IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 March 2023, accepted 30 March 2023, date of publication 3 April 2023, date of current version 12 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3264260

RESEARCH ARTICLE

Modified Earthworm Optimization With Deep Learning Assisted Emotion Recognition for Human Computer Interface

FADWA ALROWAIS[1], NOHA NEGM[2], MAJDI KHALID[3], NABIL ALMALKI[4], RADWA Marzouk[5], ABDULLAH MOHAMED[6], MESFER AL DUHAYYIM[7], AND AMANI A ALNEIL[8]

[1]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[2]Department of Computer Science, College of Science and Art at Mahayil, King Khalid University, Riyadh 12372, Saudi Arabia
[3]Department of Computer Science, College of Computer and Information Systems, Umm Al-Qura University, Makkah 21955, Saudi Arabia
[4]Department of Special Education, College of Education, King Saud University, Riyadh 12372, Saudi Arabia
[5]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[6]Research Centre, Future University in Egypt, New Cairo 11845, Egypt
[7]Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia
[8]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia

Corresponding author: Mesfer Al Duhayyim (m.alduhayyim@psau.edu.sa)

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups Project under grant number (RGP2/96/44). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R77), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors extend their appreciation to the researchers supporting project number (RSPD2023R521), King Saud University, Riyadh, Saudi Arabia. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

ABSTRACT Among the most prominent field in the human-computer interface (HCI) is emotion recognition using facial expressions. Posed variations, facial accessories, and non-uniform illuminations are some of the difficulties in the emotion recognition field. Emotion detection with the help of traditional methods has the shortcoming of mutual optimization of feature extraction and classification. Computer vision (CV) technology improves HCI by visualizing the natural world in a digital platform like the human brain. In CV technique, advances in machine learning and artificial intelligence result in further enhancements and changes, which ensures an improved and more stable visualization. This study develops a new Modified Earthworm Optimization with Deep Learning Assisted Emotion Recognition (MEWODL-ER) for HCI applications. The presented MEWODL-ER technique intends to categorize different kinds of emotions that exist in the HCI applications. To do so, the presented MEWODL-ER technique employs the GoogleNet model to extract feature vectors and the hyperparameter tuning process is performed via the MEWO algorithm. The design of automated hyperparameter adjustment using the MEWO algorithm helps in attaining an improved emotion recognition process. Finally, the quantum autoencoder (QAE) model is implemented for the identification and classification of emotions related to the HCI applications. To exhibit the enhanced recognition results of the MEWODL-ER approach, a wide-ranging simulation analysis is performed. The experimental values indicated that the MEWODL-ER technique accomplishes promising performance over other models with maximum accuracy of 98.91%.

INDEX TERMS Human-computer interaction, artificial intelligence, deep learning, emotion recognition, earthworm optimization algorithm.

I. INTRODUCTION

Since people have become more informed, they need a high level of computer intelligence [1], [2]. Human-computer interaction (HCI) is not limited to original hardware-related communication. Certain smarter communication techniques are appearing gradually in the life of people like a sequence of more intellectual techniques relevant to voice recognition [3], face recognition, and gesture recognition. Intellectual mechanisms can help establish interactions between computers and

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

humans [4], [5]. The advent of more convenient communication techniques is becoming a major advancement trend in the current domain of HCI. The objective of HCI development was naturally to make computers adapt and serve the requirements of individuals [6].

People-centred instead of compelling persons to adapt to computers. Emotions had a main role during the interaction. Detection of facial emotions will be helpful in several tasks like social robots, criminal justice systems, security monitoring, customer satisfaction identification, smart card applications, e-learning, etc [7], [8]. The core blocks in the conventional emotion recognition mechanism were classifying the emotions, detecting faces, and extracting the features [9].

The growth of machine learning (ML), particularly deep learning (DL), poses novel difficulties for human-centred models. DL presently changes work in numerous domains, like computer vision (CV), brain-machine interfaces, and natural language processing (NLP) [10], [11]. Such disciplines went from handcrafted techniques to data-driven methods to build new mechanisms. In classy iterative processes, ML methods are fine-tuned and trained, which can be possible just because the assessment was cheaper [12]. Contrary to HCI, these techniques do not undergo human-centred methodologies since the advanced solution will be valued through simple metrics. HCI necessitates users to govern the solution quality, which had proved to be affluent since this means conducting a user study [13], [14]. While utilizing ML, refining a solution need to train a novel method, because data-driven methods are unable to change the way that handcrafted ones can [15]. To Train, an ML method there comes a need for data, which HCI generally means that the design-solution step of UCDs unexpectedly necessitates studies for the collection of data.

Though several models are available in the literature, it is still needed to improve the recognition performance for HCI applications. Owing to a continual deepening of the model, the number of parameters of DL models also increases quickly which results in model overfitting. At the same time, different hyperparameters have a significant impact on the efficiency of the CNN model. Particularly, hyperparameters such as epoch count, batch size, and learning rate selection are essential to attain effectual outcomes. Since the trial and error method for hyperparameter tuning is a tedious and erroneous process, metaheuristic algorithms can be applied.

This study develops a new Modified Earthworm Optimization with Deep Learning Assisted Emotion Recognition (MEWODL-ER) for HCI applications. The presented MEWODL-ER technique intends to categorize different kinds of emotions that exist in the HCI applications. To do so, the presented MEWODL-ER technique employs the GoogleNet model to extract feature vectors and the hyperparameter tuning process is performed via the MEWO algorithm. The design of automated hyperparameter adjustment using the MEWO algorithm helps in attaining an improved emotion recognition process and shows the novelty of the work. Finally, the quantum autoencoder (QAE) model is implemented for the identification and classification of emotions related to the HCI applications. To reveal the enhanced recognition results of the MEWODL-ER algorithm, a wide-ranging simulation analysis is performed.

