IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 January 2024, accepted 22 January 2024, date of publication 29 January 2024, date of current version 15 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3359436

RESEARCH ARTICLE

Bioinspired Image Processing Enabled Facial Emotion Recognition Using Equilibrium Optimizer With a Hybrid Deep Learning Model

AHMAD A. ALZAHRANI
Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
e-mail: aalzahrani8@kau.edu.sa

This research work was funded by Institutional Fund Projects under grant no. (IFPDP-202-22). The authors gratefully acknowledge technical and financial support provided by the Ministry of Education and Deanship of Scientific Research (DSR), King Abdulaziz University (KAU), Jeddah, Saudi Arabia.

ABSTRACT Owing to the unpredictable nature of human facial expression, Facial emotion recognition (FER) from facial images becomes a tedious process. FER is a field within artificial intelligence (AI) and computer vision (CV) that concentrates on developing algorithms and technologies to interpret and analyze emotional expressions shown on human faces. The major intention of this area of research is to enable computers and machines to automatically classify and detect emotions, including sadness, happiness, anger, etc., based on facial expressions and cues. Deep learning (DL) systems namely recurrent neural network (RNN) and convolutional neural network (CNN) can be used for the task of automatically classifying and detecting emotions from human facial expressions. Furthermore, Bioinspired Image Processing Enabled FER combines principles from biological systems with image processing techniques to better the robustness and accuracy of FER. This technique is inspired by natural processes to enhance the interpretation and understanding of human emotion conveyed by facial expressions. This study designs a new Bioinspired Image Processing Enabled Facial Emotion Recognition using an Equilibrium Optimizer with Hybrid Deep Learning (BIPFER-EOHDL) model. The primary objectives of the BIPFER-EOHDL technique lie in the effectual and automated identification of facial expressions using a hyperparameter-tuned DL algorithm. In the presented BIPFER-EOHDL technique, the median filtering (MF) approach can be applied for image pre-processing. Besides, the BIPFER-EOHDL method applies the EfficientNetB7 model for the process of feature extraction. Meanwhile, the hyperparameter selection of the EfficientNetB7 model takes place by the use of the EO algorithm. Finally, the multi-head attention bi-directional long short-term memory (MA-BLSTM) model is exploited for the recognition and classification of facial emotions. A wide-ranging simulation analysis was performed to validate the higher FER outcomes of the BIPFER-EOHDL methodology. The simulation results stated that the BIPFER-EOHDL technique accomplishes better FER results than other recent approaches.

INDEX TERMS Non-verbal communication, facial emotion recognition, deep learning, equilibrium optimizer, artificial intelligence.

I. INTRODUCTION

Facial expression plays a vital part in non-verbal communication. Nonverbal signs are classified as facial expressions of a non-communicative environment [1]. It is usual and imitates emotions as well as many mental actions, physical signs, and social contacts. Facial expression recognition (FER) is commonly utilized in many applications such as elderly care, human–computer communication, security monitoring, customer satisfaction detection, the criminal justice

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

system, medical diagnostics, smart card applications as well as enlarged law enforcement services from smart cities [2]. In the current study, vision sensor-based FER has more attention and great potential in actual FER detection [3]. The researchers mainly concentrated on 7 basic expressions such as disgust, neutral, surprise, anger, fear, happiness, and sadness in vision-based FER. The FER is classified into dual sub-categories namely conventional and deep learning (DL) based techniques [4]. Emotions establish an essential feature of human behaviour that enhances the method of communication. Generally, humans show their characteristic situations in many ways like facial expressions and body language [5]. Facial expressions are a direct and significant channel of non-verbal communication that creates an entire emotional language that can immediately express a huge extent of human emotional conditions, feelings, and arrogances in numerous mental tasks [6]. The exact clarification and analysis of the expressive content of human facial expressions are vital for understanding human behaviour. Facial expressions are strong, and certainly well-known for human beings to interconnect emotions, understanding, and purposes to control connections and communication with other persons [7].

However, analysis of human facial features and detection of their emotional state are considered highly challenging as well as complex tasks [8]. The chief trouble comes from the non-uniform nature of the human face and many restrictions connected to facial pose, shadows, lighting, and orientation situations. Certainly, humans can discover and read faces as well as facial expressions without any effort. An exact and strong facial expression detection by computer methods is a high challenge [9]. DL techniques tested a stream of techniques to attain sturdiness and higher performance when evaluated to basic machine learning (ML) detection models like support vector machines (SVM) and multi-layer perceptron NN (MLP-NN). Then, humans function in a variety of circumstances so human behavior analysis requires being strong and DL approaches provide the required strength as well as scalability on novel kinds of data [10].

This study designs a new Bioinspired Image Processing Enabled Facial Emotion Recognition using an Equilibrium Optimizer with Hybrid Deep Learning (BIPFER-EOHDL) model. The primary objectives of the BIPFER-EOHDL method lie in the effectual and automated identification of facial expressions using a hyperparameter-tuned DL algorithm. In the presented BIPFER-EOHDL technique, the median filtering (MF) approach can be applied for image preprocessing. Besides, the BIPFER-EOHDL technique applies the EfficientNetB7 model for the feature extraction process. Meanwhile, the hyperparameter selection of the EfficientNetB7 model takes place by using the EO algorithm. Finally, the multi-head attention bi-directional long short-term memory (MA-BLSTM) model is exploited for the recognition and classification of facial emotions. To perform the higher FER outcomes of the BIPFER-EOHDL methodology, a wide-ranging experimental analysis has been performed. The proposed model accomplishes enhanced performance over other models with a maximum accuracy of 99.05%. The key contribution of the study is given below.

- An automated BIPFER-EOHDL technique comprising MF-based preprocessing, EfficientNetB7 feature extractor, EO-based hyperparameter tuning, and MA-BLSTM-based classification has been developed. To the best of our knowledge, the BIPFER-EOHDL technique never existed in the literature.
- Employ the EfficientNetB7 model, known for its effective consumption of computation resources, enabling robust feature extraction from facial images. This helps to capture subtle emotional cues while retaining computation efficacy.
- Applying the EO algorithm for hyperparameter tuning of the EfficientNetB7 model enables fine-tuning the network to the certain requirements of FER data. This contributes to the better effectiveness and accuracy of the model.
- The use of the MA-BLSTM model provides strong abilities for classifying and recognizing facial expressions. This architecture captures the contextual and temporal data within the image sequences, resulting in better detection performance.
- The BIPFER-EOHDL obtains a remarkable accuracy of 99.05%, which outperforms other techniques in facial emotion detection. This could have resulted in more precise and reliable emotional analysis in different applications.

II. LITERATURE REVIEW

Chakravarthi et al. [10] designed a hybrid DL approach (i.e., CNN-LSTM with ResNet-152 model) to classify emotions based on EEG signals. The activity in the brain seems to hold a specific characteristic which modified from one person to another, as well as from one emotional state to another emotional state. Mukhopadhyay et al. [11] presented a novel FER method utilizing local binary patterns (LBP), completed LBP (CLBP), and local ternary patterns (LTP). The CNNs method trained on images from prolonged datasets of JAFEE, Cohn-Kanade (CK+) and FER2013 changed into LTP, LBP, and CLBP image features. In [12], an effective FER was developed by employing a new DL Neural Network-regression activation (DR) classification algorithm. Primarily, the Gamma-HE model was used for pre-processing, and then facial facts were removed employing the Pyramid HOG (PHOG) based Supervised Descent (SMD) model. The facial portions are divided by utilizing the Viola-Jones Algorithm (VJA). Modified Monarch Butterfly Optimization (MMBO) technique utilized to pick essential features from extraction. Bentoumi et al. [13] projected a unique model for the recognition of emotion. The projected method depends on the suggestion of a pre-trained CNN approach (VGG16, ResNet50) with an MLP classifier. The pretrained approach of CNN has been employed as feature extraction. This approach adjusts unique design by totaling a

global average pooling layer (GAP) deprived of any finetuned network limits. Additionally, the study proposed an initial stopping standard.

