Journal of Big Data

RESEARCH Open Access

Group based emotion recognition from video sequence with hybrid optimization based recurrent fuzzy neural network

Velagapudi Sreenivas[1*], Varsha Namdeo[1] and E. Vijay Kumar[2]

*Correspondence: velagapudisreenivas@gmail.com
[1] Department of Computer Science and Engineering, SRK University, Bhopal, Madhya Pradesh 462026, India
Full list of author information is available at the end of the article

Abstract

Group-based emotion recognition (GER) is an interesting topic in both security and social area. In this paper, a GER with hybrid optimization based recurrent fuzzy neural network is proposed which is from video sequence. In our work, by utilizing the Neural Network the emotion recognition (ER) is performed from group of people. Initially, original video frames are taken as input and pre-process it from multi user video data. From this pre-processed image, the feature extraction is done by Multivariate Local Texture Pattern (MLTP), gray-level co-occurrence matrix (GLCM), and Local Energy based Shape Histogram (LESH). After extracting the features, certain features are selected using Modified Sea-lion optimization algorithm process. Finally, recurrent fuzzy neural network (RFNN) classifier based Social Ski-Driver (SSD) optimization algorithm is proposed for classification process, SSD is used for updating the weights in the RFNN. Python platform is utilized to implement this work and the performance of accuracy, sensitivity, specificity, recall and precision is evaluated with some existing techniques. The proposed method accuracy is 99.16%, recall is 99.33%, precision is 99%, sensitivity is 99.93% and specificity is 99% when compared with other deep learning techniques our proposed method attains good result.

Keywords: Group level emotion recognition, Frames, Face detection, Feature extraction, Feature selection and classification

Introduction

One of the major and fundamental issue is emotion recognition during the development of an interactive computer system [1–3]. Recognition of facial emotion/expression is essential, because nowadays it place its wide applications in various sectors like psychological distress and pain detection [4]. Some fields like psychology, sociology, and automatic expression recognition, therefore provided a considerable importance for this emotion recognition process to create a highly user affable software and user agents in these fields. This process of automatic facial expression recognition (FER) has exhibited its large implications in the human computer interaction (HCI) field [5]. Recently, the affective computing is considered as the most significant study field in HCI, which highly intends to improve the human–machine interaction by clearly recognizing the emotion

states from the human faces. The study required for the ER states has been analyzed from various information sources like speech signals [6], biological signals [7], images and video sequences [8, 9]. In this proposed method, the expressions required for ER are gathered from video sequence. Number of experimental analysis are carried out in this automatic FER, as it implies huge practical importance in driver fatigue surveillance, medical treatment, sociable robotics, and also in few HCI systems. Through the field of machine learning and computer vision, a number of FER systems are largely explored to gather the information regarding the expressions from the facial representations [10].

Researchers have analysed number of FER techniques having highest accuracy, but still it remains a challenging task of capturing the emotions from the faces of different nationalities and different angles [11]. Facial expression is identified as the most beneficiary source for capturing the human behavior and attitude [12]. A report provided by psychological study illustrates that the features found around the eyes, mouth, and nose are highly essential for FER purpose. In various literary works, number of different techniques have been developed for the purpose of FER in still images and video data, whereas for classification purpose different types of classifiers like decision tree (DT), Support Vector Machine (SVM), Bayesian network, neural networks (NN), K-Nearest Neighbour (KNN), and rule-based classifiers are used [13]. The ultimate aim of this classifiers is to give a perfect trade-off between the recognition accuracy and processing time [14]. Technically, this ER process is categorized into two different groups they are (i) Individual ER, (ii) Group-level ER. However, the studies found in individual based ER are found mature, but the research process in this group level ER is still in its early stages. Major challenges exhibited by this group level ER is the detection of faces from group and the aggregation of emotions content from this group through video or image [15].

Group based ER (GER) [16] is gaining a huge attention in the affective computing field. This GER is largely required for various applications like improvement in e-classroom teaching on the basis of students face expression, analyzing the crowd behavior, advertisements, typical photos best representing events, and so on. Rapid improvement has been encountered in this field due to recently occurred challenges such as EmotiW [17, 18]. Major challenging task of this classification process is to given the group image into Happy, Sad, Angry and Fear. This classification process is accurately achieved by extracting the emotions details from all the faces in the image [19, 20]. GER is now grown as a trending topic because the demands for recognizing the emotion from the group of people are gaining a maximum interest in both the social media and security arena. The World Wide Web (WWW) and multimedia are advanced a lot, therefore the users upload the number of videos and images in the internet over the social networking platforms. Due to the difference in indoor and outdoor settings, the recognition of emotions from the faces of multiple persons becomes somewhat a difficult task. Therefore, in this method even more challenging task is considered which is to classify the emotions that are shared by a group of people as Happy, Sad, Angry and Fear. This classification process is normally referred as group-level emotion recognition. While compared to other emotion recognition process, this GER is found more challenging as it depends on single image (lack of temporal information), moreover the human faces are often occurring at low resolution (lack of facial details).

In this paper we introduced the RFNN with SSD optimization algorithm for the classification purpose. The advantages of SSD is it reduces the complexity and increases the convergence speed, thus we use SSD in RFNN to get the better accuracy performance. As well as here we use modified sea lion optimization algorithm for feature selection. The existing papers does not have any feature selection techniques for GER which reduces the accuracy. Modified sea lion optimization is the combination of sea lion optimization with GSO optimization algorithm. For finding the best features, first we update the position by using GSO algorithm. GSO position is updated in sea lion optimization to get the optimal feature. The ultimate aim of this technique is to improve a system that can automatically classify the emotions from the group of peoples as Happy, Sad, Angry and Fear. Few other objectives of this process are:

- To correctly detect the face from crowded image.
- To design a technique that can effectively classify the emotion with highest accuracy and less error value.
- To reduce the dimension of extracted features using a novel adaptive algorithm.
- To improve the accuracy with less misclassification error.

The rest of the paper is organized as follows: "Related work" section gives some analysis of related work. Pre-processing, feature extraction, feature selection and classification is briefly discussed in "Proposed method for group based emotion recognition" section. The performance analysis of proposed method is explained in results and discussion which is in "Results and discussion" section. The conclusion of the work is present in "Conclusion" section.

Related work

For image based group emotion recognition, the influence of mid-level and low-level features was investigated by Balaji et al. [21]. They hypothesize that the objects and faces of humans are the major information sources and therefore it can serve as the mid-level features. Hence, using pre-trained Deep Net models they detect the objects and faces. With different encoding methods the information from different layers was extensively investigated for attain the richest feature vectors. To yield the best feature, vector of locally aggregated descriptors (VLAD) and Fisher vector encoding schemes were investigated.

Surace et al. [22] implemented a solution based on a new combination of Bayesian classifier and deep neural networks. The deep learning methods works on bottom-up techniques which evaluating emotions conveyed by isolated faces. A global emotion integrating top down features was estimated by Bayesian classifier which obtained by a scene descriptor. They tested the framework on the dataset in the wild challenge 2017 for the emotion recognition to validate the system.

In the category of group level emotion recognition Abbas et al. [23] presented the proposed solution for the emotion recognition. To classify a group of people emotion as positive, negative and neutral was one of the main objectives. For classification, their method incorporates both facial and image context information which extracted from an image. To predict facial emotions in an image from detected faces, they utilized

convolutional neural networks (CNN). In order to perform group level emotion recognition, different techniques are explored by training and combining these two deep neural networks.