II. RELATED WORKS

Jain et al. [16] examined a Hybrid CNN and RNN approach for FER in images. The projected network infrastructure comprises Convolutional layers then RNN that integrated approach extracting the connection in facial images and with the recurrent network, the temporal dependencies that occur in an image were assumed that the classifier. Li et al. [17] examined a new emotion detection infrastructure to fuse emotional features in brain EEG signals and the equivalent audio signal in emotion detection on the DEAP database. The authors utilized CNN for extracting EEG features and BiLSTM-NNs for extracting audio features. Then, the authors integrate the multi-modal features as DL infrastructure for recognizing arousal and valence levels.

Tzirakis et al. [18] introduce a novel approach for continuous emotion detection in speech. This method which has trained end-to-end was contained CNN that extracts features in the raw signal, and stacked on top of 2-layer LSTM, which considers the context data in the information. The authors in [19], an innovative EEG-oriented emotion recognition technology has been introduced. The usage of 3D Convolutional Neural Networks (3D-CNN) is studied with the help of multichannel EEG data for recognizing emotion. Then, a 3D data representative was devised from the multichannel EEG signal that is exploited as input datasets for the presented architecture.

The author in [20], proposed a lightweight A-MobileNet framework. In the initial phase, the attention mechanism has been developed into the MobileNetV1 framework for enhancing the local feature extraction of facial expressions. Next, the integration of softmax loss and centre loss is used for optimizing the model parameter for increasing inter-class distance and decreasing intra-class distance. Sahoo et al. [21] advise the TLEFuzzyNet architecture, a 3-phase pipeline for emotion detection. The former involves feature extraction of Mel spectrograms and extraction by data augmentations of the speech signal, afterwards the usage of three pre-trained TL-CNN modules such as GoogleNet, ResNet18, and Inception_v3 whose predictive score was fed into the third phase. Later, the author assigns Fuzzy Rank via a modified Gompertz function that provides the last prognosis score afterwards considering the individual score from the three CNN architectures.

Alsubai [22] presents a DNA-RCNN (Deep Normalized Attention-based Residual CNN) for extracting the suitable features dependent upon the discriminative representation of features. The presented NN also discovers alluring features with the presented attention elements that lead to consistent results. At last, the classification was executed by the presented M-RF (modified-RF) with empirical loss function.

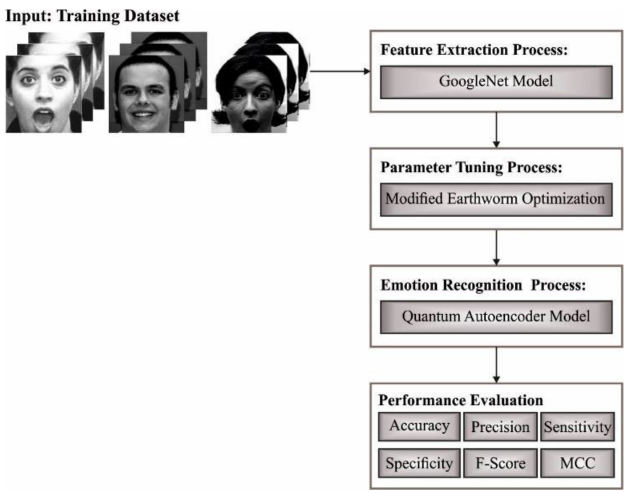

FIGURE 1. **Working process of the MEWODL-ER technique.**

Hossain and Muhammad [23] examine an emotion recognition method utilizing a DL technique in emotional Big Data. Afterwards, this Mel-spectrogram was provided to CNN. The resultants of 2 CNNs are fused utilizing 2 consecutive ELMs. The resultant of fusion was provided as SVM for last classifier of emotions.

In [24], either verbal or non-verbal sounds in a speech are assumed for emotional recognition of real-world conversations. Lastly, an order of CNN-based feature vectors to a whole dialog turn is provided as attentive LSTM-based sequence-to-sequence method to output an emotional sequence as recognition outcome. Zhong et al. [25] present a regularized graph neural network (RGNN) for EEG-based emotion detection. RGNN assumes the biological topology amongst various brain areas for capturing either local or global relations amongst distinct EEG channels. Specially, the authors present the inter-channel relation from EEG signals using an adjacency matrix in GNN but the connection and sparseness of adjacency matrix can be simulated as neuroscience models of human brain organization.

III. THE PROPOSED MODEL

In this study, a new MEWODL-ER algorithm has been presented for emotion classification in the HCI applications. The presented MEWODL-ER technique is intended for the identification of various types of emotions that exist in the HCI applications. In the presented MEWODL-ER technique, three stages of operations are involved namely GoogleNet feature extraction, MEWO hyperparameter tuning, and QAE emotion recognition. Fig. 1 defines the Working process of the MEWODL-ER technique.

A. FEATURE EXTRACTION PROCESS

To produce a useful set of feature vectors, the GoogleNet method is exploited in this study. CNN is a neural network (NN) model extensively applied in image datasets. It is composed of neurons that learn weight and bias whereas input comes through and reaches output [26]. CNN follows the basic assumption of NN as it involves a fully-connected layer and loss function. But the main dissimilarity of CNN was that not each neuron has full connectivity, except the latter fully-connected layer. This is due to full connectivity in every parameter being inefficient and might result in overfitting. Furthermore, CNN has distinct layer composition in its framework as it is encompassed fully connected, pooling, and convolutional layers. By stacking this layer, we could formulate different CNN models.

