Received: January 3, 2024. Revised: March 1, 2024.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Predicting the Facial Expression Recognition Using Novel Enhanced Gated Recurrent Unit-based Kookaburra Optimization Algorithm

Anand M[1] **Babu S[1]***

[1]*Department of Computing Technologies, SRM Institute of Science and Technology, Kattankulathur-603203, India*
* Corresponding author's Email: babus@srmist.edu.in

Abstract: Research on real-time Facial Expression Recognition (FER) and human emotion prediction has been ongoing. The desire of humans to enable cohabitation between humans as well as machines can only be realized if a machine is able to comprehend, display, and respond to a human's emotional state. If an Artificial Intelligence (AI) makes judgements on the basis of emotions rather than just algorithms and logic, it will be able to really breathe life without air. Although there have been several attempts to date to do this, it is still believed to be in the early stages of development. Hence, this paper plans to perform the FER prediction with the help of novel intelligent deep learning technology. Initially, the data is gathered from two standard benchmark sources such as FER 2013 and EMOTIC dataset. Next, the pre-processing of the gathered images is accomplished by face detection, rotation rectification, and Local Binary Pattern (LBP). From the pre-processed images, the feature extraction is performed by the Centralized Binary Pattern (CBP) method. These extracted features undergo the final prediction phase that is done by the novel Enhanced Gated Recurrent Unit (EGRU), in which the parameters of GRU are tuned by the nature inspired optimization algorithm referred as Kookaburra Optimization Algorithm (KOA) with the consideration of error minimization as the major objective function. This novel EGRU-KOA predicts the final outcome with consideration of various facial expressions such as sad, surprise, disgust, contempt, happy, and fear respectively. The novelty of the proposed model is demonstrated by comparing it with various conventional models in terms of distinct analysis respectively. The proposed EGRU-KOA in terms of accuracy is 2.78%, 5.60%, 1.63%, and 4.28% better than DBN+QPSO, GDP, CNN, and WKELM respectively. Similarly, the proposed EGRU-KOA with respect to MSE is 84.82%, 87.41%, 88.51%, and 86.82% better than DBN+QPSO, GDP, CNN, and WKELM respectively.

Keywords: Facial expression recognition prediction, Novel enhanced gated recurrent unit, Kookaburra optimization algorithm, Centralized binary pattern, Local binary pattern.

1. Introduction

The FER emerged as one among the most popular fields in computer science two decades ago [1]. There are now a lot more FER-oriented apps. They consist of, still are not restricted to, pattern recognition, animation, computer vision, and Human-Computer Interaction (HCI). As long as people don't display their emotions in a similar manner, machine learning approaches do indeed have difficulty recognizing facial expressions [2]. Within the field of machine learning, deep learning represents a relatively recent field of study that has the ability to accurately categorize the emotions on human faces into several classifications. The identification of fundamental emotions is essentially the foundation of facial expression investigation research [3].

Prior research has produced a number of techniques, most of which centre on the extraction of active characteristics from face images. The previous FER approaches are divided into two categories: appearance-oriented and geometric feature-oriented, depending on the feature extraction framework [4]. In the first scenario, facial geometry and face landmarks are used to determine the location of the characteristics that are retrieved [5]. These

techniques took advantage of features of the face image like form, the placement of the eyes, eyebrow, nose, and mouth, as well as the separation among facial landmark point pairs [6]. Even though these techniques have demonstrated impressive results in the FER, they are prone to face misalignments because of incorrect facial landmark point recognition and tracking in a variety of challenging situations, such as occlusions, poor light, and low resolution [7]. Image filters were applied to the entire face or to specific areas of the face in appearance-oriented approaches [8].

Furthermore, numerous significant features might be missed or missed by conventional methodologies since they rely on manually created features created by programmers [9]. Deep learning approaches have gained popularity recently as a result of their several advantages, including their capability to capture and learn features automatically, their ability to be robust to natural changes in the information as well as their generalizability—the capability to utilize the similar method for a variety of applications—and their scalability, which allows effectiveness to increase as more information is gathered [10].

The paper contribution is shown as below:

- To perform the FER prediction with the help of novel intelligent deep learning technology.
- To accomplish the pre-processing by the face detection, rotation rectification, and LBP and to perform the feature extraction by the CBP.
- To perform the final prediction phase that is done by the novel EGRU, in which the parameters of GRU are tuned by the nature inspired optimization algorithm referred as KOA with the consideration of error minimization as the major objective function.
- To predict the final outcome with consideration of various facial expressions such as sad, surprise, disgust, contempt, happy, and fear respectively using the same novel EGRU-KOA.

The paper organization is given as follows: Section 1 is the introduction of FER. Section 2 is a literature survey. Section 3 is proposed methodology with dataset collection, pre-processing, feature extraction by CBP, prediction by novel EGRU, and KOA algorithm. Section 4 is the results. Section 5 is the conclusion.

1.1 Problem statement

Since there are large variances in the physiognomy of faces with regard to an individual's identity, ambient lighting circumstances, and head attitude, automatic facial emotion identification describes a challenging issue. The job might be more difficult when complex emotion recognition is considered. At the moment, a major obstacle impeding the progress of research on automatic compound emotion identification is the scarcity of extensive publicly available labelled datasets in the region. Nevertheless, because of recent developments in the study of complex emotions, there exists certain efforts to identify more accurate as well as detailed facial manifestations of emotion. Yet, during the past several years, the computer vision and deep learning groups have begun to pay greater attention to the numerous intricate and complicated facial expressions that people can make by combining various fundamental ones.

