Computational Intelligence in Electrical Engineering
Vol. 17, No. 1, 2026
pp. 59-70
Research Paper

Facial Emotion Recognition Using Deep Learning Based on Optimal Image Interpolation by the Chimp Algorithm

Elham Askari, [1*], Sara Motamed[2]

[1] Assistant Professor, Department of Computer Engineering, FSh.C., Islamic Azad University, Fouman, Iran
askary.elham@iau.ac.ir
[2] Assistant Professor, Department of Computer Engineering, FSh.C., Islamic Azad University, Fouman, Iran
sara.motamed@iau.ac.ir

Abstract :
Facial emotion recognition plays an essential role in human-computer interaction, psychological assessment, and intelligent surveillance. However, achieving high performance on large, unconstrained datasets such as AffectNet remains challenging due to variations in pose, illumination, occlusion, and subtle inter-class differences. In this study, we propose a hybrid deep learning framework that combines optimal image interpolation based on the Chimp Optimization Algorithm, Early Biologically Inspired Features, and a Long Short-Term Memory network for facial emotion recognition. We used a curated, balanced subset of approximately 20000 images from AffectNet, covering seven basic emotions, for evaluation. To ensure reliable assessment, the proposed method was validated using 5-fold cross-validation. In this framework, ChOA optimizes interpolation parameters to improve feature quality, EBIF extracts biologically inspired descriptors, and the LSTM models dependencies among the extracted features. Experimental results demonstrate that the proposed approach achieves an average cross-validation accuracy of 84.3% and consistently outperforms several baseline deep learning models under the identical evaluation protocol. These findings suggest that integrating evolutionary optimization with biologically inspired feature extraction can enhance facial emotion recognition on complex real-world datasets.
Keywords: Emotion, Face, Chimp Algorithm, LSTM Network, Early Biologically Inspired Feature Model.

1. Introduction

Facial expressions represent one of the most important forms of nonverbal communication and provide valuable information about human emotional states. The human face conveys rich cues about internal reactions to environmental stimuli, making facial analysis an essential component for understanding human behavior and interaction [1]. Consequently, enabling computers and intelligent systems to automatically recognize

* Submission date:11, 03, 2025
Acceptance date: 07, 10, 2026
Corresponding author: Elham Askari, Department of Computer Engineering, Fouman and Shaft Branch, Islamic Azad University, Fouman, Iran
This is an open access article under the CC BY-NC-ND/4.0/ License (https://creativecommons.org/licenses/by-nc-nd/4.0/).
https://doi.org/10.22108/isee.2026.144644.1728

emotions through machine vision has become an important research topic in artificial intelligence and human–computer interaction.

Emotions such as anger, happiness, sadness, fear, surprise, envy, jealousy, and shame are fundamental components of human emotional experience. These emotional states are often accompanied by psychological and physiological responses that manifest through facial expressions, body language, and other observable behaviors [2]. For instance, emotions such as fear or sadness may arise when an individual perceives internal or external threats and can be associated with physiological symptoms including increased heart rate, sweating, facial flushing, and changes in breathing patterns [3]. Because facial expressions reflect these emotional states, they provide an effective and widely used medium for automatic emotion recognition systems.

Facial emotion recognition[i] focuses on analyzing facial images or video sequences to identify emotional states using computational models. Traditional FER approaches relied primarily on handcrafted features and classical machine learning techniques, while more recent studies increasingly utilize deep learning architectures to automatically learn discriminative representations from data [4]. The scientific study of facial expressions can be traced back to Darwin's work in 1872, where emotional expressions in humans and animals were systematically examined [5]. Since then, facial expression analysis has attracted significant attention in fields such as psychology, neuroscience, cognitive science, and computer vision [6–7].

Several early studies proposed emotion recognition approaches based on statistical feature extraction methods. For example, Lakshmi et al. used skin color detection to locate facial regions in video sequences and applied Principal Component Analysis[ii] for feature extraction and emotion classification [8]. Murthy and Jadon also employed PCA-based feature extraction to recognize six basic emotions using the JAFFE and Cohn–Kanade datasets, reporting recognition rates of 75% and 77.5%, respectively [9]. Manal Abdullah et al. improved facial image recognition by dividing images into smaller feature subsets, which reduced processing time while increasing classification accuracy [10]. Other studies emphasized the importance of facial emotion recognition in applications such as security systems, authentication technologies, psychological analysis, and human–computer interaction [11–12].

In addition to traditional machine learning techniques, biologically inspired models and emotional learning algorithms have also been explored. Qolipour et al. introduced an emotional learning algorithm capable of supporting rapid decision-making processes [13]. Mehrabian and Lucas proposed a control strategy based on the Brain Emotional Learning[iii] model, which mimics mechanisms of the human limbic system to improve stability and adaptability in intelligent systems [14]. Beheshti and Hashim further developed computational models inspired by emotional processing in the brain, demonstrating the potential of biologically inspired approaches for intelligent control and decision-making applications [15].