In an image, an emotional detection for a group of people was described by Shamsi et al. [24]. They utilized the combination of both top-down and bottom-up method. Using face-level predictions they create heat maps and train ConvNets on them. By constructing heat maps for emotions intensity they perform face level emotion detection and join individual images. At last, an ensemble of both top down and bottom up methods was created.

GER is considered as a challenging issue, due to some obstructions like image quality, actor's variance, variation in head and body pose, different outdoor and indoor settings, and occlusions. Less number of research works are carried out in this GER process, but now it is gaining high popularity. This is due to the availability of large amount of data in social networks, most particularly the images of group of people those who are participating in social events. Most essential clue for identifying the individual behaviour is facial expression analysis, but this analysis process is found challenging due to the existence of few variations in illuminations, different facial tones, and face poses. A large number of techniques are introduced in existing techniques for emotion recognition, but none of them attains required accuracy. Thus, we introduced the RFNN with SSD optimization algorithm for classification purpose. The advantages of SSD is it reduces the complexity and increases the convergence speed, thus we use SSD in RFNN to get the better accuracy performance. As well as here we use modified sea lion optimization algorithm for feature selection. In previous works, they did not use the feature selection technique for group based emotion recognition. Feature selection techniques are utilized here, thus will select best features and it will remove to add more number of unwanted features.

Proposed method for group based emotion recognition

Figure 1 shows the basic work flow for facial emotion recognition. At first, a group of images showing different emotions like Happy, Sad, Angry and Fear are applied in this method for pre-processing step. Additionally, in this pre-processing step both illumination and normalization are included. Then, the MLTP, GLCM and LESH are applied for feature extraction approach. These three techniques are applied from the pre-processed image to extract the features. An optimal feature selection process is done by modified sea lion optimization algorithm. Then, with this optimal features the emotion recognition is done by RFNN classifier using SSD.

In our work, emotion recognition from group of people is performed by utilizing the Neural Network. From the multi user video data, initially, original video frames are taken as input and pre-process it. During pre-processing, the input image is denoised by Fast averaging peer group (FAPG) filter. The face from that pre-processed image is detected using Viola Jones classifier. Then, an algorithm named as eye-center detection is introduced along with this face detection algorithm to improve the recognition result. The features from this detected face is extracted for recognition purpose. Here, the MLTP, GLCM and LESH based features are extracted and carried out as the features in the further processing steps. After extracting the features, certain features are selected using Modified Sea-lion

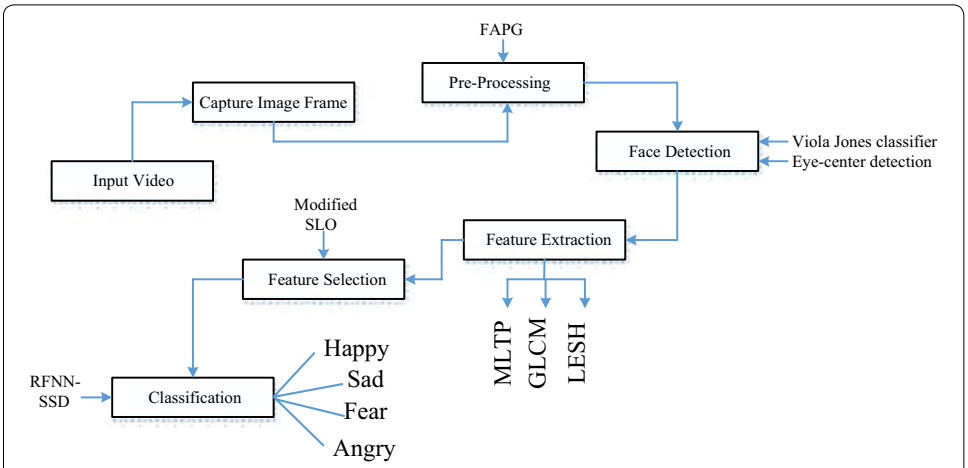

Fig. 1 Proposed work flow

optimization algorithm process. Finally, recurrent fuzzy neural network (RFNN) classifier using Social Ski-Driver (SSD), optimization algorithm is proposed for updating the weights in the RFNN.

Pre-processing

The process of pre-processing is it suppresses unwanted distortions and improve the image data for further processing. Initially, the face image is given as an input for FAPG to remove the noise and also to control the redundant information. The features of each frames gets enhanced because of this pre-processing.

Fast averaging peer group filter

In digital images, noise reduction is considered as the problematic task [25]. The corrupted images are colour images by means of impulsive noise. To overcome this problem, the fast denoising FAPG filter is used. This filter can restore the images while maintaining the tiny images and edges also it outperforms on the high noise contamination levels. Consider the colour image as the two-dimensional array, which consists of N amount of pixels $X_j = (X_{j1}, X_{j2}, X_{j3})$ through the index value (j=1,...N) which indicates the pixel position on the image domain. For q=1, 2, 3 the vector components $X_{jq} \in [0, 1]$ denote the colour channel values computed into integer domain in a given colour space. Also, the indexes are assigned to pixels that belongs to the local filtering window W, whereas X_1 indicates the central pixel and $X_2, \ldots X_n$ represents the neighbouring pixels and n signifies the size of window.

The two main steps are associated with this filter: Peer group size determination and Pixel replacement. The peer group size or else the amount of close neighbours (CN) of central pixel X_1 is evaluated in the first step. In a given colour space, when standardized Euclidean distance $\rho(X_i, X_1)$ is lower than the pre-defined threshold value d then a pixel $X_i \neq X_1$ going to W is a CN of X_1. The size of peer group is expressed as:

$$m_k = \neq \{X_j \in W : \|X_k - X_j\| < d\} \quad (1)$$

Therefore, peer group size can be represented as a pixel distortion measure produced by noise method. Pixel is treated as corrupted when value of m is low otherwise it will not be corrupted. The next step is pixel replacement. Filtering process is done after the calculation of all m values of the image pixels. Proper selection of two filter parameters such as d and γ play a significant role in noise reduction. First one is threshold parameter (d) which is stated as $0 \leq d \leq 1$. Hence, d = 0 means two identical pixels. The maximum Euclidean distance is d = 1 in colour space. The weights $we_i, \quad i = 2, \ldots n$ of the pixels X_i can be evaluated using

$$we_i = \frac{\mu_i}{\sum_{i=2}^{n} \mu_i}, \ \mu_i = m_i^{\gamma} \tag{2}$$

From the above equation, the second parameter is γ. If $\gamma > 0$, this affects the quality of results. The parameter γ gives the chance to manage the degree of membership of neighbouring pixels. The differences in neighbouring pixels the peer group sizes are reduced if $0 < \gamma < 1$. Also, if $\gamma > 1$ the differences in neighbouring pixels the peer group sizes are increased.

Detection by Viola-Jones and eye center detection

One of the significant step is face detection which found in practically all emotion recognition methods. The face detection based on this steps like nose, mouth, eyes etc. To form an expression these component plays a major role. Using Voila and Jones algorithm [26] the face region has been detected in the proposed algorithm. Four stages are presented in Voila-Jones algorithm: haarfeatures selection, creating an integral image, Adaboost training, and cascading classifiers. Some common features are in human faces which are extracted. For face detection, to select the Viola-Jones algorithm has some main reason which are: (1) it is very robust with very low false positive rate and very high detection rate and (2) it tends to be utilized for real-time scenes because of its proficiency of preparing 25 frames for each second. Thereafter, a novel eye center detection algorithm [27] (Algorithm 1) has been proposed to locate the eye centres in the faces. In such a way the algorithm is utilized to calculate L that it can handle the head deflection also.

$$L = \sqrt{\left((RE\,(x, Center) - LE(x, Center))^2 + \left(RE(y, Center) - LE\,(y, Center)\right)^2\right)} \tag{3}$$

The input video is shown in Fig. 2 and face detection by Voila & Jones and Eye Center Detection is shown in Fig. 3.

