# Facial Expression Recognition and Classification Using Optimized EfficientNet-B7

Jenu Prabha
Department of Electronics and Communication Engineering
Noorul Islam Centre for Higher Education, India
jenuj5j@gmail.com

Vinod Kumar
Department of Electronics and Communication Engineering
Noorul Islam Centre for Higher Education, India
rsvinodkumar69@yahoo.co.in

**Abstract:** *Humans use their faces to express their emotions and intentions in a simple and natural way. Face expressions are the essential components of nonverbal communication. In human-computer interaction and affective computing, facial expression recognition has various applications. There are several methods devised for recognition and classification of facial expression; still, the accurate recognition is the challenging task. Hence, in this research an automatic facial expression recognition and classification based on deep learning is introduced. Initially the input image is collected from facial expression recognition dataset. Then, the collected image is fed into pre-processing using Gaussian filtering which is used for noise reduction. Then the pre-processed images are given to feature extraction phase using Gray-Level Co-Occurrence Matrix (GLCM). GLCM is used to extract texture features for the facial expression recognition. Then the EfficientNet-B7 is utilized for the recognition and classification of facial expression due to the enhanced outcome with faster inference and smaller size. The proposed IC_EfficientNet method combines the gannet's ability to capture food with the coot bird's ability to forage. The optimization technique achieves higher convergence rates due to this hybridization, which improves the EfficientNet-B7 model's parameter tuning. In comparison to existing techniques, the hybrid Improved Coot (IC) algorithm balances exploration and exploitation, leading to a quicker and more effective optimization process. The proposed IC_EfficientNet provides better results compared to existing methods such as Deep Neural Network (DNN), MultiLayer Perceptron (MLP) neural network, Facial Detection using a Convolutional Neural Network (FD-CNN) and Convolutional Neural Network (CNN). Thus, the proposed IC_EfficientNet provides the better outcome in terms of Accuracy, Specificity, Precision, Recall, F1-Measure, and MSE acquired the better outcome of 99.13, 98.80, 97.80, 99.13, 98.44, and 0.87 respectively.*

**Keywords:** *Facial expression, recognition, optimization, efficientnet, feature extraction, deep learning.*

*Received September 24, 2024; accepted March 12, 2025*
*https://doi.org/10.34028/iajit/22/3/5*

## 1. Introduction

Facial Expression Recognition (FER) has generated significant interest in the field of computers, encompassing not only Computer Vision (CV) and Human-Computer Interaction (HCI) [7]. Numerous researchers have been investigating the area for more than 20 years due to technological advancements and the goal of achieving machine-human communication [11]. FER is the process of identifying human affective states from facial muscle movements that result from automatic reactions brought on by shifts in emotional states [26]. From a psychological perspective, there are six fundamental emotional states that humans can experience: sadness, happiness, fear, surprise, rage, and disgust [12, 29].

For effective communication with others, face expressions are crucial criteria. Most of the time, verbal and nonverbal communication occurs. The expressions on one's face are a form of nonverbal communication [13, 31]. Facial expressions provide subtle cues for more extensive communication. Paralanguage, body language, facial expressions, gestures, and Eye contact are all examples of non-verbal communication between people and animals [8, 25]. Making eye contact during a conversation is essential as it facilitates the exchange of thoughts and emotions. Making eye contact with others controls participation, facilitates conversation, and forges connections. Face expressions include a fear, surprise, disgust, grin, sadness, and rage [3, 15]. Humans need emotions because they affect how they view and comprehend the world. Emotions are a fundamental aspect of human existence. Numerous techniques have been developed over the past three decades to make it easier to analyse emotions, ranging from manual techniques such using questionnaires created by psychologists to techniques utilising computers [4, 6]. Currently, there are numerous uses for emotion identification by computers. One such application is the development of smart offices, and smart homes which uses physiological signals to recognise emotions [32]. Additionally, facial detection methodology is widely utilised today in security-related applications, educational services, and consumer services [21].

The intricacy of human emotions presents a number of difficulties for FER. Accurate recognition can be challenging because to variations in facial characteristics caused by age, gender, ethnicity, and individual differences [2, 24]. Intensity of expression also varies, and subtle expressions are frequently more difficult to pick up on. The distortion of facial characteristics caused by environmental elements such as background, lighting, and occlusions make recognition more difficult [1]. Immediate emotional processing is another requirement for FER systems, necessitating quick and effective computing. Expressions signify various things in different cultures, which adds another level of intricacy [19]. Furthermore, emotions are fluid and subject to sudden changes, which makes it challenging to record and decipher live expressions [27, 33]. To overcome these challenges many authors have conducted extensive work as described in section 2.

The organization of the research are: Section 2 details the literature review along with the problem statement. Section 3 elaborates the proposed facial expression recognition methodology and its outcome is detailed in section 4. Finally, section 5 concludes the work.

## 2. Literature Review

Some of the literatures based on the facial expression recognition and classification are detailed in this section. The expression recognition using the ensemble strategy was designed by Moung *et al*. [14] by considering the InceptionV3, ResNet-50 and CNN. The image resizing was performed initially to make the images fit for the channel. Followed by, the normalization based on zero-mean was performed for the reduction of computation burden. The normalized data is fed into the ensemble classifier for detecting the expression of the faces. Here, the ensemble approach acquired the better outcome for the expressions like neutral, surprise and happy; still it degrades its performance while analysing the sad, anger and disgust expressions.

MobileNet based facial expression recognition was modelled by Nan *et al*. [16] for processing the real time processing with lightweight computation. The devised model incorporates the attention module for providing more attention towards the local features. Here, the attention based on both the temporal and spatial attributes were accomplished through the convolutional block based attention approach. The enhanced accuracy was accomplished by the model and hence the devised model was applicable for the mood classification task. Besides, the over-fitting issues of the method were solved by the incorporation of the dropout layer. Still, the misclassification occurs for the non-familiar expressions in daily life due to the shortage significant attributes.

An ensemble with InceptionV1, ResNet-50 and EfficientNet-B0 was designed by Yu *et al*. [30] for the recognition of the facial expression. The devised model resized the image and then normalized for obtaining the image appropriate to fed into the recognition module. Followed by, the data augmentation was performed for enriching the classifier to elevate the recognition accuracy. Here, the focal loss function was considered for the loss minimization in data learning to enhance the generalization capability. The better outcome was accomplished through the F-measure to depict the effectiveness of the designed approach. However, the computational and time complexity of the model was higher.

The multi-head attention module based RegNet was devised by Phan *et al*. [20], wherein the multi-head attention module of the transformer model was replaced by the RegNet for enhancing the classification accuracy. Here, the most relevant feature extraction through the multi head attention enhances the classification accuracy. Still, the failure in incorporating the additional pre-processing stages leads to misclassification due to the poor generalization capability.

A deep learning based approach was designed by Keshri *et al*. [9], wherein the face object was detected initially from the input image for the detection of expression of the face. In this, the automatic feature extraction and classification was performed using the CNN model and has the capability of processing high dimensional data. Hence, the method has been applicable for clinical and industrial applications for detecting the expression of the face. Still, the misclassification occurs due to the artefacts present in the image.

Another deep learning model based on CNN was designed by Saeed *et al*. [22] to identify the expressions of the face and accomplished enhanced recognition accuracy. The devised model enriches the data through the data augmentation strategy that elevates the generalization capability of the deep learning model that helps to enhance the accuracy of recognition. Still, the misclassification rate of the model was higher. A hybrid deep learning model designed by Talaat *et al*. [28] utilized CNN and autoencoder for detecting the facial expression recognition for processing the real world applications. In this, the autoencoder was utilized for capturing the attributes and then, the CNN was utilized for recognizing the facial expression based on the attributes. The superior experimentation outcome illustrates the applicability of real world processing; still, the computational complexity affects the performance. A machine learning based model was designed by Saeed [23] utilized random forest for identifying the facial expression, wherein the HoG based features was considered for classifying the classes. The inefficiency in handling the larger dataset was considered as the challenging aspect of the model. The short description of the literature review is depicted in Table 1.