Despite these developments, many facial emotion recognition systems still struggle to achieve high accuracy on complex and unconstrained datasets. In particular, certain emotions such as fear or disgust are often misclassified due to subtle visual differences and limited discriminative features. This limitation is significant because reliable recognition of all emotional states is important for applications in medical diagnosis, behavioral analysis, and mental health monitoring, where accurate identification of emotional patterns can assist in detecting disorders such as panic attacks or anxiety-related conditions [16].

To address these challenges, this study proposes a hybrid deep learning framework for facial emotion recognition that combines optimal image interpolation, biologically inspired feature extraction, and deep sequential learning. In the proposed approach, informative points are first extracted from facial images through interpolation techniques, and the optimal interpolation parameters are determined using the Chimp Optimization Algorithm[iv]. The Early Biologically Inspired Feature[v] model is then applied to generate discriminative facial descriptors, which are subsequently classified using a Long Short-Term Memory[vi] neural network.

The main contributions of this work are summarized as follows. First, a hybrid FER framework is proposed that integrates interpolation-based feature localization, EBIF

descriptors, and an LSTM classifier to improve recognition performance on complex facial datasets. Second, the Chimp Optimization Algorithm is employed to optimize interpolation parameters and select discriminative facial points, which enhances the quality of extracted features while reducing irrelevant information. Third, the proposed system combines biologically inspired features with deep learning to improve the recognition of both distinct and subtle emotional expressions. Finally, extensive experiments are conducted on a curated subset of approximately 20000 images from the AffectNet dataset representing seven basic emotions, and the performance of the proposed framework is evaluated using 5-fold cross-validation. The experimental results demonstrate that the proposed method achieves an average cross-validation accuracy of 84.3% and outperforms several baseline deep learning models under comparable conditions.

2. Materials and Method

The proposed framework consists of four main stages: image preprocessing, optimized interpolation using the chimp optimization algorithm, early biologically inspired feature extraction, and final classification using a long short-term memory network. The flowchart depicting the proposed method is illustrated in Figure (1). To ensure reproducibility, all processing steps, parameters, and settings are described in detail below.

2.1. Database

In this study, a curated subset of approximately 20000 images was selected from the AffectNet database for facial emotion recognition. The selected data covered the seven basic emotion categories: neutral, happy, sad, angry, fear, surprise, and disgust. To ensure a balanced class distribution, 2850 images were selected for each emotion category, yielding a total of 19950 samples (hereafter referred to as a subset of 20000 images for brevity). The dataset was curated by retaining high-quality images with clearly visible facial landmarks and unambiguous emotional expressions, while maintaining an equal distribution across all seven classes to mitigate label noise and class imbalance inherent in large-scale datasets.

2.2. Preprocessing

All images were resized to 128×128 pixels and converted to grayscale. Intensity normalization was performed using min–max normalization, and noise reduction was applied through histogram equalization. Following the procedure in previous FER studies, each face image was divided horizontally into five equal regions; only the middle three regions were retained, as they contain the most expressive facial components (eyes, eyebrows, and mouth).

2.3. Proposed Model

The proposed model comprises three main stages. In the first stage, the input image undergoes pre-processing to normalize the intensity and contrast while removing image noise. This is achieved by applying histogram equalization. Additionally, the image is horizontally divided into five parts, and the middle three parts are considered for further analysis. The second stage involves interpolating the image between the three selected parts using the Chimp optimization algorithm. This optimization process refines the obtained important points and calculates the distances between them. Furthermore, image features are extracted using the EBIF biological model, which is known for its effectiveness in capturing relevant facial information. In the third stage, the image is classified and assigned to an emotion-related class by combining the features extracted from the EBIF model with the values derived from the interpolation points using a Long Short-Term Memory network. The LSTM network, a type of recurrent neural network, is utilized for its ability to handle sequential data and capture temporal dependencies.

2.3.1 Image Interpolation

Bilinear interpolation is a generalization of linear interpolation to interpolate linear bivariate functions using linear interpolation first in one direction and then applying it in the other direction. Bilinear interpolation is one of the basic resampling techniques in machine vision and image processing. In the following, the algorithm used for image interpolation is explained.