The current architecture accomplished improved performance in GoogLeNet. It is the conqueror of ILSVRC 2014 with the topmost 5 error rate of 6.7%. First, the Inception model was implemented which is parallelly added to consecutively stack up layers. The inception model is comprised of 1×1, 3×3, and 5×5 convolutional layers and 3×3 max-pooling layers. To prevent overfitting and minimize the spatial size, a pooling layer has been implemented. The 1×1 convolution layer was utilized beforehand 3×3 and 5×5 convolution layers for reducing dimension. Computational efficiency and improved performance can be accomplished by using the Inception model. It is made up of 9 Inception models and 22 convolution layers, resulting in an overall of 100 layers with fewer parameters. The presented architecture makes use of the ReLU function following every convolution layer to enrich the non-linearity. The 5×5 filters are used for extracting features in the convolution layer.

For optimal adjustment of the hyperparameters, the MEWO algorithm is used. EWO is a "nature-inspired evolutionary algorithm" i.e., inspired by the reproductive method of earthworms (EWs) to find solutions for the optimization problem [27]. EWO is a metaheuristic optimization algorithm that can efficiently search the solution space to determine the optimal solution in a reasonable time and with a high degree of accuracy. In addition, it is simple to design, has fast convergence, and can handle large-scale optimization problems. It employs a diversity maintenance mechanism that helps to maintain a diverse set of solutions during the optimization process. This mechanism helps to prevent premature convergence and ensures that the algorithm explores different regions of the solution space.

The following rules are used as follows: (i). a similar collection of DNA strands might be seen in children and parents of the same EW. (ii) Number of EWs from the early generation was in the best physical condition. (iii) There are only 2 different modes of reproduction presented to EWs in populations, and each EW can able to produce offspring by using any of these methods. A summary of "EWO" is given as follows

Reproduction 1

Hermaphrodites are often seen in the EW family which shows that each possesses female and male genitalia inside the body. This implies that a single-parent EW can able to produce completely independent offspring EW and it can be mathematically expressed in the following:

$$u_{i1,k} = u_{max,k} + \mathrm{u}_{\mathrm{min,k}} - \alpha u_{i,k} \tag{1}$$

The abovementioned formula details the procedure that should be followed for producing the *kth* component of infant EW $i1$ with EW i as a parent. $u_{i1,k}$ and $u_{i,k}$ denotes the $k-th$ component of EW $i1$ and i, correspondingly. The operational limitation of *the kth* component of every EW is represented as $u_{\max,k}$ and $u_{\min,k}$, correspondingly. The "similarity factor", where the value ranges from zero to one, determines the quantity of genetic substance passed from parent EWs to progenies.

Reproduction 2

The "uniform crossover" "single-point crossover", and "multipoint crossover", operators are the upgraded version of crossover operators. Consider M as the number of young EWs that can be 1, 2, or 3 in each circumstance. It is probable for the number of EW parents, represented as N. N can be fixed as 2, and M can be fixed as 1 utilizing a uniform crossover. The selection technique initiates by the spinning of the wheel that leads to the selection of 2 parent EWs, Pl and P2 as follows:

$$P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \tag{2}$$

Here, both parent produces two children u_{12} and u_{22}, correspondingly. A number that can be unpredictable within [0, 1] interval was produced (rand), and *the kth* component of u_{12} and u_{22} are formed by the following instruction:

If $rand > 05$,

$$\begin{cases} u_{12,k} = P_{1,k} \\ u_{22,k} = P_{2,k} \end{cases} \tag{3}$$

Otherwise,

$$\begin{cases} u_{12,k} = P_{2,k} \\ u_{22,k} = P_{1,k} \end{cases} \tag{4}$$

Consider $rand1$ as another integer between 0-1 generated at dictates of chance.

$$u_{i2} = \begin{cases} u_{12}\, for\ rand\ 1 < 0.5 \\ u_{22}\ else \end{cases} \tag{5}$$

The EW u_i' for the succeeding generation might be calculated as follows:

$$u_i' = \beta u_{i1} + (1-\beta)\, u_{i2} \tag{6}$$

Now β represents the "proportional factor". It was used to change the percentage of u_{i1} and u_{i2} such that the efficiency of universal and confined search might be retained in an optimum state:

$$\beta^{t+1} = \gamma \beta^t \tag{7}$$

From the equation, t represents the generation that occurs presently. Initially, if t was equivalent to 0, β was equivalent to 1. The "cooling factor of cooling schedule in the simulated annealing" matches with the factor value called "γ".

It is essential to escape from local optima, to find a solution. Consequently, Cauchy Mutation (CM) was performed. It has made "EWO" more efficient to find what you are seeing.

$$W_k = \left(\frac{N_{pop}}{\sum_{i=1} u,\, k} \right) / N_{pop} \tag{8}$$

where W_k signifies the "weight vector" of the *kth* population and *Nop* describe the magnitude of the population.

The *kth* element of the ultimate EW is represented by:

$$u''_{i,k} = u_i' + W_k * Cd \tag{9}$$

Cd represents a randomly generated number that might be selected from the "Cauchy distribution" if it was anticipated that $\tau = 1$. In such cases, τ can be represented as a "scaling parameter".

