Multi-Pose Facial Expression Recognition Using Hybrid Deep Learning Model with Improved Variant of Gravitational Search Algorithm

Yogesh Kumar
Department of Computer Science and Engineering, GD Goenka University, India
yogesh.kumar1@gdgu.org

Shashi Kant Verma
Department of Computer Science and Engineering, G B Pant Institute of Engineering and Technology, India
shashikant.verma@gbpec.ac.in

Sandeep Sharma
Department of Information Technology, Ajeenkya DY Patil University, India
tek.learn@gmail.com

Abstract: *The recognition of human facial expressions with the variation of poses is one of the challenging tasks in real-time applications such as human physiological interaction detection, intention analysis, marketing interest evaluation, mental disease diagnosis, etc. This research work addresses the problem of expression recognition from different facial poses at the yaw angle. The major contribution of the paper is the proposal of an autonomous pose variant facial expression recognition framework using the amalgamation of a hybrid deep learning model with an improved quantum inspired gravitational search algorithm. The hybrid deep learning model is the integration of the convolutional neural network and recurrent neural network. The applicability of the hybrid deep learning model can be considered as significant if the feature set is efficiently optimized to have the discriminative features respective to each expression class. Here, the Improved Quantum Inspired Gravitational Search Algorithm (IQI-GSA) is utilized for the selection and optimization of features. The IQI-GSA method is significant for optimizing the features compared to quantum-behaved binary gravitation search algorithm for handing the local optima and stochastic characteristics. Comparing with state-of-art techniques, the proposed framework exhibits the outperformed recognition rate for experimentation on Karolinska Directed Emotional Faces (KDEF) and Japanese Female Facial Expression (JAFFE) datasets.*

Keywords: *Deep learning, emotion recognition, quantum computing, swarm intelligence, convolutional neural networks, recurrent neural networks.*

Received January 9, 2021; accepted August 15, 2021
https://doi.org/10.34028/iajit/19/2/15

1. Introduction

Facial expression recognition is an autonomous task to determine human intentions from the facial components without any verbal interactions. It has also been noted that there is 48% greater impact of facial expressions during conveying a message in comparison with verbal communication [13].

The present work has proposed an ensemble approach of Improved Quantum-Inspired Gravitational Search Algorithm (IQI-GSA) and Hybrid Deep Neural Networks (HDCR-NN) for multi-pose facial expression recognition. The IQI-GSA [19] is employed to optimize the features the extracted features. The IQI-GSA is an amalgamation of Quantum Computing (QC) and Gravitational Search Algorithm (GSA) [17]. The improved variant of quantum computing with GSA IQI-GSA is adapted to avoid trapping of individual GSA in local optima and maintain the convergence rate at the later stage [19].

The HDCR-NN [4] is a composition of the convolutional neural network and the recurrent neural network. It is adapted for the classification of facial expressions. The wide acceptability of convolutional neural networks [24] to automatically learn the feature attributes and recurrent neural networks [25] to utilize sequential information is the motivation to hybrid with its attributes. The experimentation evaluation for the proposed methodology is conducted on the datasets of Karolinska Directed Emotional Faces (KDEF) and the front pose dataset of Japanese Female Facial Expression (JAFFE).

The work related to the concept has mainly been explored by numerous researchers [1, 2, 8, 18, 23]. In the existing techniques, the consideration of multi-pose expression recognition is still a challenging issue as the recognition rate declines with the consideration of yaw angles of the face. The present work addresses the issue by introducing the ensemble approach for the recognition of multi-pose facial expressions.

The rest of the paper is organized as follows: section 2 delivers the datasets and pre-processing steps of datasets. Section 3 brings the proposed ensemble approach of IQI-GSA and HDCR-NN for the recognition of facial expressions. Section 4 presents the analysis and discussion of the results. The paper ends with the concluding observations and future scope discussed in section 5.

2. Datasets and Pre-Processing

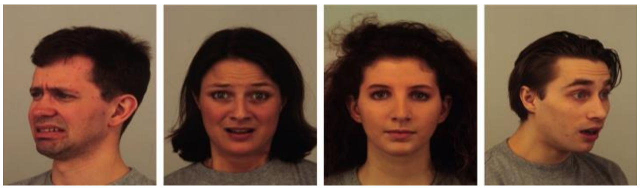
The section describes the considered datasets and the pre-processing steps.

