Iranian Journal of Fuzzy Systems

Volume 22, Number 4, (2025), pp. 57-75

Original paper

University of Sistan and Baluchestan

Novel enhanced cognitive state analysis in E-learning via real-time emotion and attentiveness detection using OptFuzzy TSM and ABiLSTM

J. Vaniya [1], M. Alizada [2], P. Nagpal [3], B. Kumar Dey [4] and G. Alesger Abbbasova [5]

[1] *Department of Information Technology, Vishwakarma Government Engineering College, Chandkheda, Gujarat Technological University, Gujarat 382424, India.*

[2] *Department of History, Western Caspian University Baku, Istiglaliyyet 31, Baku, Azerbaijan, SC ID60190589*

[3] *Faculty of Management Studies, CMS Business School, JAIN (Deemed to be University), Bangalore, Karnataka, 560009, India.*

[4] *Department of Commerce, Shaheed Bhagat Singh College, University of Delhi, New Delhi, 110017, India.*

[5] *Department of Social Sciences, Azerbaijan University of Architecture and Construction, Ayna Sultanova, 11, Baku, Azerbaijan, SC ID 60108958*

jignesh.vaniya@vgecg.ac.in, elizade.meleyke@wcu.edu.az, dr.pooja_nagpal@cms.ac.in, biplabkumar@sbs.du.ac.in, gulara.abbasova@azmiu.edu.az

Abstract

The emotional state of an online learner has drawn a lot of attention. Accurately predicting a student's emotional state can improve learning outcomes through designated mediation. Still, keeping an eye on and sustaining student's attention in online classes is challenging because there isn't any immediate supervision. To identify these challenges based on the learner's emotional states, this paper presents a novel, efficient, Optimized Fuzzy approach and signifies solutions to inspire the learner. The Improved Multi-Task Cascaded Convolutional Networks (IMTCNN) are used to identify the face region in real-time. Different emotions are classified by analyzing extracted facial expressions using an Optimized Takagi-Sugeno and Mamdani fuzzy systems (Fuzzy TSM) approach. With the Enhanced Mother Optimization Algorithm (EMO), the hyperparameters in the classification approach are optimized. The proposed method determines whether learners are attentive or inattentive during online learning sessions by computing an Attention-based bi-directional Long-Short Term Memory (ABiLSTM) to predict cognitive states. To improve learning efficiency and productivity, users receive real-time feedback. The proposed approach can give instructors ongoing feedback, allowing them to modify the way they teach and keep students interested and engaged. With recognition rates of over 98.21

Keywords: E-learning, cognitive state, dual Mamdani and neuro-fuzzy inference system, Archerfish hunting optimization algorithm (AHOA), improved position enhancement faster network (IPEFNet).

1 Introduction

E-learning platforms cannot often capture a learner's sentiments and modify the course material. The efforts of the teacher and their instructional assistants can be rendered ineffective if students lose interest or become disinterested during crucial lectures or courses [13]. Several interactive technologies for learning were recently developed to address this issue. Therefore, it is essential to create a system that can precisely evaluate an individual's mental condition in real time within a virtual classroom. Here, the student's mental attitude during their study sessions is referred to as their "cognitive state." Assessing the cognitive state of the learner is essential in an adaptive educational situation in order to ascertain their level of involvement [18].

Learning effectiveness will decline if the student disengages, so appropriate feedback is given immediately. For online learners to receive a well-rounded education, it is crucial to integrate aspects of traditional classroom settings [3]. For more than 150 years, scientists have been interested in using facial expressions to communicate emotional responses.

Corresponding Author: J. Vaniya
Received: October 2024; Revised: July 2025; Accepted: August 2025.
https://doi.org/10.22111/ijfs.2025.49950.8829

Facial feature analysis is a popular method for analyzing facial expressions [15] and describing the emotions behind them. It can distinguish between the learner's cognitively attentive and inattentive states. The automatic identification of emotions from facial expressions in computer vision has gained attention due to the need for more sophisticated human-machine interfaces [12]. Real-time automatic emotion recognition is difficult and requires complex algorithms to recognize and understand facial emotions. However, advancements in the field of computer vision have enabled the development of more efficient and effective Facial Emotion Recognition (FER) systems [2, 14]. Through precise recognition and interpretation of a student's facial expressions, this kind of system can dynamically modify the course content to improve student performance and sustain interest [1]. This would make it possible for teachers to create more individualized and successful learning experiences catered to each student's needs and preferences. It is possible to modify the content of online lectures to increase learners' engagement with the system by taking advantage of the strong correlation our results showed between learners' expression and their degree of interest [7, 9].