The MEWO algorithm is designed by the integration of the EWO algorithm with the elite oppositional-based learning (EOBL) concept. The OBL is an optimized approach utilized for improving the diversity of optimized techniques and improving their created solutions [28]. In EOBL, to all the solutions their opposite solution can be defined according to existing elite solutions. Moreover, as in the initialized stage in every optimized system, the solution is arbitrarily created. EOBL is utilized for searching in either direction of novel primary solution or novel created opposite solution and proceeds the fittest solutions to the next iterations. Simulated by these determining, the concept of EOBL is combined in the EWO technique initialized for improving their diversity and searching ability. The basic concept of EOBL is expressed based on the subsequent formulas, considering the elite solution is $x_e = [x_{e1}, x_{e2}, x_{e3}, \ldots, x_{edim}]$ that is defined in the primary created n solutions as the fittest solution betwixt these n solutions, at present to all the solutions x_j its elite opposition solution $\tilde{x}$ is defined utilizing the subsequent Eq. (10)

$$\tilde{x}_{ij} = k\left(lb_j + ub_j\right) - x_{ej} \tag{10}$$

whereas $j = 1, 2, 3, \ldots$, dim and $i = 1, 2, 3, \ldots, n$

In which, k during this case is an arbitrary value on the interval of zero and one, the upper ub_j and low-value lb_j utilized from elite opposite solutions computation and this equation to define their values are (11-12):

$$lb_j = \min\left(x_{i,j}\right) \tag{11}$$

$$ub_j = \max\left(x_{i,j}\right) \tag{12}$$

At this time, to make sure that novel opposite values can be possible and inside the boundary of searching space the subsequent in Eq. (13) is utilized after result $\tilde{x}_{i,j}$

$$\tilde{x}_{i,j} = rand\left(lb_j, ub_j\right)\left[if\, \tilde{x}_{i,j} < x_{\min} OR \tilde{x}_{i,j} > x_{\max}\right]$$
$$i = 1, 2, 3, \ldots, n, \quad j = 1, 2, 3, \ldots, \dim \tag{13}$$

In which $x_{i,j}$ implies the j^{th} value from the vector of existing j^{th} solution of problem populations, $\tilde{x}$ signifies the elite opposite solution of $x_{i,j}$, lb_j implies the minimal value of j^{th} dimensional from the searching space, ub_j denotes the maximal value of j^{th} dimensional from the searching space, rand (lb_j, ub_j) signifies the arbitrary value on the interval of

$[lb_j, ub_j]$, the maximal and minimal bounds of $\tilde{x}$ is $[x_{\min}, x_{\max}]$ that is the constraints when a novel value of $\tilde{x}_{i,j}$ designates jump out of the boundary,n denotes the population size, and dim implies the problem dimensional. Therefore, EOBL is embedded in the initialized stage for obtaining a fitter solution than the primarily created solution.

B. EMOTION RECOGNITION PROCESS

At the final stage, the QAE method is applied in this study. QAE follows the same principle as a conventional autoencoder (AE) but instead of employing the process to neurons, state vector can be applied. The quantum state autoencoder is a circuit that takes a statevector as input and it provides a reduced version of that state vector. AEs are NNs employed in several applications of unsupervised learning [29]. It is learning for mapping input vector x to compressed hidden vector z using encoding. This hidden vector provides a decoding which recreates inputs. Representing the encoding and decoding networks as $E(\Theta_E, x)$ and $D(\Theta_D, z)$ with Θ_E and Θ_D representing the learnable parameter of the corresponding network.

$$z = E\left(\Theta_E, x\right), \hat{x} = D\left(\Theta_D, z\right), \tag{14}$$

whereas $\hat{x}$ implies the recreated resultant vector. An entire network was trained using gradient descent for reducing a faithful distance L, betwixt regenerated outcome $\hat{x}$ and input vector x. For sample L is the RMSE,

$$L\left(x, \hat{x}\right) = \sqrt{\frac{\sum_{i=1}^{i=n}(\hat{x}^i - x^i)^2}{n}}, \tag{15}$$

In which, $\hat{x}^i$ and x^i signify the i^{th} element of recreated and input vector correspondingly, and n is its dimensional. A faithful encoder takes a better latent dimensional $k < n$, with k being the intrinsic dimensional of the database. This dimensionality decrease is vital in several applications of AEs that learn trivial mapping for reconstructing the resultant vectors $\hat{x}$. It separates the circuit into 3 blocks; a state preparation which encodes typical inputs as to the quantum state, a unitary evolution circuit which develops the input state, and the measurement and post-processing part which measures the developed state and processes the attained observables more.

There are several instances of state preparation in the literature that has their merits in several applications. It makes the states utilize an angle encoder that encrypts real-valued observable ϕ_j as rotation angles in addition to $the x$-axis of the Bloch sphere

$$|\Phi\rangle = \otimes_{i=1}^{n} R_x(\phi_j)|0\rangle = \otimes_{j=1}^{n}(\cos\frac{\phi_j}{2}|0\rangle - i\sin\frac{\phi_j}{2}|1\rangle), \tag{16}$$

In which $R_x = e^{-i\frac{\phi_j}{2}\hat{\sigma}_x}$ implies the rotational matrix. The count of qubits needed n, is similar to the dimensional of the input vector. With Θ signifying group of variables, develops the prepared state $|\Phi\rangle$ to the last state $|\Psi\rangle$,

$$|\Psi\rangle = U\left(\Theta\right)|\Phi\rangle. \tag{17}$$

FIGURE 2. Sample images.

TABLE 1. Details of the dataset.