2.1. Research Datasets

The image based natural datasets of JAFFE and KDEF are considered for experimentation. The sample images of the JAFFE and KDEF datasets are illustrated in Figures 1 and 2 respectively. In Figure 1, the images illustrate the facial expressions in the order: anger, happy, and sad. In Figure 2, the images illustrate facial expressions in the order: disgust (-45°), fear (0°), neutral (0°), and surprise (+45°).

a) Anger. b) Happy. c) Sad.

Figure 1. Sample Images of JAFFE dataset.



a) Disgust (-45°). b) Fear (0°). c) Neutral (0°). d) Surprise (+45°).

Figure 2. Sample Images of KDEF dataset.

JAFFE [12] is a dataset of 10 female Japanese subjects expressing seven emotion classes. There are a total of 213 images defining the facial expression with the front pose. KDEF [11] is a dataset of 70 models for the five yaw angles (-90, -45, 0, +45, and +90). In the current work, the facial expression images with angles -45, 0, and +45 are utilized for the experimentation as the full side pose images (-90 and +90 degree face images) carry lesser feature attributes to detect facial expressions. This remains the 2940 images possessing Front Pose (FTP) and Half Side Pose (HSP).

2.2. Pre-Processing

The dataset pre-processing constitutes the sub-steps of face extraction and facial features extraction. The Viola-Jones algorithm is used for the face extraction from the input dataset image. The algorithm evaluates the Haar-like features after considering the image as a rectangular grid [21]. The cascade structure extracts the face by eliminating negative regions of non-facial components and adding the positive components of the face.

The extracted face data is utilized to extract the facial features by transforming the information into feature vectors. The appearance-based feature components are extracted using the hybrid approach of Local Binary Patterns (LBP) [15] and the Gabor filter (GF) [9], termed the Hybrid LBP and GF (LGBP) approach.

The detained process of face extraction using the Viola-Jones algorithm and facial features extraction using the hybrid LGBP approach is described in our existing research work [6].

3. Proposed Ensemble Approach for Multi-Pose Facial Expression Recognition

The section describes the proposed ensemble approach of IQI-GSA and HDCR-NN. The ensemble approach is adapted to attain a higher recognition rate by using the IQI-GSA for feature optimization and HDCR-NN for the classification of facial expressions.

3.1. Features Selection and Optimization using IQI-GSA

The IQI-GSA is an improved variant of QIGSA, which is an amalgamation of quantum computing and gravitational search algorithm [19]. The GSA was introduced by Rashedi *et al*. [17] in the category of swarm intelligence with masses as the swarm agents that obey Newton's laws of motion and gravity. The Quantum Inspired Gravitational Search Algorithm (QIGSA) [14] is a variant of GSA with the integration of quantum computing attributes.

The concept of QIGSA is improved to avoid the trapping of the mass agents in the local optima. In QIGSA, the diversity of the solution space decreases with the later iterations of the mass agents which can lead to the trapping of the agents in the local search space, and this trapping can create an imbalance between exploration and exploitation [19]. The improved QIGSA IQI-GSA addresses this problem by modifying the mean best position value $XMbest_i$ of the mass agents.

The process of IQI-GSA begins with the initialization of uniformly distributed masses and random positions in the search space. The fitness of the considered agents is evaluated, and the first K agents with higher mass values and best fitness values are considered. The position of these K agents is termed as *Xkbest*. The value of the best position of the agents is also evaluated and termed as *XAbest*. The *XAbest* value is compared with the agent's fitness value, and if the current position is found to be better than *XAbest*, then the *XAbest* value is equalized to the current value. Further, the movement of the masses can be analyzed [14] as illustrated in Equation (1).

$$\begin{cases} X_i(t+1) = XBest_i + \delta.|XBest_i - X_i(t)|.\ln\left(\frac{1}{R}\right) & if\ S \geq 0.5 \\ X_i(t+1) = XBest_i - \delta.|XBest_i - X_i(t)|.\ln\left(\frac{1}{R}\right) & if\ S < 0.5 \end{cases} \quad (1)$$

Where, R and S are random values that lie within [0, 1] and are evaluated as per the uniform probability distribution. The design parameter δ is the coefficient of contraction-expansion and $XBest_i$ is the updated best position of mass agents. The evaluation of the position of the mass agents is updated on the basis of the mean best position value (*XMbest*) and the current best position of agents (*XAbest*) values as illustrated in Equation (2).

