

Optimization Assisted Convolutional Neural Network for Facial Emotion Recognition

Amit Sarkar

Sanjay Ghodawat Group of Institutes
Hatkanangale, Maharashtra, India
sarkar.amit2k@gmail.com

Abstract: Facial Expression Recognition (FER) is an important type of visual information that can be used to understand a human's emotional situation. FER has attained a significant interest in human-computer interaction, autopilot, medical healing as well as various face expression dependent areas, and it is enormously used in most research areas. Hence, this paper intends to develop an intelligent facial emotion recognition model by following two major processes namely (a) Feature extraction and (b) Classification. Initially, the input image is subjected to extract Local Binary Pattern (LBP) based features. Further, the extracted features are classified using a Convolutional neural network (CNN). Moreover, the weights of CNN are optimally tuned by the Improvised Steering angle and Gear-based ROA (ISG-ROA) algorithm. Finally, the superiority of the ISG-ROA method is compared over existing methods and its improvement is proved effective.

Keywords: Facial Emotion; Expressions; Rider Optimization; Steering angle; Convolutional Neural Network; Local Binary Pattern.

Nomenclature

Abbreviations	Descriptions
CNN	Convolutional Neural Network
C-RNN	Convolutional-Recurrent NN
DA	Dragon Fly algorithm
DLP	Deep Locality-Preserving
FER	Facial Emotion Recognition
FF	Fire Fly algorithm
FFNN	Feed Forward NN
ISG-ROA	Improvised Steering angle and Gear-based ROA
JA	Jaya algorithm
LBP	Local Binary Pattern
PSO	Particle Swarm Optimization
MBO	Monarch Butterfly Optimization
NN	Neural Network
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
RAF-DB	Real-world Affective Face Database
RNN	Recurrent Neural Network
ROA	Rider Optimization Algorithm
STRNN	Spatial-Temporal Recurrent NN

1. Introduction

The face emotion-based interaction process has recently received much interest and attractiveness. The human face is a major part in processing the expression based on emotions, and also identification in social life [6]. Generally, humans keep on changing their emotions based on mental or physical instances occasionally. Based on different human emotions, psychology has described a collection of six facial expressions as a common expression which serves as a foundation for the other expressions, namely, happiness, surprise, sadness, anxiety, anger, and disgust. Such feelings are expressed in the human faces through the gestures of the facial muscles, and the head, lips, nose, and mouth that serves as the essential facial characteristics [7]. Even though humans can recognize people's facial expressions, they get stuck in ambiguity while interacting with a vast number of unfamiliar faces [8].

The facial expressions were playing a key role in many fields of research and thus large quantities of approaches were established so far. Such approaches are generally characterized as methods based on the features, and methods based on appearance. One of the main topics is feature-based methods, where the characteristics related to human facial emotions are extracted from its geometric relations in facial features like nose, head, mouth, and eyes [9]. Many facial feature extraction techniques mostly based on the performance of facial detections with no reliability for realistic applications [10]. This would be the main requirement of facial recognition systems to provide sufficient memory and fast processing speed to overcome human shortcomings such as the variance of the pose, occlusion of features, and facial emotions [11].

The conventional facial recognition system with PCA is really a simple face recognition approach and data compression method. Still, the lighting conditions are not sensitive [12]. The LDA has been one of the commonly utilized projection techniques which would be effective in mapping high-dimensional measurements into the low-dimensional space. LDA's major drawback emerges with such a limited sample whenever the samples are low in dimension [13]. Moreover, AdaBoost lowers the computational time and also improves the speed of detection. The new deep learning approach involves CNNs that concern object recognition as well as detection that also appears to become a complicated process in FER [14]. CNN will evaluate the automated processing of distinct levels of data representation involving higher facial emotions [15]. Nevertheless, due to the insufficiency of data sets, it could not be implemented in the FER region.

The main contribution of the paper is to develop an intelligent facial emotion recognition model by following two major processes such as Feature extraction and Classification. This paper deals with the following sections: Section 2 shows the reviews on FER. Feature extraction based on Local binary pattern is determined in Section 3. The optimization assisted CNN for FER depicts in Section 4. Moreover, Section 5 elaborates on the determining of optimal weights via the ISG-ROA algorithm. Section 6 portrays the results and their discussions and the conclusion of this work is shown in Section 7.