Table 1. Short description of literature review.

| Reference | Techniques used | Benefits | Challenges |
|---|---|---|---|
| [14] | Ensemble approach | The misclassification rate of the positive classes of emotions is higher. | While analysing the negative classes, the method accomplished degraded performance. |
| [16] | MobileNet model | Computation burden of the model was minimal due to the lightweight design. | Failure in considering the significant attributes limits the outcome. |
| [30] | Multi-modal ensemble approach | The enhanced accuracy was accomplished by the model due to the consideration of multi-model features. | High computational complexity. |
| [20] | Hybrid tranformer-RegNet model | Accomplished better test and validation result. | The method accomplished degraded performance compared to the traditional transformer based model. |
| [9] | Deep learning method | Recognized expressions more accurately qualitatively and hence can applicable for identifying the feelings of the audience. | Failed to analyse the performance of the model. |
| [22] | Deep learning model | Accomplished minimal error rate. | Higher misclassification rate. |
| [28] | Hybrid deep learning | Accomplished better outcome that assist real world processing | High computational complexity. |
| [23] | Machine learning | The experimental results indicate that the designed method accurately and effectively identifies facial expressions. | The classification accuracy is lower for real world processing |

## 2.1. Problem Statement and Motivation

Numerous developments in facial expression identification and classification are highlighted in the literature currently in publication, however there are also significant research gaps. Although deep learning and ensemble models such as InceptionV3, ResNet-50, and MobileNet exhibit encouraging outcomes, many approaches suffer from excessive computational complexity, incorrect negative emotion classification, and poor real-world dataset management. Furthermore, problems like overfitting and the inability to identify unfamiliar expressions persist even as attention mechanisms increase accuracy. Though they perform better, hybrid models (such as CNN with autoencoders or multi-head attention with RegNet) have substantial computational overhead and standardisation problems. Furthermore, techniques such as CNN-based models and Random Forest show limited accuracy and scalability in practical settings. To overcome these limitations, IC_EfficientNet is proposed. The proposed approach improves facial expression recognition by increasing accuracy and reducing information loss by combining the IC algorithm with EfficientNet-B7. Gaussian filtering and GLCM are two efficient feature extraction algorithms that lower processing complexity. By striking a balance between exploration and exploitation, the algorithm's hybrid design ensures quicker convergence and resilience to frequent problems like misclassification.

The novelty of this work lies in the introduction of the IC algorithm, which combines the gannet's food-gathering behaviour with the coot bird's foraging behaviour to improve the rate at which parameters are optimised for face expression identification. Through the use of the advanced deep learning architecture EfficientNet-B7, the study increases recognition rates while guaranteeing quicker inference and a smaller model size. Furthermore, the effective capture of important attributes is made possible by the use of sophisticated feature extraction techniques like GLCM and SIFT, which lower computing complexity. By properly tackling typical issues in facial expression recognition, like high misclassification rates, the suggested technique greatly advances the field by achieving amazing performance metrics when compared to existing methods like DNN, MLP, FD_CNN, and CNN.

The major contributions of the research are:

- **Design of IC Algorithm**: the proposed Improved Coot (IC) algorithm is designed by integrating the foraging behaviour of the coot and the capturability behaviour of the Gannet to obtain the global best solution with enhanced rate of convergence.
- **Design of IC_EfficientNet for Facial Expression Recognition**: the EfficientNet-B7 is utilized for the recognition of the facial expression, wherein the adjustable parameters are modified using the proposed IC algorithm for enhancing the recognition rate.

# 3. Proposed Facial Expression Recognition and Classification

The proposed facial expression recognition is depicted in Figure 1. For this, the input data is taken from the publically available dataset. Followed by, the artefacts from the image is removed using the Gaussian filter. From the filtered outcome, the significant attributes are extracted using the Gray-Level Co-Occurrence Matrix (GLCM) and Scale-Invariant Feature Transform (SIFT) features. The extracted features are fed into the proposed Improved Coot based EfficientNet-B7 (IC_EfficientNet). In this, the EfficientNet B7 is utilized for recognizing the facial expression and classification, wherein its adjustable parameters are modified using the proposed IC algorithm. The proposed IC algorithm is designed by integrating the foraging behaviour of the water bird Coot and the food capturing behaviour of the Gannet to obtain the global best solution. Thus, the optimally tuned EfficientNet enhances the recognition accuracy through the less information loss.

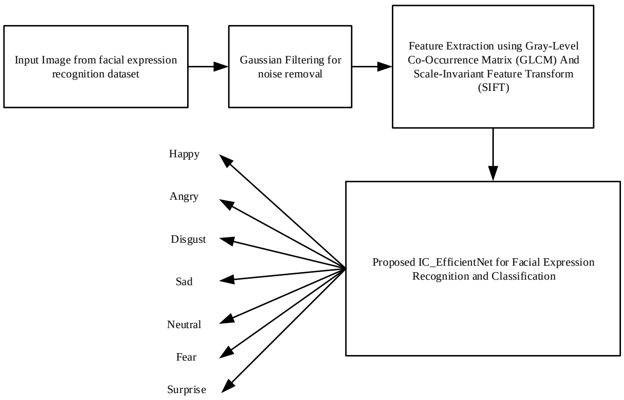



Figure 1. Work flow of proposed facial expression recognition and classification.

## 3.1. Data Acquisition

Let the input image utilized for the processing of the proposed IC_EfficientNet for the recognition of the facial expression is taken from the dataset. Here, $E$ refers to the dataset with $t$ images totally and is represented in Equation (1),

$$E = \{E_1, E_2, \ldots E_i, \ldots E_t\} \tag{1}$$

where, $t$ refers to the total image count in dataset and the image utilized for processing the proposed method is indicated as $E_i$.

## 3.2. De-Noising using Gaussian Filter

The de-noising of the input image by smoothing the input image is devised using the Gaussian filtering approach. Here, the Gaussian distribution is followed for the noise reduction using the smoothing function $\sigma$. Let us consider $E_i(k,l)$ as the input image and the outcome based on the Gaussian Filtering is written in Equation (2):

$$F_G(k,l) = S(k,l) * \frac{e^{\frac{-(k^2+l^2)}{2\sigma^2}}}{\sum_k \sum_l e^{\frac{-(k^2+l^2)}{2\sigma^2}}} \tag{2}$$

where, the smoothening process is notated as $S(k, l)$ and the outcome of the Gaussian Filter is notated as $F_G(k, l)$. By maintaining the image's boundary details, the Gaussian filtering method tries to improve the image's quality through the smoothening process.

## 3.3. Feature Extraction

The most informative features for the facial expression recognition is extracted using the GLCM and SIFT based features.

### 3.3.1. GLCM

The Gray-Level Co-Occurrence Matrix (GLCM) texture features are extracted using six various measures like correlation, energy, angular second moment, dissimilarity, homogeneity and contrast from the filtered image. The texture details of the image is extracted through the GLCM as it offers the better outcome with minimal complexity and processing time.

- **Correlation**: the correlation is measured for the gray scale image based on the linear dependencies and is outlined in Equation (3):

$$F_1 = \sum_{a=0}^{R-1} \sum_{b=0}^{R-1} \frac{(x - Avg_x)(y - Avg_y)}{\sqrt{(Var_a)(Var_b)}} \tag{3}$$

where, the correlation factor is notated as $F1$, the variance is defined as $Var$, the mean is defined as $Avg$, *x and y* are the gray level values, *a and b* are the kernels utilized for extracting the features.

- **Energy**: the measure of texture uniformity by taking the square root of angular second moment. The better energy is measured for the input, when its windows are arranged orderly and is expressed in Equation (4):

$$F2 = \sqrt{\sum_{a=0}^{R-1} \sum_{b=0}^{R-1} X(a,b)^2} \tag{4}$$

- **Angular Second Moment**: the textual uniformity is measured through the angular second moment for identifying the variations in pixel values of image. It is measured in Equation (5):

$$F3 = \sum_{a=0}^{R-1} \sum_{b=0}^{R-1} X(a,b)^2 \tag{5}$$

- **Dissimilarity**: dissimilarity is utilized for analysing the local variations of the image using the linear measure and is outlined in Equation (6):

$$F4 = X(a,b)z|a-b| \tag{6}$$

- **Homogeneity**: the homogeneity is utilized to calculate the inverse difference moment, and a greater value indicates that the image's components are more similar. It is outlined in Equation (7):

$$F4 = \sum_{a=0}^{R-1} \sum_{b=0}^{R-1} \frac{X(a,b)}{1+(a-b)^2} \tag{7}$$

- **Contrast**: by measuring the contrast, the local variations concerning the image are evaluated and is outlined in Equation (8):

$$F5 = \sum_{a=0}^{R-1} \sum_{b=0}^{R-1} (a-b)^2 \tag{8}$$

where, the correlation, energy, angular second moment, dissimilarity, homogeneity and contrast are mentioned as $F1$, $F2$, $F3$, $F4$, $F5$, and $F6$. Here, the normalized value is notated as $X(a, b)$, the sum of the kernels *a and b* is equal to 1, and the gray level number is defined as $R$.