Fig. 2 Input video

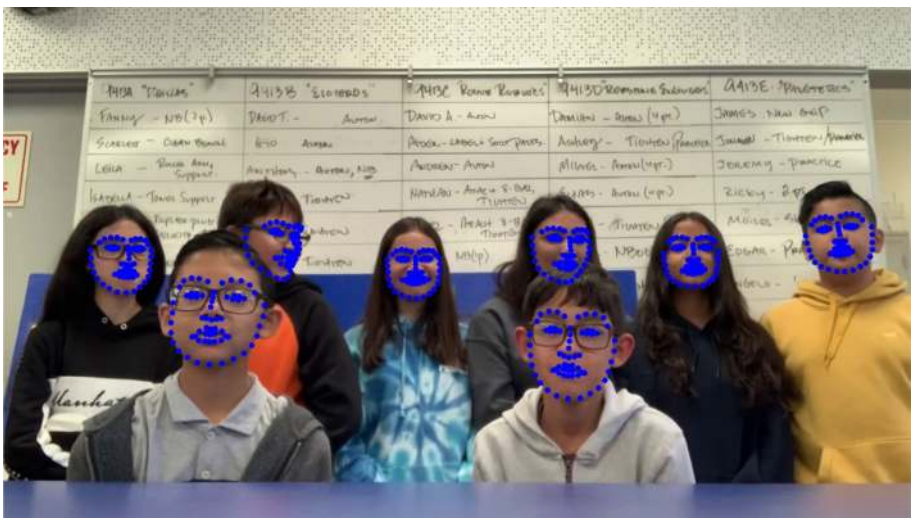

Fig. 3 Face detection by Voila and Jones and eye center detection

Input: face Image
Output: Extract the Eyes using Right Eye (RE) of face region in query image I_Q
Step 1: EyeBox N = EyeBox 1, EyeBox 2
Step 2: if then EyeBox(2, location) Eye Box (1, location)
Step 3: LE= EyeBox2
Step 4: RE=EyeBox1
Step 5: else
Step 6: RE = EyeBox2
Step 7: LE = EyeBox1
Step 8: end if
Step 9: Find the L distance between right and left eye
Step 10: By using equation (1), L can be evaluated
Step 11: L= Height = Node Box, Width = $\frac{2.25*L}{4}$
Step 12: For each face area
Step 13: Compute Nose Box
Step 14: Compute Mouth Region
Step 15: Compute Forehead
Step 16: Compute Right Cheek
Step 17: Compute Left Cheek

Feature extraction

The feature extraction is necessary to perform from the object detected image. This features are very useful to accurately recognize the emotion from the video. Here, three different feature extraction methods are applied they are MLTP, GLCM and LESH.

Multivariate local texture pattern

From a 3×3 neighbourhood, the local texture pattern model extracts local texture information. Let 'X' denoted by local neighbourhood. At that point the connection between centre pixel and its eight neighbours is characterized as below:

$$P\left(g_i, g_c\right) = \begin{cases} 0 & if\ g_i \prec \left(g_c - \Delta_g\right) \\ 1 & if\ \left(g_c - \Delta_g\right) \leq g_i \leq \left(g_c + \Delta_g\right) \\ 9 & if\ g_i \succ \left(g_c + \Delta_g\right) \end{cases} \tag{4}$$

where, Δ_g is a small positive value with the centre pixel which is set to communicate the closeness of neighbouring pixel. In the neighbourhood, the output level assigned to the ith pixel is $p\left(g_i, g_c\right)$. The MLTP output is shown in Fig. 4.

Gray-level co-occurrence matrix

We extract features from image using the GLCM. Feature extracted from GLCM are correlation, entropy, contrast, variance, average, angular second moment, dissimilarity, homogeneity, energy, mean, standard deviation and maximum probability.

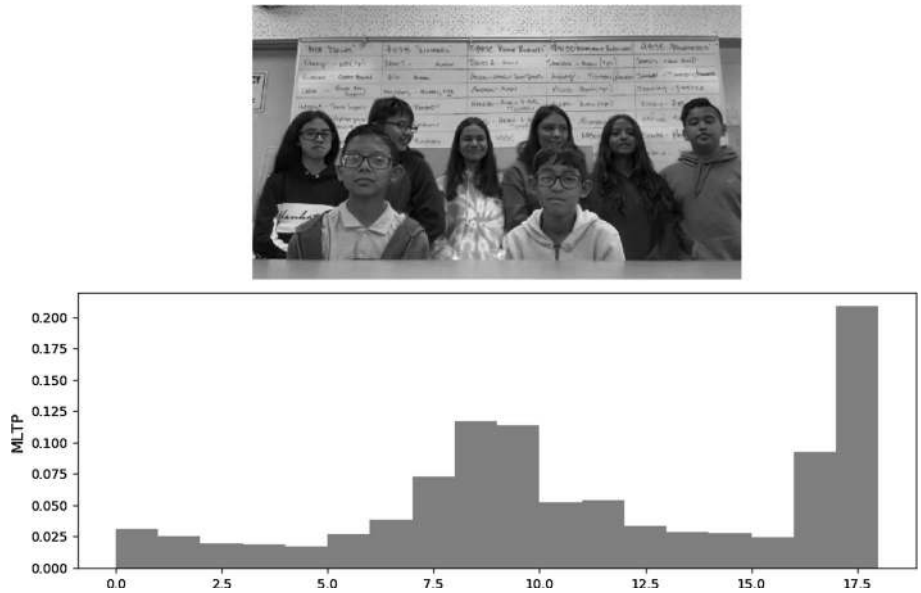

Fig. 4 MLTP output

Table 1 Gray-level co-occurrence matrix

Correlation	$\frac{\sum_i \sum_j (i\,j)\,P\,(i,j) - \mu_x\,\mu_y}{\sigma_x \sigma_y}$
Entropy	$-\sum_i \sum_j \rho\,(i\,,j)\,\log\,(\rho\,(i\,,j))$
Contrast	$\sum_{i=1}^{M}\,\sum_{k=1}^{M}\,(i\,-\,k)^2\,*\,p_i k$
Variance	$\sum_i \sum_j (i\,-\,\mu)^2\,p\,(i\,,j)$
Average	$\sum_{i=2}^{2Ng} ip_{x+y}\,(i)$
Angular second moment	$\sum_i \sum_j \{p(i,j)\}^2$
Maximum probability	$\max\,\left(p\,(j)_{ik}\right)_{i,k}$

Local energy based shape histogram

For the image of interest, LESH feature extraction method is primarily depended upon the idea of calculating the local energy pattern histogram (Table 1). Using phase congruency scheme, the local energies are calculated with different orientations.

Local energy is computed as follow

$$E\,=\,\frac{\sum_n W\,(x)\,[A_n\,(x)\,(\cos\,(\phi_n\,(x)\,-\,\varphi\,(x))\,-\,|\sin\,\,(\phi_n\,(x)\,-\,\varphi\,(x))|)\,-\,T]}{\sum_n A_n\,(x)\,+\,\varepsilon} \tag{5}$$

where, T represents the noise cancellation factor, W represents the sine of phase deviation and factor.