2. Literature Review

2.1 Related Work

In 2019, Zhang et al. [1] have developed an STRNN based on face image in facial emotion recognition systems. The suggested STRNN contained a multidimensional RNN framework defined by the co-occurrence patterns throughout feelings and emotions and collecting long-range contextual signs of spatial components of every other time period of different directions through it. Subsequently, the adopted approach was found with better consistency than conventional methods.

In 2019, Kim et al. [2] have introduced a new technique depending on hierarchical deep learning in FER. The facial images with neural emotions have been generated and extracted by utilizing the autoencoder methodology. In addition, the proposed approach required the extraction of CNN with LBP characteristics and facial expressions from geometric image adjustments. At last, the mixture of dynamic appearance characteristics and static geometric characteristics has provided a much more efficient and accurate outcome.

In 2018, Neha et al. [3] have formulated hybrid C-RNN approach with RNN and CNN methods. CNN methods have been used specifically for the extraction of features and removal of the regression surface. Two separate sources of signals were connected to a spatial-temporal dependence model throughout the suggested model, and also the facial emotions were established for each image. Compared to conventional approaches, the experimental outcomes achieved with high efficiency and better results in the proposed method.

In 2018, Shui-Hua et al. [4] have suggested the FER system to solve the current FER-systems issues. The stationary wavelet entropy approach eliminated the features and hence the single hidden layer FFNN was exploited as a classifier through this work. In addition, JA was implemented in this research which prohibited the training of the classifier at the optimum locals. Compared to other conventional methods, the results of the proposed approach offered the highest accuracy.

In 2019, Li et al. [5] have presented the deep features of modern DLP-CNN technologies involving inter-class scattering and locality nearest. A new facial expression database as well as RAF-DB included several facial images with varying emotions, racial groups, and ages, and each image was recognized from distinct annotators. In the end, the output results of the proposed method have shown improved performance than other existing methods.

3. Feature extraction based on Local Binary Pattern

3.1 Local Binary Pattern

The input image I is subjected to a feature extraction phase, in which the LBP features are extracted. The image pixels were marked by decimal numbers throughout the LBP operator and are matched with neighboring pixels and the center pixel values are removed by the labeling process. The resultant negative value is encoded as 0 and hence the positive, as well as zero values, are encoded with 1. To access the binary numbers, in clockwise directions from top-left the binary codes were rotated and hence known as Local Binary Patterns or LBP codes. In Eq. (2), I_{s_n} and I_{s_m} indicates the intensity of center pixel to its neighbor n , and the center pixel in an image. The pixel in the LBP descriptor is represented by $B_l(\bullet)$. Here, the neighbor count indicated by C_n . The LBP descriptor function was determined using $f_{B(n,m)}$ in Eq. (2). Thus, the extracted features of LBP is represented by f_B .

$$B_l(I_{s_m}) = \sum_{n=0}^{C_n} f_{B(n,m)} 2^{n-1} \quad (1)$$

$$f_{B(n,m)} = \begin{cases} 1, & \text{if } I_{s_n} - I_{s_m} \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

4. Optimization Assisted CNN for Facial Emotion Recognition

4.1 Optimized CNN Model

The extracted feature (f_B) is supplied as an input to optimized CNN for optimal weight tuning [20]. “CNN is a well-known classifier consisting of 3 layers which include a fully connected layer, pooling layer, and convolution layers”.

Since the convolution layer contains numerous convolution kernels to use the feature maps for calculating various optimal weights. Using multiple kernels the complete feature map was established. However, the q^{th} layers similar to t^{th} feature map, the features obtained in the location of (a,b) is determined by $P_{a,b,t}^q$ in Eq. (3). Likewise, the t^{th} filter values were provided in q^{th} layer. The bias term and the optimal weight vector is expressed as w_y^q and o_y^q , correspondingly. Moreover, the weight w_y^q is optimally tuned by the ISG-ROA approach. The t^{th} layer with location (a,b) is linked with input patches as demonstrated in $h_{a,b}^q$. The nonlinearities would be presented in CNN by using an activation function that helps predict a multi-layer network with non-linear features. Suppose that the nonlinear activation function seems to be determined by $G(\bullet)$, the convolutional features as $P_{a,b,t}^q$ and hence the activation value $(k_{a,b,t}^q)$ was expressed in Eq. (4). Nevertheless, the pooling layers will shift-variance as demonstrated by reducing the resolution in feature maps using Eq. (5). For each feature map, the pooling function was expressed as $\text{pool}(\cdot)$ and the local neighborhood at a nearby location (a,b) in each feature map $(k_{a,b,t}^q)$ was demonstrated by $R_{a,b}$.

$$P_{a,b,t}^q = w_t^q h_{a,b}^q + o_t^q \quad (3)$$

$$k_{a,b,t}^q = G(P_{a,b,t}^q) \quad (4)$$

$$l_{a,b,t}^q = \text{pool}(k_{a,b,t}^q), \forall (u,v) \in R_{a,b} \quad (5)$$

Accordingly, the CNN loss function is determined as Eq. (6). The total constraints (θ) in CNN connected to N input-output relations indicated as $\{l^{(y)}, m^{(y)}\}; y \in [1, \dots, N]\}$.