Symbols and Notations

Symbols and Notations	**Description**
(My, Mz)	Initial location in the face image
(My', Mz')	Coordinate following rotation transformation
θ	Rotation angle
O	Count of neighbouring pixels
$T(*)$	Signature function
h_d and h_o	Values associated with the centre as well as neighbouring pixels
$t(y)$	Symbol function
N	Neighbourhood count
S	Radius
h_n	Neighbourhood pixel
h_d	Centre pixel
E	Threshold constant
$1 - a_u$	Input gate
a_u	Update gate
g_u	Forget gate
n	Count of decision variables
s	Random count in the interval [0, 1]
LB_e	Lower bound
UB_e	Upper bound
Y	KOA population matrix
Y_j	Candidate solution
$y_{j,e}$	Decision variable
G_j	Fitness function
G	Vector
DQ_j	Group of potential prey
$TDQ_{j,e}$	e^{th} dimension of chosen prey for j^{th} kookaburra

J	Random count from group {1, 2}
Y_j^{Q1}	Novel proposed location
u	Method's iteration counter
U	Maximum count of algorithm iterations

2. Related work

A FER framework on the basis of Quantum Particle Swarm Optimization (QPSO) and Deep Belief Networks (DBNs) [11]. The framework consists of four stages: the input image was pre-processed by cropping the Region Of Interest (ROI) to obtain the necessary region and eliminate non-important portions; the ROI was divided into numerous blocks and the integral image was utilized to describe the superior (major effective) blocks; the image down sampling method was used to minimize the size associated with the novel sub image in order to enhance the system effectiveness; fourth, the emotion's class was recognized utilizing the DBN. Rather than adjusting DBN parameters manually, QPSO was employed to automatically optimize the values related to the DBN parameters. Still, it does not merge various techniques for enhancing the accuracy associated with the system.

This work presents a unique face descriptor, Gradient Direction Pattern (GDP), for FER [12]. With the use of gradients, GDP encoded the format of the face picture in a more compact fashion so that the FER was resilient to position fluctuations. Additionally, edge feature maps extracted from facial images filtered with Gaussian and Gabor filters at various orientations were computed for GDPs. These edge feature maps made the identification framework resilient to scale, light, noise, and orientational fluctuations. Initially, the face image was separated into tiny sections, and GDPs were extracted from them. Following the concatenation of these patterns, GDP Histograms (GH) were computed histograms. It did not demonstrate the effectiveness on various datasets.

The selection describes a basic CNN architecture with two convolutional layers and two pooling layers [13]. Two types of filtering were applied to the input face images exhibiting emotions in order to increase effectiveness: a LoG filter and a Gaussian filter. The CNN method achieved 100% recognition accuracy while classifying seven different emotions on filtered face images from the Extended Cohn-Kanade (CK+) dataset, which included 981 images. This suggests that the suggested technique was successful. It did not find the facial patches and describe the facial action units.

By concatenating auto encoder features, CNN feature outputs, and handcrafted features like Speed Up Robust Feature (SURF), Scale-Invariant Feature Transform (SIFT), and Oriented Fast Rotated Brief (ORB), measured by the Bag of Visual Words (BOVW), a novel method was suggested in this paper referred as CNNCraft-net, which combines the benefits of CNN and conventional methods to recognize eight facial expressions for static RGB images [14]. Multiple measures, including loss, accuracy, precision, F-measure, and recall, were employed for the comparison study. The suggested method was assessed using the highly unbalanced FER2013 and AffectNet datasets; for AffectNet, it achieved accuracy of 65% for seven expressions, 61.9% for eight expressions, and 69% for FER2013. Still, the methodology was not enhanced for accepting distinct sizes of images.

In order to enable the neural network to continuously acquire the machine's visual information as well as the brain's cognitive knowledge, a unique brain-machine linked learning strategy was provided for face emotion identification [15]. The suggested technique jointly trains the methods in the cognitive and visual domains using Electro Encephalo Gram (EEG) inputs and images. There are two types of interaction channels in every domain method: private and common. Reverse engineering allows a methodology to decipher the brain's cognitive processes, as EEG readings could represent brain activity. The familiar channel in the visual domain could approach the cognitive procedure in the cognitive domain by decoding the EEG signals caused by the face emotion images. But, it cannot be used for working with human computer interaction methodologies.

The iCV-MEFED data set, which comprises labels evaluated by psychologists for 50 classes of compound emotions were made available [16]. The work was difficult since complex facial expressions from several categories share a lot of identities. Furthermore, at the 2017 FG workshop, a challenge on the basis of the suggested iCV-MEFED data set was arranged. In this study, further in-depth tests on the suggested data set were conducted and analyzed the top three winning strategies. Studies show that, in contrast to the seven fundamental emotions, pairings of compound emotions—such as shockingly joyful vs delightfully surprised—were harder to identify. On the other hand, it was anticipated that the suggested data set will aid in paving the path for more study on compound facial emotion identification. It did not consider incorporating complementary and dominant emotions for attaining compound FER.

With a focus on the granularity associated with the emotion concepts, it was aimed to describe the semantic richness problem in the FER methodology [17]. Specifically, the inspiration was drawn from previous psycho-linguistic research that selected 135 emotion names from hundreds of English emotion phrases in a prototypicality rating study. A large-scale 135-class FER image collection was gathered on the basis of the 135 emotion categories, examined the related facial expressions, and then introduced a facial emotion identification system. Comprehensive assessments were performed on the dataset credibility and the corresponding baseline classification method in order to illustrate the feasibility of motivating FER research at a more detailed level. The issue's significance as well as the efficacy of the solution are demonstrated by the quantitative and qualitative outcomes. It cannot be implemented on vast scale dataset for evaluating the efficiency of the system.