Class	Labels	No. of Samples
Anger	An	45
Contempt	Co	18
Disgust	Di	59
Fear	Fe	25
Happy	Ha	69
Neutral	Nu	593
Sad	Sa	28
Surprise	Su	83
Total Number of Sample Images		920

The last measurement step contains the measurement of observables on the last state $|\Psi\rangle$. While measurement in the quantum process was inherently probabilistic, it is measured several times (called shots) for obtaining an accurate outcome. To accomplish this, it requires quantum hardware which is prepare a huge count of pure same input states $|\Phi\rangle$ to all the data points.

IV. EXPERIMENTAL VALIDATION

The proposed model is simulated using Python 3.6.5 tool on PC i5-8600k, GeForce 1050Ti 4GB, 16GB RAM, 250GB SSD, and 1TB HDD. The parameter settings are given as follows: learning rate: 0.01, dropout: 0.5, batch size: 5, epoch count: 50, and activation: ReLU.

In this section, the emotion classification performance of the MEWODL-ER method is tested with the use of the dataset [30], containing 920 instances as demonstrated in Table 1. The dataset contains a total of 8 classes. Fig. 2 illustrates the sample images. For experimental validation, the dataset is split into 80:20 and 70:30 of training/testing dataset.

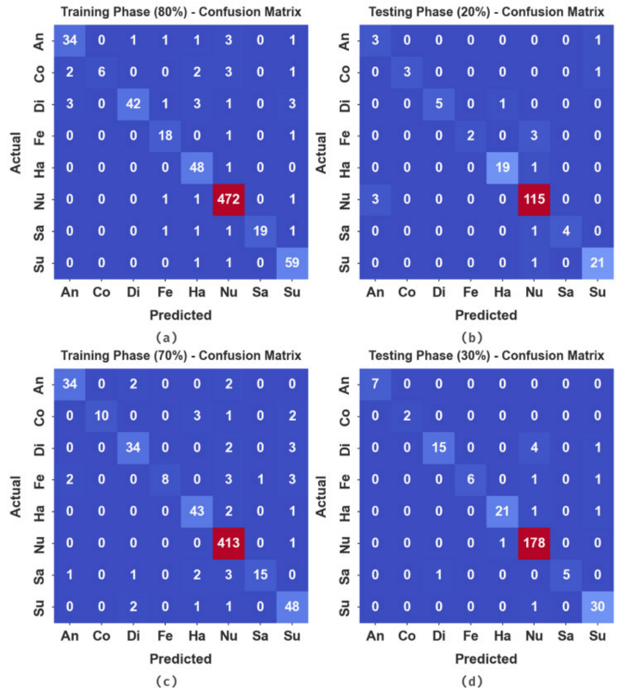

FIGURE 3. Confusion matrices of MEWODL-ER algorithm (a-b) TR and TS databases of 80:20 and (c-d) TR and TS databases of 70:30.

TABLE 2. Emotion classification outcome of MEWODL-ER system under 80:20 of TR/TS databases.

Labels	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	F_{score}	MCC
Training Phase (80%)						
An	98.37	87.18	82.93	99.28	85.00	84.17
Co	98.91	100.00	42.86	100.00	60.00	65.11
Di	98.37	97.67	79.25	99.85	87.50	87.18
Fe	99.18	81.82	90.00	99.44	85.71	85.40
Ha	98.64	84.21	97.96	98.69	90.57	90.14
Nu	98.10	97.72	99.37	95.79	98.54	95.84
Sa	99.46	100.00	82.61	100.00	90.48	90.64
Su	98.64	88.06	96.72	98.81	92.19	91.57
Average	98.71	92.08	83.96	98.98	86.25	86.25
Training Phase (20%)						
An	97.83	50.00	75.00	98.33	60.00	60.21
Co	99.46	100.00	75.00	100.00	85.71	86.36
Di	99.46	100.00	83.33	100.00	90.91	91.03
Fe	98.37	100.00	40.00	100.00	57.14	62.72
Ha	98.91	95.00	95.00	99.39	95.00	94.39
Nu	95.11	95.04	97.46	90.91	96.23	89.32
Sa	99.46	100.00	80.00	100.00	88.89	89.19
Su	98.37	91.30	95.45	98.77	93.33	92.43
Average	98.37	91.42	80.16	98.42	83.40	83.21

The emotion classification outcomes of the MEWODL-ER method are seen in Fig. 3. The outcomes indicated the MEWODL-ER model has properly recognized eight different kinds of emotions.

In Table 2 and Fig. 4, the emotion classification outcomes of the MEWODL-ER method are studied under 80% of TR and 20% of TS databases in terms of different measures such

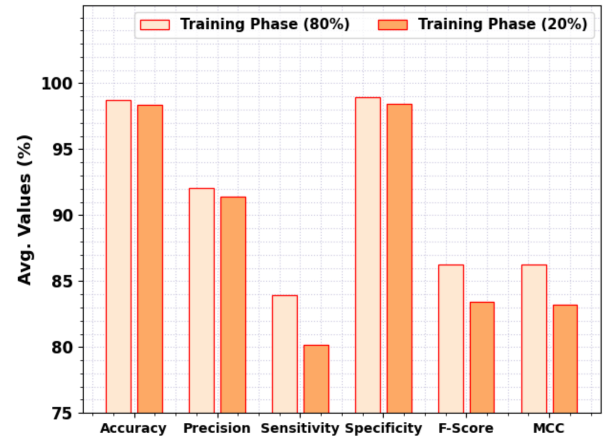

FIGURE 4. Average outcome of MEWODL-ER system under 80:20 of TR/TS databases.