It is assumed that the value of the unknown function f should be found at the point (x, y), and it is assumed that the value of f at four points Q11=(x1,y1), Q12=(x1,y2), Q21=(x2,y1), and Q22=(x2,y2) is known. The linear interpolation for x is as follows:

$$f(x, y_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}),$$
$$f(x, y_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}), \quad (1)$$

In the following, the interpolation will be done for y, and the desired value will be obtained:

$$
\begin{aligned}
f(x,y) &\approx \frac{y_2-y}{y_2-y_1}f(x,y_1)+\frac{y-y_1}{y_2-y_1}f(x,y_2),\\
&=\frac{y_2-y}{y_2-y_1}\left(\frac{x_2-x}{x_2-x_1}f(Q_{11})+\frac{x-x_1}{x_2-x_1}f(Q_{21})\right)\\
&+\frac{y-y_1}{y_2-y_1}\left(\frac{x_2-x}{x_2-x_1}f(Q_{12})+\frac{x-x_1}{x_2-x_1}f(Q_{22})\right)\\
&=\frac{1}{(x_2-x_1)(y_2-y_1)}\big(f(Q_{11})(x_2-x)(y_2-y)\\
&\quad+f(Q_{21})(x-x_1)(y_2-y)\\
&+f(Q_{12})(x_2-x)(y-y_1)+f(Q_{22})(x-x_1)(y\\
&\quad-y_1)\\
&=\frac{1}{(x_2-x_1)(y_2-y_1)}\begin{bmatrix}x_2-x & x\end{bmatrix}\\
&-x_1\begin{bmatrix}f(Q_{11}) & f(Q_{12})\\ f(Q_{21}) & f(Q_{22})\end{bmatrix}\begin{bmatrix}y_2-y\\ y-y_1\end{bmatrix}
\end{aligned}
\tag{2}
$$

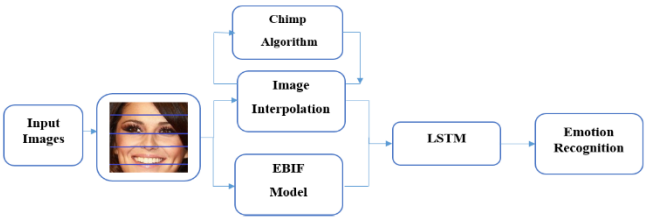


Figure (1): Flowchart of the proposed model

Between the two lines of the image, 70 points are estimated. Next, to optimally determine the location of the points, the chimp algorithm will be used. To determine the optimal locations for these points, the chimpanzee optimization algorithm is employed. The chimp algorithm is a computational algorithm inspired by the behavior and learning of chimpanzees. It is commonly used for solving optimization problems, agent-based search, and multi-objective optimization.

2.3.2. Chimp Algorithm

The chimpanzee algorithm is designed to model the complex behaviors and learning methods observed in chimpanzees, which are among the animals most closely related to humans. This algorithm aims to solve computational problems by emulating the problem-solving approaches of chimpanzees. The following steps are typically followed in the chimpanzee algorithm [17, 18]:

1- Determining the search space: Define the range of possible values for each parameter in the problem. This search space encompasses all potential values that the parameters can take.

2-Creating an initial population: Generate an initial population of pairs of parameter values based on the defined search space. This population is typically created randomly to explore a wide range of possibilities.

3-Calculation of the objective function: Evaluate the performance of the system or model by calculating the objective function for each pair of parameter values in the population. The objective function usually represents a measure of the system's performance, such as an error metric.

4-Selection of the next generation: Select individuals from the current population based on their objective function values. Individuals with better objective function values are more likely to be selected for the next generation.

5-Reproduction and mutation: Reproduce selected individuals to create new individuals for the next generation. This reproduction process involves combining the parameter values of selected individuals and introducing random mutations to explore new areas of the search space.

6-Repeat steps 3 to 5: Iterate through steps 3 to 5 until termination conditions are met. Termination conditions can include a maximum number of iterations or achieving a desired level of accuracy or performance.

7-Analysis and selection of the best pair: After the optimization process concludes, analyze the results and select the pair of parameter values that achieved the best objective function value. This pair represents the optimal solution or configuration for the given problem.

In the proposed framework, the Chimp Optimization Algorithm was applied to determine the optimal interpolation points between the selected facial regions. The algorithm searches for parameter combinations that maximize the discriminative capability of the extracted facial features.

The parameters of the ChOA were set as follows: a population size of 30 candidate solutions and a maximum of 50 iterations. The search space dimension was defined as 20. The optimization process was guided by an objective function that maximizes the ratio between inter-class distance and intra-class compactness in the feature space. Convergence was assumed when no significant improvement in the objective function was observed for five consecutive iterations.

After the optimization stage, linear interpolation was performed between the corresponding key points of the selected facial

regions. The ChOA algorithm selected the subset of interpolation points that maximized the objective function. As a result, the final representation consisted of 60 optimized distance-based features, which were then used as input to the subsequent feature extraction and classification stages.

2.3.3. Early Biologically Inspired Feature Model

The face is one of the most important biometric modalities for emotion and identity recognition because it contains rich structural information and hidden patterns that significantly contribute to recognition accuracy. In this work, we employ the Early Biologically Inspired Features Model to extract discriminative facial representations. EBIF simulates early stages of the human visual processing system through a set of biologically inspired filtering and pooling operations, enabling the capture of essential spatial and textural characteristics from facial regions.