#### 3.3.2. SIFT based Feature Extraction

A method of feature vector extraction approach called SIFT is used for extracting the significant features. Such feature points can withstand geometric transformations as well as normal signal processing procedures. Hence, the SIFT based feature extraction is utilized for facial expression recognition. Following is a description of SIFT based feature point extraction procedure.

- **Scale-Space Extrema Detection**: the key-points are identified using the difference of Gaussian function, wherein the orientation and scale are considered as invariant. For this, the different scale values concerning the Gaussian function is convolved with the scale space. It is expressed in Equation (9):

$$A(i, j, Var) = G(i, j, Var) * P(i, j) \tag{9}$$

where, the ordinate and the abscissa of the image is defined as $j$ and $i$ respectively and * refers to the convolution operator. The different scales are indicated as $Var=\{Var_1, Var_2,\ldots, Var_n\}$ and the input image is indicated as $P(i, j)$.

- **Localization and Orientation Assignment of Key-Point**: the position and size concerning the key-points are identified in this stage. During the detection process, the poor contrast, unstable critical feature points are removed. Each key-point needs to have an orientation assigned to it, which requires the estimation of gradient orientations among the neighbour key-points. Afterward, a gradient orientation histogram is created from the number of each orientation, which is then counted. The key-point's orientation is represented by its orientation at its peak of the orientation histogram.

- **Generation of Descriptor**: the orientation, scale, and position of each key-point are established after the detection of key-point. As a result of this geometric data, key-point descriptions are formed.

#### 3.3.3. Feature Concatenation

The features extracted using the GLCM and SIFT are combined together to form the feature vector. It is indicated in Equation (10),

$$F = \{F1, F2, F3, F4, F5, F6, F_{SIFT}\} \tag{10}$$

Here, the concatenated feature is indicated as $F$.

#### 3.3.4. Architecture of EfficientNet-B7

By learning the relevant data, neural networks are frequently used to accomplish numerous tasks, such as classification, recognition, and detection, with greater accuracy. The neural network named EfficientNet has eight different iterations and uses the concept of compound scaling. By balancing the necessary features in this case, compound scaling plays the role of improving recognition and classification accuracy. The main distinction between the various EfficientNet variants' architectures is their variation in feature maps. Accurate recognition is typically improved by a deeper neural network. EfficientNet-B7 is therefore used in the proposed facial expression recognition and classification method for obtaining the greater recognition accuracy. Figure 2 illustrates EfficientNet-B7, which consists of seven blocks of channels, strides, and filters.

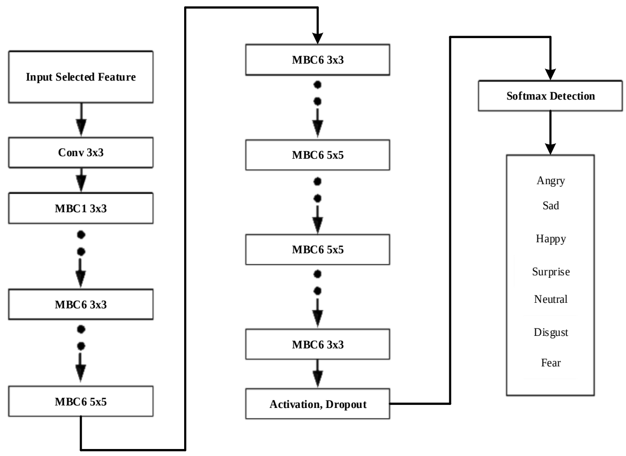



Figure 2. EfficientNet-B7 for intrusion detection.

EfficientNet-B7 depth, width and resolution parameters like, $Dep$, $Wid$ and $Res$ respectively for compound scaling and it is given in Equations (11), (12), and (13).

$$Dep = \alpha^{v} \tag{11}$$

$$Wid = \beta^{v} \tag{12}$$

$$Res = \gamma^{v} \tag{13}$$

where, $\alpha\beta^2\gamma^2 \approx 2$ and $\alpha \geq 1$, $\beta \geq 1$, $\gamma \geq 1$. Here, $\alpha$, $\beta$ *and* $\gamma$ refers to the scaling coefficients and $v$ denotes the compound coefficient. The depth, width, and resolution of the EfficientNet-B7 are determined by the number of layers, channels, and related parameters. The EfficientNet-B7, which includes squeeze and excitation optimisation that is built using the Mobile inverted Bottleneck Convolution (MBC) as its building blocks. In Figure 3, the MBC with squeeze and excitation optimisation is shown.

The expansion of the MBC point-wise convolution takes place, which elevates the number of channels; thus the expansion takes place. Followed by, the depth-wise convolution takes place that process the features with many channels. The global pooling is performed at the squeezing process and then the reduction of the number of channels is employed using the point-wise convolution. Then, the feature size is regenerated into the original format through the excitation process. The computation overhead of the EfficientNet-B7 is minimized through the MBC using the squeeze and excitation network. The fine-tuning of the weights of the EfficientNet-B7 using the proposed IC approach enhances the accuracy of recognition of facial expression.

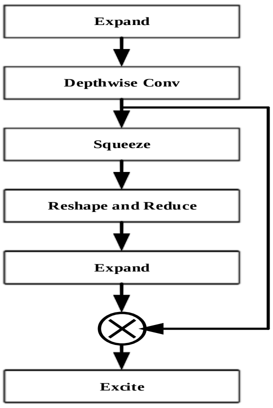



Figure 3. Architecture of MBC.

### 3.3.5. Improved Coot Algorithm

A medium sized bird that lives in the water named Coot [17] associated to the Rallidae family. The foraging behaviour of the bird is considered in the optimization for solving the optimization issues. The birds lives in a swarm and utilizes various motions for identifying and capturing the prey, the prey may be a small fish. The bird's moves in a chain manner for obtaining the prey, wherein the leader bird leads the group and the remaining members follows the leader. The various motions utilized by the birds for getting the prey are:

- **Arbitrary Search**: the bird's moves arbitrarily in the water surface for identifying the prey that corresponds to explore more area in the search space.
- **Motion in chain manner**: in this chain based movement, the birds move towards the prey one after another to form the chain to capture the prey.
- **Position adjustment by members**: the members of the group transform their motion based on the movement of the leader to move towards the prey.
- **Position adjustment by Leader**: the leader of the swarm transforms the motion concerning the prey to capture the prey.

The mobility of the bird to capture the prey is utilized for tuning the adjustable parameters of the EfficientNet-B7. Here, the capturing the prey is considered as the solution; still, there is a chance for escaping the prey from the coot. Thus, the capturability of the Gannet [18] is hybridized with the coot for minimizing the escaping capability of the prey. The Gannet is a bird that lives in the sea shore and lakes that utilizes high capturability with powerful eye sight for capturing the prey that minimizes the chance of escaping. Hence, the convergence of the solution is accomplished faster using the hybrid IC algorithm. Also, the balanced diversification and intensification capability limits the local optimal trapping and provides the global best solution. In the proposed IC algorithm, the capturing the prey is the solution utilized for tuning the parameters of the deep learning model.

- **Mathematical Modelling**

The mathematical modelling of the proposed IC algorithm comprises of:

- Initialization.
- Computation of Fitness.
- Exploration.
- Exploitation.
- Fitness Re-computation.
- Stopping Criteria.