This article used facial expressions and EEG to categorize three category emotions using the DBN [18]. 15 deaf participants' signals during their viewing of the moving movie clips were recorded. The approach segments the EEG signals into five frequency bands using a 1-s window that does not overlap, and next it extracts the Differential Entropy (DE) characteristic. Facial expression images as well as the DE feature of EEG provide multimodal input for subject-dependent emotion identification. But, it cannot be applied for FER on video slices that in turn stand out as a major drawback.

The vectorized landmark and landmark curvature, two recently described geometric properties, are used in this research's facial emotion identification approach [19]. These characteristics were taken from face landmarks connected to certain facial muscle movements. The approach that is being discussed combined Support Vector Machine (SVM) with a Genetic Algorithm (GA) oriented classification to solve a multi-attribute optimization issue including parameter and feature selection. The enlarged CK+ dataset and the Multimedia Understanding Group (MUG) dataset were used for experimental assessments. The test accuracy for the 7-class CK+, 8-class CK+, and 7-class MUG was 96.56%; the validation accuracy was 95.58%, 93.57%, and 96.29%, correspondingly. The recall, F1-score, and overall accuracy were around 0.95, 0.96, and 0.97. It does not adopt various approaches for detecting the landmark because appropriate detection of facial landmark described a critical pre-constraint to superior FER.

For the purpose of recognizing complicated emotions, a bionic two-system methodology was suggested [20]. The arrangement resembles the way the human brain processes information and makes decisions in response to challenges. A quick compound sensing module was System I. System II was a more integrated cognitive decision module that operates at a slower pace. System I has two branches: one for physiological measurement, which was an image-only implementation for practicality, and one for facial expression feature description, which includes action units, basic emotion, and valence arousal detection. To make sure the selected time includes the emotion occurrence, System II uses a decision module having segmentation. It then uses reinforcement learning to iteratively optimize the emotion information in a particular segment. It cannot be implemented in 3D faces with consideration of both the side and frontal views associated with the human faces for broadening the potential capabilities of the FER approach.

2.1 Drawbacks of conventional techniques

Categorizing images of human faces into emotional groups describes a difficult task. For this purpose, deep learning works well majority of the time. Owing to its feasible applications, FER represents one among the major significant research issues in Artificial Intelligence (AI) and computer vision. Numerous researches have been introduced for FER, whether they focus on end-to-end AI or use handcrafted (Craft) features having conventional machine learning approaches. Affective computing relies heavily on emotion recognition. In recent years, affective computing researchers have been more interested in fine-grained emotion analysis, like compound facial expression of emotions. The data sets utilized in current research on compound emotions are small in quantity and have uneven distributions of information; labels are automatically generated by machine learning-oriented methods, which may result in errors.

3. Proposed methodology

3.1 Proposed model

The proposed FER prediction model consists of various phases such as data collection, pre-processing, feature extraction, and prediction. FER 2013 and the EMOTIC dataset are two common benchmark sources from which the data is first collected. Then, face detection, rotation rectification, and LBP complete the pre-processing of the collected images. The CBP approach is used to extract features from the pre-processed images. These extracted

features proceed through the final prediction step, which is carried out by the innovative EGRU. Error minimization is the primary objective function, and the parameters of GRU are adjusted using the nature-

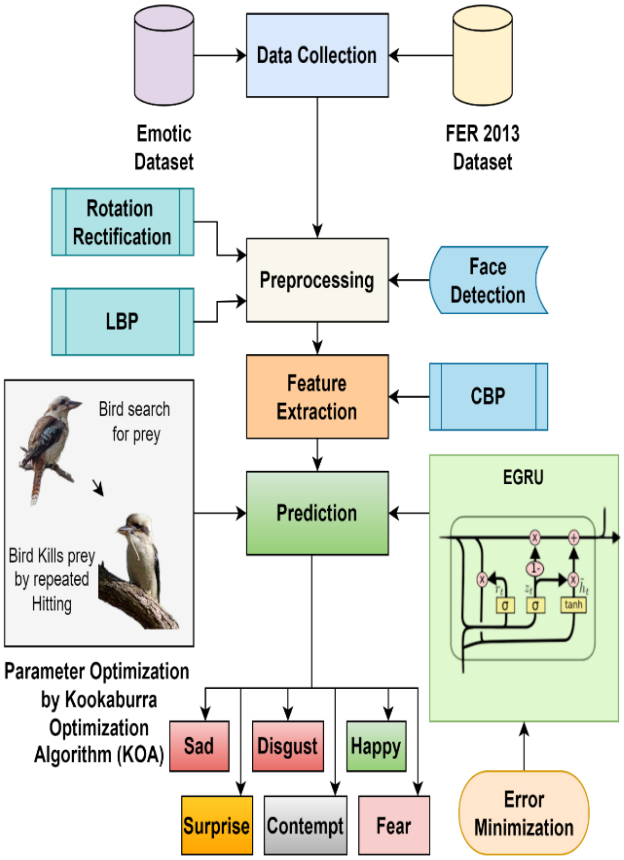

Figure. 1 Proposed novel FER prediction model

inspired optimization method known as KOA. This new EGRU-KOA forecasts the result by considering several facial emotions, including sadness, surprise, disgust, contempt, happiness, and fear, in that order. The proposed FER prediction model is shown in Fig. 1.

3.2 Dataset collection

The dataset used for the proposed FER prediction model is gathered from two standard benchmark sources such as FER 2013 and the EMOTIC dataset. The description of these two datasets is described below.

Facial Expression Recognition 2013 (FER-2013): The 2013 Facial Expression Recognition dataset (FER2013) was provided by Yang et al. in 2013. This dataset can be accessed via Kaggle. The dataset's example images are shown in Fig. 2. The FER-2013 dataset contains grayscale images having a resolution of 48 pixels by 48 pixels, and every face has been allocated an emotion class.