In addition to standard EBIF processing, distance-based measurements derived from key facial structures are also incorporated to emphasize critical geometric variations across different expressions. Using the EBIF framework, a total of 256 descriptive features are extracted from each preprocessed facial image. These descriptors form a compact yet informative representation of facial expression patterns and serve as the input feature vector for the subsequent optimization stage using the Chimp Optimization Algorithm [19, 20].

The EBIF model is an improvement over the BIM model, specifically designed for enhancing emotional state recognition. It consists of four layers: Image Layer, S1 Layer, C1 Layer, and Local L2-norm Layer. The functions of each layer are explained below.

-Image layer: In this layer, the images are initially converted to grayscale and then pyramidally scaled down in N=6 (in this article, N=6) levels, from large to small. Each point in the higher-level image is a factor of √2 smaller than the current image. The pyramid structure is created from K to K+1 (K=1, 2, ..., N-1). After the grayscale conversion, the images form a six-level pyramid, which serves as the S1 layer database.

-S1 layer: The S1 layer employs a Gabor filter bank to extract facial features. Gabor filters are specifically chosen due to their resistance to rotation and robustness in detecting facial regions. They also compensate for photometric issues such as lighting variations and image noise [18-20]. When properly configured, Gabor filters exhibit excellent performance in detecting both texture features and texture edge features. Another key advantage of Gabor filters is their high spatial and frequency selectivity, allowing for localized and adjustable response. The Gabor filter used in the proposed model employs a convolution kernel that combines a complex exponential and a Gaussian function. The formula for the Gabor filter in this model is given below.

$$G(x,y) = exp(-\frac{X^2+\gamma^2Y^2}{2\sigma^2})COS(\frac{2\pi}{\lambda}X) \quad (3)$$

The equations for the Gabor filter response to an image are as follows:

$$X = x \times cos(\theta) + y \times sin(\theta) \quad (4)$$

$$X = -x \times sin(\theta) + y \times cos(\theta) \quad (5)$$

These equations control the orientations of the filter. The filter bank consists of M orientations, where M is equal to 9. The orientations are evenly distributed in the range [0, π). Each filter has a size of 5x5, and the values of x and y range from 2 to 2-change. The aspect ratio, width, and length of the trend are set to 1, 3, and 5, respectively.

Equation (6) represents the response of the Gabor filter to an image.

$$R=G\times I \quad (6)$$

The Gabor filter responses play a crucial role in object localization and distinguishing its structure across different orientations.

C1 layer: The C1 layer operates based on the complex cells of the visual cortex. In this layer, the equal Gabor pyramid is divided into sub-pyramids with specific directions. The size of the bottom-level sub-pyramids is 4x4, while the top-level sub-pyramids have a size of 3x3, following the basic BIM model. Reduction is performed using MaxPooling in this layer. Essentially, a combination of pyramids is used, and instead of selecting the maximum response, the limit and standard deviation are used to represent each sub-pyramid. The advantage of the C1 layer is that each element in this layer, with its Gabor filter response, is locally utilized for edge and position filtering. It allows for fine control over the position of neighboring elements and different orientations.

L2 local normalization: In the L2 local normalization layer, the brightness variance, which is a crucial feature for face recognition, is calculated. To ensure the robustness of the EBIF model against variations in lighting conditions, normalization is applied first. The feature vector f for each location in the C1 layer is calculated using

the L2-norm as follows.

$$\tilde{f} = \frac{f}{\sqrt{||f||^2 + \xi}} \quad (7)$$

In the EBIF model, there is a constant that is used to soften the features of the vector at different positions within the C1 layer. This constant helps to shape the response of the subsets in the feature representation of existing objects. All N-1 subsets consist of multi-constrained features that contain information about local normalization and multi-orientations. These features have the ability to capture subtle variations, such as lip features, and exhibit tolerance in the final EBIF feature response [19].

2.3.4. LSTM Network

The descriptors obtained from the EBIF and optimized interpolation stages were organized as a spatially structured sequence before being supplied to the LSTM. As described in the preprocessing stage, each facial image was divided into five horizontal regions, of which the three middle regions containing the main expression-related components were retained. For each retained region, candidate points were extracted using bilinear interpolation and optimized using ChOA, resulting in 20 distance-based features per region (60 geometric features in total per image).