The local histogram h is expressed as below:

$$h_{r,b}\,=\,\sum W_r\,\times\,E\,\times\,\delta_{LB} \tag{6}$$

where, δ_{LB} represents the Kronecker's delta, local energy is computed by E and W_r is represented as Gaussian weighting function of region r, orientation label map is represented by L and current bin is represented as b.

Gaussian weighting function of region r is given as follow.

$$W_r = \frac{1}{\sqrt{2\pi\sigma}} e^{\left[(x - r_{x0})^2 + (y - r_{y0})^2\right]/\sigma^2} \tag{7}$$

Feature selection by modified sea lion optimization

To improve the performance of classifier, feature selection is done which choose the relevant subset of features from the high dimensional dataset. The feature extracted by the above process occupies huge dimension which may reduce the effective performance of this proposed approach. In this method, a modified sea lion optimization algorithm [28] is applied for feature selection to avoid such issue. More relevant features are identified by this selection process and remove the unwanted features from the extracted features.

Modified sea lion optimization algorithm

One of the most intelligent animals is sea lion, which lives in tremendous colonies and it have a huge number of individuals. Within them there are numerous subgroups of their own hierarchy. It can move about these subgroups ordinarily in their lives. The navigations of sea lions rely upon their age, sexual orientation and capacity they have for the entire colony. The sea lions have three phases (i) using the whiskers the sea lion tracking and chasing the prey (ii) once it finds the prey the sea lion calling different individuals to joined their subgroup and surrounding the prey. And then the prey is attacked by sea lion.

Detecting and tracking phase

To recognize shape, position and size of features, the sea lions are utilized their whiskers. This encourages sea lion to detect the existing feature and to recognize their position when the whiskers course is on the other way of water waves. The position of the features is recognized by the sea lions and consider different individuals that will join its subgroup to locate the ideal features.

$$\vec{Dist} = \left| \vec{2B} \,.\, \vec{P(t)} - \vec{SL(t)} \right| \tag{8}$$

where, the distance between the target feature is indicated by $\vec{Dist}$, the position vector of the target features and sea lion is represented as $\vec{SL}(t)$ and $\vec{P}(t)$, t is denoted as the current iteration and random vector is represented as $\vec{B}$.

The sea lions move towards the target features to be nearest at the next iterations. The mathematical model of this behaviour is in below equation.

$$\vec{SL}(t+1) = \vec{P}(t) - \vec{Dist} \,.\, \vec{C} \tag{9}$$

where, $\vec{C}$ is decreased linearly from 2 to 0 and $(t + 1)$ define the next iteration.

Finding best feature

Sea lions can detect the targets prey position and surround them. The hunting system is conducted by the leader (the best search agent), who finds prey and inform them

concerning others. Typically target prey is measured the finest results for the current candidate.

$$\overrightarrow{SL}(t+1) = \left|\overrightarrow{P}(t) - \overrightarrow{SL}(t)\right| \cdot \cos(2\pi m) + \overrightarrow{P}(t) \tag{10}$$

For finding the best features, first we update the position by using galactic swarm optimization (GSO) algorithm [29]. GSO position is updated in sea lion optimization to get the optimal feature.

$$y^{(i)} \leftarrow y^{(i)} + v^{(i)} \tag{11}$$

where, $y^{(i)}$ is represented as the position of GSO. $\left|\overrightarrow{P}(t) - \overrightarrow{SL}(t)\right|$ denotes the distance among the finest optimum solution (target feature) and the search agent, random number is represented by m and $\cos(2\pi m)$ is utilized to represent the features behaviour.

Recurrent fuzzy neural network classifier using social ski-driver for classification

In numerous fields, fuzzy neural networks have been used and RFNN is a notable fuzzy neural network. RFNN [30] has four layers such as input, membership, fuzzy and output layer. Let the input and output is $o_i^{(k)}$ and $u_i^{(k)}$ at the ith node in the layer k respectively. The following procedure is RFNN which is described in detail (Table 2).

Layer 1

First layer is called as input it has N number of nodes every layer is resembles with a parameter. Climate data input parameters like average humidity, evaporation per day, rain quantity, average of temperature and number of daylight hours.

$$O_i^{(1)} = u_i^{(1)} = x_i(t), \quad \text{where } i = 1 \div N \tag{12}$$

Layer 2

Membership layer is known as second layer. In fuzzy data, nodes in second layer will be changed to the crisp data using membership functions like Gauss function. In this layer, neural nodes is represented as $N \times M$, where fuzzy rules is denoted by M. Three parameters are in each and every nodes that is θ_{ij}, σ_{ij} and m_{ij}.

$$O_{ij}^{(2)} = \exp\left[-\frac{\left(u_{ij}^{(2)} - m_{ij}\right)^2}{\left(\sigma_{ij}\right)}\right], \quad \textit{where } i = 1 \div N \text{ and } j = 1 \div M \tag{13}$$

Table 2 Hyper-parameters

Hyper parameters	Ranges
Convolution layer	3
Max pooling layer	3
Padding layer	1

where, m_{ij} is represented as the center of Gauss distribution function and σ_{ij} is represented as the variance of Gauss distribution function.

$$u_{ij}^{(2)}(t) = O_i^{(1)} + \theta_{ij} O_{ij}^{(2)}(t-1), \quad \text{where } i = 1 \div N \text{ and } = 1 \div M \tag{14}$$

where, O_{ij} represents the weight of recurrent nodes. In this layer, the nodes from input has the factor $O_j^{(2)}(t-1)$. The residual data of the earlier learning step is denoted by this layer. This factor signifies the residual data of the earlier learning step. Therefore, in Eq. (13) after replacing $u_{ij}^{(2)}$ by Eq. (12), we get Eq. (14).

$$O_{ij}^{(2)} = \exp\left[-\frac{\left[x_i(t) + \theta_{ij} O_{ij}^2(t-1) - m_{ij}\right]^2}{(\sigma_{ij})^2}\right] \tag{15}$$

Layer 3

Fuzzy rules is represented as the third layer. Fuzzy conclusion is given by connecting layer 3 and Layer 4. In this layer, each node resembles with an AND operation. Every AND operation expression is explained as below.

$$O_j^{(2)} = \prod_{i=1}^{N} O_{ij}^{(2)} \tag{16}$$

$$O_j^{(2)} = \prod_{i=1}^{N} \exp\left[-\frac{\left[x_i(t) + \theta_{ij} O_{ij}^2(t-1) - m_{ij}\right]^2}{(\sigma_{ij})^2}\right] \quad \text{where, } j = 1 \div M \tag{17}$$

Layer 4

Output layer is present in fourth layer which includes P nodes. P will be set to 1 in this model; this is the river runoff value. The fourth layer node is liable for changing fuzzy to crisp.

$$y_k = O_k^{(4)} = \sum_{j=1}^{M} u_{jk}^{(4)} w_{jk} = \sum_{j=1}^{M} O_{jk}^{(3)} w_{jk} \tag{18}$$

$$y_k = \sum_{j=1}^{M} w_{jk} \prod_{i=1}^{N} \exp\left[-\frac{\left[x_i(t) + \theta_{ij} O_{ij}^2(t-1) - m_{ij}\right]^2}{(\sigma_{ij})^2}\right] \quad \text{where, } k = 1 \div P \tag{19}$$

We use SSD [31] optimization algorithm to train RFNN after the defining process of RFNN. From various meta-heuristic algorithms, the behaviour of SSD was inspired. Its name tributes to the way that its stochastic exploration by one way or another resembles the ways that ski-drivers take downhill. Various parameters are in SSD, the description of these parameters is given below.

