	0.8417	0.8914
	Recall	0.8361	0.8311	0.8236	0.8417	0.8914

Comparative discussion: Table 1 portrays the comparative discussion of developed method. When considering the training data of 90%, accuracy measured by CDMML, LEMHI-CNN, TPOEM, GAN and proposed Taylor-CSO based Deep GAN is 0.8355, 0.8012, 0.8205, 0.8320 and 0.8846, respectively. The precision achieved by CDMML, LEMHI-CNN, TPOEM, GAN, and proposed Taylor-CSO based Deep GAN is 0.8456, 0.8502, 0.8457, 0.8404 and 0.8996 for 90% of training data. When it is considered the illumination pixel intensity as '75', the accuracy measure computed by the CDMML, LEMHI-CNN, TPOEM, GAN and proposed Taylor-CSO based Deep GAN is 0.7789, 0.7972, 0.7849, 0.8345 and 0.8856, respectively. However, the precision measure achieved by CDMML, LEMHI-CNN, TPOEM, GAN, and proposed Taylor-CSO based Deep GAN is 0.8511, 0.8356, 0.8236, 0.8417, and 0.8914 for 100% of feature size. When considering the feature size as 100%, the recall achieved by CDMML, LEMHI-CNN, TPOEM, GAN, and proposed Taylor-CSO based Deep GAN is 0.8361, 0.8311, 0.8236, 0.8417 and 0.8914.

CONCLUSION

A robust and efficient recognition approach is designed to classify the facial expressions using proposed Taylor-CSO based Deep GAN. However, the proposed Taylor-CSO is derived by the integration of Taylor series and CSO algorithm. At first, the video frames are extracted from the video sequences and the video frames are effectively pre-processed using ROI extraction module. Thereafter, the face detection process is done to detect the face objects using Viola Jones algorithm and the features are extracted using IILBP, which is the modification of LBP descriptor. The feature acquired from the IILBP is the feature matrix that is used to perform FER process using Deep GAN such that the training procedure of deep learning classifier is done by the optimization algorithm named Taylor-CSO. The proposed method effectively increases the training speed and minimizes the computational issues. However, the developed approach obtained the accuracy, precision and recall of 0.8846, 0.8996 and 0.8952 for the training data. The future dimension of research would be the consideration of some other deep learning classifiers for increasing the efficiency of FER system. Moreover, the training process can be done by employing some other optimization algorithm.

REFERENCES

01. Stankowski, J., D. Karwowski, T. Grajek, K. Wegner and J. Siast *et al*., 2014. Bitrate distribution of syntax elements in the HEVC encoded video. International Conference on Signals and Electronic Systems (ICSES), September 11-13, 2014, IEEE, pp: 1-4.
02. Zhang, S., X. Pan, Y. Cui, X. Zhao and L. Liu, 2019. Learning affective video features for facial expression recognition via hybrid deep learning. IEEE Access, 7: 32297-32304.
03. Dhavalikar, A.S. and R.K. Kulkarni, 2014. Face detection and facial expression recognition system. International Conference on Electronics and Communication Systems (ICECS), February 13-14, 2014, IEEE, Coimbatore, India, pp: 1-7.
04. Cohn, J.F. and P. Ekman, 2005. Measuring Facial Action. In: Harrigan, J.A., R. Rosenthal and K.R. Scherer (Eds.)., The New Handbook of Methods in Nonverbal Behavior Research, Oxford University Press, New York, pp: 9–64.
05. Li, L., Y. Yuan, M. Li, H. Xu, R. Li and S. Lu, 2019. Subject independent facial expression recognition. Proceedings of the 2019 International Conference on Image, Video and Signal Processing, February, 2019, ACM Press, pp: 85-92.
06. Liu, X., B.V.K.V. Kumar, P. Jia and J. You, 2019. Hard negative generation for identity-disentangled facial expression recognition. Pattern Recognit., 88: 1-12.
07. Puttaswamy, M.R., 2020. Improved deer hunting optimization algorithm for video based salient object detection. Multimedia Res., Vol. 3.
08. Gan, Y.S., S.T. Liong, W.C. Yau, Y.C. Huang and L.K. Tan, 2019. OFF-ApexNet on micro-expression recognition system. Signal Process. Image Commun., 74: 129-139.