EMOTIC dataset: The EMOTIC dataset consists of images of people in their natural environments that have been annotated with details on their apparent emotional states. The collection consists of 23, 571 images and 34, 320 human annotations. Certain

Figure. 2 Sample images of FER-2013 dataset

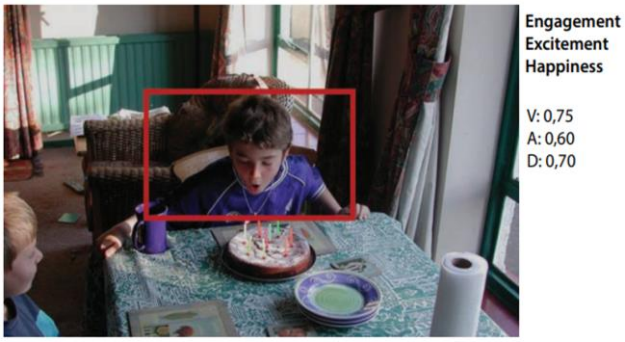

Figure. 3 Sample image of EMOTIC dataset including their comments

images were obtained from the Google search engi ne. In order to do this, the researcher created a collection of questions that had words for a variety of places, activities, social circumstances, and mental states. The images show people in a variety of settings, with different groups, places, and activities. Fig. 3 shows one annotated image from the EMOTIC dataset.

Amazon Mechanical Turk (AMT) was employed to annotate the images. The annotation task involved labelling images according to the annotators' perceptions of the subjects' moods. The visuals are additionally commented using the continuous dimensions Arousal (A), Valence (V), and Dominance (D), as seen in Fig. 3. The dataset was divided into three groups following the first round of annotations, with every group involving a similar proportion of emotional categories throughout validation (10%), training (70%), and testing (30%) images. Furthermore, two annotators are utilized for testing images and four annotators are employed for validation. After removing certain noisy images, three annotators are employed for the testing group.

3.3 Pre-processing

The pre-processing of the gathered images of the proposed novel FER prediction model is accomplished by face detection, rotation rectification, and LBP. Each of these techniques is explained below.

Face detection: The main problem with FER is face detection. Even when a face image is chosen from a benchmarking facial expression dataset, it still contains excessive background data that is uncorrelated to expression recognition. Therefore, accurate face detection results—which must ideally exclude uncorrelated background information—are crucial to exact FER. Current research uses the widely utilized Viola-Jones methodology for face identification.

Rotation rectification: Even for images of the identical subject, facial images in benchmarking datasets and actual surroundings change over time. These changes shall impact on the accuracy of FER as they are not connected to facial expressions. In order to rectify this problem, the rotation transformation matrix with the following definition is used to align the face region using rotation rectification:

$$(My', Mz', 1) = [My, Mz, 1]\ [cos\ cos\ \theta \\ sin\ sin\ \theta\ \ 0\ \ -sin\ sin\ \theta\ \ \ cos\ cos\ \theta\ \ 0\ 0\ 0\ 1\ ] \quad (1)$$

Here, the initial location in the face image is represented by (My, Mz), and the coordinate (y, z) following rotation transformation is represented by (My', Mz'). The line segment that travels from one eye centre to the other forms the rotation angle denoted by θ. Zero is on the horizontal axis. To decrease the dimension, all identified face areas are rescaled to 72 × 72 after rotation rectification. Although a smaller face region might speed up FER even more, it shall also result in the loss of facial data, particularly for that which is gleaned from facial LBP images.

LBP: One popular descriptor for gathering texture data about a target is LBP. A pixel's LBP coding is determined by comparing its value to that of nearby pixels. The LBP value of a pixel may be computed using the following formula once it has been successfully encoded via LBP coding:

$$LBP = \sum_{o=1}^{O} \quad T(h_o - h_d) * 2^o \quad (2)$$

Here, O is the count of neighboring pixels and $T(*)$ is the signature function. The values associated with the centre as well as neighboring pixels are indicated by the symbols h_d and h_o, correspondingly. The LBP value related to every pixel may be calculated to produce an LBP facial image. Face areas connected to expressions, like the eyes, lips, and eyebrows, stand out more in LBP images than in grayscale ones.

3.4 Feature extraction by CBP

The pre-processed images undergo the feature extraction phase that is accomplished with the help of CBP approach. Accurately extracting the texture aspects related to the eyes, lips, and eyebrows is crucial since these areas contain the majority of the facial expression data. To increase discrimination, the CBP operator considers the centre pixel's function and assigns it the highest weight. The definition of the CBP code value is

$$CBP(N, S) = \sum_{n=0}^{\frac{N}{2}-1} \quad t\left(h_n - h_{n+\left(\frac{N}{2}\right)}\right) 2^n + \\ t\left(h_d - \frac{1}{N+1}\left(\sum_{n=0}^{N-1} \quad h_n + h_d\right)\right) 2^{\frac{N}{2}} \quad (3)$$

Here, the definition of the symbol function $t(y)$ is

$$t(y) = \{1\ when\ |y| \geq E\ \ 0\ when\ |y| < E \quad (4)$$

The terms neighborhood count and radius, correspondingly, are denoted by N and S in (3). The value of the neighborhood pixel is $h_n(n = 1,2,3,\cdots,7)$ and the centre pixel is h_d . E describes the threshold constant in (4).

By comparing the difference among the closest neighbor points as well as the centre pixels, CBP may better capture gradient data. This allows it to capture delicate texture by comparing the difference among "neighbor point pairs." Furthermore, because the current LBP Operator's symbol function has changed, CBP is impervious to white noise.