Position of the agent

To calculate the objective function, the position of the agents is utilized at that particular location, where dimension of search space is represented by n.

Previous best position

Using fitness function, the fitness value is designed for all agents. For each agent the fitness value is linked with its current position, and finally the best position is kept.

Mean global solution

Agents move to the global perspective, which speaks to the best three average solutions.

$$M_i^t = \frac{X_\alpha + X_\beta + X_\gamma}{3} \tag{20}$$

where, best three solutions are represented as X_α , X_β and X_γ

Velocity of the agents

By adding the velocity V_i, the agent positions are updated.

$$X_i^{t+1} = X_i^t + V_i^t \tag{21}$$

where, V_i is the velocity of X_i.

Input: Coefficient of RFNN individual structure and SSD algorithm.
Output: Best RFNN individual fulfil one of criteria.
Step 1: At first, Initialize the SSD algorithm. Linking weights of each RFNN individuals are chance in range [0,1].
Step 2: While ending conditions are not fulfilled do
Step 3: For each RFNN individual ith do
Step 4: If ending conditions are not fulfilled then
Step 5: Break out for loop
Step 6: End if
Step 7: End for
Step 8: If ending conditions are not fulfilled then
Step 9: //Make the generation from SSD
Step 10: Velocity and Position
Step 11: End if
Step 12: End While

Results and discussion

The emotion recognition from group video is extensively applied in various applications. The whole work is done by Python and the performance is also carried out. The dataset is not available in publicly so we collected the video from YouTube. In this group video, the kids are introduced by their self and talking about their inventions. The dataset has size—6s video, source—YouTube, count—8 number of kids and

statistics of classes—4 expressions. The accuracy, sensitivity, specificity, recall and precision is evaluated for four different emotions (Happy, sad, Fear and Angry) and the performance is compared with existing deep learning methods such as CNN [24], DNN + Bayesian [22], SVM [21] and CNN + VGG-16 [23]. The emotion recognition is shown in Fig. 5. The performance using GLCM, LESH and MLTP is shown in Fig. 6.

Figure 7 shows the specificity performance with proposed and exsisting techniques. Here, in this graph the RFNN + SSD is denoted as orange color, CNN + VGG-16 is indicated as green color, SVM is denoted as ash color, DNN + Bayesian is denoted as brown color and CNN is denoted as pink color. Here, the specificity performance is taken as the average of 100% which includes the proposed and existing techniques. In that, the specificity performance of our proposed RFNN + SSD is very high i.e. 26.5% and the specificity performance of remaining existing deep learning methods like CNN + VGG-16 is 19.1%, SVM is 18.7%, DNN + Bayesian is 19.1% and the CNN is 16.6%. Thus, we know our proposed method achieved more specificity than others.

Figure 8 shows the sensitivity performance with proposed and existing techniques. Here, the sensitivity performance is taken as the average of 100% which includes the