TABLE 3. Emotion classification outcome of MEWODL-ER system under 70:30 of TR/TS databases.

Labels	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	F_{score}	MCC
Training Phase (70%)						
An	98.91	91.89	89.47	99.50	90.67	90.10
Co	99.07	100.00	62.50	100.00	76.92	78.68
Di	98.45	87.18	87.18	99.17	87.18	86.35
Fe	98.60	100.00	47.06	100.00	64.00	68.11
Ha	98.60	87.76	93.48	99.00	90.53	89.82
Nu	97.67	96.72	99.76	93.91	98.22	94.96
Sa	98.76	93.75	68.18	99.84	78.95	79.38
Su	97.83	82.76	92.31	98.31	87.27	86.24
Average	98.49	92.51	79.99	98.72	84.22	84.21
Training Phase (30%)						
An	100.00	100.00	100.00	100.00	100.00	100.00
Co	100.00	100.00	100.00	100.00	100.00	100.00
Di	97.83	93.75	75.00	99.61	83.33	82.77
Fe	99.28	100.00	75.00	100.00	85.71	86.28
Ha	98.91	95.45	91.30	99.60	93.33	92.77
Nu	97.10	96.22	99.44	92.78	97.80	93.66
Sa	99.64	100.00	83.33	100.00	90.91	91.12
Su	98.55	90.91	96.77	98.78	93.75	92.99
Average	98.91	97.04	90.11	98.85	93.11	92.45

as accuracy ($accu_y$), precision ($prec_n$), sensitivity ($sens_y$), specificity ($spec_y$), F-score (F_{score}), and Mathew Correlation Coefficient (MCC). The outcomes indicated that the MEWODL-ER methodology has proficiently recognized different emotions. For example, on 80% of the TR database, the MEWODL-ER method has reached an average $accu_y$ of 98.71%, $prec_n$ of 92.08%, $sens_y$ of 83.96%, $spec_y$ of 98.98%, F_{score} of 86.25%, and MCC of 86.25%. Furthermore, on 20% of the TS database, the MEWODL-ER system has gained an average $accu_y$ of 98.37%, $prec_n$ of 91.42%, $sens_y$ of 80.16%, $spec_y$ of 98.42%, F_{-score} of 83.40%, MCC of 83.21%.

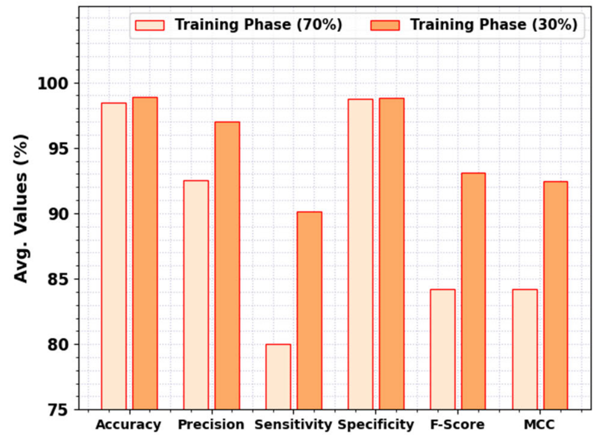

FIGURE 5. Average outcome of MEWODL-ER system under 70:30 of TR/TS databases.

FIGURE 6. TACC and VACC analysis of the MEWODL-ER system.

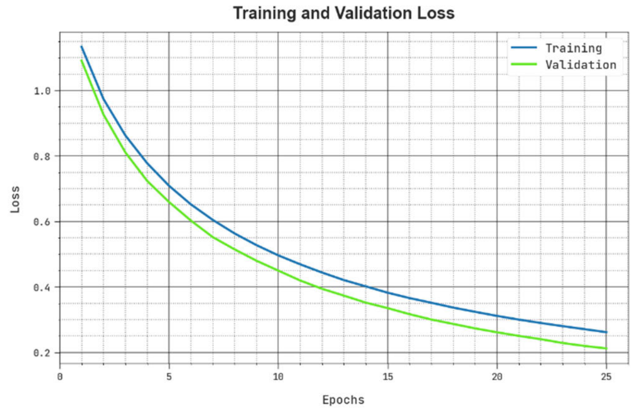

FIGURE 7. TLS and VLS analysis of the MEWODL-ER system.

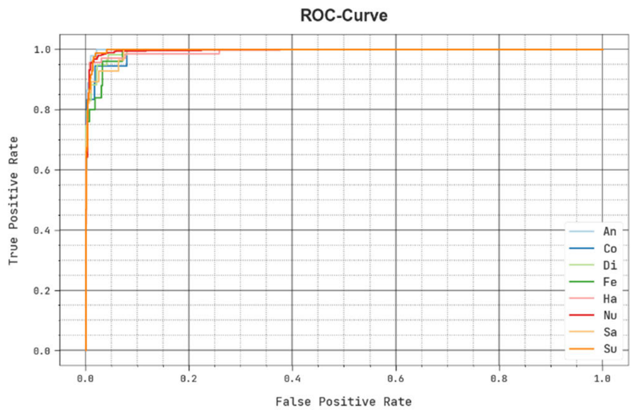

FIGURE 8. ROC analysis of the MEWODL-ER system.