- **Initialization**: the setting of birds in the search area arbitrarily is devised in the initialization phase, which is mentioned as $\vec{a} = \left\{ \vec{a}_1, \vec{a}_2, ......\vec{a}_x \right\}$ and the number of maximal iteration of the algorithm also assigned initially, which is indicated as $Max_t$. Here, the expression for the arbitrary initialization is written in Equation (14):

$$F(t) = p(1,u) * (K - L) + L \tag{14}$$

where, the bounds of the search area is defined as *K* and *L* that corresponds to its upper and lower limits. The solution estimated by the IC algorithm has *u* dimensions and *p* refers to the random number. *F*(*t*) defines the position of the bird in the search area. Here, the bounds are defined in Equation (15):

$$L = [L_1, L_2, .....L_u], \quad K = [K_1, K_2, ....K_u] \tag{15}$$

- **Computation of Fitness**: each bird in the search area is analysed by computing the fitness to identify the feasibility using the mean square error, which defines the closeness of the solution obtained by the birds in the search space compared to the target solution. Thus, the minimal fitness value is considered in the proposed IC algorithm and is written in Equation (16):

$$Fit = \frac{1}{Q} \sum_{i=1}^{Q} (A_{so\ln} - P_{so\ln})^2 \tag{16}$$

where, the fitness is defined as *Fit*, the target solution is mentioned as $P_{soln}$, the estimated solution is mentioned as $A_{soln}$, and the total number of birds population in the search space is defined as *Q*.

- **Exploration**: in the search space, the birds moves arbitrarily in all the directions to find the global best solution. The searching of prey in all the directions limits the algorithm from trapping at local solutions. The search area utilized by the birds at this stage is written in Equation (17):

$$Q = p(1,u) * (K - L) + L \tag{17}$$

After exploring the search area, the solution updated by the bird is evaluated in Equation (18):

$$F(t+1)_{Coot} = F(t) + G \times B2 \times (Q - F(t)) \tag{18}$$

where, $B2$ refers the random number with range [0,1], and motion of birds arbitrarily in all the direction of search area is defined as $G$ and is evaluated in Equation (19):

$$G = 1 - D \times \left(\frac{1}{Max_t}\right) \tag{19}$$

where, iteration indicates the current processing stage is defined as $D$.

Here, the prey's escaping capability from the birds eye sight is limited by incorporating the capturability of the Gannet, which is expressed in Equation (20):

$$F(t+1)_{Gannet} = t * \beta * (F(t) - F_{best}(t)) + F(t) \tag{20}$$

The factor $\beta$ depends on the capturability of the Gannet and is evaluated in Equation (21),

$$\beta = H * |F(t) - F_{best}(t)| \tag{21}$$

where, the best Gannet in the current iteration is indicated as $F_{best}(t)$, the capturability is indicated as $H$. Then, the expression for the capturability of the Gannet is expressed in Equations (22), (23), (24), and (25),

$$H = \frac{1}{J * t_2} \tag{22}$$

where,

$$t_2 = 1 + \frac{t_1}{Max_t} \tag{23}$$

$$J = \frac{Y * e^2}{N} \tag{24}$$

$$N = 0.2 + (2 - 0.2) * B_5 \tag{25}$$

Here, the random number is indicated as $B_5$ that has the range of [0, 1]. The velocity and mass of the Gannet is indicated as $Y$ and $e$, which has the value of 1.5m/s and 2.5 kg. In this, the sudden turning capability of Gannet reduces the escaping capability of the prey and catches it to make the convergence faster. Thus, the solution acquired by hybridizing the capturability of the Gannet and foraging behaviour of the Coot is expressed in Equations (26) and (27),

$$F(t+1) = 0.5[F(t+1)_{Gannet}] + 0.5[F(t+1)_{Coot}] \tag{26}$$

$$\begin{aligned} F(t+1) = {} & 0.5\left(t * \beta * \left(F(t) - F_{best}(t)\right) + F(t)\right) \\ & + 0.5\left(F(t) + G \times B2 \times (Q - F(t))\right) \end{aligned} \tag{27}$$

Using the hybridized behaviour the proposed IC algorithm explores more area and capture the global best solution with fast convergence rate.

- **Motion in chain manner**: the birds will form the chain movement while they move towards the solution, wherein the distance between the birds is measured through the averaging the solutions accomplished at present and past iteration. It is defined in Equation (28):

$$F(t) = \frac{1}{2}[F(t-1) + F(t)] \tag{28}$$

- **Position Adjustment by Members**: the birds in the swarm follows the leader to get the food, wherein the best leader is identified using the index estimation based on Equation (29):

$$E = 1 + (gMODQ) \tag{29}$$

where, $E$ is the index and the bird with index $E$ is notated as $g$. After estimating the index, birds follow the best leader and the solution updation takes place and it is given in in Equation (30):

$$F(t) = W(n) + 2 \times B1 \times \cos(2B\pi) \times W(n) - F(t) \tag{30}$$

where, $W(n)$ notates the leader, $B$ notates random number [-1, 1], and [0, 1] is the range of $B1$, random number.

- **Exploitation**: the leaders in the swarm adjust their location to capture the prey and hence each members of the swarm also adjusts their motions based on the leader. Here, the position of the leader updation is written in Equation (31):

$$W(t) = \begin{cases} M \times B3 \times \cos(2B\pi) \times F_{best} - W(t) + F_{best} & B4 < 0.5 \\ M \times B3 \times \cos(2B\pi) \times F_{best} - W(t) - F_{best} & B4 \geq 0.5 \end{cases} \tag{31}$$

where, the random numbers $B3$ and $B4$ has the range of [0, 1], and the position of the best candidate solution is indicated as $F_{bes}$. The term $M$ is defined in Equation (32),

$$M = 2 - t \times \left(\frac{1}{Max_t}\right) \tag{32}$$

This stage involves finding a solution to the optimization problems and determining whether it is apt by assessing its efficacy by considering the fitness.

- **Identifying the Feasibility**: after revising the position, the feasibility is calculated through Equation (16).
- **Termination**: the attainment of $t_{max}$ or the global best solution stops the termination of IC algorithm. The pseudo-code for IC is provided in Algorithm (1).

*Algorithm 1: Pseudo-Code for IC Algorithm.*

*Step 1: Initialization of search agents*
*Step 2: Choose a leader randomly*
*Step 3: Evaluate the fitness*
*Step 4: $F_{best}$ is measured*
*Step 5: While ($t<Max_t$)*
*Step 6: {*
*Step 7: Estimate G and M*
*Step 8: Based on equation (29) evaluate index leader*
*Step 9: If (P>0.5)*
*Step 10: Based on equation (30) update indexed leader*
*Step 11: Else*
*Step 12: If (p<0.5, t~=1)*
*Step 13: Based on equation (28) updatelocation of chain movement*
*Step 14: Else*

*Step 15: Based on equation (27) updatelocation of exploration phase*
*Step 16: End*
*Step 17: End*
*Step 18: Based on equation (31) update location of exploitation phase*
*Step 19: }*
*Step 20: End*
*Step 21: t=t+1*
*Step 22: stop*

Thus, the best solution accomplished by the IC algorithm is utilized for solving the optimization issues.

## 4. Result and Discussion

The implementation of the proposed IC_EfficeintNet for the facial expression recognition and classification is performed in Python programming language with Windows 10OS PC with 8GBRAM.

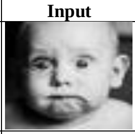
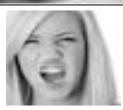
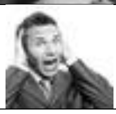
### 4.1. Dataset Description

The facial expression recognition dataset from Kaggle is utilized for the processing of the proposed methodology. The dataset comprises of seven various expressions with training and validation files.

### 4.2. Experimental Analysis

The experimental outcome of the IC EfficientNet is depicted in Table 2.

Table 2. Experimental outcome.

| Expressions | Input | Pre-processing | Output |
|---|---|---|---|
| **Surprise** |  |  |  |
| **Angry** |  |  |  |
| **Sad** |  |  |  |
| **Neutral** |  |  |  |
| **Disgust** |  |  |  |
| **Happy** |  |  |  |
| **Fear** |  |  |  |

### 4.3. Comparative Assessment

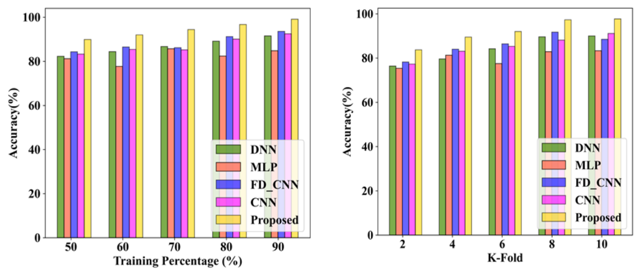
The IC_EfficientNet is compared with conventional facial expression recognition and classification methods like DNN [10], MLP [5], FD_CNN [22], and CNN [9]. The outcomes of the methods are evaluated using various measures like accuracy, specificity, precision, recall, F1-Measure, and MSE by varying the training percentage and K-Fold value.