Our work presents a novel cognitive state detection system that performs exceptionally well in learner engagement prediction and real-time monitoring. Through the integration of sophisticated techniques for both facial emotion analysis and feature optimization, the system attains remarkable accuracy and offers prompt feedback, thereby augmenting learner engagement and instructional efficacy. This innovative method provides a more dependable and flexible solution for online learning environments, significantly outperforming current approaches. The significant key contributions of this research are as follows,

• Improved Multi-Task Cascaded Convolutional Networks (IMTCNN) are used to improve facial emotion detection and the accurate real-time identification and alignment of facial regions.

• For effective feature extraction, the Improved Res2Net model captures significant facial features, which are then selected using the Improved Binary Crayfish Optimization Algorithm (IBCOA).

• Optimized Takagi-Sugeno and Mamdani Fuzzy Systems (Fuzzy TSM) are employed to classify emotional states accurately, with hyperparameters optimized by the Enhanced Mother Optimization Algorithm (EMO), providing precise classification and tuning for better emotion recognition.

• To predict cognitive states and provide feedback, Attention-based Bi-directional Long-Short Term Memory (ABiLSTM) is utilized, offering real-time prediction and feedback that enhances learner engagement and supports adaptive instructional strategies.

• To ensure high efficiency, the entire system is optimized for real-time processing, combining advanced detection and classification techniques with a streamlined feedback mechanism. This allows for quick adaptation and accurate predictions during online learning sessions, making the approach highly effective in maintaining and enhancing learner engagement.

Paper Organization: This research's remaining sections are organized as follows: Part 2 compiles the existing studies conducted in the field. Part 3 discussed the datasets and methods used in this investigation. The efficiency of the proposed method is compared with related attempts in Part 4, and the advantages and disadvantages are listed along with suggestions for improvement in Section 5.

2 Review of existing related works

Offering students immediate support according to their attention levels is critical to a successful e-learning environment. Many researchers are attempting to identify students' learning states by tracking their mental states in real time. Determining whether students are attentive or inattentive is crucial in ensuring that they remain engaged during online courses and that the retention rate rises. There are numerous ways to monitor students' attentiveness in an online learning environment.

Gupta, et al. [5] proposed a multimodal system using facial expressions, eye-blink counts, and head movements via live video to assess student engagement in e-learning, employing VGG-19, ResNet-50, and facial landmark techniques. Rathi, et al. [16] introduced a Deep LSTM model optimized using Squirrel Search and Rider Optimization (SS-ROA) to classify learner states like frustration and involvement based on interaction logs. Kukkar, et al. [8] developed the SAPP model combining a 4-layer LSTM, Random Forest, and Gradient Boosting to predict academic performance. Ma, et al. [10] applied a neutrosophic cognitive diagnosis with NS similarity and collaborative filtering to predict learner success while addressing data sparsity. Benabbes, et al. [4] used BiLSTM with FastText embeddings to detect learner mindsets in forum discussions, followed by unsupervised clustering. Ma, et al. [11] presented a fuzzy cloud cognitive diagnostic framework (FC-CDF) using normal cloud models to assess uncertainty in skill proficiency. Zuo, et al. [19] introduced GUGEN, leveraging global and user-based graphs for enhanced POI recommendations. Huang, et al. [6] proposed XKT, a knowledge tracing model based on MIRT and cognitive learning theory, integrating multi-feature embedding for rich semantic understanding. Shi, et al. [17] developed UGRIE, a unified framework for image emotion

classification that blends BART and CLIP models via a flexible natural language template for improved performance. Table 1 summarizes these related works.