In Table 3 and Fig. 5, the emotion classification results of the MEWODL-ER technique are studied under 70% of TR and 30% of TS databases. The outcomes exhibited the MEWODL-ER approach has proficiently recognized different emotions. For example, on 70% of the TR database, the MEWODL-ER approach has obtained an average $accu_y$ of 98.49%, $prec_n$ of 92.51%, $sens_y$ of 79.99%, $spec_y$ of 98.72%, F_{score} of 84.22%, and MCC of 84.21%. Moreover, on 30% of the TS database, the MEWODL-ER process has attained an average $accu_y$ of 98.91%, $prec_n$ of 97.04%, $sens_y$ of 90.11%, $spec_y$ of 98.85%, F_{score} of 93.11%, MCC of 92.45%.

The TACC and VACC of the MEWODL-ER technique are inspected on emotion classification performance in Fig. 6. The results depict that the MEWODL-ER algorithm has illustrated superior performance with improved values of TACC and VACC. Particularly, the MEWODL-ER technique has the highest TACC outcomes.

The TLS and VLS of the MEWODL-ER approach are tested on emotion classification performance in Fig. 7. The figure exhibited that the MEWODL-ER approach has demonstrated superior performance with the minimum values of TLS and VLS. Seemingly, the MEWODL-ER approach has minimum VLS outcomes.

A clear ROC inspection of the MEWODL-ER approach under the test database is given in Fig. 8. The figure designated that the MEWODL-ER algorithm has improved value of ROC values under all classes.

TABLE 4. Comparative analysis of the MEWODL-ER system with other existing approaches.

Methods	Precision	Sensitivity	Specificity	Accuracy	F1 score
MEWODL-ER	97.04	90.11	98.85	98.91	93.11
VGG19	84.74	82.29	98.05	96.64	83.34
Resnet50	92.20	89.23	98.15	97.69	91.20
MobileNet	94.62	88.67	97.53	98.72	91.78
Inception V3	78.37	79.74	96.56	94.50	75.56
SVM	92.55	88.16	92.58	92.63	92.54

In Table 4, a detailed comparative analysis of the MEWODL-ER method with existing methods is provided [31]. The outcome shows that the MEWODL-ER system has reached greater results than other existing models. Based on $prec_n$, the MEWODL-ER model has gained a higher $prec_n$ of 97.04% while VGG19, Resnet50, MobileNet, Inception V3, and SVM models have resulted in lower $prec_n$ of 84.74%, 92.20%, 94.62%, 78.37%, and 92.55%. Also, based on $sens_y$, the MEWODL-ER method has acquired a higher $sens_y$ of 90.11% while VGG19, Resnet50, MobileNet, Inception V3, and SVM methods have lower $sens_y$ of 82.29%, 89.23%, 88.67%, 79.74%, and 88.16%. Moreover, based on $spec_y$, the MEWODL-ER approach has obtained a higher $spec_y$ of 98.85% while VGG19, Resnet50, MobileNet, Inception V3, and SVM methodologies have resulted in a lower $spec_y$ of 98.05%, 98.15%, 97.53%, 96.56%, and 92.58%. Eventually, based on $accu_y$, the MEWODL-ER method has reached a higher $accu_y$ of 98.91% while VGG19, Resnet50, MobileNet, Inception V3, and SVM techniques have lower $accu_y$ of 96.64%, 97.69%, 98.72%, 94.50%, and 92.63%. Therefore, the MEWODL-ER model accomplished enhanced emotion classification results over other existing techniques. The enhanced performance of the proposed model is due to the inclusion of MEWO based optimal hyperparameter tuning process.

V. CONCLUSION

In this study, a novel MEWODL-ER method was introduced for emotion classification in the HCI applications. The presented MEWODL-ER algorithm is intended for the identification of various types of emotions that exist in the HCI applications. The proposed model follows a three stage process namely GoogleNet feature extraction, MEWO based hyperparameter tuning, and QAE classification. The design of automated hyperparameter adjustment using the MEWO algorithm helps in attaining an improved emotion recognition process. To show the enhanced recognition results of the MEWODL-ER approach, a wide-ranging simulation analysis is done. The experimental values indicated that the MEWODL-ER technique accomplishes promising performance over other models with higher accuracy of 98.91%. In the future, the efficiency of the MEWODL-ER algorithm will be boosted by hybrid DL classifiers.

REFERENCES

[1] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: A review," *IEEE Access*, vol. 7, pp. 117327–117345, 2019.

[2] S. Kwon, "A CNN-assisted enhanced audio signal processing for speech emotion recognition," *Sensors*, vol. 20, no. 1, p. 183, Dec. 2019.

[3] M. M. Taghi Zadeh, M. Imani, and B. Majidi, "Fast facial emotion recognition using convolutional neural networks and Gabor filters," in *Proc. 5th Conf. Knowl. Based Eng. Innov. (KBEI)*, Feb. 2019, pp. 577–581.

[4] M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, "Human emotion recognition using deep belief network architecture," *Inf. Fusion*, vol. 51, pp. 10–18, Nov. 2019.

[5] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Müller-Putz, "Domain adaptation techniques for EEG-based emotion recognition: A comparative study on two public datasets," *IEEE Trans. Cogn. Devel. Syst.*, vol. 11, no. 1, pp. 85–94, Mar. 2019.

[6] H. Meng, T. Yan, F. Yuan, and H. Wei, "Speech emotion recognition from 3D log-mel spectrograms with deep learning network," *IEEE Access*, vol. 7, pp. 125868–125881, 2019.

[7] E. Kanjo, E. M. G. Younis, and C. S. Ang, "Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection," *Inf. Fusion*, vol. 49, pp. 46–56, Sep. 2019.