3.5 Prediction by novel EGRU

The extracted features undergo the final prediction phase that is performed by the EGRU. Another well-liked RNN methodology is GRU. GRU forgets, synchronizes the writes, and connects the state explicitly. The following describes the formulation:

$$s_u = \sigma(X_s t_{u-1} + V_s y_u + c_s) \quad (5)$$

$$a_u = \sigma(X_a t_{u-1} + V_a y_u + c_a) \quad (6)$$

$$\widetilde{t_u} = tanh\ tanh\left(X_t(s_u \odot t_{u-1} + V_t y_u + c_t)\right) \quad (7)$$

$$t_u = a_u \odot t_{u-1} + (1 - a_u) \odot \widetilde{t_u} \quad (8)$$

GRU gives up certain expressiveness and selectively overwrites by making the forget gate equal to 1 minus the write gate, as opposed to selective forgets and selective writes. The input gate is determined by $1 - a_u$, and the update gate, a_u, is the identical as the forget gate from the framework LSTM, g_u. This works since it converts t_u into a weighted average of t_{u-1} and $\widetilde{t_u}$ elementwise, which is bounded if and only if t_{u-1} and $\widetilde{t_u}$ are bounded. As an alternative to LSTM, GRU fared better than LSTM on almost entire tasks, with the exception of language modelling using the naïve initialization. Hence, to overcome the limitations, the parameters of GRU are tuned by KOA with the consideration of error minimization as the fitness or objective function, therefore referred as EGRU. This EGRU overcomes the computational complexity as well as the initialization and language modelling problems. Finally, this proposed novel ERU predicts the final outcome into one of six emotions such as sadness, surprise, disgust, contempt, happiness, and fear respectively.

3.6 KOA algorithm

The KOA algorithm is selected here for the proposed FER prediction model to optimize the parameters of the GRU model that in turn can lead to error minimization as the major fitness or the objective function. KOA mimics the kookaburra's instinctive behavior. The way kookaburras seek as well as kill their prey serves as the basic model for KOA. The KOA theory is outlined, and two stages of its mathematical modelling are described: (i) exploration, which simulates prey hunting, and (ii) exploitation, which simulates kookaburra behavior, which ensures that their prey is slain.

On the basis of a random search in the problem-handling space, the KOA technique represents a population-oriented optimizer that may offer appropriate solutions for optimization issues in an iterative procedure. Every kookaburra in the KOA population represents a potential solution to the issue that may be represented by a vector since they are arranged in the problem-handling space such that everyone chooses values for the decision variables according to where it is located. Eq. (9) may be used to represent the KOA population matrix, which is made up of kookaburras together. Eq. (10) is used to randomly initialize the kookaburras' location at the start of the KOA implementation.

$$Y = \left[Y_1 \vdots Y_j \vdots Y_O\right]_{O\times n} = \left[\begin{matrix} y_{1,1} & \vdots & y_{j,1} & \vdots & y_{O,1} \\ \cdots & \ddots & \cdots & \therefore & \cdots \\ y_{1,e} & \vdots & y_{j,e} & \vdots & y_{O,e} \\ \cdots & \therefore & \cdots & \ddots & \cdots \\ y_{1,n} & \vdots & y_{j,n} & \vdots & y_{O,n} \end{matrix}\right]_{O\times n} \quad (9)$$

$$y_{j,e} = LB_e + s \cdot (UB_e - LB_e) \quad (10)$$

Here, O shows the count of kookaburras, n shows the count of decision variables, s shows a random count in the interval [0, 1], LB_e and UB_e represents the lower as well as upper bounds associated with the e^{th} decision variable, correspondingly. Y shows the KOA population matrix, Y_j shows the j^{th} kookaburra (candidate solution), and $y_{j,e}$ shows its e^{th} dimension in search space (decision variable).

The fitness function associated with the issue may be assessed by considering that every kookaburra's location in the problem-handling space represents a potential solution for the issue that corresponds to every kookaburra. Eq. (11) allows for the vector representation of the group of evaluated values for the problem's fitness function.

$$G = \left[G_1 \vdots G_j \vdots G_O\right]_{O\times 1} = \left[G(Y_1) \vdots G(Y_j) \vdots G(Y_O)\right]_{O\times 1} \quad (11)$$

In this case, G_j shows the evaluated fitness function on the basis of the j^{th} kookaburra, and G shows the vector associated with the evaluated fitness function.

The location of remaining kookaburras, which have a superior fitness function value, is taken into consideration as the prey position in KOA model for every kookaburra in order to imitate the hunting scheme of kookaburras. Eq. (12) is then used to identify the possible prey group for every kookaburra on the basis of the comparison of the fitness function values.

$$DQ_j = \{Y_l : G_l < G_j \ and\ l \neq j\}, in\ which\ j = 1,2,\cdots,O\ and\ l\epsilon\{1,2,\cdots,O\} \quad (12)$$

In this case, Y_l shows the kookaburra that has a higher fitness function value than the j^{th} kookaburra, G_l shows the fitness function value, and DQ_j shows the group of potential prey for the j^{th} kookaburra. It is considered in the KOA model that every kookaburra chooses its target at random and attacks it. Eq. (13) is used to determine the kookaburra's novel location on the basis of the simulation of its progress towards the prey in the

hunting scheme. In this instance, Eq. (14) states that the novel location may take the place of the associated kookaburra's prior location if the value associated with the goal function is enhanced there.

$$y_{j,e}^{Q1} = y_{j,e} + s \cdot \left(TDQ_{j,e} - J \cdot y_{j,e}\right), j = 1,2,\cdots,O, and\ e = 1,2,\cdots,n \quad (13)$$

$$Y_j = \{Y_j^{Q1}, G_j^{Q1} < G_j\ \ Y_j, otherwise \quad (14)$$

Here, s shows a random count having a normal distribution in the interval [0, 1], $TDQ_{j,e}$ shows the e^{th} dimension of chosen prey for j^{th} kookaburra, J shows a random count from group {1, 2}, O shows the count of kookaburra, and n shows the count of decision variables. Y_j^{Q1} shows the novel proposed location associated with the j^{th} kookaburra on the basis of the initial stage of KOA, $y_{j,e}^{Q1}$ shows its e^{th} dimension. G_j^{Q1} shows its fitness function value.