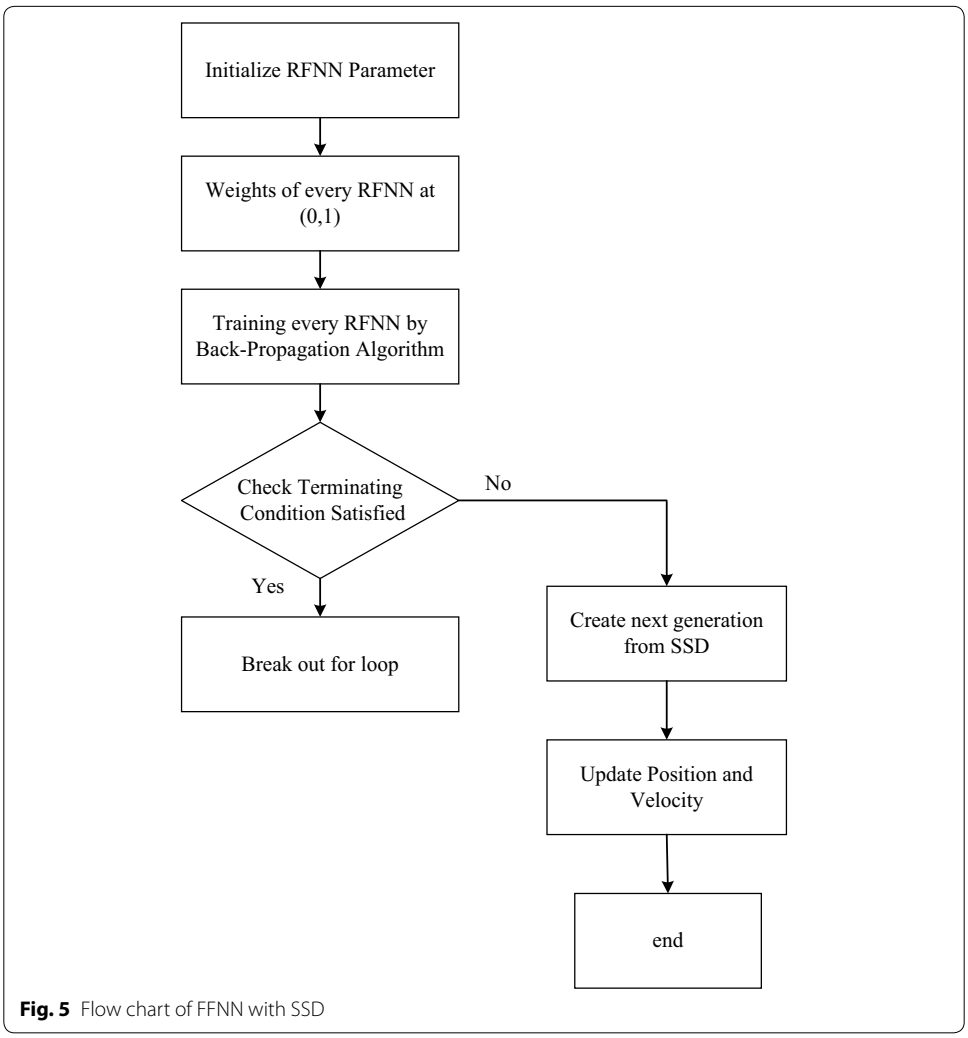

Fig. 5 Flow chart of FFNN with SSD

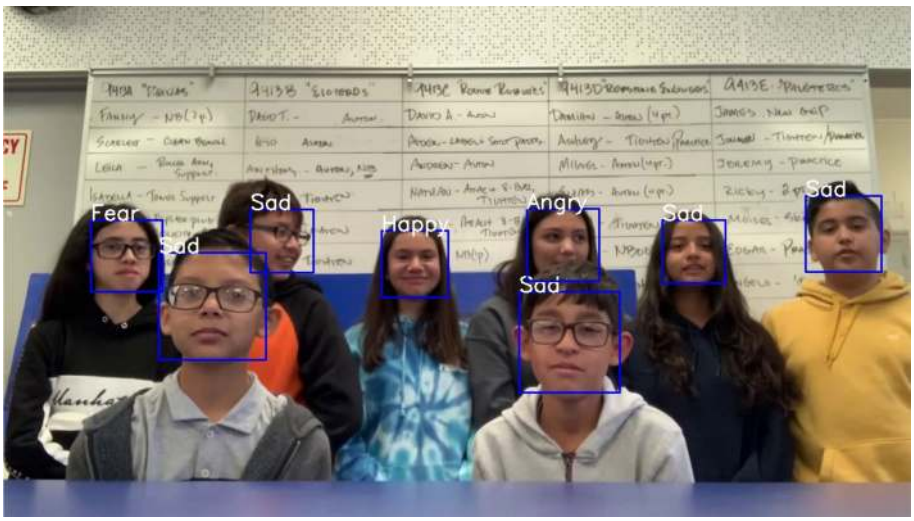

Fig. 6 Emotion recognition

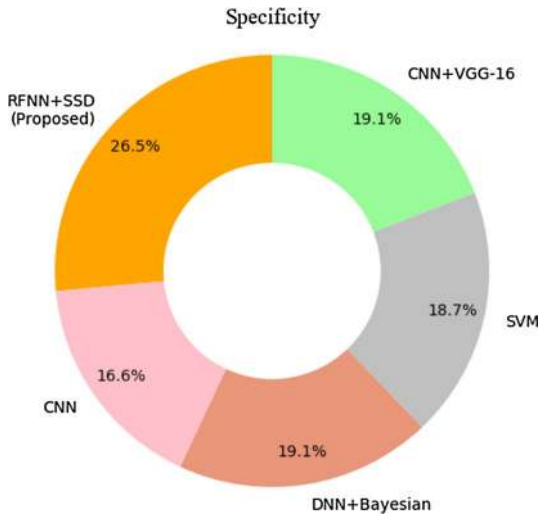

Fig. 7 Specificity performance

proposed and existing techniques. In that, the sensitivity performance of our proposed RFNN + SSD is very high, i.e. 27.0% and the sensitivity performance of remaining existing deep learning methods like CNN + VGG-16 is 20.4%, SVM is 19.6%, DNN + Bayesian is 18.2% and the CNN is 14.8%. Thus, we know the performance of our proposed method achieved more sensitivity than others.

Figure 9 shows the recall performance with proposed and existing techniques. Here, the recall performance is taken as the average of 100% which includes the proposed and existing techniques. In that, the recall performance of our proposed RFNN + SSD is very high i.e. 27.1% and the recall performance of remaining existing deep learning methods like CNN + VGG-16 is 20.4%, SVM is 19.6%, DNN + Bayesian is 18.2% and

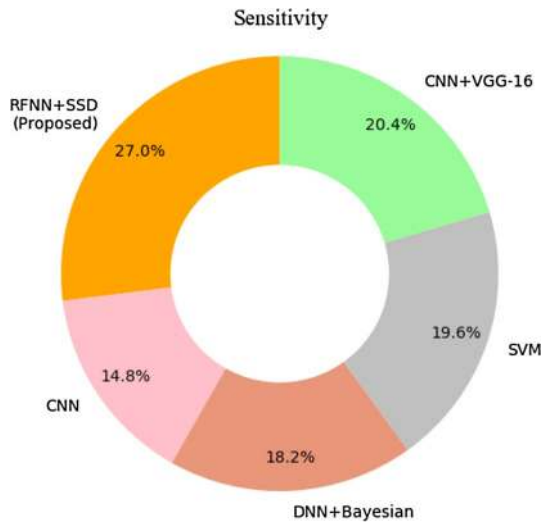

Fig. 8 Sensitivity performance

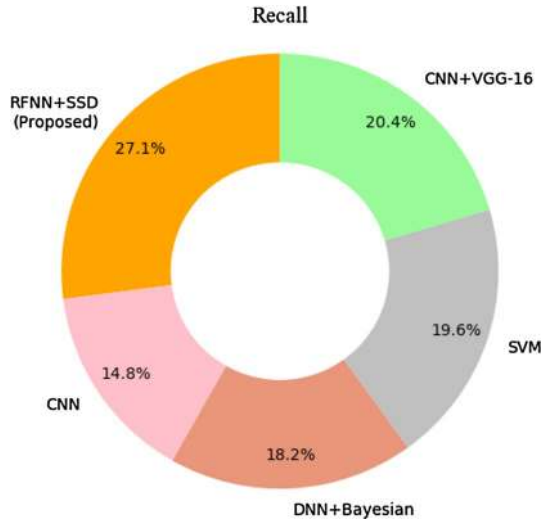

Fig. 9 Recall performance

the CNN is 14.8%. Thus, we know the performance of our proposed method achieved more recall than others.

Figure 10 shows the precision performance with proposed and exsisting techniques. Here, the precision performance is taken as the average of 100% which includes the proposed and existing techniques. In that, the precision performance of our proposed RFNN + SSD is very high i.e. 25.8% and the precision performance of remaining existing deep learning methods like CNN + VGG-16 is 19.5%, SVM is 19.5%, DNN + Bayesian is 19.0% and the CNN is 16.1%. Thus, we know the performance of our proposed method achieved more precision than others.

Figure 11 shows the accuracy performance with proposed and existing methods. Here, the accuracy performance is taken as the average of 100% which includes the

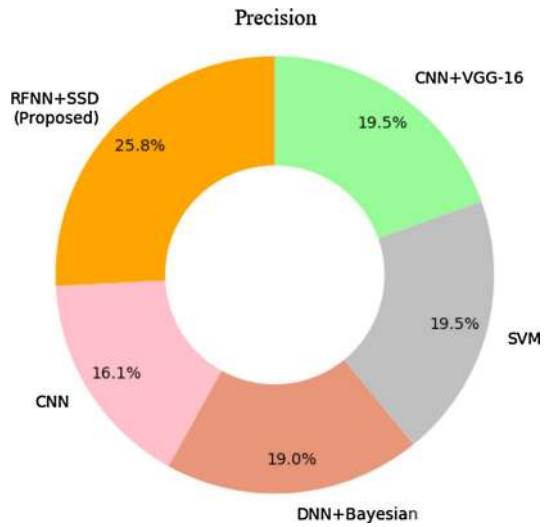

Fig. 10 Precision performance

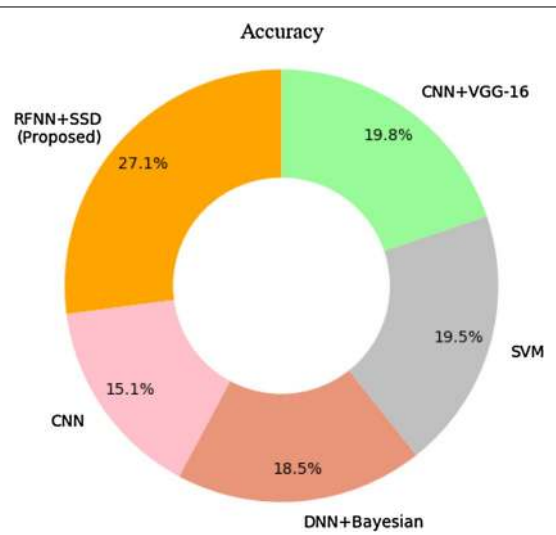

Fig. 11 Comparison of accuracy

proposed and existing techniques. In that, the accuracy performance of our proposed RFNN + SSD is very high i.e. 27.1% and the accuracy performance of remaining existing deep learning methods like CNN + VGG-16 is 19.8%, SVM is 19.5%, DNN + Bayesian is 18.5% and the CNN is 15.1%. Thus, we know the performance of our proposed method achieved more accuracy than others.

Figure 12 shows the proposed method performance. Here, the accuracy of our proposed method is 99.16%, recall is 99.33%, specificity is 99%, precision is 99% and the sensitivity is 99.93%.