[8] P. Li, H. Liu, Y. Si, C. Li, F. Li, X. Zhu, X. Huang, Y. Zeng, D. Yao, Y. Zhang, and P. Xu, "EEG based emotion recognition by combining functional connectivity network and local activations," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2869–2881, Oct. 2019.

[9] J. X. Chen, P. W. Zhang, Z. J. Mao, Y. F. Huang, D. M. Jiang, and Y. N. Zhang, "Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 44317–44328, 2019.

[10] E. H. Houssein, A. Hammad, and A. A. Ali, "Human emotion recognition from EEG-based brain–computer interface using machine learning: A comprehensive review," *Neural Comput. Appl.*, vol. 34, pp. 12527–12557, May 2022.

[11] Q. Gao, Y. Yang, Q. Kang, Z. Tian, and Y. Song, "EEG-based emotion recognition with feature fusion networks," *Int. J. Mach. Learn. Cybern.*, vol. 13, no. 2, pp. 421–429, Feb. 2022.

[12] A. A. Alnuaim, M. Zakariah, P. K. Shukla, A. Alhadlaq, W. A. Hatamleh, H. Tarazi, R. Sureshbabu, and R. Ratna, "Human–computer interaction for recognizing speech emotions using multilayer perceptron classifier," *J. Healthcare Eng.*, vol. 2022, pp. 1–12, Mar. 2022.

[13] K. Tzafilkou, A. A. Economides, and N. Protogeros, "Mobile sensing for emotion recognition in smartphones: A literature review on non-intrusive methodologies," *Int. J. Hum.–Comput. Interact.*, vol. 38, no. 11, pp. 1037–1051, Jul. 2022.

[14] P. R. Bhise, S. B. Kulkarni, and T. A. Aldhaheri, "Brain computer interface based EEG for emotion recognition system: A systematic review," in *Proc. 2nd Int. Conf. Innov. Mech. Ind. Appl. (ICIMIA)*, Mar. 2020, pp. 327–334.

[15] M. La Mura and P. Lamberti, "Human-machine interaction personalization: A review on gender and emotion recognition through speech analysis," in *Proc. IEEE Int. Workshop Metrol. Ind. 4.0 & IoT*, Jun. 2020, pp. 319–323.

[16] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, and M. Zareapoor, "Hybrid deep neural networks for face emotion recognition," *Pattern Recognit. Lett.*, vol. 115, pp. 101–106, Nov. 2018.

[17] Z. Li, G. Zhang, J. Dang, L. Wang, and J. Wei, "Multi-modal emotion recognition based on deep learning of EEG and audio signals," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2021, pp. 1–6.

[18] P. Tzirakis, J. Zhang, and B. W. Schuller, "End-to-end speech emotion recognition using deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 5089–5093.

[19] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. Wahby, "EEG-based emotion recognition using 3D convolutional neural networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 1–9, 2018.

[20] Y. Nan, J. Ju, Q. Hua, H. Zhang, and B. Wang, "A-MobileNet: An approach of facial expression recognition," *Alexandria Eng. J.*, vol. 61, no. 6, pp. 4435–4444, Jun. 2022.

[21] K. K. Sahoo, I. Dutta, M. F. Ijaz, M. Wozniak, and P. K. Singh, "TLEFuzzyNet: Fuzzy rank-based ensemble of transfer learning models for emotion recognition from human speeches," *IEEE Access*, vol. 9, pp. 166518–166530, 2021, doi: 10.1109/ACCESS.2021.3135658.

[22] S. Alsubai, "Emotion detection using deep normalized attention-based neural network and modified-random forest," *Sensors*, vol. 23, no. 1, p. 225, Dec. 2022.

[23] M. S. Hossain and G. Muhammad, "Emotion recognition using deep learning approach from audio-visual emotional big data," *Inf. Fusion*, vol. 49, pp. 69–78, Sep. 2019.

[24] K.-Y. Huang, C.-H. Wu, Q.-B. Hong, M.-H. Su, and Y.-H. Chen, "Speech emotion recognition using deep neural network considering verbal and nonverbal speech sounds," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 5866–5870.

[25] P. Zhong, D. Wang, and C. Miao, "EEG-based emotion recognition using regularized graph neural networks," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1290–1301, Jul. 2022.

[26] Y. Seo and K.-S. Shin, "Hierarchical convolutional neural networks for fashion image classification," *Expert Syst. Appl.*, vol. 116, pp. 328–339, Feb. 2019.

[27] H. Moayedi and B. Le Van, "The applicability of biogeography-based optimization and earthworm optimization algorithm hybridized with ANFIS as reliable solutions in estimation of cooling load in buildings," *Energies*, vol. 15, no. 19, p. 7323, Oct. 2022.

[28] M. Tubishat, M. A. Abushariah, N. Idris, and I. Aljarah, "Improved whale optimization algorithm for feature selection in Arabic sentiment analysis," *Appl. Intell.*, vol. 49, no. 5, pp. 1688–1707, 2019.

[29] V. S. Ngairangbam, M. Spannowsky, and M. Takeuchi, "Anomaly detection in high-energy physics using a quantum autoencoder," *Phys. Rev. D, Part. Fields*, vol. 105, no. 9, May 2022, Art. no. 095004.

[30] Accessed: Oct. 14, 2022. [Online]. Available: http://www.jeffcohn.net/Resources/

[31] M. K. Chowdary, T. N. Nguyen, and D. J. Hemanth, "Deep learning-based facial emotion recognition for human–computer interaction applications," *Neural Comput. Appl.*, to be published, doi: 10.1007/s00521-021-06012-8.