$$XBest_i = \frac{r_1 XMbest_i + r_2 XAbest_i}{r_1 + r_2} \quad (2)$$

Where, r_1 and r_2 are the random numbers that lie within the range of [0, 1] and are evaluated with a uniform probability distribution. The modified formulation of $XMbest_i$ as per IQI-GSA is described in Equations (3) and (4).

$$XMbest_i = \frac{\sum_{p=1}^{K} \frac{M_i}{distance_{i,p}} random_p . Xkbest_p}{\sum_{q=1}^{K} \frac{M_i}{distance_{i,q}} random_q} \quad (3)$$

Where,

$$distance_{i,q} = \left\| X_i - Xkbest_q \right\| \quad (4)$$

Where, '*random*' is considered as a randomized vector equalized to the *Xkbest* set of values that are evaluated randomly within the value range of [0, 1]. This '*random*'vector helps to break out the trapping of agents in local optima by retaining the stochastic characteristics.

Furthermore, the fitness value of all the mass agents is evaluated. The fitness function [3] for any agent *i* can be evaluated as illustrated in Equation (5).

$$fit_i^{itr} = \frac{DSC_i^{itr}}{DDC_i^{itr}} \quad (5)$$

Where, DSC_i^{itr} and DDC_i^{itr} indicate the summation of the Euclidean distance of selected features that indicate the same facial expression class and different expression classes respectively. The term *itr* refers to the iteration number. The values of DSC_i^{itr} and DDC_i^{itr} are calculated as illustrated in Equations (6) and (7) respectively.

$$DSC_i^{itr} = \sum_c \sum_{g,h \in c,\ g \neq h} \left\| F_{g,i}^{itr} - F_{h,i}^{itr} \right\| \quad (6)$$

$$DDC_i^{itr} = \sum_{c,c'} \sum_{g \in c,\ s \in c',\ c \neq c'} \left\| F_{g,i}^{itr} - F_{s,i}^{itr} \right\| \quad (7)$$

Where, c and c' are considered to describe the expression classes of images g^{th}, h^{th} (for c class), and s^{th} (for c' class) respectively. The terms $F_{g,i}^{itr}$, $F_{h,i}^{itr}$ and $F_{s,i}^{itr}$ refer to feature set of g^{th}, h^{th} and s^{th} images respectively.

The features are optimized on the basis of IQI-GSA from the set of extracted features obtained using the LGBP approach. The IQI-GSA method optimized the features by reducing the dimensionality of the feature vectors. The final decision to select the optimized features is conducted based on the binarization criteria expressed in Equation (8).

$$x_i(t+1) = \begin{cases} 1, if\ r < |\beta_i(t+1)|^2 \\ 0,\ otherwise \end{cases} \quad (8)$$

Where, $x_i(t + 1)$ is the updated position of any agent *i* whose initial position was $x_i(t)$. $|\beta_i|^2$ is the probability of quantum state as the q-bit value of 1. The position of the agent *i* changes on the basis of random variable *r* that lies in [0, 1].

The determination of value 1 of the position vector indicates the optimized feature and 0 indicates the feature is inefficient to consider.

3.2. Classification using HDCR-NN

The classification of the facial expressions is conducted using the HDCR-NN [4]. Recent research studies of deep learning models have presented significantly more results than the other techniques [10, 22]. In the current classification process with HDCR-NN, the Deep Convolutional Neural Network (DCNN) fixes all the parameters among the selected optimized features along with the elimination of the regression layer and the DRNN combines the sequential information attained with the previous network.

DCNN is a network of numerous convolutional, pooling, and fully connected layers that map the input and result outcomes in a matrix format. The convolution layer handles the huge data of images by restricting the connections between input and hidden layers. A separate filter map (weight vectors) is produced with respect to each convolution unit. At each convolution layer, the facial feature map is determined between local patches and weight vectors. The training of the network is conducted with the Rectified Linear Function (ReLU) activation function of unit v_{nc}^{xy} as described in Equation (9).

$$v_{nc}^{xy} = max\left(b_{nc} \sum_m \sum_{h-1}^{H} \sum_{w-1}^{W} W_{(c-1)m}^{hw}\, v_{(c-1)m}^{(x+h)(y+w)}, 0\right) \quad (9)$$

Where (x, y) describe the position of the pixel in the c^{th} layer of n^{th} filter, b_{nc} illustrate the bias, and $W_{(c-1)m}^{hw}$ depict the weight connection among the units (h,w) from the previous layer $(c$-1).