#### 4.3.1. Accuracy Based Comparison

The correct facial expression classifications employed by the methods are measured through the accuracy measure and it is given in Equation (33):

$$FRC_{Acc} = \frac{FRC_{tp} + FRC_{tn}}{FRC_{tp} + FRC_{tn} + FRC_{fp} + FRC_{fn}} \tag{33}$$

Where, the true positive is indicated as $FRC_{tp}$, true negative is indicated as $FRC_{tn}$, false positive is indicated as $FRC_{fp}$, false negative is indicated as $FRC_{fn}$ and the accuracy is indicated as $FRC_{Acc}$. The accuracy measure of facial expression recognition and classification methods concerning the training percentage and K-Fold is depicted in Figure 4.




a) Training percentage. b) K-Fold.

Figure 4. Accuracy measure in terms of training percentage and K-Fold.

Here, the accuracy measured by the IC_EfficientNet is 93.60, which is 9.91%, 7.62%, 7.87%, and 2.61% improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods with 70% of learning data. Likewise, with K-Fold value of 6, the accuracy measured by the IC_EfficientNet is 88.53, which is 11.66%, 5.08%, 2.46%, and 7.72% improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods. The elaborated assessment of accuracy is presented in Table 3.

Table 3. Accuracy measure based comparison.

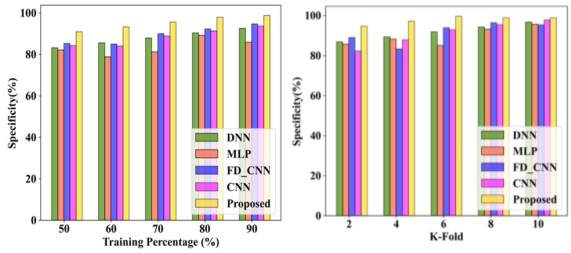
| Methods | DNN | MLP | FD_CNN | CNN | Proposed |
|---|---|---|---|---|---|
| **Training Percentage** | | | | | |
| 50 | 82.27 | 84.41 | 86.75 | 89.10 | 91.49 |
| 60 | 81.24 | 77.74 | 85.72 | 82.42 | 84.81 |
| 70 | 84.33 | 86.47 | 86.24 | 91.16 | 93.60 |
| 80 | 83.30 | 85.44 | 85.21 | 90.13 | 92.52 |
| 90 | 89.91 | 92.05 | 94.38 | 96.74 | 99.13 |
| **K-fold** | | | | | |
| 2 | 76.35 | 79.62 | 84.24 | 89.57 | 90.00 |
| 4 | 75.42 | 81.26 | 77.54 | 82.86 | 83.31 |
| 6 | 78.21 | 84.03 | 86.35 | 81.69 | 88.53 |
| 8 | 77.28 | 83.11 | 85.30 | 88.06 | 91.09 |
| 10 | 83.69 | 89.50 | 91.96 | 97.30 | 97.67 |

#### 4.3.2. Specificity Based Comparison

The correct recognition of the face by the facial expression methods is measured through the specificity measure and it is given in Equation (34):

$$FRC_{Spc} = \frac{FRC_{tn}}{FRC_{tn} + FRC_{fp}} \tag{34}$$

Where, the specificity is indicated as $FRC_{Spc}$. The specificity measure of facial expression recognition and classification methods concerning the training percentage and K-Fold is depicted in Figure 5.




a) Training percentage. b) K-Fold.

Figure 5. Specificity measure in terms of training percentage and K-Fold.

Here, the specificity measured by the IC_EfficientNet is 93.68, which is 10.09%, 10.34%, 5.07%, and 2.54% improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods with 80% of learning data. Likewise, with K-Fold value of 8, the specificity measured by the IC_EfficientNet is 97.80, which is 15.90%, 10.21%, 5.06%, and 2.47% improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods. The elaborated assessment of specificity is presented in Table 4.

Table 4. Specificity measure based comparison.

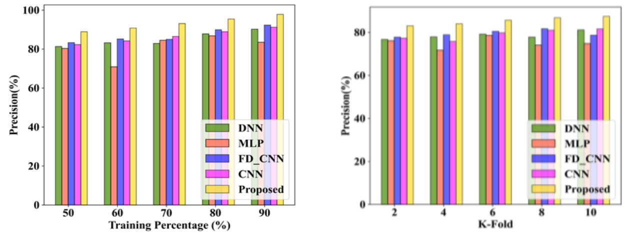
| Methods | DNN | MLP | FD_CNN | CNN | Proposed |
|---|---|---|---|---|---|
| Training Percentage | | | | | |
| 50 | 83.18 | 85.52 | 87.88 | 90.26 | 92.63 |
| 60 | 82.14 | 78.83 | 81.20 | 89.21 | 85.95 |
| 70 | 85.27 | 85.04 | 90.02 | 92.34 | 94.72 |
| 80 | 84.22 | 84.00 | 88.93 | 91.30 | 93.68 |
| 90 | 90.86 | 93.19 | 95.56 | 97.93 | 98.80 |
| K-fold | | | | | |
| 2 | 86.81 | 89.28 | 91.76 | 94.23 | 96.71 |
| 4 | 85.72 | 88.19 | 85.02 | 93.14 | 95.62 |
| 6 | 88.99 | 83.25 | 93.94 | 96.41 | 95.34 |
| 8 | 82.25 | 87.81 | 92.85 | 95.38 | 97.80 |
| 10 | 94.63 | 97.10 | 99.58 | 98.80 | 98.80 |

### 4.3.3. Precision Based Comparison

The correct recognition of the face by the facial expression methods is measured through the precision measure from the positive classes and it is given in Equation (35):

$$FRC_{\mathrm{Pre}} = \frac{FRC_{tp}}{FRC_{tp} + FRC_{fp}} \tag{35}$$

where, the precision is indicated as $FRC_{\mathrm{Pre}}$. The Precision measure of facial expression recognition and classification methods concerning the training percentage and K-Fold is depicted in Figure 6.




a) Training percentage. b) K-Fold.

Figure 6. Precision measure in terms of training percentage and K-Fold.

Here, the precision measured by the IC_EfficientNet is 91.22, which is 9.79%, 7.66%, 5.15%, and 2.59% improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods with 80% of learning data. Likewise, with K-Fold value of 8, the precision measured by the IC_EfficientNet is 81.63, which is 5.35%, 7.12%, 2.20%, and 0.75%improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods. The elaborated assessment of Precision is presented in Table 5.

Table 5. Precision measure based comparison.

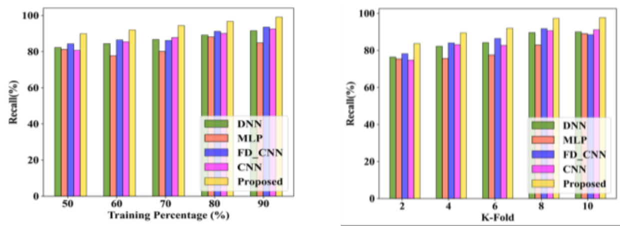
| Methods | DNN | MLP | FD_CNN | CNN | Proposed |
|---|---|---|---|---|---|
| Training Percentage | | | | | |
| 50 | 81.27 | 83.22 | 82.95 | 87.85 | 90.21 |
| 60 | 80.25 | 70.91 | 84.50 | 86.83 | 83.55 |
| 70 | 83.31 | 85.25 | 84.97 | 89.87 | 92.29 |
| 80 | 82.29 | 84.23 | 86.53 | 88.86 | 91.22 |
| 90 | 88.87 | 90.81 | 93.10 | 95.44 | 97.80 |
| K-fold | | | | | |
| 2 | 76.69 | 77.87 | 79.21 | 77.82 | 81.10 |
| 4 | 76.12 | 71.70 | 78.58 | 74.11 | 74.81 |
| 6 | 77.84 | 78.90 | 80.46 | 81.65 | 78.67 |
| 8 | 77.27 | 75.82 | 79.83 | 81.02 | 81.63 |
| 10 | 82.96 | 83.97 | 85.64 | 86.83 | 87.36 |

### 4.3.4. Recall Based Comparison

The correct recognition of the face by the facial expression methods is measured through the precision measure from the negative classes and it is given in Equation (36):

$$FRC_{\mathrm{Recall}} = \frac{FRC_{tp}}{FRC_{tp} + FRC_{fn}} \tag{36}$$

where, the precision is indicated as $FRC_{\mathrm{Recall}}$. The recall measure of facial expression recognition and classification methods concerning the training percentage and K-Fold is depicted in Figure 7.




a) Training percentage. b) K-Fold.