Since, the y^{th} input data, the output of CNN and the linked target values can be represented as $l^{(y)}$, $z^{(y)}$ and $m^{(y)}$ correspondingly.

$$\text{Loss} = \frac{1}{N} \sum_{q=1}^q q(\theta; m^{(y)}, z^{(y)}) \quad (6)$$

Pooling layer: “Pooling layers in CNN carries the down sampling operations with the results obtained from the convolutional layers. Currently, the two renowned types of pooling such as max-pooling (where the higher value is observed) and average pooling (where average pooling is observed)”.

Fully connected layer: “Fully connected layers can be worked with flattened inputs. Normally, the results of the pooling layer were supplied as input to the fully connected layer and hence the name portrays as all inputs are connected to all layers. Typically, fully connected layers happen at the edge of the CNN framework”.

Solution Encoding and Objective Function

The weights of CNN are optimally tuned for enhancing the performance of the FER detection rate. The solutions given for encoding are determined as Fig. 1, where w indicates the weight and ‘M’ represents the total weight counts. The objective function (OF) of this work is expressed in Eq. (7), where, e indicates the error (loss) value.

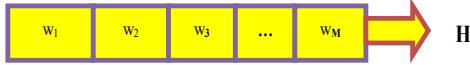
$$OF = \text{Min}(e) \quad (7)$$


Fig. 1. Solution encoding

5. Determining Optimal Weights via ISG-ROA Algorithm

5.1 ISG-ROA Algorithm

ROA [21] is modeled based on the rider group that travels towards the destination in order to win the race.

Initialization: The position of the rider was assumed in Eq. (8), here, X and Y represents the count of total co-ordinates and riders. $H^{\alpha}(c,d)$ indicates the position of c^{th} rider α . The riders “B, F, G and A”, i.e., bypass, followers, over takers, and attackers” were taken in this work. Consider the angles related to the ‘position, steering, and coordinate of the vehicle of c^{th} rider such as β_c , $M_{c,d}^{\alpha}$ and ϕ ’. Moreover, the vehicle constraints like an accelerator, brake, and gear of c^{th} rider are determined using b_c , V_c and t_c correspondingly. The gear V_c gathered values between 0 and 4, in which accelerator b_c and the brake values t_c range between 0 and 1.

$$H^{\alpha} = \{H^{\alpha}(c,d)\}; (1 \leq c \leq Y); (1 \leq d \leq X) \quad (8)$$

Evaluate the success rate: The rider success rate is modified as well as the higher rate of the rider is considered to be the greater one.

Update the position of leading rider: “The leading rider is indeed a rider who is far closer to the target and it is important to remember that the leading rider's position changes with time depending on the success rate. As a consequence, the success rate of all riders is determined and after the iteration the leading rider is predetermined”.

Model of the riders' position: “The rider's position is changed based on tuning these constraints in predetermined approach and hence the procedure is executed till the off-time, α_{off} .” The bypass rider for standard ROA is expressed as per Eq. (9).

$$H_{c,d}^{\alpha+1}(B) = \theta [H_{\kappa,d}^{\alpha} * \rho(d) + H_{\lambda,d}^{\alpha} * (1 - \rho(d))] \quad (9)$$

In Eq. (9), “ θ indicates an arbitrary number obtaining the values between 0 and 1, κ denotes the arbitrary number attaining the values between 1 and Y , λ refers to the arbitrary number obtained between 1 and X . Consecutively, η depicts the arbitrary number value between 0 and 1 with the size $(1 \times X)$. Simultaneously, the follower's position is added based on the leader as expressed in Eq. (10). Where, j refers to the coordinate selector and leaders' position as H^L , L denotes the index of a leader in $H^L(L,j)$, $M_{c,j}^{\alpha}$ indicates the steering angle of c^{th} rider with j^{th} coordinate and ∂_c^{α} refers to the distance covered by c^{th} rider”.