In parallel, the EBIF framework extracted 256 descriptive features, comprising 255 region-wise features and one scale-invariant global interocular-distance feature. To establish spatial correspondence, the 255 region-wise EBIF features were divided into three equal subsets of 85 features and concatenated with their corresponding 20-dimensional geometric descriptors as follows:

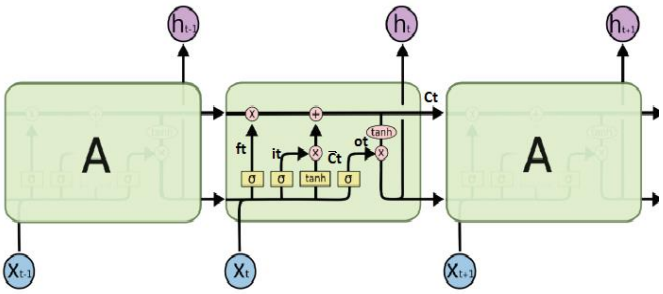


Figure (2): Architecture of long short-term memory network

Region 1 (eyebrow and eye region): 85 EBIF features + 20 distance-based features = 105 features

Region 2 (nose region): 85 EBIF features + 20 distance-based features = 105 features

Region 3 (mouth region): 85 EBIF features + 20 distance-based features = 105 features

Consequently, each facial image was represented by an input matrix $X \in R^{3\times105}$, where the three rows correspond to the retained facial regions arranged in a fixed top-to-bottom spatial hierarchy. The LSTM input is therefore not interpreted as an arbitrary or temporal sequence, but as a spatially structured sequence capturing inter-regional anatomical dependencies.

LSTM is a recurrent neural architecture capable of modeling dependencies within meaningfully ordered sequences while mitigating the vanishing and exploding gradient problems of conventional recurrent neural networks [21]. In the proposed framework, it learns spatial dependencies across the ordered facial regions. The architecture and information flow of the employed LSTM are illustrated in Figure (2). Finally, the additional global feature ($g \in R$) was concatenated with the LSTM output representation before the dense classification layer, yielding a comprehensive 316-dimensional feature representation ($3\times105+1=316$).

The central component of an LSTM block is the memory cell or cell state (C_t). Unlike other parts of the LSTM block, the cell state does not have an activation function applied to it. Instead, information can be added or subtracted directly to the cell state, which helps mitigate the issue of vanishing gradients during backward propagation [21].

Within an LSTM block, there are three gates: the forget gate, the input gate, and the output gate. Each gate consists of a sigmoid activation layer and

a pointwise multiplication operation, which control the flow of information within the cell. The following equations are used to calculate the values of each gate:

$$f_t = \sigma(W_f . [h_{t-1}, x_t] + b_f) \quad (8)$$

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad (9)$$

$$o_t = \sigma(Wo . [h_{t-1}, x_t] + b_o) \quad (10)$$

$$\widetilde{C}_t = tanh(W_c . [h_{t-1}, x_t] + b_c) \quad (11)$$

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C}_t \quad (12)$$

$$h_t = o_t \times tanh(C_t) \quad (13)$$

The value of f_t determines how many cells of the previous state will be forgotten, and i_t determines what new information will be stored in the state of the cell. o_t decides what to see in the output and calculates the hidden output value of this block using the value obtained for the state of the cell. Finally, the output vector from this layer is used as the input of the next step [22, 23]. The hyperparameters of the LSTM were defined in Table (1).

Table (1): The hyperparameters of LSTM

Parameters	Numbers
Layer 1	64 LSTM units
Layer 2	32 LSTM units
Dropout	0.3
Dense layer	7-unit softmax output
Loss function	categorical cross-entropy
Optimizer	Adam (learning rate 0.0005)
Batch size	64
Epochs	60 (with Early Stopping, patience=10)

The selection of each component in the proposed framework is based on their complementary roles in facial emotion recognition. ChOA was chosen as the optimization engine because it offers an effective balance between exploration and exploitation and superior performance in escaping local optima compared to classical evolutionary algorithms like GA. EBIF was selected for its capability to extract biologically inspired descriptors, which are robust against local facial variations. Finally, the LSTM network was utilized to capture temporal dependencies among the extracted features, enabling the framework to model emotional expressions with greater accuracy. This combination ensures a task-oriented design that prioritizes feature discriminability over brute computational complexity.

2.3.5. Overfitting Mitigation Strategy

To address the potential risk of overfitting associated with the curated subset of the AffectNet dataset, several regularization techniques were employed:

a) Data Augmentation: To artificially increase the diversity of the training samples, we applied a comprehensive augmentation pipeline including random rotation (±15°), horizontal flipping, and brightness adjustment. This ensured the model remains invariant to minor spatial and lighting variations.

b) Dropout Regularization: A dropout rate of 0.3 was integrated into the LSTM layers to prevent co-adaptation of neurons and enhance the generalization capability of the network.

c) Early Stopping: The training process was monitored using the validation fold of the current cross-validation split. An early stopping mechanism was implemented with a patience of 10 epochs to terminate training once the validation loss ceased to improve, thereby preventing the model from over-adjusting to the training noise.

d) Feature Selection via ChOA: By using the Chimp Optimization Algorithm to select only the most discriminative interpolation points, we reduced the dimensionality of the input space. This reduction in parameter space significantly lowers the complexity of the model, which is a fundamental step in avoiding overfitting in deep learning frameworks.