Table 1: Summary of Related works

References	Strengths	Limitations
Gupta et al. [5]	Multiple modalities used (facial expressions, eye blinking, head movements), leveraging deep learning techniques	Limited to facial expressions and basic head movements, no real-time emotion analysis
Rathi et al. [16]	Integration of Squirrel Search and Rider Optimization with Deep LSTM for dynamic state prediction	Relies on interaction logs, lacks real-time emotion analysis
Kukkar et al. [8]	Predicts academic performance using multi-layered deep learning techniques like LSTM, RF, and GB	Focused on academic predictions, does not address emotional or engagement states
Ma et al. [10]	Uses Neutrosophic Cognitive Diagnosis and collaborative filtering to predict student success	Not focused on real-time engagement or emotional state of learners
Benabbes et al. [4]	BiLSTM with FastText embeddings for better clustering and state detection in forum data	Lacks real-time interaction capability and emotional state detection
Ma et al. [11]	Combines fuzzy cloud models for skill proficiency prediction with uncertainty representation	Does not address emotional state detection or real-time engagement
Zuo et al. [19]	Focuses on both global and local perspectives for POI recommendations, leveraging trajectory learning	Not related to emotional state detection or e-learning environments
Huang et al. [6]	Improves knowledge tracing accuracy and interpretability using MIRT and a neural network approach	Does not focus on emotional state or engagement monitoring
Shi et al. [17]	Unified generative framework for emotion classification, capable of handling multiple emotion models	Focuses on image emotion recognition, not on real-time learning contexts

2.1 Research gap

Most current e-learning models process emotional and cognitive feedback separately, lacking a unified approach to fully capture student engagement. This separation hinders real-time understanding and support. Additionally, standardized techniques overlook individual learner differences, resulting in generic, less effective feedback. Many existing systems also struggle with real-time responsiveness due to high processing costs and latency, making them impractical for live sessions. Their reliance on costly, external sensors like EEGs further limits scalability and accessibility. These models often ignore context-such as task complexity or learning preferences-leading to inaccurate predictions and ineffective interventions. Subtle emotional cues like confusion or hesitation are frequently missed, and complex setups limit deployment across diverse platforms. To address these gaps, this study proposes a real-time, context-aware, and non-intrusive framework that integrates emotional and cognitive signals, adapts to learning environments, and emphasizes efficiency, personalization, and scalability for enhanced e-learning outcomes.

3 Overview of proposed methodology

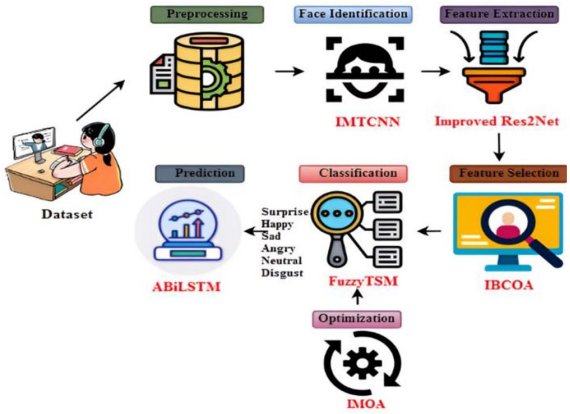


Figure 1: Overall Framework for the proposed methodology

Understanding student engagement, comprehension, and emotional reactions in e-learning is critical to enabling personalized learning experiences and better results. However, problems such as data sparsity, imprecise detection of subtle cognitive states, and restricted flexibility to different learning contexts are frequently encountered by existing approaches. Therefore, we introduced a novel optimized approach to predict and classify the cognitive state in this research. The below sections discuss every phase in detail. The overflow of the proposed study is shown in Figure 1.

3.1 Preprocessing

In cognitive state prediction, preprocessing enhances the quality and relevance of data, thereby laying the basis for efficient analysis and modelling. Ensuring that the dataset is accurate and consistent involves cleaning the data to remove errors, missing values, and outliers. Preprocessing well improves model performance, lowers complexity, and produces predictions and insights into cognitive states that are more accurate.