In [14], a saliency map and DL-based FER model for facial images have been presented. The generative adversarial network (GAN), CLAHE, bilateral filter, and saliency map are employed for data preprocessing. A DCNN trained to utilize Nadam optimizer to identify facial emotion. Saurav et al. [15] project a new deep combined CNN method termed Emotion Network (EmNet). The EmNet technique contains dual physically parallel DCNN methods and their combined variant, jointly optimized employing a joint-optimization model. Sultana et al. [16] propose a CNN-built face emotion classification method. CNN is mainly employed for classification. Then, Pooling, Convolution, and Dropout layers effort on the pre-processed image to remove its features. An entirely associated layer with a classification algorithm is employed. The research employed various classifiers, optimizers, and DL models.

AlEisa et al. [17] project a Henry Gas Solubility Optimizer with DL-Based FER (HGSO-DLFER) model for HCI. Additionally, the MobileNet model is employed for feature vector generation, and the HGSO model selects its hyperparameter image. The HGSO-DLFER method employs an autoencoder (AE) classification algorithm for the detection of facial emotions with a Nadam optimizer. Kumari and Bhatia [18] developed an efficient DL-based technique. Primarily, the CLAHE was executed. Next, an adapted joint trilateral filter was performed to improve images for noise extraction. At last, an efficient DCNN was designed. In addition, the Adam optimizer has been exploited to enhance the cost function of DCNN.

III. THE PROPOSED MODEL

In this article, we have established a novel BIPFER-EOHDL algorithm. The main purpose of the BIPFER-EOHDL method lies in the effectual and automated identification of facial expressions using a hyperparameter-tuned DL model. It involves four main procedures such as MF-based preprocessing, EfficientNetB7-based feature extraction, EO-based parameter tuning, and MA-BLSTM-based classification. Fig. 1 depicts the working flow of the BIPFER-EOHDL system. The figure highlighted that the input images are initially preprocessed using the MF technique. Then, the EfficientNetB7 model is employed for the comprehensive and efficient extraction of discriminatory features from facial images, improving the model's capability to detect nuanced emotional expressions. Later, the EO model can be utilized for hyperparameter tuning of the EfficientNetB7 network. The usage of EO ensures optimum configuration, which maximizes the model's accuracy in facial emotion detection. Lastly, the MA-BLSTM model is used for the detection and classification of facial emotion. This sophisticated model considered context data and sequential dependency in facial expression.

A. PREPROCESSING

Primarily, the MF approach can be applied for image preprocessing. The MF algorithm for image preprocessing'' is a conventional approach used in image processing for reducing noise and enhancing the quality of image [19]. It works by substituting all the pixel values with the median value from the local neighborhood of the pixel. Particularly, this technique is effective at eliminating salt-and-pepper noise that is isolated pixels with remarkably low or high-intensity values. MF is a nonlinear filtering model and is frequently used before applying increasingly sophisticated image processing techniques to enhance the reliability and accuracy of outcomes. It is used in different applications, namely object recognition, image denoising, and feature extraction, where noise reduction is paramount for accurate interpretation and analysis of images.

B. FEATURE EXTRACTION

In this work, the BIPFER-EOHDL method applies the EfficientNetB7 model for the feature extraction process. The first stream, pretrained EfficientNet uses the knowledge sharing from the CNN-based pre-trained model on larger ImageNet data [20]. This excels in tasks such as image classification and object recognition owing to impressive performance and efficacy in feature extraction. The pre-trained model used to resolve the problem can be defined as transfer learning (TL). The TL model provides numerous benefits including enhancement of NN performance, the lessening of data requirement, and reduction of training time. CNN is commonly up-scaled to improve performance while implementing classification tasks on benchmark databases. However, the scaling procedure of the convolution model has been performed randomly. With regard to the width and depth of the network, some models are scaled. The scaling process demands a considerable time and needs manual control. On the other hand, EfficientNet applies a technique called ''compound coefficient'' for scaling models efficiently and straightforwardly. The major benefit of the EfficientNet model is its reliable scalability. EfficientNet uniformly scales each network dimension using a constant ratio through the compound coefficient method. The EfficientNetB7 architecture contains 33 layers and applies the compound coefficient model for adjusting to width, depth, and resolution of the model. This results in improved performance outcomes, but the computational cost remains comparatively lower. The MBConv (inverted bottleneck block), mentioned before in MobileNetV2, acts as the building block in EfficientNet. In these blocks, the application of Depthwise Separable Convolution (DWConv) is observed. The preliminary step includes expanding the channel through pointwise convolution (1×1 Conv). Then, a 3×3 DWConv is used for a considerably lower number of parameters. Lastly, a 1×1 Conv is utilized to reduce the channel count, which enables the integration of the preliminary and last phases. EfficientNet uses the Squeeze and Excitation (SE) block

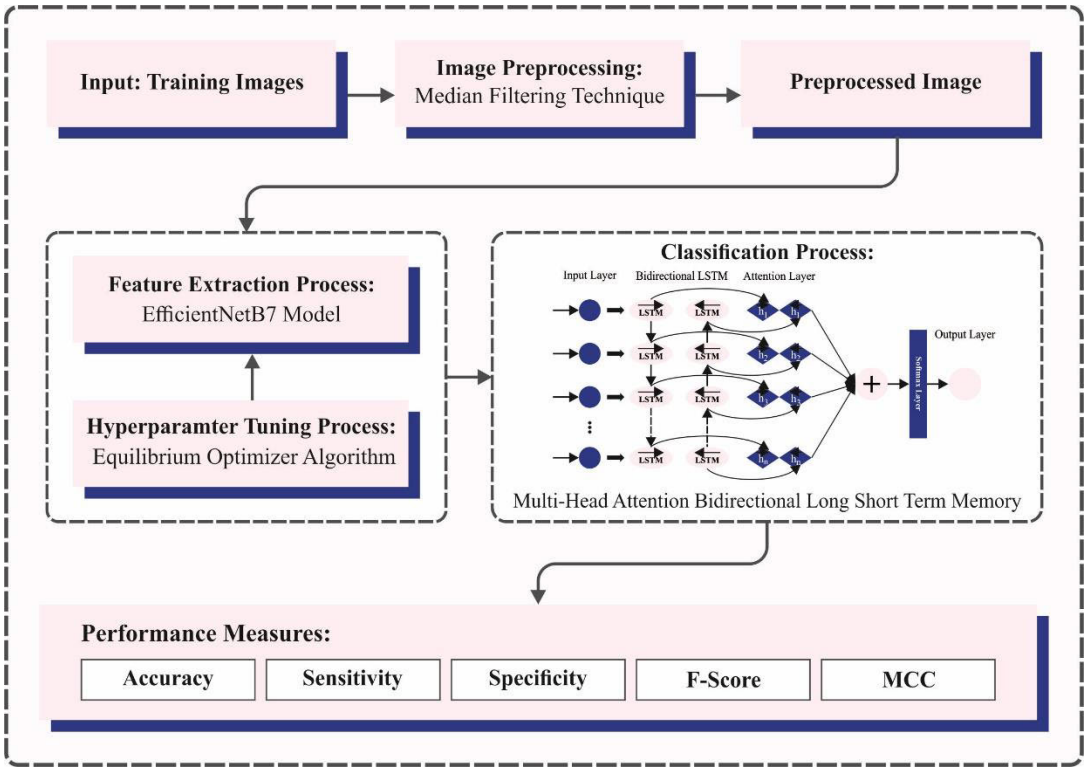

FIGURE 1. Workflow of BIPFER-EOHDL algorithm.

along with MBConv blocks, such that the network dynamically allocates a maximum weight to the central channels, thereby making the features effective for the FER task on hand.