$$H_{c,d}^{\alpha+1}(F) = H^L(L,j) + [\cos(M_{c,j}^{\alpha}) * H^L(L,j) * \partial_c^{\alpha}] \quad (10)$$

In [24], the rider covered distance is taken by multiply velocity and α_{off} . The position of over takers is determined using Eq. (11). Where “ $H_{c,j}^{\alpha}$ represents the position of c^{th} rider in j^{th} coordinate and D_c^{α} denotes the directional indicator of c^{th} the rider in α ” and hence the evaluated depending as the success rate in Eq. (12).

$$H_{c,j}^{\alpha+1}(O) = H_{c,j}^{\alpha} + [D_c^{\alpha} * H^L(L,j)] \quad (11)$$

$$D_c = \left[\frac{2}{1 - \log(F_c)} \right] - 1 \quad (12)$$

In Eq. (12), F_c depicts the success rate of c^{th} rider in α . The higher success rate is determined between Y riders and hence the value ranges between 0 and 1. Hence, the evaluation of coordinate selector changes was obtained in the positions of leading rider and c^{th} rider. The attacker's position is expressed as per Eq. (13).

$$H_{c,d}^{\alpha+1}(A) = H^L(L, j) + \left[\cos(M_{c,j}^{\alpha}) * H^L(L, j) + \delta_c^{\alpha} \right] \quad (13)$$

As per the proposed ISG-ROA model “the gear is determined if $A_j^{\alpha+1}(i) = 0$ and the steering angle is obtained if $A_j^{\alpha+1}(i) = 1$. Since the distance is determined between each solution with the leading solution as expressed using Eq. (14). Consequently, the sorting of solutions are depending on the minimum distance, and the five initial solutions are taken as the best solutions. The steering angle is evaluated if it fulfills the condition $A_j^{\alpha+1}(i) = 1$, and it fits into the best solutions, and gear is attained if it fulfills the condition $A_j^{\alpha+1}(i) = 0$, and it is present in the best solutions. This updating process provided better alteration at greater convergence rate”. Here, LS indicates the leading solution and CS represents the current solution.

$$Dis = Dis\ tan\ ce(LS, CS) \quad (14)$$

Activity Counter: If the success rate of the rider at time $\alpha + 1$ goes above the rate, it will be 1, that is determined at the time α and 0 is allocated for another case of lagging value” as expressed in Eq. (15).

$$A_j^{\alpha+1}(i) = \begin{cases} 1, & \text{if } t_{\alpha+1}(i) > t_{\alpha}(i) \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

Steering Angle: The activity counter is updated as expressed in Eq. (16).

$$T_{i,k}^{\alpha+1} = \begin{cases} T_{i+1,k}^{\alpha}, & \text{if } A_j^{\alpha+1}(i) = 1 \\ T_{i-1,k}^{\alpha}, & \text{if } A_j^{\alpha+1}(i) = 0 \end{cases} \quad (16)$$

Gear: Based on activity counter at $\alpha + 1$ the vehicle gear is updated and hence the higher gear value is determined as per Eq. (17).

$$E_i^{\alpha+1} = \begin{cases} E_i^{\alpha} + 1, & \text{if } A_j^{\alpha+1}(i) = 1 \ \& \ E_i^{\alpha} \neq |E| \\ E_i^{\alpha} - 1, & \text{if } A_j^{\alpha+1}(i) = 0 \ \& \ E_i^{\alpha} \neq 0 \\ E_i^{\alpha}, & \text{otherwise} \end{cases} \quad (17)$$

Re-evaluate the success rate: “The rider updates the position depending on the success rate of individual rider and the rider with higher success rate is declared as the leading rider”.

Update the rider constraints at the termination of the position update: “The parameters to be updated consist of the accelerator, gear, ride off-time, brake, and steering angle together with the activity counter, which is updated at the termination of the iteration”.

Termination: “The phases of optimization are continued until the termination. At last, the optimal solution is attained by the optimization denoting the particular users”.

6. Results and Discussion

6.1 Simulation Procedure

The simulation of the proposed method in FER was evaluated and the corresponding results were obtained using “four famous facial expression datasets, i.e., Oulu-CASIA, JAFFE, CK+, and FER2013 datasets”. The simulation outcomes have shown that the ISG-ROA scheme was better than other traditional models for all four datasets. Finally, the ISG-ROA model was compared with traditional methods like PSO [16], FF [17], MBO [18], and DA [19] and the results were obtained.