In the pooling layer, the dimensionality of the feature map is reduced and maximum activation of the local receptive field $u(x, y)$ is determined with pooling units using Equation (10).

$$a_j = max_{nxn}\, v_{nc} u(x,y)) \quad (10)$$

The fully connected layer is the last layer of DCNN. In the proposed DCNN network, the number of input layer neurons depends on the number of feature vectors. The output of the DCNN is redirected to the DRNN network and evaluated as a single output layer. The training network of the DCNN is a supervised network with back propagation learning mechanism. The network is learned using Equations (11) and (12).

$$\Delta w_{ij}(t) = -\eta \frac{\partial E_p(t)}{\partial w_{ij}(t)} \quad (11)$$

$$E_p = \frac{1}{2} \sum (d_{pk} - s_{pk})^2 \quad (12)$$

The learning rate is specified by η and output neurons as k. E_p refers to the network error for p^{th} pattern. d_p and s_p represents the preferred, and calculated neural output for p^{th} training vector. The network testing is conducted by allocating the utmost value of output neurons to the class of each obtainable pattern.

The output of the DCNN network is directed to the DRNN network as the input layers. The DRNN network considers these inputs and generates the sequential output. In the DRNN network, the Elman network structure is utilized which consists of three types of layers in a sequential manner. The three-layer network maintains the input, hidden, and output layer network structure. The classification evaluation with DRNN considers the time t, activation functions (σ_h) for the hidden layer and σ_y for the output layer). The evaluation of the hidden layer vector (h_t) and output layer vector (y_t) is illustrated in Equations (13) and (14) respectively.

$$h_t = \sigma_t(W_h x_t + W_{rec} h_{t-1}) \quad (13)$$

$$y_t = \sigma_y(W_{out} h_t) \quad (14)$$

Where, W_h is the weight of the hidden layer matrix, W_{rec} is the weight of the recurrent matrix, W_{out} is the weight of the output layer matrix as the resultant matrix parameter, and x_t is the input vector.

In the process of HDCR-NN, the classification is conducted by considering the 200-dimensional vectors while passing the image from the fully connected layers. The batch size of 32 is set for experimentation. The total number of 150 hidden units and 6 hidden layers are considered to attain higher classification performance [4]. It proceeds with the time t, with the number of frames N, and the utilization of extracted feature vectors to pass from DCNN. The output frame vector is passed into the DRNN and final classified labeled results are obtained.

4. Experimentation and Discussion

The proposed system is evaluated using the recognition rate parameter. The formulation of considered performance measures is illustrated in Equations (15).

$$Recognition\ Rate = \frac{TP}{TP+FN} \quad (15)$$

Where, TP and FN indicate the true positive and false negative values of the evaluated results, respectively.

4.1. Results Evaluation

The experimentation of the proposed ensemble approach is conducted with the JAFFE and KDEF datasets.

The JAFFE dataset is divided into 60:40 ratios for the training and testing respectively. This division employs the consideration of 129 images out of a total of 213 images to train the system and the rest of 84 images for testing. The classification results as the confusion matrix for the JAFFE dataset are presented in Table 1.

In confusion matrix tables, the rows and columns are Surprise (SU), Sad (SA), Disgust (DI), Happy (HA), Neutral (NE), Anger (AN), Fear (FE), and Total (TT).

Table 1. Confusion matrix results of proposed ensemble approach for experimentation on JAFFE dataset.

	AN	DI	FE	HP	NE	SA	SU	TT
AN	12	00	00	00	00	00	00	12
DI	00	11	01	00	00	00	00	12
FE	00	01	10	00	00	00	00	11
HA	00	00	00	12	00	00	01	13
NE	00	00	00	00	12	01	00	13
SA	00	00	01	00	00	11	00	12
SU	00	00	00	00	00	00	11	11
TT	12	12	12	12	12	12	12	84

Table 2. Confusion matrix results of proposed ensemble approach for experimentation on front pose images of KDEF dataset.