C. PARAMETER TUNING USING THE EO MODEL

The EO model has been instrumental in the BIPFER-EOHDL method by serving as the hyperparameter tuning engine for the EfficientNetB7 model. EO, based on the natural equilibrium principle, dynamically adjusts the model's hyperparameter to strike an optimum balance. It enhances configurations, namely layer-specific parameters and learning rate, fostering a harmonious setting that improves the model's capability to capture nuanced features for FER. In the meantime, the hyperparameter selection of EfficientNetB7 is performed by applying the EO model. The EO algorithm is based on the formula of mass balance in a control volume from the physical principles, trying to find out the equilibrium state of the system [21]. Initialization, equilibrium pooling, and concentration update are the three stages of EO.

Step 1: Initialization. The position of all the particles is considered as a concentration of control volume (C), but a group of particles is arbitrarily produced amongst the boundaries as follows:

$$C_{i,j} = c_{\min,j} + r\left(c_{\max,j} - c_{\min,j}\right),\ i = 1,K,nj = 1,K,d \tag{1}$$

In Eq. (1), the location in *the*j^{th} dimension of the i^{th} particles are $C_{i,j}$, a random number between [1, 0] is r, the boundaries of all the particles at j^{th} dimension are $c_{\min,j}$ and $c_{\max,j}$, correspondingly. The fitness value is assessed and after the generation of the initialized particle, each particle is arranged to create an equilibrium pooling.

Step2: Equilibrium pooling and candidate. Many particles are needed to higher the population diversity to search for the last equilibrium state of the system. Thus, an equilibrium pool is created. Next, four particles are sorted with the optimum fitness value chosen as the candidate from the equilibrium pooling. In addition, the average location of the abovementioned 4 particles has been evaluated and concurrently stored from the equilibrium pooling. Based on the above mechanism, the five candidates are updated after each iteration as follows:

$$C_{eq,pool} = \left\{C_{eq(1)}, C_{eq(2)}, C_{eq(3)}, C_{eq(4)}, C_{eq(ave)}\right\} \tag{2}$$

In Eq. (2), the equilibrium pooling is $C_{eq,pool}, C_{eq(i)}$ $(i = 1, 2, 3, 4)$ are the 4 candidates with optimum fitness value, and the average location of four candidates is $C_{eq(ave)}$ that is formulated by Eq. (3):

$$C_{eq(ave)} = \frac{C_{eq(1)} + C_{eq(2)} + C_{eq(3)} + C_{eq(4)}}{4} \tag{3}$$

The candidates are randomly selected from the equilibrium pooling as optimum particles in the existing iteration. The candidate in equilibrium pooling has a similar possibility to be chosen, which provides better population diversity.

Step 3: Concentration update. Two essential terms must be considered for updating the concentration of particles,

generation rate (G) and exponential term (F). The F term is used for controlling the balance between exploitation and exploration, as follows:

$$F = a_1 sign\,(r_1 - 0.5)\left[\exp\left(-r_2 t_{EO}\right) - 1\right] \tag{4}$$

In Eq. (4), the random vectors within [1, 0] are $r1$ and $r2$, a constant $a1$ and $a2$ controls the exploration and exploitation ability, and the coefficient of EO is denoted as t_{EO} that can be updated at each iteration:

$$t_{EO} = (1 - T/M_{iter})^{(a_2 T/M_{iter})} \tag{5}$$

In Eq. (5), T and M_{iter} are the existing and the maximal iterations. If $a1$ is large, then exploration will be enhanced. Likewise, if $a2$ is large, then the exploitation of EO will be strengthened. $a1$ and $a2$ are fixed to 2 and 1, correspondingly.

In the EO algorithm, the generation rate (G) is also a crucial term for the concentration update process and is used to transmit accurate solutions with improving exploitation. This can be mathematically modelled as follows:

$$G = -P\left(C_{eq} - r_2 C_i\right) F, i = 1, K, n \tag{6}$$

$$P = \begin{cases} 0.5 r_{d1} \cdot u & r_{d2} \geq GP \\ 0 & r_{d2} < GP \end{cases} \tag{7}$$

Here a candidate selected in the equilibrium pool can be represented as C_{eq}, the location of i^{th} particles are C_i, the random integer within [1, 0] are $rd1$ and $rd2$,u is a unit vector, and GP refers to the generation possibility that affects exploitation and exploration that is equivalent to 0.5:

$$C_i^{new} = C_{eq} + \left(C_i - C_{eq}\right) F + \frac{(1 - F)\, G}{r_3 V} \tag{8}$$

In Eq. (8), the updated location of i^{th} particles is C_i^{new}, the random vector at (0, 1) interval is represented as $r3$,$andV$ is equivalent to 1. Later, check the boundary of the upgraded position and evaluate the fitness value, then apply the memory-saving model to obtain the best particles from the upgraded solution. The steps involved in EO are demonstrated in Fig. 2.

The EO method derives an FF to reach better efficacy of classification. It describes a positive integer to signify the high efficiency of the candidate result. The decrease in classifier error rate has been regarded as the FF.

$$\begin{aligned} fitness\,(x_i) &= Classifier\ Error\ Rate\,(x_i) \\ &= \frac{No.\ of\ misclassified\ samples}{Total\ No.\ of\ samples} * 100 \end{aligned} \tag{9}$$

D. CLASSIFICATION USING MA-BLSTM

Finally, the MA-BLSTM approach has been exploited for the classification and detection of facial emotions. Incorporating the attention model enables to focus on applicable facial features, which capture intricate spatial dependency essential for accurate emotion classification. The BLSTM improves the model's capability to grasp temporal relationships, which capture the dynamic evolution of emotion over time. Further, the multi-head attention model refines feature extraction by simultaneously taking various perspectives into account. This allows the model to discern subtle nuances in facial expression, which provides a context-aware and more nuanced approach to FER than classical architecture.

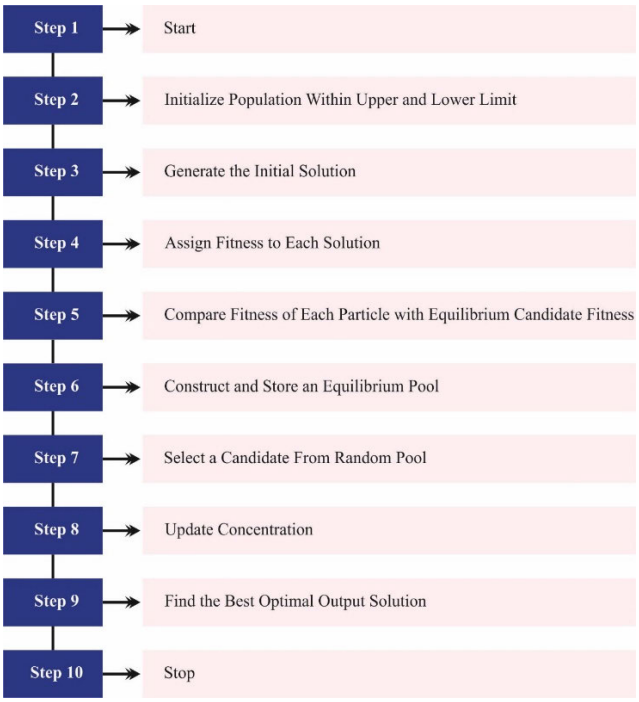

FIGURE 2. Steps involved in EO.