Figure 13 shows the performance of proposed RMSE, FNR and FPR. The proposed method contains less root mean square error (RMSE), false negative rate (FNR) and

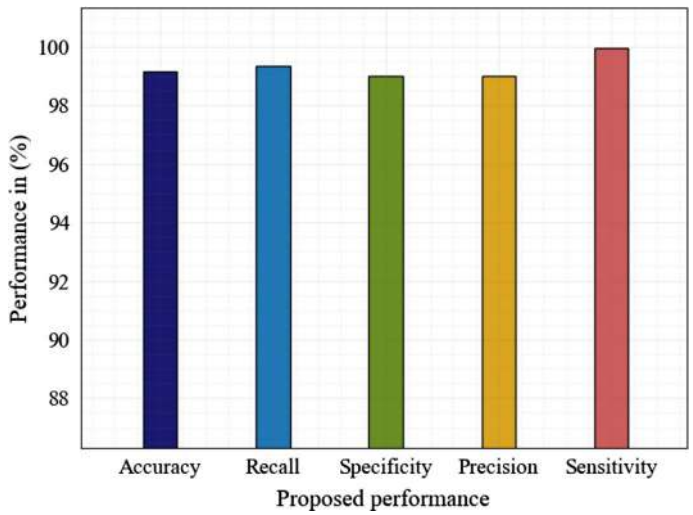

Fig. 12 Performance of proposed methods

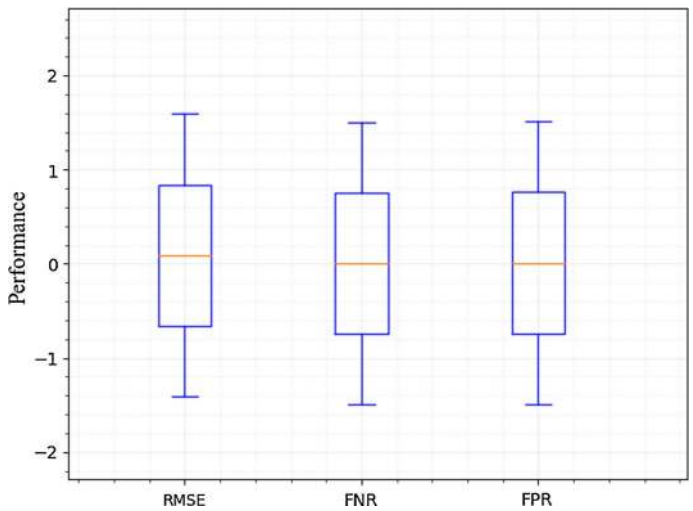

Fig. 13 Proposed RMSE, FNR and FPR

false positive rate (FPR). The value of proposed RMSE is 0.091%, FPR is 0.01% and the FNR is 0.0066%.

Figure 14 shows the proposed performance of accuracy, recall, specificity, precision and sensitivity for low resolution images. The proposed accuracy value for low resolution image is 87.8%, recall value of low resolution image is 88%, specificity value for low resolution is 87.7%, precision for low resolution is 87.7% and sensitivity value for low resolution is 88%.

Figure 15 shows the proposed performance of RMSE, FPR and FNR for low resolution images. The proposed performance of RMSE value for low resolution is 0.35%, the FPR value for low resolution image is 0.13% and the FNR value for low resolution image is 0.124%.

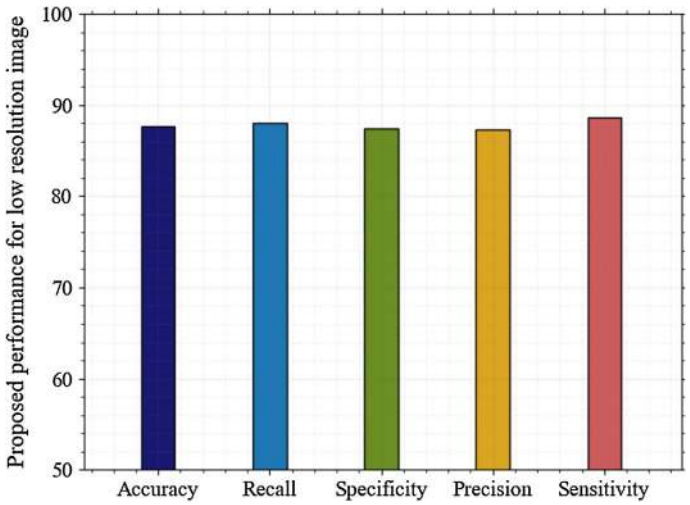

Fig. 14 Low resolution for accuracy, recall, specificity, precision and sensitivity

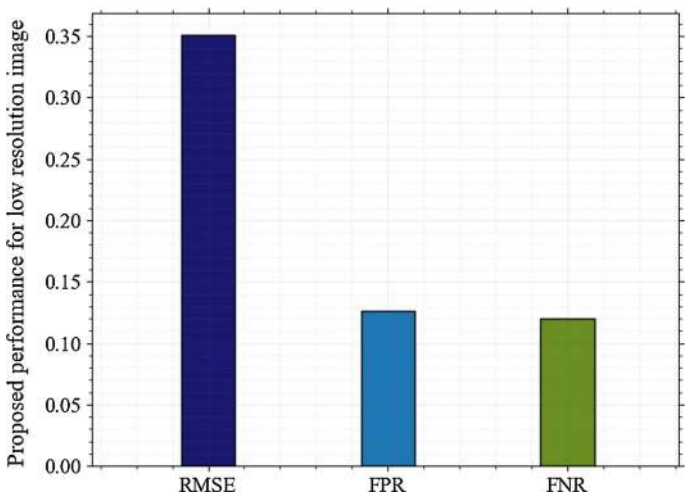

Fig. 15 Low resolution for RMSE, FPR and FNR

Conclusion

In this proposed work, an emotion recognition from group video is taken as the major research topic. In different fields, group video based emotion recognition is gaining huge significance. In different manner, various emotions are shown by human, therefore it is difficult for machines to accurately recognize the human emotions. So here, a group based emotion recognition from video sequence with hybrid optimization based recurrent fuzzy neural network is proposed. In our work, emotion recognition from group of people is performed by utilizing the Neural Network. From the multi user video data, initially, original video frames are taken as input and pre-process it. From this pre-processed image, extract the facial features using MLTP, GLCM, and LESH and carried out as the features in the further processing steps. After extracting the features, certain features are selected using Modified Sea-lion optimization algorithm process.

Finally, RFNN classifier using SSD, optimization algorithm is proposed for updating the weights in the RFNN. The whole work is done by Python platform and the performance of accuracy, sensitivity, specificity, recall and precision is evaluated with some existing techniques. The proposed method accuracy is 99.16%, recall is 99.33%, precision is 99%, sensitivity is 99.93% and specificity is 99% when compared with other deep learning techniques our proposed method attains good result. For low resolution image, the proposed method accuracy is 87.8%, recall is 88%, specificity is 87.7%, precision is 87.7% and sensitivity is 88%. Here, the data is taken from YouTube, when we compared with more data's the complexity will be increase. So in future we will increase the number of data and reduce the complexity by using different machine learning techniques.