	AN	DI	FE	HP	NE	SA	SU	TT
AN	28	01	00	00	00	00	00	29
DI	00	26	00	00	00	02	00	28
FE	00	00	26	00	00	00	01	27
HA	00	00	00	28	00	00	00	28
NE	00	00	00	00	27	01	00	28
SA	00	01	01	00	01	25	00	28
SU	00	00	01	00	00	00	27	28
TT	28	28	28	28	28	28	28	196

Further, the experimentation results for the KDEF dataset are evaluated. Among the total 2940 images, 980 images possess a front pose (0 degrees), 1960 images possess a half side pose (-45 and +45 degrees), and the combined category experiment considers all the 2940 images. For all three experiments, the training and testing ratio of 80:20 is considered. This remains the testing data of 196 images of the front pose, 392 images of the half side pose, and 588 images for the combined front and half side pose. The confusion matrix results for these categories of the KDEF dataset are illustrated in Tables 2, 3, and 4.

Table 3. Confusion matrix results of proposed ensemble approach for experimentation on half side pose images of KDEF dataset.

	AN	DI	FE	HP	NE	SA	SU	TT
AN	50	04	01	00	00	03	01	59
DI	04	47	02	01	00	02	00	56
FE	02	02	44	01	02	04	04	59
HA	00	01	00	54	00	00	00	55
NE	00	01	02	00	53	01	00	57
SA	00	01	02	00	01	46	00	50
SU	00	00	05	00	00	00	51	56
TT	56	56	56	56	56	56	56	392

Table 4. Confusion matrix results of proposed ensemble approach for experimentation on half side pose images of KDEF dataset.

	AN	DI	FE	HP	NE	SA	SU	TT
AN	78	05	01	00	00	04	00	88
DI	04	74	02	00	00	03	00	83
FE	02	02	71	00	00	01	05	81
HA	00	00	00	83	01	00	00	84
NE	00	01	02	01	82	04	00	90
SA	00	02	02	00	01	72	01	78
SU	00	00	06	00	00	00	78	84
TT	84	84	84	84	84	84	84	588

The confusion matrix results illustrated in Tables 1, 2, 3, and 4 for JAFFE and KDEF datasets are utilized to calculate the recognition rate for these datasets with different available poses. The recognition rate for these datasets is evaluated in Table 5.

The confusion matrix and performance assessment results evaluated for the JAFFE dataset (refer to Tables 1 and 5) indicate that the anger, happy, and neutral facial expression classes have attained a 100% recognition rate. The other expression classes have some confusion with other expressions. The proposed ensemble approach is calculated with an average recognition rate of 94.05%.

The evaluated results illustrated in Tables 2 and 5 for KDEF FTP indicate that the anger and happy emotion classes are classified with 100% recognition rate. The other expression classes are noted with some confusion of expressions. The proposed ensemble approach has attained an average recognition rate of 95.41%.

The results calculated in Tables 3 and 5 for the HSP image category of the KDEF dataset signify that the performance of the ensemble approach reduces with the variation of the pose to +45 and -45 degree angles. The major confusion in the fear expression class is noted with almost all the expression classes except the happy class. The proposed ensemble approach has attained an average recognition rate of 88.01%.

The combined category results of the KDEF dataset shown in Tables 4 and 5 indicate that the combined category achieved superior performance to the individual category of half-side pose images but inferior to the front pose images. The average recognition rate of 91.50% has been attained by the proposed ensemble approach.

Table 5. Recognition rate of different datasets using proposed ensemble approach.

Dataset	Anger (%)	Disgust (%)	Fear (%)	Happy (%)	Neutral (%)	Sad (%)	Surprise (%)
JAFFE (Front Pose)	100	91.67	83.34	100	100	91.67	91.67
KDEF (Front Pose)	100	92.86	92.86	100	96.43	89.29	96.43
KDEF (Half Side Pose)	89.29	83.93	78.57	96.43	94.64	82.14	91.07
KDEF (Combined Front and Half Side Pose)	92.86	88.10	84.52	98.81	97.62	85.71	92.86

4.2. Comparisons

The comparison of the proposed ensemble approach with state-of-art techniques is conducted on the basis of the recognition rate. For the JAFFE dataset, the comparison with the existing techniques of Principal Component Analysis with Deep Convolutional Neural Network (PCA+DCNN) [6], Kernel based Principal Component Analysis Network (K-PCANet) [20], and Kernel based Linear Discriminant Analysis Network (K-LDANet) [20] is illustrated in Table 6.