LSTM is an improved version of RNN to overcome challenges such as gradient exploding and vanishing by introducing memory states and multiple gate control functions [22]. LSTM is mainly composed of memory cells, forget, input, and output gates. Three gating modules are used for updating *thec* value of the memory unit. The forget gate f_t takes the data from the prior cell state c_{t-1} and existing c_t, passes over an activation function, and decides which data that discarded or retained. The input gate stores the x_t data in the present state c_t for updating the storage unit. The output gate o_t have the existing state c_t as an input to produce a novel h_t hidden layer to decide the values of subsequent cell states. Finally, LSTM contains two outputs (viz., h_t existing output value and c_t cell state) and three input values (viz., x_t existing input value, x_t resultant value of previous time, and c_{t-1} cell layer).

$$f_t = \sigma\left(W_f x_t + W_f h_{t-1} + b_f\right) \tag{10}$$

$$i_t = \sigma\left(W_i x_t + W_i h_{t-1} + b_i\right) \tag{11}$$

$$o_t = \sigma\left(W_o x_t + W_o h_{t-1} + b_o\right) \tag{12}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_c x_t + W_c h_{t-1} + b_c) \tag{13}$$

$$h_t = o_t \odot tanh\left(c_t\right) \tag{14}$$

Here $\odot$ shows the scalar product of 2 vectors, $\sigma(\cdot)$ refers to the sigmoid function, and W and b represent the weight matrix and bias, correspondingly.

The Bi-LSTM enhances upon LSTM and includes 2 single LSTMs with similar input but opposite directions of data communication. The benefits of Bi-LSTM are the capability to simultaneously extract features from sequential variables by taking the past and the present datasets which enhances classifier performance. The backward and forward LSTM have a similar input dataset but are calculated from opposite directions. Once the backward LSTM calculates from time t to 1, then the forward LSTM calculates from time 1 to t, enabling the latent output state of backward and forward LSTMs to be attained and saved in the memory cell to be outputted by the Bi-LSTM.

$$\vec{h}_t = \vec{LSTM}\left(x_t, h_{t-1}\right) \tag{15}$$

$$\overleftarrow{h}_t = \overleftarrow{LSTM}\left(x_t, h_{t+1}\right) \tag{16}$$

$$H_t = \left(\vec{h}_t, \overleftarrow{h}_t\right) \tag{17}$$

Now the hidden layer of Bi-LSTM at t he t time step is H_t and $LSTM(\cdot)$ represents the defined LSTM computation .

The attention module is a weighted sum of sequences that assigns computation resources to comparatively significant data that address the problems of data overload caused by the NN learning multiple parameters. The multi-head attention (MHA) module extracts essential data and decreases the loss of prior data compared to the classical attention module. The presented method has an MHA module that allocates dissimilar weights to the output of Bi-LSTM. The MHA module is positioned among the input and output layers and produces connection weight dynamically by applying the attention module to highlight the input part with the maximum effect on the output.

A series of outcomes in the Bi-LSTM represented as $X = [x_1, \cdots, x_n]$ is inputted to the MHA module. Linear conversion is used for obtaining three sets of vector series:

$$Q = XW_q \tag{18}$$

$$K = XW_k \tag{19}$$

$$V = XW_v \tag{20}$$

Now query vector, key vector, and value vector series are Q, K, and V correspondingly, and learnable parameter matrices are W_q, W_k, and W_v. Next, the correlation or similarity among the output as well as input features can be evaluated. The scaled dot-product module has been employed for obtaining the attention weight distribution:

$$ATT\left(Q, K, V\right) = soft\max\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{21}$$

In Eq. (21), the dimensionality of the input dataset is d. The output of the MHA mechanism is given by

$$H = Norm\left[X + ATT\left(Q, K, V\right)\right] \tag{22}$$

In Eq. (22), $Norm(\cdot)$ represents normalization and H denotes the output set of the attention module.

TABLE 1. Details on database.

Classes	No. of Instances
Anger	45
Contempt	18
Disgust	59
Fear	25
Happy	69
Neutral	593
Sad	28
Surprise	83
Total Instance	920

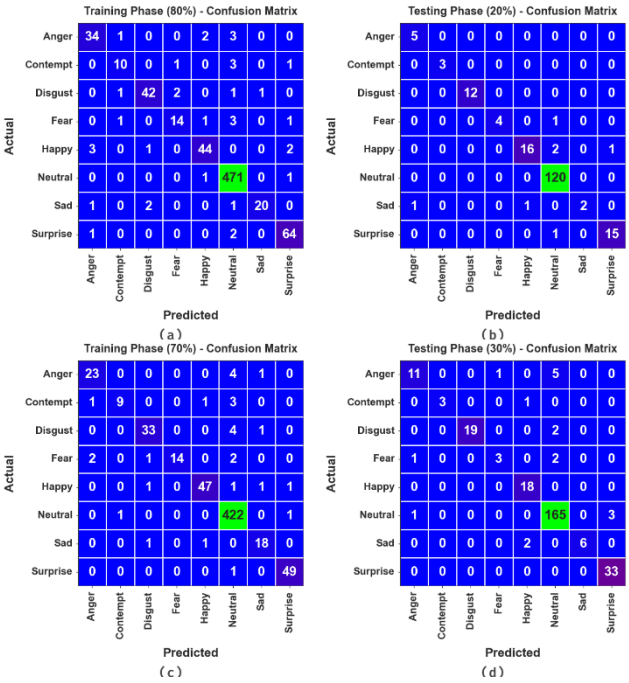

FIGURE 3. Confusion matrices of (a-c) TRPH of 80% and 70% and (b-d) TSPH of 20% and 30%.

IV. EXPERIMENTAL VALIDATION

The FER outcomes of the BIPFER-EOHDL algorithm are inspected on the Extended Cohn-Kanade (CK+) Database [23], which contains 920 instances with 8 classes as represented in Table 1.

The confusion matrices obtained by the BIPFER-EOHDL method under 80:20 and 70:30 of TR phase (TRPH)/TS phase (TSPH) are demonstrated in Fig. 3. The results indicate the effectual recognition and detection of all eight classes.

The FER outcomes of the BIPFER-EOHDL methodology are investigated under 80:20 of TRPH/TSPH is examined in Table 2 and Fig. 4. The outcomes showed that the BIPFER-EOHDL method reaches effectual performance under all classes.

TABLE 2. FER outcome of BIPFER-EOHDL system with 80:20 of TRPH/TSPH.