Abbreviations

MLTP: Multivariate local texture pattern; GLCM: Grey level co-occurrence matrix; LESH: Local energy-based shape histogram; RFNN: Recurrent fuzzy neural network; FAPG: Fast averaging peer group; SSD: Social ski-driver; FER: Facial expression recognition; HCI: Human computer interaction; DT: Decision tree; SVM: Support vector machine; NN: Neural networks; KNN: K-nearest neighbors; ER: Emotion recognition; GER: Group based emotion recognition; WWW: World Wide Web; TNVPF: Temporal NVP fusion; GECV: Group-level emotion on crowd videos; AAM: Active appearance models; PCA: Principal component analysis; LDA: Linear discriminant analysis; CNN: Convolutional neural network.

Acknowledgements

Not applicable.

Authors' contributions

SV has found the proposed algorithms and obtained the datasets for the research and explored different methods discussed. VN and VKE contributed to the modification of study objectives and framework. Their rich experience was instrumental in improving our work. All authors contributed to the editing and proofreading. All authors read and approved the final manuscript.

Funding

Authors did not receive any funding for this study

Availability of data and materials

The Datasets are not available in publicly so we collected the videos from youtube.

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no Competing interests.

Author details

[1] Department of Computer Science and Engineering, SRK University, Bhopal, Madhya Pradesh 462026, India. [2] Department EEE, SRK University, Bhopal, Madhya Pradesh 462026, India.

Received: 17 March 2020 Accepted: 13 July 2020
Published online: 03 August 2020

References

1. Khorrami, P., Le Paine, T., Brady, K., Dagli, C. and Huang, T.S., 2016, September. How deep neural networks can improve emotion recognition on video data. In 2016 IEEE international conference on image processing (ICIP) (pp. 619-623). IEEE.
2. Kahou SE, Pal C, Bouthillier X, Froumenty P, Gülçehre Ç, Memisevic R, Vincent P, Courville A, Bengio Y, Ferrari RC, Mirza M. December. Combining modality specific deep neural networks for emotion recognition in video. In: Proceedings of the 15th ACM on International conference on multimodal interaction. 2013, pp. 543–50.
3. Walecki R, Rudovic O, Pavlovic V, Pantic M. Variable-state latent conditional random fields for facial expression recognition and action unit detection. In: 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG), vol. 1. IEEE 2015, pp. 1–8.
4. Lee J, Kim S, Kiim S, Sohn K. Spatiotemporal Attention Based Deep Neural Networks for Emotion Recognition. In 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE. 2018, pp. 1513–7.
5. Gupta O, Raviv D, Raskar R. Illumination invariants in deep video expression recognition. Pattern Recogn. 2018;76:25–35.

6. Yan J, Zheng W, Xu Q, Lu G, Li H, Wang B. Sparse kernel reduced-rank regression for bimodal emotion recognition from facial expression and speech. IEEE Trans Multimedia. 2016;18(7):1319–29.
7. Bernal G, Maes P. Emotional beasts: visually expressing emotions through avatars in VR. In: Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems. 2017, pp. 2395–402.
8. Mavridou I, McGhee JT, Hamedi M, Fatoorechi M, Cleal A, Ballaguer-Balester E, Seiss E, Cox G, Nduka C. FACETEQ interface demo for emotion expression in VR. In: 2017 IEEE virtual reality (VR). IEEE. 2017, pp. 441–2
9. Fonnegra RD, Díaz GM. Deep learning based video spatio-temporal modeling for emotion recognition. In: International conference on human–computer interaction. Cham: Springer. 2018, pp. 397–408
10. Li S, Deng W. Deep facial expression recognition: a survey. arXiv preprint arXiv:1804.08348. 2018.
11. Lv Y, Feng Z, Xu C. Facial expression recognition via deep learning. In: 2014 International conference on smart computing. IEEE. 2014, pp. 303–8.
12. Fridlund AJ. Human facial expression: an evolutionary view. New York: Academic Press; 2014.
13. Hossain MS, Muhammad G, Alhamid MF, Song B, Al-Mutib K. Audio-visual emotion recognition using big data towards 5G. Mobile Netw Appl. 2016;21(5):753–63.
14. Sajjad M, Zahir S, Ullah A, Akhtar Z, Muhammad K. Human behavior understanding in big multimedia data using CNN based facial expression recognition. Mobile Netw Appl. 2019;9:1–11.
15. Smith ER, Seger CR, Mackie DM. Can emotions be truly group level? Evidence regarding four conceptual criteria. J Pers Soc Psychol. 2007;93(3):431.
16. Lakshmy V, Murthy OR. Image based group happiness intensity analysis. In: Computational vision and bio inspired computing. Cham: Springer. 2018, pp. 1032–40.
17. Dhall A, Goecke R, Ghosh S, Joshi J, Hoey J, Gedeon T. From individual to group-level emotion recognition: Emotiw 5.0. In: Proceedings of the 19th ACM international conference on multimodal interaction. 2017, pp. 524–8.
18. Dhall A, Kaur A, Goecke R, Gedeon T. Emotiw 2018: audio-video, student engagement and group-level affect prediction. In: Proceedings of the 20th ACM international conference on multimodal interaction. 2018, pp. 653–6.
19. Nagarajan B, Oruganti VRM. Group Emotion recognition in adverse face detection. In: 2019 14th IEEE international conference on automatic face and gesture recognition (FG 2019). IEEE. 2019, pp. 1–5.
20. Jangid M, Paharia P, Srivastava S. Video-based facial expression recognition using a deep learning approach. In: Advances in computer communication and computational sciences. Singapore: Springer. 2019, pp. 653–60.
21. Balaji B, Oruganti VRM. Multi-level feature fusion for group-level emotion recognition. In: Proceedings of the 19th ACM international conference on multimodal interaction. 2017, pp. 583–6.
22. Surace L, Patacchiola M, BattiniSönmez E, Spataro W, Cangelosi A. Emotion recognition in the wild using deep neural networks and Bayesian classifiers. In: Proceedings of the 19th ACM international conference on multimodal interaction. 2017, pp. 593–7.
23. Abbas A, Chalup SK. Group emotion recognition in the wild by combining deep neural networks for facial expression classification and scene-context analysis. In: Proceedings of the 19th ACM international conference on multimodal interaction. 2017, pp. 561–8.
24. Shamsi SN, Rawat BPS, Wadhwa M. Group affect prediction using emotion heatmaps and scene information. In: Proceedings of 2018 IEEE winter applications of computer vision workshops (WACVW). 2018, pp. 77–83.
25. Malinski L, Smolka B. Fast averaging peer group filter for the impulsive noise removal in color images. J Real-Time Image Proc. 2016;11(3):427–44.
26. Wang YQ. An analysis of the Viola-Jones face detection algorithm. Image Processing Line. 2014;4:128–48.
27. Ibrahim FN, Zin ZM, Ibrahim N. Eye center detection using combined Viola-Jones and neural network algorithms. In: 2018 international symposium on agent, multi-agent systems and robotics (ISAMSR). IEEE. 2018, pp. 1–6.
28. Masadeh R, Mahafzah BA, Sharieh A. Sea lion optimization algorithm. Sea. 2019;10:5.
29. Nguyen BM, Tran T, Nguyen T, Nguyen G. Hybridization of galactic swarm and evolution whale optimization for global search problem. IEEE Access. 2020;8:74991–5010.
30. Pratama M, Lu J, Lughofer E, Zhang G, Er MJ. An incremental learning of concept drifts using evolving type-2 recurrent fuzzy neural networks. IEEE Trans Fuzzy Syst. 2016;25(5):1175–92.
31. Tharwat A, Gabel T. Parameters optimization of support vector machines for imbalanced data using social ski driver algorithm. Neural Comput Appl. 2019. https://doi.org/10.1007/s00521-019-04159-z.

Publisher's Note