6.2 Comparative Analysis using Oulu-CASIA Datasets

The comparative analysis of the ISG-ROA method over the traditional models using Oulu-CASIA datasets is given in Table 1. From the analysis, the recognition rate attained by the ISG-ROA method over the other schemes achieved better outcomes in Oulu-CASIA datasets. The recognition rate of the ISG-ROA method is 94.63% better than other traditional algorithms such as PSO, FF, MBO, and DA

is 93.06%, 88%, 86.95%, and 94.63%. Thus, the betterment of the ISG-ROA method is proved from the results.

Table 1: Comparison of ISG-ROA model over conventional models using Oulu-CASIA datasets

Methods	Recognition Rates (%)
PSO [16]	93.06
FF [17]	88.00
MBO [18]	86.95
DA [19]	88.75
ISG-ROA	94.63

6.3 Comparative Analysis using JAFFE Datasets

Table 2 demonstrates the comparative analysis of the ISG-ROA method over other conventional schemes using JAFFE datasets. Here, the analysis based on the recognition rate attained by the ISG-ROA method over the other schemes achieved higher outcomes in using JAFFE datasets. The recognition rate of traditional algorithms such as PSO, FF, MBO, and DA is 95.8%, 91.8%, 91.8%, and 95.7% and hence the ISG-ROA method achieved better outcomes as 98.52%. Consequently, the improvement of the ISG-ROA method is achieved from the results.

Table 2: Comparison of ISG-ROA model over conventional models using JAFFE datasets

Methods	Recognition Rates (%)
PSO [16]	95.8
FF [17]	91.8
MBO [18]	91.8
DA [19]	95.7
ISG-ROA	98.52

6.4 Comparative analysis using CK+ datasets

The comparative analysis of the ISG-ROA method over other traditional schemes using CK+ datasets is given in Table 3, from which the betterment of the ISG-ROA can be observed. "The CK+ dataset contains 593 image sequences taken from 123 subjects and 327 sequences are labeled with six facial expressions such as anger, disgust, fear, happiness, sadness, and surprise". Here, for all the four traditional schemes like PSO, FF, MBO, and DA obtained 97.3%, 98.5%, 96.10%, and 95.29% and the ISG-ROA method achieved a higher recognition rate of 98.68% using CK+ datasets.

Table 3: Comparison Of Isg-Roa Model Over Conventional Models Using CK+ datasets

Methods	Recognition Rates (%)
PSO [16]	97.3
FF [17]	98.5
MBO [18]	96.10
DA [19]	95.29
ISG-ROA	98.68

6.5 Comparative Analysis using FER2013 Datasets

"FER2013 is a dataset collected in real-world that includes 28,709 training images and 3,589 test images. Here, the FER2013 dataset included pictures of different postures, unbalanced illumination, and occlusion". The comparative analysis of the ISG-ROA method over the traditional models using FER2013 datasets is given by Table 4. From the analysis, the recognition rate attained by the ISG-ROA method is superior to the other schemes outcomes in FER2013 datasets. The recognition rate of the ISG-ROA method is 75.82% better than other traditional algorithms such as PSO, FF, MBO, and DA is 71.1%, 75.2%, 73.73%, and 71.33%, respectively. Finally, the improvement of the ISG-ROA method is achieved using FER2013 datasets.

Table 4: Comparison Of Isg-Roa Model Over Conventional Models Using FER2013 datasets

Methods	Recognition Rates (%)
PSO [16]	71.1
FF [17]	75.2
MBO [18]	73.73
DA [19]	71.33
ISG-ROA	75.82

6. Conclusion

This paper indented to establish a facial emotion recognition model by following two major processes: Feature extraction and Classification. Here, the input image was given to the feature extraction process, where the LBP features were extracted. Then, the extracted features were given to the classification process, for which the CNN model was deployed. Further, the weight on CNN was optimally tuned by the ISG-ROA algorithm. From the analysis, the recognition rate of the ISG-ROA method was 94.63% better than other traditional algorithms such as PSO, FF, MBO, and DA was 93.06%, 88%, 86.95% and 94.63% in Oulu-CASIA datasets. Finally, the recognition rate of the ISG-ROA method is 75.82% better than other traditional algorithms such as PSO, FF, MBO, and DA is 71.1%, 75.2%, 73.73%, and 71.33%, using FER2013 datasets. Thus, the betterment for the presented scheme was determined effectively.