Classes	$Accu_y$	$Sens_y$	$Spec_y$	$F1_{Score}$	MCC
TR Phase (80%)					
Anger	98.51	85.00	99.28	86.08	85.29
Contempt	98.91	66.67	99.58	71.43	71.07
Disgust	98.91	89.36	99.56	91.30	90.75
Fear	98.78	70.00	99.58	75.68	75.31
Happy	98.64	88.00	99.42	89.80	89.09
Neutral	97.96	99.58	95.06	98.43	95.57
Sad	99.32	83.33	99.86	88.89	88.75
Surprise	98.91	95.52	99.25	94.12	93.53
Average	98.74	84.68	98.95	86.96	86.17
TS Phase (20%)					
Anger	99.46	100.00	99.44	90.91	91.03
Contempt	100.00	100.00	100.00	100.00	100.00
Disgust	100.00	100.00	100.00	100.00	100.00
Fear	99.46	80.00	100.00	88.89	89.19
Happy	97.83	84.21	99.39	88.89	87.85
Neutral	97.83	100.00	93.75	98.36	95.25
Sad	98.91	50.00	100.00	66.67	70.32
Surprise	98.91	93.75	99.40	93.75	93.15
Average	99.05	88.50	99.00	90.93	90.85

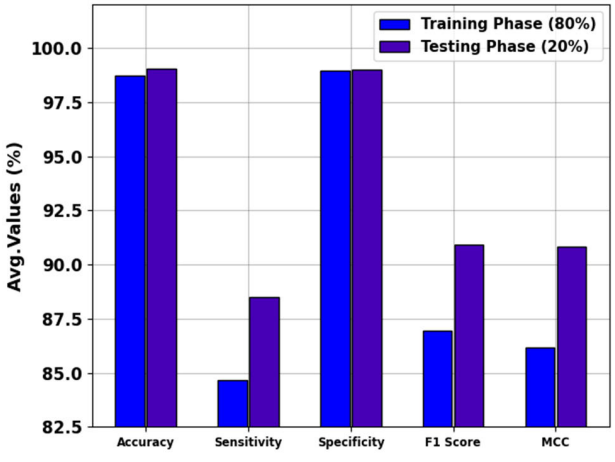

FIGURE 4. Average of BIPFER-EOHDL algorithm with 80:20 of TRPH/TSPH.

With 80% of TRPH, the BIPFER-EOHDL technique attains an average $accu_y$ of 98.74%, $sens_y$ of 84.68%, $spec_y$ of 98.95%, $F1_{score}$ of 86.96%, and MCC of 86.17%. Likewise, with 20% of the TSPH, the BIPFER-EOHDL method obtains an average $accu_y$ of 99.05%, $sens_y$ of 88.50%, $spec_y$ of 99%, $F1_{score}$ of 90.93%, and MCC of 90.85%.

The FER result of the BIPFER-EOHDL system is investigated under 70:30 of the TRPH/TSPH is examined in Table 3 and Fig. 5. The outcomes inferred that the BIPFER-EOHDL system attains effective outcomes under all classes. With 70% of TRPH, the BIPFER-EOHDL algorithm reaches an average $accu_y$ of 98.87%, $sens_y$ of 85.83%, $spec_y$ of 98.86%, $F1_{score}$ of 88.89%, and MCC of 88.27%. On the other hand, with 30% of the TSPH, the BIPFER-EOHDL methodology achieves an average $accu_y$ of 98.37%, $sens_y$ of 81.60%, $aspec_y$ of 98.51%, a $F1_{score}$ of 85.49%, and MCC of 84.73%.

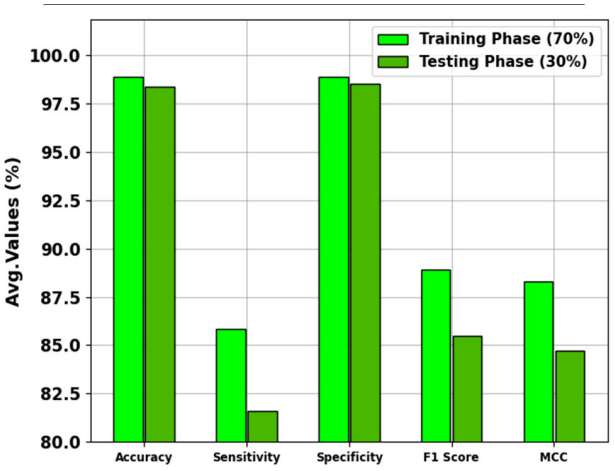

FIGURE 5. Average of BIPFER-EOHDL algorithm with 70:30 of TRPH/TSPH.

TABLE 3. FER outcome of BIPFER-EOHDL method with 70:30 of TRPH/TSPH.

Classes	$Accu_y$	$Sens_y$	$Spec_y$	$F1_{Score}$	MCC
TR Phase (70%)					
Anger	98.76	82.14	99.51	85.19	84.60
Contempt	99.07	64.29	99.84	75.00	75.64
Disgust	98.76	86.84	99.50	89.19	88.57
Fear	99.22	73.68	100.00	84.85	85.50
Happy	99.07	92.16	99.66	94.00	93.52
Neutral	97.36	99.53	93.18	98.03	94.14
Sad	99.22	90.00	99.52	87.80	87.43
Surprise	99.53	98.00	99.66	97.03	96.78
Average	98.87	85.83	98.86	88.89	88.27
TS Phase (30%)					
Anger	97.10	64.71	99.23	73.33	72.55
Contempt	99.64	75.00	100.00	85.71	86.44
Disgust	99.28	90.48	100.00	95.00	94.75
Fear	98.55	50.00	99.63	60.00	60.56
Happy	98.91	100.00	98.84	92.31	92.04
Neutral	95.29	97.63	91.59	96.21	90.06
Sad	99.28	75.00	100.00	85.71	86.28
Surprise	98.91	100.00	98.77	95.65	95.15
Average	98.37	81.60	98.51	85.49	84.73

To compute the efficiency of the BIPFER-EOHDL approach with 80:20 of TRPH/TSPH, we have created $accu_y$ curves for TRA and TSPHs, as illustrated in Fig. 6. These curves provide valued insights as the method learning growth and its ability to generalization. While it increases the epoch counts, an obvious improvement in both TRA and TES $accu_y$ curves becomes evident. This improvement shows the model's capacity to further define patterns from both the TRA and TES data.

Fig. 7 also represents an outline of the BIPFER-EOHDL algorithm with 80:20 of TRPH/TSPH, the loss outcomes