Figure 7. Recall measure in terms of training percentage and K-Fold.

Here, the recall measured by the IC_EfficientNet is 84.81, which is 4.22%, 8.34%, 5.59%, and 7.95% improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods with 60% of learning data. Likewise, with K-Fold value of 4, the recall measured by the IC_EfficientNet is 88.96, which is 15.22%, 15.00%, 12.84%, and 6.86% improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods. The elaborated assessment of recall is presented in Table 6.

Table 6. Recall measure based comparison.

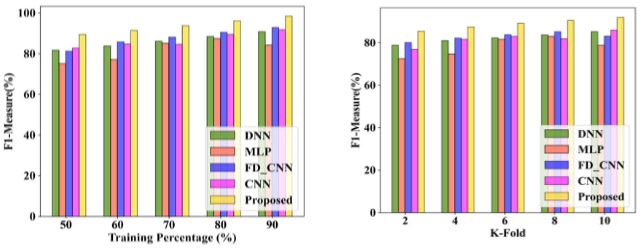
| Methods | DNN | MLP | FD_CNN | CNN | Proposed |
|---|---|---|---|---|---|
| **Training Percentage** | | | | | |
| 50 | 82.27 | 84.41 | 86.75 | 89.10 | 91.49 |
| 60 | 81.24 | 77.74 | 80.07 | 78.07 | 84.81 |
| 70 | 84.33 | 86.47 | 86.24 | 91.16 | 93.60 |
| 80 | 80.73 | 85.44 | 87.77 | 90.13 | 92.52 |
| 90 | 89.91 | 92.05 | 94.38 | 96.74 | 99.13 |
| **K-fold** | | | | | |
| 2 | 76.35 | 82.18 | 84.24 | 89.57 | 90.00 |
| 4 | 75.42 | 75.61 | 77.54 | 82.86 | 88.96 |
| 6 | 78.21 | 84.03 | 86.35 | 91.69 | 88.53 |
| 8 | 74.71 | 83.11 | 82.73 | 90.63 | 91.09 |
| 10 | 83.69 | 89.50 | 91.96 | 97.30 | 97.67 |

### 4.3.5. F1-Measurebased Comparison

The harmonic mean between the precision and recall is measured through F1-Measure and it is given in Equation (37):

$$FRC_{F1-M} = 2\frac{FRC_{\Pr e} * FRC_{Recall}}{FRC_{\Pr e} + FRC_{Recall}} \tag{37}$$

Where, the precision is indicated as $FRC_{F1\text{-}M}$=. The F1-Measure measure of facial expression recognition and classification methods concerning the training percentage and K-Fold is depicted in Figure 8.




a) Training percentage. b) K-Fold.

Figure 8. F1-measure measure in terms of training percentage and K-fold.

Here, the F1-Measure measured by the IC_EfficientNet is 84.16, which is 10.79%, 8.36%, 10.79%, and 8.01% improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods with 60% of learning data. Likewise, with K-Fold value of 4, the F1-Measure measured by the IC_EfficientNet is 78.82, which is 7.96%, 5.23%, 9.22%, and 7.39% improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods. The elaborated assessment of F1-Measure is presented in Table 7.

Table 7. F1-Measure measure based comparison.

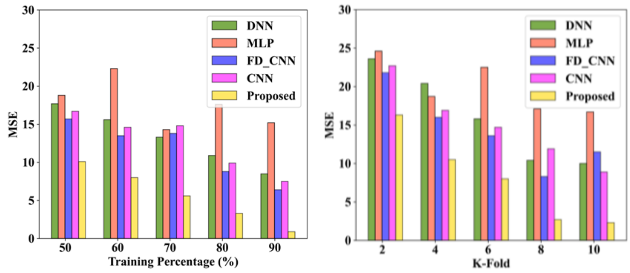
| Methods | DNN | MLP | FD_CNN | CNN | Proposed |
|---|---|---|---|---|---|
| **Training Percentage** | | | | | |
| 50 | 81.75 | 83.78 | 86.10 | 88.44 | 90.82 |
| 60 | 75.08 | 77.12 | 75.08 | 77.42 | 84.16 |
| 70 | 81.23 | 85.83 | 88.14 | 90.48 | 92.92 |
| 80 | 82.77 | 84.80 | 84.55 | 89.46 | 91.85 |
| 90 | 89.37 | 91.40 | 93.71 | 96.05 | 98.44 |
| **K-fold** | | | | | |
| 2 | 78.84 | 80.96 | 82.29 | 83.73 | 85.18 |
| 4 | 72.54 | 74.70 | 81.55 | 72.99 | 78.82 |
| 6 | 80.15 | 82.20 | 83.76 | 85.22 | 83.11 |
| 8 | 76.93 | 81.58 | 83.03 | 81.91 | 85.89 |
| 10 | 85.35 | 87.36 | 89.05 | 90.52 | 91.87 |

### 4.3.6. MSE Based Comparison

The error measured between the original outcome and the recognized outcome is measured through MSE and it is given in Equation (38):

$$MSE = \frac{1}{Q}\sum_{i=1}^{Q}(A_{soln} - P_{soln})^2 \tag{38}$$

Where, the MSE is defined as *MSE*, the target solution is mentioned as $P_{soln}$, the estimated solution is mentioned as $A_{soln}$, and the total number of birds population in the search space is defined as *Q*. The MSE measure of facial expression recognition and classification methods concerning the training percentage and K-Fold is depicted in Figure 9.




a) Training percentage. b) K-Fold.

Figure 9. MSE measure in terms of training percentage and K-fold.

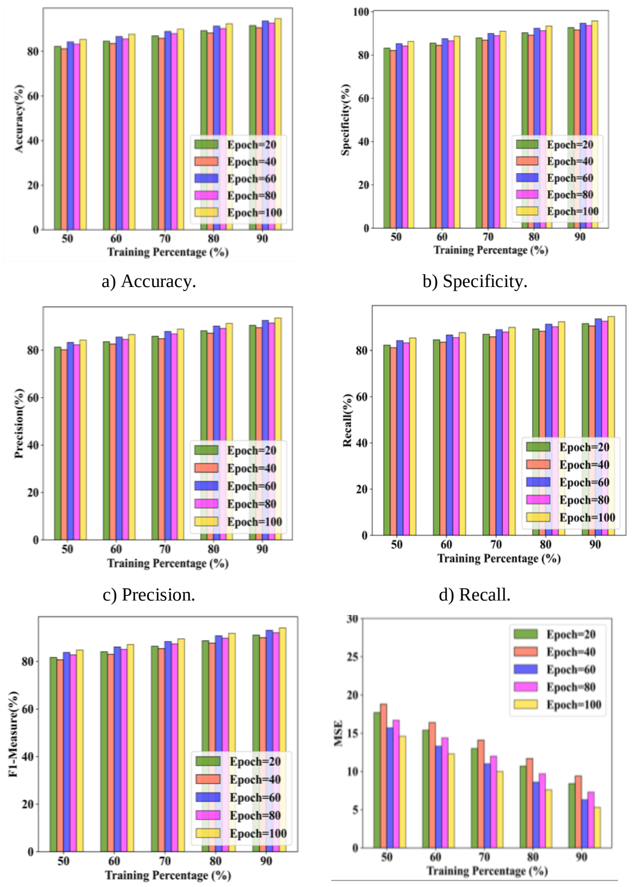
Here, the MSE measured by the IC_EfficientNet is 7.48, which is 55.20%, 48.61%, 49.41%, and 24.20% improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods with 80% of learning data. Likewise, with K-Fold value of 8, the MSE measured by the IC_EfficientNet is 8.91, which is 760.80%, 47.29%, 39.43%, and 25.38% improved outcome compared to the conventional DNN, MLP, FD_CNN and CNN methods. The elaborated assessment of MSE is presented in Table 8.