09. Wang, W. and F. Chang, 2014. Facial expression recognition based on feature block and local binary pattern. In: Bio-Inspired Computing-Theories and Applications, Pan, L., G. Paun, M.J. Pérez-Jiménez and T. Song (Eds.), Springer, Berlin Heidelberg, ISBN: 978-3-662-45048-2 pp: 447-451.
10. Richhariya, B. and D. Gupta, 2019. Facial expression recognition using iterative universum twin support vector machine. Appl. Soft Comp., 76: 53-67.
11. Ma, L. and K. Khorasani, 2004. Facial expression recognition using constructive feedforward neural networks. IEEE Trans. Syst. Man Cybernet. Part B: Cybernet., 34: 1588-1595.
12. Cortes, C. and V. Vapnik, 1995. Support-vector networks. Mach. Learn., 20: 273-297.
13. Michel, P. and R.E. Kaliouby, 2003. Real time facial expression recognition in video using support vector machines. Proceedings of 5th International Conference on Multimodal Interfaces, 5th International Conference, ACM Press, 258-264.
14. Burton, A.M., V. Bruce and P.J.B. Hancock, 1999. From pixels to people: A model of familiar face recognition. Cognit. Sci., 23: 1-31.
15. Lyons, M., S. Akamatsu, M. Kamachi and J. Gyoba, 1998. Coding facial expressions with Gabor wavelets. Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition, April 14-16, 1998, Institute of Electrical and Electronics Engineers (IEEE), pp: 200-205.
16. Zhang, S., X. Zhao and B. Lei, 2012. Facial expression recognition based on local binary patterns and local fisher discriminant analysis. WSEAS Trans. Signal Process., 8: 21-31.
17. Yan, H., 2018. Collaborative discriminative multi-metric learning for facial expression recognition in video. Pattern Recognit., 75: 33-40.
18. Hu, M., H. Wang, X. Wang, J. Yang and R. Wang, 2019. Video facial emotion recognition based on local enhanced motion history image and CNN-CTSLSTM networks. J. Visual Commun. Image Represent., 59: 176-185.
19. Cruz, E.A.S., C.R. Jung and C.H.E. Franco, 2018. Facial expression recognition using temporal POEM features. Pattern Recognit. Lett., 114: 13-21.
20. Makhmudkhujaev, F., M. Abdullah-Al-Wadud, M.T.B. Iqbal, B. Ryu and O. Chae, 2019. Facial expression recognition with local prominent directional pattern. Signal Process.: Image Commun., 74: 1-12.
21. Fernandez, P.D.M., F.A.G. Pena, T.I. Ren and A. Cunha, 2019. FERAtt: Facial expression recognition with attention net. Computer Vision and Pattern Recognition, February 2019. https://openaccess.thecvf.com/content_CVPRW_2019/papers/MBCCV/Fernandez_FERAtt_Facial_Expression_Recognition_With_Attention_Net_CVPRW_2019_paper.pdf
22. Rathod, P., L. Gagnani and K. Patel, 2014. Facial expression recognition: Issues and challenges. Int. J. Enhanced Res. Sci. Technol. Eng., 3: 108-111.
23. Samadiani, N., G. Huang, B. Cai, W. Luo, C.H. Chi, Y. Xiang and J. He, 2019. A review on automatic facial expression recognition systems assisted by multimodal sensor data. Sensors, Vol. 19. 10.3390/s19081863
24. Mangai, S.A., B.R. Sankar and K. Alagarsamy, 2014. Taylor series prediction of time series data with error propagated by artificial neural network. Int. J. Comput. Appl., 89: 41-47.
25. Meng, X., Y. Liu, X. Gao and H. Zhang, 2014. A new bio-inspired algorithm: Chicken swarm optimization. Proceedings of the 5th International Conference on Advances in Swarm Intelligence, October 17-20, 2014, Hefei, China, pp: 86-94.
26. Subramanyam, T.C., J.B. Subrahmanyam and T. Ram, 2019. An adaptive chicken swarm algorithm to solve optimal power flow problem considering FACTS device. J. Comput. Mech. Power Syst. Control, 2: 38-47.
27. Kumar, C.A. and R. Vimala, 2020. Load balancing in cloud environment exploiting hybridization of chicken swarm and enhanced raven roosting optimization algorithm. Multimedia Res., 3: 45-55.
28. Chaudhari, M.N., M. Deshmukh, G. Ramrakhiani and R. Parvatikar, 2019. Face detection using viola jones algorithm and neural networks. Fourth International Conference on Computing Communication Control and Automation, April 25, 2019, IEEE, pp: 1-6.
29. Usman, M., S. Latif, M. Asim, B.D. Lee and J. Qadir, 2020. Retrospective motion correction in multishot MRI using generative adversarial network. Sci. Rep., 10: 1-11.
30. Livingstone, S.R. and F.A. Russo, 2018. The ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PLoS ONE, Vol. 13. 10.1371/journal.pone.0196391