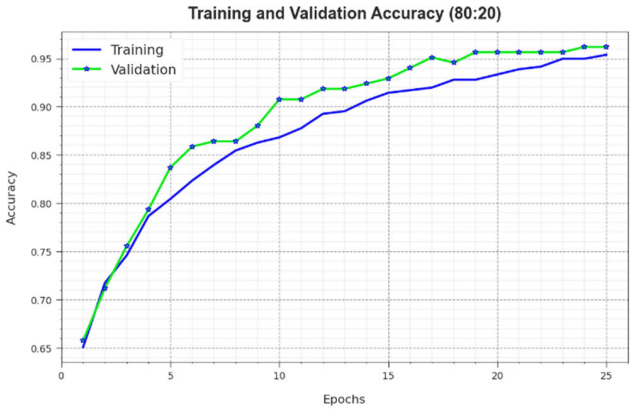

FIGURE 6. **Accu$_y$ curve of BIPFER-EOHDL algorithm at 80:20 of TRPH/TSPH.**

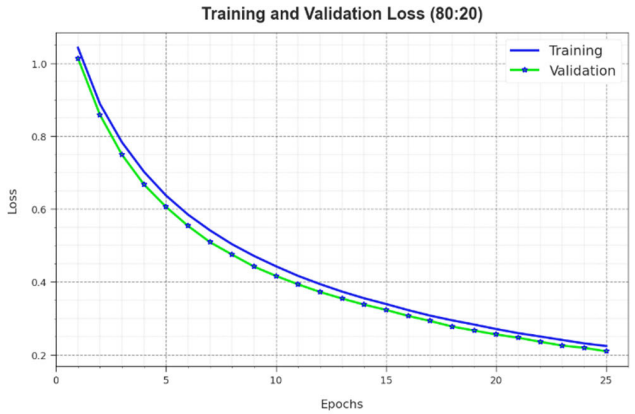

FIGURE 7. **Loss curve of BIPFER-EOHDL method at 80:20 of TRPH/TSPH.**

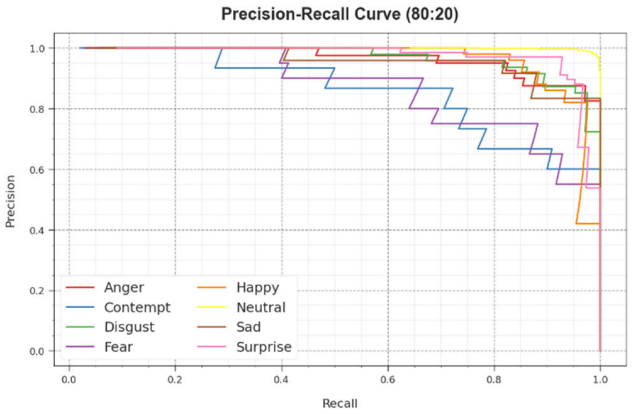

FIGURE 8. **PR curve of BIPFER-EOHDL technique with 80:20 of TRPH/TSPH.**

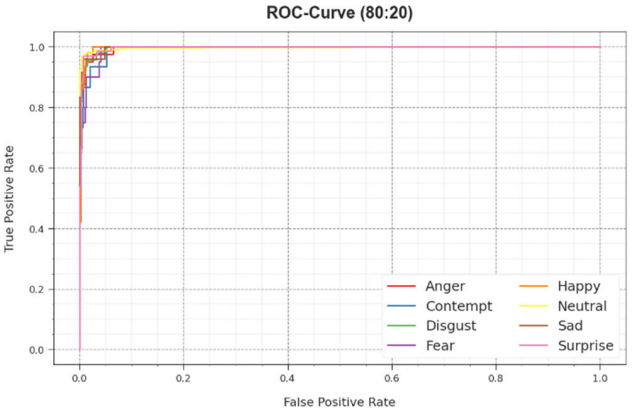

FIGURE 9. **ROC curve of BIPFER-EOHDL method with 80:20 of TRPH/TSPH.**

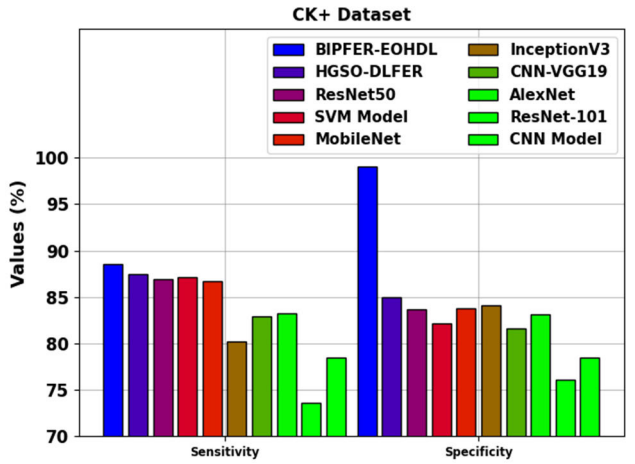

FIGURE 10. **Sens$_y$ and spec$_y$ outcome of the BIPFER-EOHDL approach under CK+ dataset.**

throughout the TRA method. The decreasing tendency in TRA loss with epochs shows that the model constantly refines its weight to minimize prediction errors on both TRA and TES data. This loss curve reflects how the model fits the TR data well. Especially, the TRA and TES losses constantly decrease, demonstrating the model's effective learning of designs present in both datasets. Furthermore, it indicates the model adaptation in decreasing inconsistencies between the original TRA and prediction classes.

The PR outcome of the BIPFER-EOHDL method with 80:20 of TRPH/TSPH, the precision of the plot against recall as defined in Fig. 8, revealing that our method achieves better precision-recall values across all classes. This graph depicts the model's capability to detect several classes, particularly excelling in correctly identifying positive samples while decreasing false positives.

Fig. 9 also contains ROC curves of the BIPFER-EOHDL algorithm with 80:20 of TRPH/TSPH, which represents the model's capability to discriminate among classes. These curves offer valuable insights into the tradeoff between true positive rate (TPR) and false positive rate (FPR) across distinct classifier thresholds and epochs. They demonstrate the model's accurate predictive outcome across various classes, further emphasizing its classification capabilities.

A comparative study of the BIPFER-EOHDL technique under the CK+ dataset is provided in Table 4 [17], [24]. Fig. 10 portrays a comparison study of the BIPFER-EOHDL

TABLE 4. Comparative outcomes of the BIPFER-EOHDL model with other approaches on CK+ dataset [17], [24].

Methods	$Sens_y$	$Spec_y$	$Accu_y$	$F1_{score}$
BIPFER-EOHDL	88.50	99.00	99.05	90.93
HGSO-DLFER	87.45	84.99	98.65	87.78
ResNet50	86.96	83.65	88.54	85.99
SVM Model	87.17	82.18	91.64	87.55
MobileNet	86.74	83.81	92.32	86.52
InceptionV3	80.23	84.06	93.74	73.82
CNN-VGG19	82.95	81.59	94.03	81.75
AlexNet	83.25	83.10	95.88	82.50
ResNet-101	73.65	76.11	93.89	74.29
CNN Model	78.51	78.53	94.76	78.08

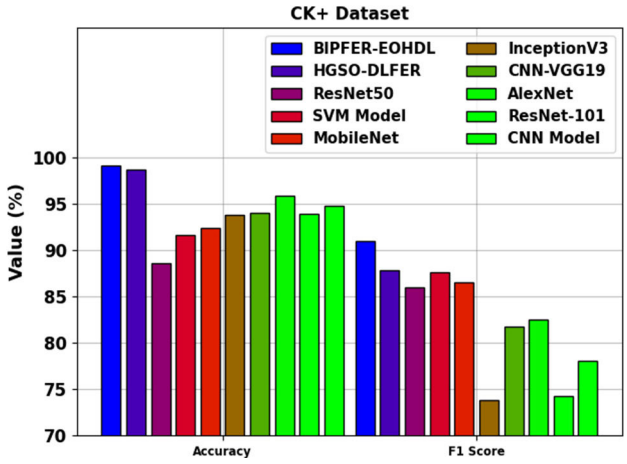

FIGURE 11. $Accu_y$ and $F1_{score}$ outcome of BIPFER-EOHDL approach under CK+ dataset.

technique in terms of $sens_y$ and $spec_y$. The figure indicates that the BIPFER-EOHDL technique exhibits higher outcomes than other approaches. Based on $sens_y$, the BIPFER-EOHDL technique offers a higher $sens_y$ of 88.50% while the HGSO-DLFER, ResNet50, SVM, MobileNet, InceptionV3, CNN-VGG19, AlexNet, ResNet-101, and CNN models have obtained lower $sens_y$ values of 87.45%, 86.96%, 87.17%, 86.74%, 80.23%, 82.95%, 83.25%, 73.65%, and 78.51%, correspondingly. Followed by, with respect to $spec_y$, the BIPFER-EOHDL system attained a superior $spec_y$ of 99% while the HGSO-DLFER, ResNet50, SVM, MobileNet, InceptionV3, CNN-VGG19, AlexNet, ResNet-101, and CNN techniques are gaining lesser $spec_y$ values of 84.99%, 83.65%, 82.18%, 83.81%, 84.06%, 81.59%, 83.10%, 76.11%, and 78.53%, correspondingly.