3. Results

All experiments were implemented in Python 3.8 using TensorFlow 2.8. The images selected in this article are selected from the AffectNet database. It is a large face dataset with about 0.4 million images. To ensure robust evaluation and mitigate the risk of overfitting, the proposed model was assessed using a 5-fold cross-validation scheme. The dataset was partitioned into five stratified subsets; in each iteration, the model was trained on four folds and evaluated on the remaining held-out validation fold. To prevent data leakage, data augmentation techniques including horizontal flipping, slight rotation, brightness adjustment, and random cropping were applied strictly to the training folds of each iteration. The corresponding validation fold was kept entirely unaugmented and unseen during both feature optimization and model training. The final evaluation metrics reported in this study represent the average performance across all five validation folds. Accuracy, precision, recall, and F-score criteria are used to evaluate the presented model. How to calculate the criteria of

accuracy, precision, recall, and F-score is shown in the following relationships, respectively:

$$Precision = \frac{T_p}{T_p + F_p} \quad (11)$$

$$Recall = \frac{T_p}{T_p + F_N} \quad (12)$$

$$Accuracy = \frac{T_p + T_N}{T_p + F_P + T_N + F_N} \quad (13)$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (14)$$

T_P algorithm classified the sample in the positive category, and the sample is also positive. F_P algorithm classifies the sample as positive, but the sample is negative. T_N algorithm classifies the sample in the negative category, and the sample is also negative. F_N algorithm classifies the sample in the negative category, but the sample is positive [24, 25]. The hyperparameters used in the proposed model are shown in Table (2).

Table (2): Hyperparameters used to adjust the proposed model

General hyperparameters of model training	Learning rate	0.0005
	Max Epochs	60
	Optimizer	Adam
	Batch size	64
	Activation	ReLU
Chimp algorithm hyperparameters	Population size	30
	Maximum iterations	50
	Mutation rate	0.5
	Elite rate	0.5

3.1. The Different Emotional States

In Table (3), the results obtained in different emotional states are shown:

Table (3): The results obtained on the data set

Labels	Accuracy	Precision	Recall	F-measure
Neutral	0.831	0.823	0.812	0.821
Surprise	0.894	0.885	0.872	0.873
Disgust	0.825	0.813	0.803	0.814
Fear	0.808	0.802	0.792	0.797
Happy	0.881	0.871	0.863	0.873
Angry	0.843	0.834	0.832	0.823
Sad	0.817	0.824	0.794	0.809
Average	**0.843**	**0.836**	**0.824**	**0.826**

As can be seen, happiness, anger, and sadness are recognized with high accuracy, and fear is recognized with acceptable accuracy, but less than other states. The results show that the proposed model has a higher average accuracy than other methods. The results are shown in Table (4).

Table (4): Comparison of the proposed model with other basic methods

Method	Accuracy
1D-CNN	68.3
2D-CNN	70.05
LSTM	72.1
Bi-LSTM	75.6
ShuflleNet	78.3
ResNet	81.9
Proposed Model	84.3

The results presented in Table (4) demonstrate that incorporating geometric interpolation alongside EBIF feature fusion significantly enhances facial expression recognition performance. Specifically, the proposed method achieved an average validation accuracy of 84.3% across the 5-fold cross-validation scheme**,** consistently outperforming all baseline models evaluated under the identical protocol while remaining in line with the intrinsic complexity of the AffectNet dataset. Although ResNet is a deeper and computationally more expensive architecture, the proposed framework achieves better performance through a task-oriented design. Specifically, ChOA-based interpolation improves input quality, EBIF extracts discriminative facial descriptors, and LSTM captures dependencies among the extracted features. This combination enables the model to focus more effectively on emotion-related facial patterns, which may explain its superior performance compared with larger generic architectures such as ResNet.

Figure (3) shows the comparison of the average efficiency of three powerful methods, Bi-LSTM, ShuffleNet, ResNet, and the proposed model, with different evaluators.

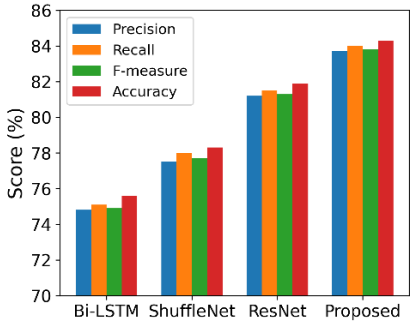


Figure (3): Evaluating the efficiency of different methods

As can be seen, the average efficiency of the proposed method is better. Figure (4) shows the average accuracy of the proposed method in recognizing different states of fear, happiness, sadness, and anger.