References

- [1] T. Zhang, W. Zheng, Z. Cui, Y. Zong and Y. Li, "Spatial-Temporal Recurrent Neural Network for Emotion Recognition," *IEEE Transactions on Cybernetics*, vol.49, no.3, pp. 839-847, March 2019.
- [2] J. Kim, B. Kim, P. P. Roy and D. Jeong, "Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure," *IEEE Access*, vol.7, pp. 41273-41285, 2019.
- [3] Neha Jain, Shishir Kumar, Amit Kumar, Pourya Shamsolmoali, Masoumeh Zareapoor, "Hybrid deep neural networks for face emotion recognition", *Pattern Recognition Letters*, vol.115, pp.101-106, Nov 2018.
- [4] Shui-Hua Wang, Preetha Phillips, Zheng-Chao Dong, Yu-Dong Zhang, "Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm" *Neurocomputing*, vol.272, pp. 668-676, Jan 2018.
- [5] S. Li and W. Deng, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition," *IEEE Transactions on Image Processing*, vol.28, no.1, pp. 356-370, Jan. 2019.
- [6] Min Hu, Haowen Wang, Xiaohua Wang, Juan Yang, Ronggui Wang, "Video facial emotion recognition based on local enhanced motion history image and CNN-CTSLSTM networks" *Journal of Visual Communication and Image Representation*, vol. 59, pp. 176-185, Feb 2019.
- [7] Xusheng Wang, Xing Chen, Congjun Cao, "Human emotion recognition by optimally fusing facial expression and speech feature" *Signal Processing: Image Communication*, vol. 84, Art.no. 115831, May 2020.
- [8] Aya Hassouneh, A. M. Mutawa, M. Murugappan, "Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods" *Informatics in Medicine Unlocked*, vol. 20, Art.no. 100372, 2020.
- [9] Zhaoli Zhang, Chenghang Lai, Hai Liu, You-Fu Li, "Infrared facial expression recognition via Gaussian-based label distribution learning in the dark illumination environment for human emotion detection" *Neurocomputing*, vol. 409, pp.341-350, 7 Oct 2020.
- [10] Jianfeng Zhao, Xia Mao, Lijiang Che, "Speech emotion recognition using deep 1D & 2D CNN LSTM networks" *Biomedical Signal Processing and Control*, vol. 47, pp. 312-323, Jan 2019.
- [11] Jing Li, Kan Jin, Dalin Zhou, Naoyuki Kubota, Zhaojie Ju, "Attention mechanism-based CNN for facial expression recognition" *Neurocomputing*, vol. 411, pp. 340-350, 21 Oct 2020.
- [12] Wang Xiaohua, Peng Muzi, Pan Lijuan, Hu Min, Ren Fuji, "Two-level attention with two-stage multi-task learning for facial emotion recognition" *Journal of Visual Communication and Image Representation*, vol. 62, pp. 217-225, July 2019.
- [13] Daizong Liu, Xi Ouyang, Shuangjie Xu, Pan Zhou, Shiping Wen, "SAANet: Siamese action-units attention network for improving dynamic facial expression recognition" *Neurocomputing*, vol. 413, pp. 145-157, 6 Nov 2020.
- [14] M. Shamim Hossain, Ghulam Muhammad, "Emotion recognition using secure edge and cloud computing" *Information Sciences*, vol. 504, pp. 589-601, Dec 2019.
- [15] Luefeng Chen, Mengtian Zhou, Wanjuan Su, Min Wu, Kaoru Hirota, "Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction" *Information Sciences*, vol. 428, pp. 49-61, Feb 2018.
- [16] M.E.H. Pedersen and A.J. Chipperfield, "Simplifying Particle Swarm Optimization", *Applied Soft Computing*, vol. 10, pp. 618-628, 2010.
- [17] A.H. Gandomi, X.-S. Yang, S. Talatahari, and A.H. Alavi, "Firefly algorithm with chaos", *Commun Nonlinear Sci Numer Simulat*, vol. 18, pp. 89-98, 2013.
- [18] Gai-Ge Wang, Suash Deb, Zhihua Cui, "Monarch butterfly optimization" *Neural Computing & Application*, 2015, DOI: 10.1007/s00521-015-1923-y.
- [19] Seyedali Mirjalili, "Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems" *Neural Computing Applications*, vol.27, no.4, pp. 1053-1073, 2015.
- [20] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision", In *Circuits and Systems, International Symposium on*, pp.253-256, 2010.
- [21] D. Binu and B. S. Kariyappa, "RideNN: A New Rider Optimization Algorithm-Based Neural Network for Fault Diagnosis in Analog Circuits," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 1, pp. 2-26, Jan. 2019, doi: 10.1109/TIM.2018.2836058.