Fig. 11 depicts the comparative outcome of the BIPFER-EOHDL algorithm interms of $accu_y$ and $F1_{score}$. The outcome implies that the BIPFER-EOHDL algorithm depicts a higher outcome than other methods. With respect to $accu_y$, the BIPFER-EOHDL methodology provides enhanced $accu_y$ of 99.05% whereas the HGSO-DLFER, ResNet50, SVM, MobileNet, InceptionV3, CNN-VGG19, AlexNet, ResNet-101, and CNN methods have obtained lesser $accu_y$ values of 98.65%, 88.54%, 91.64%, 92.32%, 93.74%, 94.03%, 95.88%, 93.89%, and 94.76%, correspondingly. Besides, based on $theF1_{score}$, the BIPFER-EOHDL methodology offers a maximum $F1_{score}$ of 90.93% whereas the HGSO-DLFER, ResNet50, SVM, MobileNet, InceptionV3, CNN-VGG19, AlexNet, ResNet-101, and CNN models have reached minimum $F1_{score}$ values of 87.78%, 85.99%, 87.55%, 86.52%, 73.82%, 81.75%, 82.50%, 74.29%, and 78.08% correspondingly.

TABLE 5. Comparative outcomes of the BIPFER-EOHDL model with other approaches on the FER-2013 dataset.

Methods	$Sens_y$	$Spec_y$	$Accu_y$	$F1_{score}$
BIPFER-EOHDL	83.42	83.48	88.50	83.50
AlexNet	61.76	64.34	65.30	62.03
ResNet-101	60.11	62.99	63.43	60.52
CNN Model	61.59	64.25	64.80	61.69
CNN-3 Model	81.63	81.79	80.79	81.73
Logistic Regression	51.66	49.65	51.55	49.66
Random Forest	63.50	64.51	62.51	60.54
SVM Model	62.52	63.58	62.58	59.53
Decision Tree	51.79	51.64	51.52	51.71
Naïve Bayes	40.57	45.59	40.55	40.51

A comparative analysis of the BIPFER-EOHDL methodology under the FER-2013 database is provided in Table 5. Fig. 12 depicts a comparative outcome of the BIPFER-EOHDL system with respect to $sens_y$ and $spec_y$. The outcome implies that the BIPFER-EOHDL algorithm outperforms superior performances than other methods. Based on $sens_y$, the BIPFER-EOHDL methodology gains increase $sens_y$ of 83.42% while the AlexNet, ResNet-101, CNN, CNN-3, LR, RF, SVM, DT, and NB approaches have acquired lesser $sens_y$ values of 61.76%, 60.11%, 61.59%, 81.63%, 51.66%, 63.50%, 62.52%, 51.79%, and 40.57%, correspondingly. Afterwards, with respect to $spec_y$, the BIPFER-EOHDL methodology achieved a higher $spec_y$ of 83.48% while the AlexNet, ResNet-101, CNN, CNN-3, LR, RF, SVM, DT, and NB algorithms reached minimal $spec_y$ values of 64.34%, 62.99%, 64.25%, 81.79%, 49.65%, 64.51%, 63.58%, 51.64%, and 45.59%, correspondingly.

Fig. 13 portrays the comparative outcome of the BIPFER-EOHDL methodology in terms of $accu_y$ and $F1_{score}$. The outcome implies that the BIPFER-EOHDL model outperforms a higher outcome than other methodologies. With respect to $accu_y$, the BIPFER-EOHDL methodology provides an improved $accu_y$ of 88.50 % whereas the AlexNet, ResNet-101, CNN, CNN-3, LR, RF, SVM, DT, and NB methods have achieved lesser $accu_y$ values of 65.30%, 63.43%, 64.80%, 80.79%, 51.55%, 62.51%, 62.58%, 51.52%, and 40.55%, correspondingly. In addition, based on $theF1_{score}$, the BIPFER-EOHDL method offers a maximal $F1_{score}$ of 83.50% whereas the AlexNet, ResNet-101, CNN, CNN-3, LR, RF, SVM, DT,

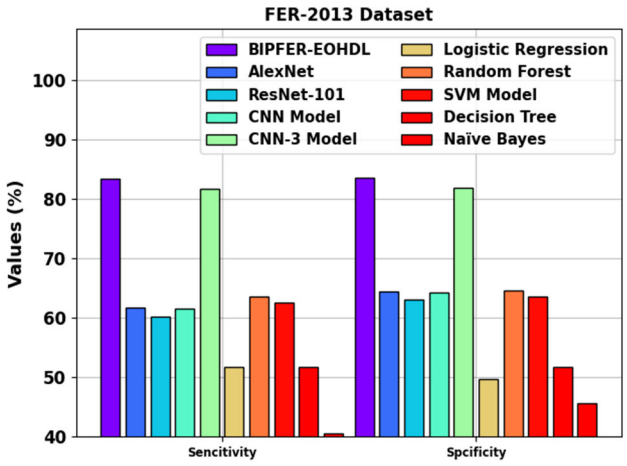

FIGURE 12. **Sens$_y$ and spec$_y$ outcome of the BIPFER-EOHDL approach under the FER-2013 dataset.**

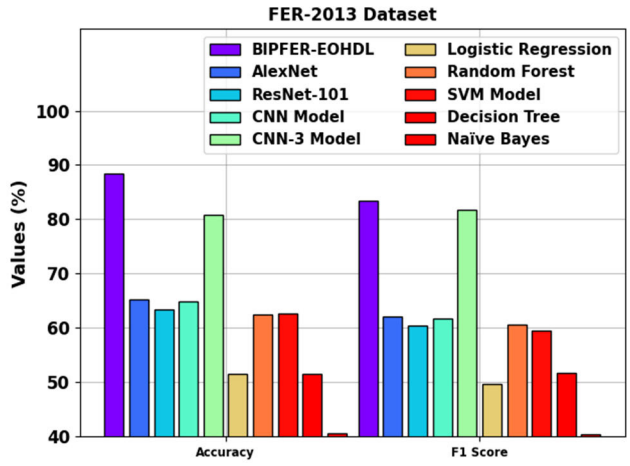

FIGURE 13. **Accu$_y$ and F1$_{score}$ outcome of BIPFER-EOHDL approach under FER-2013 dataset.**

and NB algorithms have reached lower $F1_{score}$ values of 62.03%, 60.52%, 61.69%, 81.73%, 49.66%, 60.54%, 59.53%, 51.71%, and 40.51% correspondingly.

In summary, the BIPFER-EOHDL technique exhibits improved performance over other techniques on the FER method.

V. CONCLUSION

In this article, we have developed a new BIPFER-EOHDL model. The main purpose of the BIPFER-EOHDL method lies in the effectual and automated identification of facial expressions using a hyperparameter-tuned DL model. It involves four main procedures such as MF-based preprocessing, EfficientNetB7-based feature extraction, EO-based parameter tuning, and MA-BLSTM-based classification. Besides, the hyperparameter selection of the EfficientNetB7 model takes place by the use of the EO algorithm. Finally, the MA-BLSTM approach has been exploited for the classification and recognition of facial emotions. To perform the higher FER outcomes of the BIPFER-EOHDL methodology; a wide-ranging experimental analysis has been performed. The simulation results stated that the BIPFER-EOHDL technique accomplishes better FER results than other recent approaches.