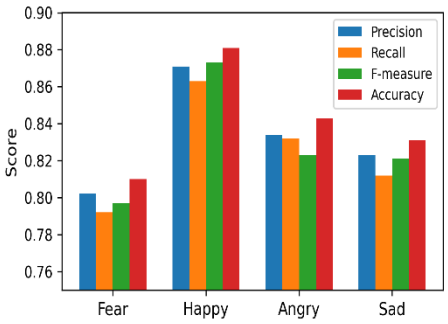


Figure (4): Comparing the accuracy of the proposed method in different situations with other methods

As can be seen, the accuracy of the proposed method is better in all compared cases.

4. Discussion

The results obtained from the proposed ChOA-LSTM model demonstrate a notable improvement in facial emotion recognition performance on the AffectNet dataset. By effectively optimizing the hyperparameters of the LSTM network, the Chimp Optimization Algorithm was able to find a set of parameters that leads to more stable convergence and higher overall accuracy. This optimization allows the model to extract and classify discriminative facial features more effectively.

The experimental findings show that the proposed method not only outperforms conventional baseline models such as ResNet50, but also achieves better performance compared to an LSTM network without ChOA-based optimization. This highlights the crucial role of ChOA in exploring the hyperparameter space and guiding the model toward a more optimal configuration. A more detailed analysis of the classification results indicates that the model achieves strong performance in recognizing more distinct emotions such as happy and neutral, while facing greater challenges in distinguishing more subtle and complex emotions such as “fear” and “disgust”. This behavior is consistent with previous studies on facial emotion recognition using large-scale datasets like AffectNet, where inter-class similarity and intra-class variation are higher for certain emotions.

Despite the improvements achieved, the proposed model can still be further enhanced. Future work may involve employing more advanced data augmentation strategies or integrating deeper and more sophisticated hybrid architectures to improve the robustness and generalization capability of the model. In addition, evaluating the proposed approach on cross-cultural or cross-dataset emotion recognition scenarios could provide further insights into its generalizability and practical applicability.

5. Conclusions

Facial emotion recognition is one of the important areas in the field of artificial intelligence, which, using deep learning, increases the ability to recognize and analyze human emotions in their facial expressions. Facial emotion recognition can be used in fields such as customer emotion recognition in marketing, patient emotion recognition in medicine, psychiatry, as well as in technologies based on human communication and robotic artificial intelligence. In this regard, in this article, an effective model based on deep learning was introduced for facial emotion recognition. First, the features and important points of the images with different emotional states were extracted based on interpolation; then the selected points were optimized with the help of the chimpanzee optimization algorithm, and the values of these distances were obtained. It is also used to extract useful features from the image using the EBIF model. The obtained values were given to the short-term long-term memory network, and the recovery accuracy was calculated. The proposed model was able to achieve 84.3% accuracy using AffectNet database images.

Acknowledgment

The authors would like to thank all those who took the time and participated in this research.

References

[1] A. Khattak, M. Z. Asghar, M. Ali, U. Batool, “An efficient deep learning technique for facial emotion recognition”, Multimedia Tools and Applications, pp.1-35, 2022. https://doi.org/10.1007/s11042-021-11298-w

[2] F. Z. Canal, T. R. Müller, J. C. Matias, G. G. Scotton, A. Junior, E. Pozzebon, A. C. Sobieranski, “A survey on facial emotion recognition techniques: A state-of-the-art literature review”, Information Sciences, Vol. 582, pp. 593-617, 2022. https://doi.org/10.1016/j.ins.2021.10.005

[3] G. S. Hayes, S. N. Mclennan, J. D. Henry, H. L. Philips, G. Terret, P. G. Rendell, R. M. Pelly, “Task characteristics influence facial emotion

recognition age-effects: A meta-analytic review", Psychology and Aging, Vol. 35, No. 2, 2020. https://doi.org/10.1037/pag0000441

[4] E. Pranav, S. Kamal, C. Chandran, M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network, in Proc 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, pp. 317-320, 2020. https://doi.org/10.1109/ICACCS48705.2020.9074302

[5] C. Darwin, P. Prodger, "The expression of the emotions in man and animals", Oxford University Press, USA, 1998.

[6] E. Marg, "Descartes' error: Emotion, reason, and the human brain", Optometry and Vision Science, Vol. 72, No. 11, pp. 847-848, 1995. https://doi.org/10.1002/j.1538-9235.1995.tb03988.x

[7] R. B. Wilbur, " Nonmanuals, semantic operators, domain marking, and the solution to two outstanding puzzles in ASL", Sign Language & Linguistics, Vol. 14, No. 1, pp. 148-178, 2011. https://doi.org/10.1075/sll.14.1.08wil

[8] V. Praseeda Lekshmi, M. Sasikumar, S. Naveen, "Analysis of facial expressions from video images using PCA", in IEEE Conf, 2008.