Table 6. Comparative analysis for the JAFFE dataset.

Technique	Recognition Rate
Proposed Ensemble Approach	94.05%
PCA+DCNN [6]	77.38%
K-PCANet [20]	68.80 %
K-LDANet [20]	62.69 %

For the KDEF dataset, the comparison of the proposed ensemble approach is conducted with the existing techniques of PCA+DCNN [6], K-PCANet [20], K-LDANet [20], QIBGSA+DCNN [7], Multi-class Support Vector Machine (SVM) [5], Convolutional Neural Network (CNN) Yu [16], CNN Tang [16], CNN Kahou [16], and CNN Image-Net [16]. The comparison results for the KDEF dataset are illustrated in Table 7.

The performance comparison illustrated in Tables 6 and 7 evidently depicts the outperformed performance of the proposed ensemble approach in comparison with existing techniques. It can also be analyzed that the recognition rate decreases with the movement of face sideways at yaw angle. The combined category results are superior to the half-side pose expressions and inferior to front pose expressions.

Table 7. Comparative Analysis on the basis of recognition rate for the KDEF dataset.

	FTP	HSP	Combined FTP and HSP
Proposed Ensemble Approach	95.41%	88.01%	91.50%
PCA+DCNN [6]	89.80%	79.85%	N.A.
K-PCANet [20]	80.20 %	N.A.	N.A.
K-LDANet [20]	78.13 %	N.A.	N.A.
QIBGSA+DCNN [7]	92.35%	84.95%	87.76%
Multi-class SVM [5]	81.94 %	78.45 %	N.A.
CNN Yu [16]	85.38 %	N.A.	80.86 %
CNN Tang [16]	83 %	N.A.	79.73 %
CNN Kahou [16]	83.68 %	N.A.	74.26 %
CNN Image-Net [16]	89.58 %	N.A.	73.86 %

5. Conclusions

In this paper, an ensemble approach of IQI-GSA and HDCR-NN has been proposed for the recognition of multi-pose facial expressions. The experiments have been performed on the images based JAFFE and KDEF datasets. The proposed ensemble approach has attained an average recognition rate of 94.05% with the JAFFE dataset and 91.64% with the KDEF dataset. The research evaluation indicates the outperformed outcomes of the proposed ensemble approach in comparison with existing techniques. It has been found that the recognition rate of facial expressions decreases with the movement of the face at some yaw angle.

In the future, the research work can be extended for the recognition of facial expressions with the availability of facial obstacles such as eye-glasses, face makeup, etc., Also, the proposed model can be utilized to detect human activities in real-time.

References

[1] Alexandre G., Soares J., and Thé G., "Systematic Review of 3D Facial Expression Recognition Methods," *Pattern Recognition*, vol. 100, no. 3, pp. 107108, 2020.

[2] Barman A. and Dutta P., "Facial Expression Recognition Using Distance and Shape Signature Features," *Pattern Recognition Letters*, vol. 145, pp. 254-261, 2021.

[3] Chakraborti T., Chatterjee A., Halder A., and Konar A., "Automated Emotion Recognition Employing a Novel Modified Binary Quantum-Behaved Gravitational Search Algorithm with Differential Mutation," *Expert Systems*, vol. 32, no. 4, pp. 522-530, 2015.

[4] Jain N., Kumar S., Kumar A., Shamsolmoali P., and Zareapoor M., "Hybrid Deep Neural Networks for Face Emotion Recognition," *Pattern Recognition Letters*, vol. 115, pp. 101-106, 2018.

[5] Jeni L., Hashimoto H., and Kubota T., "Robust Facial Expression Recognition Using Near Infrared Cameras," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 16, no. 2, pp. 341-348, 2012.

[6] Kumar Y., Verma S., and Sharma S., "Multi-Pose Facial Expression Recognition using Appearance-based Facial Features," *International Journal of Intelligent Information and Database Systems*, vol.13, no. 2-4, pp.172-190, 2020.

[7] Kumar Y., Verma S., and Sharma S., "Quantum-Inspired Binary Gravitational Search Algorithm to Recognize the Facial Expressions," *International Journal of Modern Physics C*, vol. 31, no. 10, pp. 2050138, 2020.

[8] Li S. and Deng W., "Deep Facial Expression Recognition: A Survey," *IEEE Transactions on Affective Computing*, 2020.