Eq. (15) is used to determine a random location in the KOA model, which simulates the behavior of kookaburras depending on their movement close to the hunting area. It is really expected that this displacement having a radius equal to $\frac{(UB_e - LB_e)}{u}$ happens at random in a neighborhood near the centre of every kookaburra. In order to improve the accuracy of the local search that aims to converge towards better answers, the radius associated with this neighborhood is originally placed to its maximum value. During subsequent iterations, this radius gets decreased. If the novel location determined for every kookaburra enhances the value of the goal function as per Eq. (16), then it substitutes the prior location.

$$y_{j,e}^{Q2} = y_{j,e} + (1 - 2s) \cdot \frac{(UB_e - LB_e)}{u}, j = 1,2,\cdots,O, e = 1,2,\cdots,n, and\ u = 1,2,\cdots,U \quad (15)$$

$$Y_j = \{Y_j^{Q2}, G_j^{Q2} < G_j\ \ Y_j, otherwise \quad (16)$$

Here, G_j^{Q2} shows its fitness function value, u shows the method's iteration counter, U shows the maximum count of algorithm iterations, Y_j^{Q2} shows the novel proposed location associated with the j^{th} kookaburra on the basis of the second stage of KOA, and $y_{j,e}^{Q2}$ shows its e^{th} dimension.

After updating every Kookaburra's position using the data from the first as well as second stages, the initial iteration associated with KOA is finished. The optimal solution found up to that point in the iteration is updated and preserved at the conclusion of each one. The method next moves on to the following iteration on the basis of the updated locations and the newly assessed values for the goal function. Kookaburra positions are updated until the final iteration of the method, which is dependent on Eqs. (12) - (16). The optimal candidate solution found during the method's iterations is ultimately offered as KOA's suggested remedy for the issue. Algorithm 1 presents the KOA implementation phases as pseudo code.

Algorithm 1: *KOA*

Begin KOA

Input problem data: fitness function (error minimization), variables, and conditions (extracted features of the proposed FER prediction model)

Set iterations U and KOA population size O

$$y_{j,e} = LB_e + s \cdot (UB_e - LB_e)$$

Evaluate the fitness function (error minimization of the proposed FER prediction model)

For $u = 1\ to\ U$

For $j = 1\ to\ O$

$DQ_j = \{Y_l: G_l < G_j\ and\ l \neq j\}, in\ which\ j = 1,2,\cdots,O\ and\ l\epsilon\{1,2,\cdots,O\}$

Select the prey for the j^{th} *KOA individual in a random manner*

$$y_{j,e}^{Q1} = y_{j,e} + s \cdot \left(TDQ_{j,e} - J \cdot y_{j,e}\right), j = 1,2,\cdots,O, and\ e = 1,2,\cdots,n$$

$$Y_j = \{Y_j^{Q1}, G_j^{Q1} < G_j\ \ Y_j, otherwise$$

$$y_{j,e}^{Q2} = y_{j,e} + (1 - 2s) \cdot \frac{(UB_e - LB_e)}{u}, j = 1,2,\cdots,O, e = 1,2,\cdots,n, and\ u = 1,2,\cdots,U$$

$$Y_j = \{Y_j^{Q2}, G_j^{Q2} < G_j\ \ Y_j, otherwise$$

End

Save the optimal candidate solution attained so far

End

Output the best quasi-optimal solution attained with the KOA (best predicted value of the proposed FER prediction model)

End KOA

Stop

Table 1. Accuracy analysis

Methods	Iterations				
	10	20	30	40	50
DBN + QPSO [11]	93.6	94.5	95.4	96.3	97.2
GDP [12]	90.1	91.0	92.9	93.8	94.6
CNN [13]	94.8	95.6	96.5	97.4	98.3
WKELM [21]	91.3	92.2	93.0	94.9	95.8
Proposed EGRU-KOA	95.4	96.3	97.2	98.0	99.9

Table 2. Accuracy analysis for various datasets

Datasets	Accuracy (%)
JAFFE	91.09
CK	94.67
AffectNet	65.01
MUG	97.09
FER-2013	98.1
Emotic dataset	99.9

Table 3. MSE analysis

Methods	Iterations				
	10	20	30	40	50
DBN+QPSO [11]	1.71	1.60	1.49	1.14	1.12
GDP [12]	1.79	1.68	1.57	1.46	1.35
CNN [13]	1.92	1.81	1.70	1.59	1.48
WKELM [21]	1.73	1.62	1.51	1.40	1.29
Proposed EGRU-KOA	0.61	0.50	0.39	0.28	0.17

4. Result

4.1 Experimental setup

Training and testing of the developed framework has been carried out on two widely used datasets. The MATLAB platform is used to perform scenario experiments. The specifics of the databases that are being examined for simulation are first investigated under this simulation research. The specifics of the simulation metrics—which are used to gauge performance—are then examined. Lastly, a thorough comparison study is presented to demonstrate how well the suggested FER approach performs.

4.2 Accuracy analysis

This section explains the accuracy analysis of the proposed EGRU-KOA for the FER prediction methodology as shown in Table 1. It can be demonstrated clearly that the accuracy of the proposed EGRU-KOA outperforms the other methods thereby revealing its superiority for the proposed FER prediction model. The proposed EGRU-KOA in terms of accuracy is 2.78%, 5.60%, 1.63%, and 4.28% better than DBN+QPSO, GDP, CNN, and WKELM respectively. Hence, it can be concluded that the proposed EGRU-KOA is better than the other conventional methods with respect to accuracy. The accuracy can be measured as follows.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{17}$$

In the Eq. (17), the terms TP, TN, FP, and FN describe the true positive, true negative, false positive, and false negative respectively. Table 2 elaborates the comparison of the distinct datasets for the FER model in terms of accuracy.