[9] G. Murthy, R. Jadon, "Effectiveness of eigenspaces for facial expressions recognition", International Journal of Computer Theory and Engineering, Vol. 1, No. 5, 2009. https://doi.org/10.1002/j.1538-9235.1995.tb03988.x

[10] M. Abdullah, M. S. Wazzan, "Optimizing face recognition using PCA", International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 3, No. 2, pp. 23-31, 2012. https://doi.org/10.5121/ijaia.2012.3203 "Optimizing face recognition using PCA", arXiv:1206, Vol. 1515, 2012. https://doi.org/10.48550/arXiv.1206.1515

[11] M. Murtaza, M. Sharif, M. Raza, J. H. Shah, "Analysis of face recognition under varying facial expression: A survey", Int. Arab J. Inf. Technol., Vol. 10, No. 4, pp. 378-388, 2013. https://iajit.org/portal/PDF/vol.10,no.4/4172.pdf

[12] J. Kumari, R. Rajesh, K. Pooja, "Facial expression recognition: A survey", Procedia Computer Science, Vol. 58, pp. 486-491, 2015. https://doi.org/10.1016/j.procs.2015.08.011

[13] M. Parsapour, "Brain Emotional Learning-based Prediction Model for the Prediction of Geomagnetic Storms", Electrical Engineering and Systems Science, 2020. https://doi.org/10.48550/arXiv.2007.15579

[14] A. Mehrabian, R. Lucas, "Emotion-learning-based intelligent robust adaptive controller for stable uncertain nonlinear systems", International Journal of Computational Intelligence, Vol. 2, No. 4, pp. 246-252, 2005. https://doi.org/10.5281/zenodo.1330449

[15] Z. Beheshti, S. M. Hashim, "A review of emotional learning and its utilization in control engineering", Int. J. Adv. Soft Comput. Appl., Vol. 2, No. 2, pp. 191-208, 2010. https://www.researchgate.net/publication/50590462

[16] M. M. Taghi Zadeh, M. Imani, B. Majidi, "Fast facial emotion recognition using convolutional neural networks and Gabor filters", IEEE 5th Conference on Knowledge Based Engineering and Innovation (KBEI), pp. 577-581, 2019. https://doi.org/10.1109/KBEI.2019.8734943

[17] M. Khishe, M. Mosavi, "Chimp optimization algorithm", Expert Systems with Applications, Vol. 149, 2020. https://doi.org/10.1016/j.eswa.2020.113338

[18] M. Khishe, M. Nezhadshahbodaghi, M. R. Mosavi, D. Martín, "A Weighted Chimp Optimization Algorithm", IEEE Access, Vol. 9, pp. 158508-158539, 2021. https://doi.org/10.1109/ACCESS.2021.3130933

[19] A. Samsonovich, K. Jóhannsdóttir, A. Chella, "Ebook: Biologically Inspired Cognitive Architectures", Proceedings of the First Annual Meeting of the BICA Society, IOS Press, 2010.

[20] E. Askari, "Fear Recognition Using Early Biologically Inspired Features Model", Journal of Information Systems and Telecommunication (JIST), pp. 12-19, 2024. https://doi.org/10.61186/jist.39051.12.45.12

[21] R. Febrian, B. M. Halim, M. Christina, D. Ramdhan, A. Chowanda, "Facial expression recognition using bidirectional LSTM – CNN", in Procedia Computer Science, pp. 39-47, 2023. https://doi.org/10.1016/j.procs.2022.12.109

[22] F. An, Z. Liu. "Facial expression recognition algorithm based on parameter adaptive initialization of CNN and LSTM", Visual Computer, Vol. 36, pp. 483–98, 2020. https://doi.org/10.1007/s00371-019-01635-4

[23] S. Sheykhivand, Z. Mousavi, T. Yousefi Rezaii, S. Alaei, "Recognition of Emotion

Provoked by Auditory Stimuli using EEG Signal Based on Deep Neural Networks", Computational Intelligence in Electrical Engineering, Vol. 13, No. 1, pp. 1-18, 2022. https://doi.org/10.22108/ISEE.2020.122304.1357 [In Persian]

[24] E. Askari, S. L. Setarehdan, A. Sheikhani, M. R. Mohammadi, M. Teshnehlab, "Modeling the connections of brain regions in children with autism using cellular neural networks and electroencephalography analysis", Artificial Intelligence in Medicine, Vol. 89, pp. 40-50, 2018. https://doi.org/10.1016/j.artmed.2018.05.003

[25] M. Ramezani, S. Motamed, M. R. Yamaghani, "Face and Speech Emotion Recognition Using GSOAR Model", Computational Intelligence in Electrical Engineering, Vol. 16, No. 1, pp. 59-74, 2025. https://doi.org/10.22108/isee.2025.140659.1675 [In Persian]

[i] FER
[ii] PCA
[iii] BEL
[iv] ChOA
[v] EBIF
[vi] LSTM