Table 8. MSE measure based comparison.

| Methods | DNN | MLP | FD_CNN | CNN | Proposed |
|---|---|---|---|---|---|
| **Training Percentage** | | | | | |
| 50 | 17.73 | 15.59 | 13.25 | 10.90 | 8.51 |
| 60 | 18.76 | 22.26 | 14.28 | 17.58 | 15.19 |
| 70 | 15.67 | 13.53 | 13.76 | 8.84 | 6.40 |
| 80 | 16.70 | 14.56 | 14.79 | 9.87 | 7.48 |
| 90 | 10.09 | 7.95 | 5.62 | 3.26 | 0.87 |
| **K-fold** | | | | | |
| 2 | 23.65 | 20.38 | 15.76 | 10.43 | 10.00 |
| 4 | 24.58 | 18.74 | 22.46 | 17.14 | 16.69 |
| 6 | 21.79 | 15.97 | 13.65 | 8.31 | 11.47 |
| 8 | 22.72 | 16.89 | 14.70 | 11.94 | 8.91 |
| 10 | 16.31 | 10.50 | 8.04 | 2.70 | 2.33 |

### 4.4. Analysis of IC_EfficientNet

The robustness of the proposed IC_EfficientNet is evaluated by varying the epoch size for various assessment measures, which is depicted in Figure 10. The analysis indicates that the increase in epoch value elevates the performance of the model. Here, the accuracy measure is depicted in Figure 10-a), the specificity measure is depicted in Figure 10-b), the precision measure is depicted in Figure 10-c), the recall measure is depicted in Figure 10-d), the F1-Measure is depicted in Figure 10-e), and the MSE is depicted in Figure 10-f).




a) Accuracy. b) Specificity.

c) Precision. d) Recall.

e) F1-measure. f) MSE.

Figure 10. Analysis of IC_EfficientNet in terms of accuracy, specificity, precision, recall, F1-measure and MSE.

### 4.5. Discussion Based on Best Case

The discussion based on the best case evaluated by the facial recognition and classification methods are depicted in Table 9. The accuracy measured by the IC_EfficientNet is 99.13 that is 9.30%, 7.14%, 4.79%, and 2.41% better outcome as compared to DNN, MLP, FD_CNN, and CNN. The specificity measured by the IC_EfficientNet is 98.80 that is 8.04%, 5.67%, 3.28%, and 0.88% better outcome as compared to DNN, MLP, FD_CNN, and CNN. Likewise, the precision measured by the IC_EfficientNet is 97.80 that is 9.13%, 7.15%, 4.80%, and 2.41% better outcome as compared to DNN, MLP, FD_CNN, and CNN. The higher recall of 99.13 is measured by the IC_EfficientNetthat is 9.30%, 7.14%, 4.79%, and 2.41% better outcome as compared to DNN, MLP, FD_CNN, and CNN. Maximal F1 Measure evaluated by the IC_EfficientNet is 98.44 that is 9.21%, 7.16%, 4.81%, and 2.42% better outcome as compared to DNN, MLP, FD_CNN, and CNN. Finally, the minimum MSE estimated by the IC_EfficientNet is 0.87 that is 91.34%, 89.01%, 84.44%, and 73.23% better outcome as compared to DNN, MLP, FD_CNN, and CNN.

Table 9. Analysis based on best case.

| Methods/Metrics | DNN | MLP | FD_CNN | CNN | Proposed |
|---|---|---|---|---|---|
| **Accuracy** | 89.91 | 92.05 | 94.38 | 96.74 | 99.13 |
| **Specificity** | 90.86 | 93.19 | 95.56 | 97.93 | 98.80 |
| **Precision** | 88.87 | 90.81 | 93.10 | 95.44 | 97.80 |
| **Recall** | 89.91 | 92.05 | 94.38 | 96.74 | 99.13 |
| **F1-measure** | 89.37 | 91.40 | 93.71 | 96.05 | 98.44 |
| **MSE** | 10.09 | 7.95 | 5.62 | 3.26 | 0.87 |

The analysis interprets the enhanced outcome of the IC_EfficientNet compared to the conventional methods. The newly devised IC algorithm for adjusting the tunable parameters of the EfficientNet-B7 optimally modifies the parameters through the balanced stages of exploration and exploitation. Besides, the effective solution updating behaviour of the Pelican optimization eliminates the solution moving away from the optimization and hence the fast convergence is obtained with global best solution. Also, the EfficientNet-B7 offers the better outcome compared to the traditional deep learning models in terms of accuracy and computation. Thus, the IC_EfficientNet accomplished superior performance.

Table 10. Statistical analysis of variance (ANOVA).

| Source of variation | DF | Sum of square | Mean square | F | PR(>F) |
|---|---|---|---|---|---|
| C(Factor1) | 4.0 | 374.8504 | 93.71260 | 34.060588 | 1.214298e-07 |
| C(Factor2) | 4.0 | 221.9704 | 55.49260 | 20.169226 | 4.320292e-06 |
| Residual | 16.0 | 44.0216 | 2.75135 | NaN | NaN |

The statistical significance of the variables affecting FER accuracy is displayed in the ANOVA Table 10. With four Degrees of Freedom (DF), Factor 1 and Factor 2 account for 374.8504 and 221.9704 of the variation, respectively. The significant differences between the groups are indicated by the large F-statistics of 34.06 for Factor 1 and 20.17 for Factor 2. The extremely low PR-values of 1.21e-07 for Factor1 and 4.32e-06 for Factor2 demonstrate that both factors have a considerable impact on the result. 44.0216 is the residual variance.

## 5. Conclusions

This research introduced a novel facial expression recognition using the IC_EfficientNet, wherein the EfficientNet-B7 is trained using the proposed IC algorithm. Here, the newly designed IC algorithm hybrids the foraging behaviour of the Coot water bird and the food capturing behaviour of the bird named

Gannet. The food capturability behaviour of the Gannet enhances the convergence rate of the algorithm with the balanced exploration and exploitation phase. Thus, the proposed method accomplished the superior outcome in facial expression recognition and classification task. Besides, the additional feature extraction techniques extracts the significant attributes that reduces the computational overhead of the model. The performance based on various metrics like Accuracy, Specificity, Precision, Recall, F1-Measure, and MSE acquired the better outcome of 99.13, 98.80, 97.80, 99.13, 98.44, and 0.87 respectively. The present deep facial expression recognition system is dedicated to addressing the following two issues:

1) Overfitting as a result of insufficient training data.
2) Interference issues caused by other variables unrelated to expression in the real world.

However, the error evaluated by the method needs to further reduced for enhancing the performance of the model. Thus, in the future a hybrid deep learning method will be devised for enhancing the outcome of the model.

## References


[1] Aslan M., "CNN Based Efficient Approach for Emotion Recognition," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 9, pp. 7335-7346, 2022. DOI:doi.org/10.1016/j.jksuci.2021.08.021

[2] Bhatia Y., Bari A., Hsu G., and Gavrilova M., "Motion Capture Sensor-Based Emotion Recognition Using a Bi-Modular Sequential Neural Network," *Sensors*, vol. 22, no. 1, pp. 403, 2022. https://doi.org/10.3390/s22010403

[3] Boughida A., Kouahla M., and Lafifi Y., "A Novel Approach for Facial Expression Recognition Based on Gabor filters and Genetic Algorithm," *Evolving Systems*, vol. 13, no. 2, pp. 331-345, 2022. doi.org/10.1007/s12530-021-09393-2

[4] Cha H. and Im C., "Performance Enhancement of Facial Electromyogram-Based Facial-Expression Recognition for Social Virtual Reality Applications Using Linear Discriminant Analysis Adaptation," *Virtual Reality*, vol. 26, no. 1, pp. 385-398, 2022. doi.org/10.1007/s10055-021-00575-6

[5] Dino H. and Abdulrazzaq M., "Facial Expression Classification Based on SVM, KNN and MLP Classifiers," *in Proceedings of the International Conference on Advanced Science and Engineering*, Duhok, pp. 70-75, 2019. DOI:10.1109/ICOASE.2019.8723728

[6] Dresvyanskiy D., Ryumina E., Kaya H., Markitantov M., Karpov A., and Minker W., "End-to-End Modeling and Transfer Learning for Audiovisual Emotion Recognition in-the-Wild," *Multimodal Technologies and Interaction*, vol. 6, no. 2, pp. 11, 2022. https://doi.org/10.3390/mti6020011