4.3 MSE analysis

The MSE analysis of the suggested EGRU-KOA for the FER prediction approach is explained in this section as given in Table 3. The advantage of the suggested EGRU-KOA for the proposed FER prediction model can be readily shown from the way its MSE beats the other approaches. The proposed EGRU-KOA with respect to MSE is 84.82%, 87.41%, 88.51%, and 86.82% better than DBN+QPSO, GDP, CNN, and WKELM respectively. Thus, it can be said that in terms of MSE, the suggested EGRU-KOA performs better than the other traditional approaches. This is one way to measure the MSE:

$$MSE = \frac{1}{o}\sum_{j=1}^{o} \left(Z_j - \hat{Z}_j\right)^2 \tag{18}$$

In the Eq. (18), the predicted values are shown by $\hat{Z}_j$, observed values are shown by Z_j, and count of data points are shown by o respectively.

4.4 MAE analysis

This section explains the MAE analysis of the proposed EGRU-KOA for the FER prediction method as shown in Fig. 4. The manner that the recommended EGRU-KOA outperforms the alternative methods' MAE clearly demonstrates the advantage of this technique for the proposed FER prediction model. Therefore, it can be concluded that the proposed EGRU-KOA outperforms the other conventional methods in terms of MAE. Here's one method for calculating the MAE:

$$MAE = \frac{\sum_{j=1}^{o} |z_j - y_j|}{o} \tag{19}$$

In the Eq. (19), total count of data points is shown by o, true value is shown by y_j, and prediction is shown by z_j respectively.

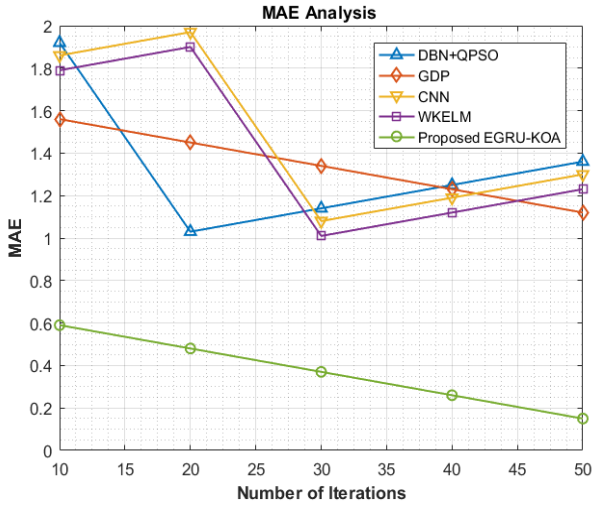

Figure. 4 MAE analysis

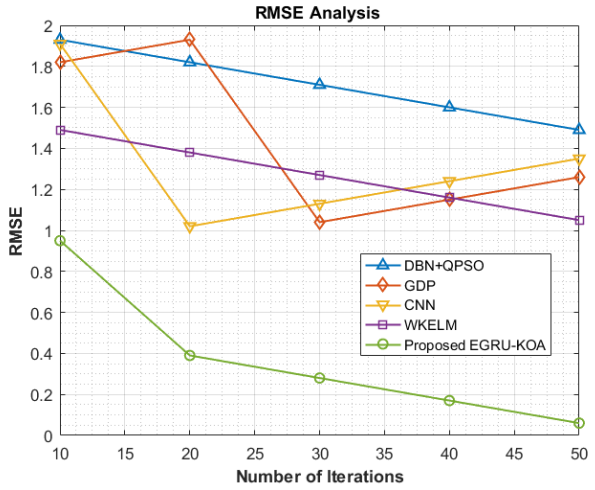

Figure. 5 RMSE analysis

4.5 RMSE analysis

The RMSE analysis of the suggested EGRU-KOA for the FER prediction technique is explained in this section as given in Fig. 5. The way in which the suggested EGRU-KOA surpasses the RMSE of the alternative approaches amply illustrates the benefit of this methodology for the suggested FER prediction model. Thus, it can be said that, in terms of RMSE, the suggested EGRU-KOA performs better than the other traditional approaches. This is one way to figure out the RMSE:

$$RMSE = \sqrt{\frac{1}{o}\sum_{j=1}^{o} \left(z_j - \hat{z}_j\right)^2} \quad (20)$$

In the Eq. (20), the count of observations is shown by o, the predicted value is shown by $\hat{z}_j$, and the actual value is shown by z_j respectively.

4.6 Discussion

A comprehensive comparative analysis is provided to show the effectiveness of the proposed FER method. In terms of accuracy, the suggested EGRU-KOA outperforms DBN+QPSO, GDP, CNN, and WKELM by 2.78%, 5.60%, 1.63%, and 4.28%, respectively. Thus, it can be said that in terms of accuracy, the suggested EGRU-KOA is superior to the other traditional approaches. In terms of MSE, the suggested EGRU-KOA outperforms DBN+QPSO, GDP, CNN, and WKELM by 84.82%, 87.41%, 88.51%, and 86.82%, respectively. Therefore, it can be concluded that the proposed EGRU-KOA outperforms the other conventional methods in terms of MSE. Overall, the proposed FER model outstands other traditional methods with respect to several considered metrics as shown in the results respectively.