[9] Littlewort G., Bartlett M., Fasel I., Susskind J., and Movellan J., "Dynamics of Facial Expression Extracted Automatically from Video," *in Proceedings of Conference on Computer Vision and Pattern Recognition Workshop*, Washington, 2006.

[10] Liu W., Wang Z., Liu X., Zeng N., Liu Y., and Alsaadi F., "A Survey of Deep Neural Network Architectures and Their Applications," *Neurocomputing*, vol. 234, pp. 11-26, 2017.

[11] Lundqvist D., Flykt A., and Öhman A., "The Karolinska Directed Emotional Faces (KDEF)," *CD ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet*, vol. 91, pp. 630-639, 1998.

[12] Lyons M., Akamatsu S., Kamachi M., and Gyoba J., "Coding Facial Expressions with Gabor Wavelets," *in Proceedings of 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, pp. 200-205, 1998.

[13] Mehrabian A., *Communication without Words*, Routledge, 2008.

[14] Moghadam M., Nezamabadi-Pour H., and Farsangi M., "A Quantum Behaved Gravitational Search Algorithm," *Intelligent Information Management*, vol. 4, no. 6, pp. 390-395, 2012.

[15] Ojala T., Pietikäinen M., and Harwood D., "A Comparative Study of Texture Measures with Classification based on Featured Distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.

[16] Puthanidam R. and Moh T., "A Hybrid Approach for Facial Expression Recognition," *in Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*, New York, pp. 1-8, 2018.

[17] Rashedi E., Nezamabadi-Pour H., and Saryazdi S., "GSA: A Gravitational Search Algorithm," *Information Sciences*, vol. 179, no. 13, pp. 2232-2248, 2009.

[18] Revina I. and Emmanuel W., "A Survey on Human Face Expression Recognition Techniques," *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 6, pp. 619-628, 2021.

[19] Soleimanpour-Moghadam M. and Nezamabadi-Pour H., "An Improved Quantum Behaved Gravitational Search Algorithm," *in Proceedings of the 20th Iranian Conference on Electrical Engineering*, Tehran, pp. 711-715, 2012.

[20] Sun Z., Hu Z., Chiong R., Wang M., and He W., "Combining the Kernel Collaboration Representation and Deep Subspace Learning for Facial Expression Recognition," *Journal of Circuits, Systems and Computers*, vol. 27, no. 08, pp. 1850121, 2018.

[21] Viola P. and Jones M., "Rapid Object Detection using a Boosted Cascade of Simple Features," *in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, pp. 511-518, 2001.

[22] Wang H., Lei Z., Zhang X., Zhou B., and Peng J., "A Review of Deep Learning for Renewable Energy Forecasting," *Energy Conversion and Management*, vol. 198, pp. 111799, 2019.

[23] Zhou L., Shao X., and Mao Q., "A Survey of Micro-Expression Recognition," *Image and Vision Computing*, vol. 105, pp. 104043, 2021.

[24] Zhou W. and Jia J., "Training Convolutional Neural Network for Sketch Recognition on Large-Scale Dataset," *The International Arab Journal of Information Technology*, vol. 17, no. 1, pp. 82-89, 2020.

[25] Zouari R., Boubaker H., and Kherallah M., "RNN-LSTM Based Beta-Elliptic Model for Online Handwriting Script Identification," *The International Arab Journal of Information Technology*, vol. 15, no. 3A, pp. 532-539, 2018.

Yogesh Kumar is working as an Assistant Professor at GD Goenka University, Gurugram, Haryana, India. He is pursuing Ph.D in Computer Science and Engineering from UTU, Dehradun with a focus on Human Computer Interface. He is member of International Association of Engineers (IAENG).

Shashi Kant Verma is an Associate Professor at G B Pant Engineering College, Pauri-Garhwal, Uttarakhand. He received the Ph. D. from Uttarakhand Technical University, Dehradun U.K. India. He has vast experience as an academician for over 10 years. His area of interest is in Computer Architecture, Microprocessors, and Signal Processing.

Sandeep Sharma is a Senior Member of IEEE. He completed his Ph D and Masters in Electronics from University of Delhi and has over 20 years of experience in Research, Industry and Academia. He is currently working as a Head, Academics, at iNurture Education Solutions Pvt Ltd, India. He has published several National and International Journals including many SCI Journals.