The accuracy and efficiency of the BIPFER-EOHDL model make it a valuable tool across sectors, which addresses the nuanced application and contributes to advancement in healthcare, security, and human-computer interaction. In security, the model is used for the analysis of real-time emotion in surveillance systems, which enhance threat recognition by finding emotional states or suspicious behaviors. In healthcare, BIPFER-EOHDL contributes to mental health assessment by monitoring the well-being of patients' emotions through facial expression, which helps to diagnose and treatment of conditions including anxiety or depression. Moreover, in human-computer interaction, the model's capability to discern user emotion could improve the adaptive interface, tailoring responses based on emotional cues for a more empathetic and personalized user experience. Future work can focus on the inspection of the proposed technique on large-scale real-time datasets.

REFERENCES

[1] N. Kumari and R. Bhatia, "Deep learning-based efficient emotion recognition technique for facial images," *Int. J. Syst. Assurance Eng. Manag.*, vol. 14, pp. 1–16, May 2023.

[2] A. F. Yaseen, A. Shaukat, and M. Alam, "Emotion recognition from facial images using hybrid deep learning models," in *Proc. 2nd Int. Conf. Digit. Futures Transformative Technol. (ICoDT)*, May 2022, pp. 1–7.

[3] R. Gill and J. Singh, "A deep learning approach for real time facial emotion recognition," in *Proc. 10th Int. Conf. Syst. Model. Advancement Res. Trends (SMART)*, Dec. 2021, pp. 497–501.

[4] R. A. Mansouri and M. Ragab, "Equilibrium optimization algorithm with ensemble learning based cervical precancerous lesion classification model," *Healthcare*, vol. 11, no. 1, p. 55, Dec. 2022.

[5] S. K. Singh, R. K. Thakur, S. Kumar, and R. Anand, "Deep learning and machine learning based facial emotion detection using CNN," in *Proc. 9th Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2022, pp. 530–535.

[6] I. Agrawal, A. Kumar, D. Swathi, V. Yashwanthi, and R. Hegde, "Emotion recognition from facial expression using CNN," in *Proc. IEEE 9th Region Humanitarian Technol. Conf. (R-HTC)*, Sep. 2021, pp. 01–06.

[7] S. Hangaragi, T. Singh, and N. Neelima, "Face detection and recognition using face mesh and deep neural network," *Proc. Comput. Sci.*, vol. 218, pp. 741–749, Jan. 2023.

[8] W. Dias, F. Andaló, R. Padilha, G. Bertocco, W. Almeida, P. Costa, and A. Rocha, "Cross-dataset emotion recognition from facial expressions through convolutional neural networks," *J. Vis. Commun. Image Represent.*, vol. 82, Jan. 2022, Art. no. 103395.

[9] A. S. Al-Ghamdi, S. M. Alshammari, and M. Ragab, "Deep learning based face mask detection in religious mass gathering during COVID-19 pandemic," *Comput. Syst. Sci. Eng.*, vol. 46, no. 2, pp. 1863–1877, 2023.

[10] B. Chakravarthi, S.-C. Ng, M. R. Ezilarasan, and M.-F. Leung, "EEG-based emotion recognition using hybrid CNN and LSTM classification," *Frontiers Comput. Neurosci.*, vol. 16, Oct. 2022, Art. no. 1019776.

[11] M. Mukhopadhyay, A. Dey, and S. Kahali, "A deep-learning-based facial expression recognition method using textural features," *Neural Comput. Appl.*, vol. 35, no. 9, pp. 6499–6514, Mar. 2023.

[12] A. Suputri Devi D and S. Ch, "An efficient facial emotion recognition system using novel deep learning neural network-regression activation classifier," *Multimedia Tools Appl.*, vol. 80, no. 12, pp. 17543–17568, May 2021.

[13] M. Bentoumi, M. Daoud, M. Benaouali, and A. Taleb Ahmed, "Improvement of emotion recognition from facial images using deep learning and early stopping cross validation," *Multimedia Tools Appl.*, vol. 81, no. 21, pp. 29887–29917, Sep. 2022.
[14] N. Kumari and R. Bhatia, "Saliency map and deep learning based efficient facial emotion recognition technique for facial images," *Multimedia Tools Appl.*, pp. 1–24, Jul. 2023, doi: 10.1007/s11042-023-16220-0.
[15] S. Saurav, R. Saini, and S. Singh, "EmNet: A deep integrated convolutional neural network for facial emotion recognition in the wild," *Int. J. Speech Technol.*, vol. 51, no. 8, pp. 5543–5570, Aug. 2021.
[16] S. Sultana, R. Mustafa, and M. S. Chowdhury, "Human emotion recognition from facial images using convolutional neural network," in *Proc. Int. Conf. Mach. Intell. Emerg. Technol.* Cham, Switzerland: Springer, 2022, pp. 106–120.
[17] H. N. AlEisa, F. Alrowais, N. Negm, N. Almalki, M. Khalid, R. Marzouk, M. M. Alnfiai, G. P. Mohammed, and A. A. Alneil, "Henry gas solubility optimization with deep learning based facial emotion recognition for human computer interface," *IEEE Access*, vol. 11, pp. 62233–62241, 2023.
[18] N. Kumari and R. Bhatia, "Efficient facial emotion recognition model using deep convolutional neural network and modified joint trilateral filter," *Soft Comput.*, vol. 26, no. 16, pp. 7817–7830, Aug. 2022.
[19] A. Shah, J. I. Bangash, A. W. Khan, I. Ahmed, A. Khan, A. Khan, and A. Khan, "Comparative analysis of median filter and its variants for removal of impulse noise from grayscale images," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 3, pp. 505–519, 2022.
[20] A. M. Islam, F. B. Masud, M. R. Ahmed, A. I. Jafar, J. R. Ullah, S. Islam, S. Shatabda, and A. K. M. M. Islam, "An attention-guided deep-learning-based network with Bayesian optimization for forest fire classification and localization," *Forests*, vol. 14, no. 10, p. 2080, Oct. 2023.
[21] C. Zhong, G. Li, Z. Meng, H. Li, and W. He, "A self-adaptive quantum equilibrium optimizer with artificial bee colony for feature selection," *Comput. Biol. Med.*, vol. 153, Feb. 2023, Art. no. 106520.
[22] J.-H. Li, X.-Y. Gao, X. Lu, and G.-D. Liu, "Multi-head attention-based hybrid deep neural network for aeroengine risk assessment," *IEEE Access*, vol. 11, pp. 113376–113389, 2023.
[23] *Extended Cohn-Kanade (CK+) Database*. Accessed: Nov. 14, 2023. [Online]. Available: http://www.jeffcohn.net/Resources/
[24] M. K. Chowdary, T. N. Nguyen, and D. J. Hemanth, "Deep learning-based facial emotion recognition for human-computer interaction applications," *Neural Comput. Appl.*, vol. 35, pp. 1–18, Apr. 2021.

AHMAD A. ALZAHRANI received the Ph.D. degree in computer science from La Trobe University, Australia, in 2014. He is currently an Associate Professor with the Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include pervasive computing and human–computer interaction.