5. Conclusion

In this research, unique intelligent deep learning technology was used to do the FER prediction. FER 2013 and the EMOTIC dataset were two common benchmark sources from which the data was first collected. Then, face detection, rotation rectification, and LBP completed the pre-processing of the collected images. The CBP approach was used to extract features from the pre-processed images. These extracted features proceed through the final prediction step, which was carried out by the innovative EGRU. Error minimization was the primary objective function, and the parameters of GRU were adjusted using the nature-inspired optimization method known as KOA. This new EGRU-KOA predicted the result by considering several facial emotions, including sadness, surprise, disgust, contempt, happiness, and fear in that order. By contrasting it with other standard models in terms of distinct analysis, respectively, the suggested model's originality was illustrated. The proposed EGRU-KOA in terms of accuracy was 2.78%, 5.60%, 1.63%, and 4.28% better than DBN+QPSO, GDP, CNN, and WKELM respectively. Similarly, the proposed EGRU-KOA with respect to MSE was 84.82%, 87.41%, 88.51%, and 86.82% better than DBN+QPSO, GDP, CNN, and WKELM respectively for the considered FER model.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Mr. Anand M- "Research proposal - construction of the work flow and model - Final Drafting - Survey of Existing works - Improvisation of the proposed model"; Dr. Babu S- "Initial Drafting of the paper - Collection of datasets and choice of their suitability - Formulation of pseudocode."

References

[1] B. C. Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information", *Sensors*, Vol. 18, No. 2, pp. 401, 2018.

[2] M. I. Georgescu, R. T. Ionescu, and M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition", *IEEE Access*, Vol. 7, pp. 64827-64836, 2019.

[3] Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial Expression Recognition: A Survey", *Symmetry*, Vol. 11, No. 10, pp. 1189, 2019.

[4] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion aware facial expression recognition using CNN with attention mechanism", *IEEE Transactions on Image Processing*, Vol. 28, No. 5, pp. 2439-2450, 2018.

[5] F. Ahmed and M.H. Kabir, "Facial expression recognition under difficult conditions: A comprehensive study on digital directional texture patterns", *International Journal of Applied Math and Computer Science*, Vol. 28, No. 2, pp. 399-409, 2018.

[6] S. Minaee and A. Abdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network", *arXiv preprint arXiv:1902.01019*, Vol. 2019, 2019.

[7] R. Vedantham, L. Settipalli, and E. S. Reddy, "Real Time Facial Expression Recognition in Video Using Nearest Neighbor Classifier", *International Journal of Pure and Applied Mathematics*, Vol. 18, No. 9, pp. 849-854, 2018.

[8] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013", *Advances in Hybridization of Intelligent Methods*, pp.1-16, 2018.

[9] Samuel Manoharan, J. "Geospatial and Social Media Analytics for Emotion Analysis of Theme Park Visitors using text mining and GIS", *Journal of information technology*, Vol. 2, No. 2, pp: 100 - 107, 2020.

[10] Manoharan, Samuel, and Narain Ponraj, "Analysis of Complex Non-Linear Environment Exploration in Speech Recognition by Hybrid Learning Technique", *Journal of Innovative Image Processing (JIIP)*, Vol. 2, No. 04 pp: 202-209, 2020.

[11] K. A. El Dahshan, E. K. Elsayed, A. Aboshoha, and E. A. Ebeid, "Recognition of Facial Emotions Relying on Deep Belief Networks and Quantum Particle Swarm Optimization", International *Journal of Intelligent Engineering and Systems*, Vol. 13, No. 4, 2020, doi: 10.22266/ijies2020.0831.09.

[12] V. Lakshmi and P. Mohanaiah, "Gradient-Based Compact Binary Coding for Facial Expression Recognition", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 6, 2021, doi: 10.22266/ijies2021.1231.34.

[13] R. Agarwal, N. Mittal, and H. Madasu, "Convolutional Neural Network Based Facial Expression Recognition Using Image Filtering Techniques", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 5, 2021, doi: 10.22266/ijies2021.1031.08.

[14] A. Hesham Mostafa, H. Abdel-Galil El-Sayed, and M. Bela, "Facial Expressions Recognition Via CNNCraft-net for Static RGB Images", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 4, 2021, doi: 10.22266/ijies2021.0831.36.

[15] D. Liu, W. Dai, H. Zhang, X. Jin, J. Cao, and W. Kong, "Brain-Machine Coupled Learning Method for Facial Emotion Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 9, pp. 10703-10717, 2023.

[16] J. Guo et al., "Dominant and Complementary Emotion Recognition from Still Images of Faces", *IEEE Access*, Vol. 6, pp. 26391-26403, 2018.

[17] K. Chen, X. Yang, C. Fan, W. Zhang, and Y. Ding, "Semantic-Rich Facial Emotional Expression Recognition", *IEEE Transactions on Affective Computing*, Vol. 13, No. 4, pp. 1906-1916, 2022.

[18] Y. Yang, Q. Gao, Y. Song, X. Song, Z. Mao, and J. Liu, "Investigating of Deaf Emotion Cognition Pattern by EEG and Facial Expression Combination", *IEEE Journal of Biomedical and Health Informatics*, Vol. 26, No. 2, pp. 589-599, 2022.

[19] X. Liu, X. Cheng, and K. Lee, "GA-SVM-Based Facial Emotion Recognition Using Facial Geometric Features", *IEEE Sensors Journal*, Vol. 21, No. 10, pp. 11532-11542, 2021.

[20] Y. -C. Wu, L. -W. Chiu, C. -C. Lai, B. -F. Wu and S. S. J. Lin, "Recognizing, Fast and Slow: Complex Emotion Recognition with Facial Expression Detection and Remote Physiological Measurement", *IEEE Transactions on Affective Computing*, Vol. 14, No. 4, pp. 3177-3190, 2023.

[21] M. Anand and S. Babu, "Multi-class Facial Emotion Expression Identification using DL based Feature extraction with Classification Models", *International Journal of Computational Intelligence Systems*, Vol. 17, No. 25, pp. 1-17, 2024.