[7] Fard A. and Mahoor M., "AD-CORRE: Adaptive Correlation-Based Loss for Facial Expression Recognition in the wild," *IEEE Access*, vol. 10, pp. 26756-26768, 2022. DOI:10.1109/ACCESS.2022.3156598

[8] Guo J., "Deep Learning Approach to Text Analysis for Human Emotion Detection from Big Data," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 113-126, 2022. DOI:10.1515/jisys-2022-0001

[9] Keshri A., Singh A., Kumar B., Pratap D., and Chauhan A., "Automatic Detection and Classification of Human Emotion in Real-Time Scenario," *Journal of IoT in Social, Mobile, Analytics, and Cloud*, vol. 4, no. 1, pp. 41-53, 2022. DOI:10.36548/jismac.2022.1.005

[10] Kim J., Kim B., Roy P., and Jeong D., "Efficient Facial Expression Recognition Algorithm based on Hierarchical Deep Neural Network Structure," *IEEE Access*, vol. 7, pp. 41273-41285, 2019. DOI:10.1109/ACCESS.2019.2907327

[11] Kim J., Kim N., and Won C., "Facial Expression Recognition with Swin Transformer," *arXiv preprint*, vol. arXiv:2203, pp. 13472, 2022. https://doi.org/10.48550/arXiv.2203.13472

[12] Li Y., Wei J., Liu Y., Kauttonen J., and Zhao G., "Deep Learning for Micro-Expression Recognition: A Survey," *IEEE Transactions on Affective Computing*. vol. 13, no. 4, pp. 2028-2046, 2022. DOI:10.1109/TAFFC.2022.3205170

[13] Liu Y., Feng C., Yuan X., Zhou L., Wang W., Qin J., and Luo Z., "Clip-Aware Expressive Feature Learning for Video-Based Facial Expression Recognition," *Information Sciences*, vol. 598, pp. 182-195, 2022. https://doi.org/10.1016/j.ins.2022.03.062

[14] Moung E., Wooi C., Sufian M., On C., and Dargham J., "Ensemble-Based Face Expression Recognition Approach for Image Sentiment Analysis," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 3, pp. 2588, 2022. DOI: http://doi.org/10.11591/ijece.v12i3.pp2588-2600

[15] Nan F., Jing W., Tian F., Zhang J., Chao K., Hong Z., and Zheng Q., "Feature Super-Resolution Based Facial Expression Recognition for Multi-Scale Low-Resolution Images," *Knowledge-Based Systems*, vol. 236, pp. 107678, 2022. https://doi.org/10.1016/j.knosys.2021.107678

[16] Nan Y., Ju J., Hua Q., Zhang H., and Wang B., "A-MobileNet: An Approach of Facial Expression Recognition," *Alexandria Engineering Journal*, vol. 61, no. 6, pp. 4435-4444, 2022. https://doi.org/10.1016/j.aej.2021.09.066

[17] Naruei I. and Keynia F., "A New Optimization

Method Based on COOT Bird Natural Life Model," *Expert Systems with Applications*, vol. 183, pp. 115352, 2021. https://doi.org/10.1016/j.eswa.2021.115352

[18] Pan J., Zhang L., Wang R., Snášel V., and Chu S., "Gannet Optimization Algorithm: A New Metaheuristic Algorithm for Solving Engineering Optimization Problems," *Mathematics and Computers in Simulation*, vol. 202, pp. 343-373, 2022. https://doi.org/10.1016/j.matcom.2022.06.007

[19] Peng X., Gu Y., and Zhang P., "Au-Guided Unsupervised Domain-Adaptive Facial Expression Recognition," *Applied Sciences*, vol. 12, no. 9, pp. 4366, 2022. https://doi.org/10.3390/app12094366

[20] Phan K., Nguyen H., Huynh V., and Kim S., "Facial Expression Classification using Fusion of Deep Neural Network in Video," *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, pp. 2507-2511, 2022. DOI:10.1109/CVPRW56347.2022.00280

[21] Rawal N. and Stock-Homburg R., "Facial Emotion Expressions in Human-Robot Interaction: A Survey," *International Journal of Social Robotics*, vol. 14, no. 7, pp. 1583-1604, 2022. https://doi.org/10.1007/s12369-022-00867-0

[22] Saeed S., Shah A., Ehsan M., Amirzada M., Mahmood A., and Mezgebo T., "Automated Facial Expression Recognition Framework Using Deep Learning," *Journal of Healthcare Engineering*, vol. 2022, pp. 1-11, 2022. https://doi.org/10.1155/2022/5707930

[23] Saeed V., "A Framework for Recognition of Facial Expression Using HOG Features," *International Journal of Mathematics, Statistics, and Computer Science*, vol. 2, pp. 1-8, 2024. https://doi.org/10.59543/ijmscs.v2i.7815

[24] Shanthi P. and Nickolas S., "Facial Landmark Detection and Geometric Feature-Based Emotion Recognition," *International Journal of Biometrics*, vol. 14, no. 2, pp. 138-154, 2022. https://doi.org/10.1504/IJBM.2022.121799

[25] Shen J., Yang H., Li J., and Cheng Z., "Assessing Learning Engagement Based on Facial Expression Recognition in MOOC's Scenario," *Multimedia Systems*, vol. 28, pp. 469-478, 2022. https://doi.org/10.1007/s00530-021-00854-x

[26] Siam A., Soliman N., Algarni A., El-Samie A., Fathi E., and Sedik A., "Deploying Machine Learning Techniques for Human Emotion Detection," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1-16, 2022. https://doi.org/10.1155/2022/8032673

[27] Swaminathan A., Vadivel A., and Arock M., "FERCE: Facial Expression Recognition for Combined Emotions Using FERCE Algorithm," *IETE Journal of Research*, vol. 68, no. 5, pp. 3235-3250, 2022. https://doi.org/10.1080/03772063.2020.1756471

[28] Talaat F., Ali Z., Mostafa R., and El-Rashidy N., "Real-Time Facial Emotion Recognition Model Based on Kernel Autoencoder and Convolutional Neural Network for Autism Children," *Soft Computing*, vol. 28, pp. 6695-6708, 2024. https://doi.org/10.1007/s00500-023-09477-y

[29] Xue F., Tan Z., Zhu Y., Ma Z., and Guo G., "Coarse-to-Fine Cascaded Networks with Smooth Predicting for Video Facial Expression Recognition," *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, pp. 2412-2418, 2022. DOI:10.1109/CVPRW56347.2022.00269

[30] Yu J., Cai Z., He P., Xie G., and Ling Q., "Multi-Model Ensemble Learning Method for Human Expression Recognition," *arXiv preprint*, vol. arXiv:2203, pp.14466, 2022. doi.org/10.48550/arXiv.2203.14466

[31] Zeng D., Lin Z., Yan X., Liu Y., Wang F., and Tang B., "Face2exp: Combating Data Biases for Facial Expression Recognition," *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, pp. 20291-20300, 2022. DOI:10.1109/CVPR52688.2022.01965

[32] Zhi J., Song T., Yu K., Yuan F., Wang H., Hu G., and Yang H., "Multi-Attention Module for Dynamic Facial Emotion Recognition," *Information*, vol. 13, no. 5, pp. 207, 2022. doi.org/10.3390/info13050207

[33] Ziade H., Ayoubi R., and Velazco R., "A Survey on Fault Injection Techniques," *The International Arab Journal on Information Technology*, vol. 1, no. 2, pp. 171-186, 2004.

**Jenu Prabha** obtained her M.E. in Communication and Networking from Ponjesly College of Engineering, Nagercoil, and later pursued an M.B.A. in Human Resources from ManonmaniamSundaranar University. With nearly four years of corporate experience, she has honed her expertise in Digital Image Processing, showcasing her dedication to advancing research in this specialized field.

**Vinod Kumar** has completed his Ph. D from the Anna University, Chennai and has completed his M. E. Communication Systems from the Mepco Schlenk Engineering College, Sivakasi. He has teaching experience for 30 years and research experience for 12 years. His specialised areas in research are Digital Image Processing. He has the credit of having published 17 research papers in Web of Science/Scopus Indexed Journals and 0 papers in UGC Care Listed, Refereed and Reviewed Journals. He Has guided 4 Ph. D scholars and he is now guiding 6 Ph.D scholars. He received 1 Awards and Recognition for his contribution in the field of research/technology.