3.2 Face region identification using improved multitask cascaded convolutional networks (IMTCNN)

MTCNN's distinctive architecture for face detection and alignment makes it perfect for face region identification in cognitive state prediction. Entire face detection, bounding box regression, and facial landmark localization are achieved by combining three convolutional network stages. This multitasking approach provides accurate facial landmarks for additional cognitive state analysis, such as identifying emotions and stress levels. It ensures high accuracy even with scale, pose, and lighting variations. In this study, we propose an improved MTCNN (IMTCNN) framework by enhancing the original MTCNN architecture to better suit our dataset. The model consists of three CNNs—P-Net, R-Net, and O-Net. P-Net processes $12\times12\times3$ images to extract approximate facial regions using 3×3 convolutional kernels, followed by pooling, normalization, and ReLU activation. R-Net and O-Net follow a similar structure, with O-Net handling $48\times48\times3$ images to perform accurate bounding box regression and facial landmark localization. Across all networks, ReLU is used as the activation function, and the final layers include fully connected and softmax layers that convert features into 256-dimensional and 16-dimensional vectors, respectively. The output layer predicts 16 key points, including facial landmarks (eyes, nose, mouth), bounding box coordinates, and face classification. Parameter settings for all three networks were selected through consistent experimental procedures to ensure optimal performance. The parameter details of O-Net are illustrated in Table 2.

3.2.1 The whole procedure of IMTCNN

Subsequently, the testing set samples from the suggested datasets are tested using the model that was trained. Figure 2 shows how the IMTCNN algorithm's process chart is used. First, the P-Net can extract the approximate facial windows from the input face image. Second, using R-Net, the precise face region is labelled. The coordinates of the five facial characteristics are then labelled using O-Net.

3.3 Feature extraction

By incorporating a multi-scale representation into every residual block, the Improved Res2Net model improves feature extraction and can continuously capture fine details and larger patterns. It efficiently learns intricate features by breaking the feature map input into smaller sections and analyzing them through convolutions with different receptive fields. This method improves the model's effectiveness in identification and categorization by enhancing its capacity to extract pertinent information from images.

3.3.1 Res2NetModel

Res2Net performs better in terms of generalization than ResNet. The residual structure of cells of the framework incorporates hierarchical small residual blocks, thereby augmenting the effective sensory field of every layer and enhancing the overall network's extracted feature efficiency. We chose the Res2Net model as the foundation network for the antler slice categorization task because of its ability to extract significant characteristics from such images, considering the small size and challenges of identifying deer antler slices.

3.3.1.1. Improved RES2NET model

Grouped Convolution

By splitting the input feature map into smaller groups and convolving each group independently using a different set of filters, the approach known as "grouped convolution" greatly reduces computing complexity. By reducing the number of parameters, this method uses memory more effectively and speeds up calculations. Usually, the input is segmented according to its channels, and different kernels are applied to each group of channels during the convolution process. Deep neural networks benefit greatly from this technique since it enables them to capture richer feature representations with fewer parameters, increasing efficiency without sacrificing performance. When performing group convolution, distinct kernels are applied to every group of feature input maps. The parameters for convolution are "W" for width, "H" for height, and "C" for the amount of channels. This divides the input image into two groups.

$$\text{Size of t?e images} = H \times W \times \frac{c}{2}, \tag{1}$$

$$\text{Size of t?e output images} = H' \times W' \times \frac{c'}{2}, \tag{2}$$

$$\text{Number of Participants} = H \times W \times C \times \frac{c'}{2}, \tag{3}$$

$$\text{Volume of Operations} = H \times W \times C \times C' \times W' \times \frac{H'}{2} \tag{4}$$

In addition to increasing computational speed and filter correlation, grouped convolution dramatically reduces parameters. One way to prevent overfitting is to reduce the number of parameters by grouping the input feature maps into four groups.

Improved grouped convolution

Improved Grouped Convolution improves feature extraction and learning by including an inverse bottleneck framework into conventional grouped convolution algorithms. The framework is composed of three stages: an initial 1×1 convolution to improve input dimensions before to the main convolutions, a 3×3 convolution for efficient feature extraction, and a 1×1 convolution for restoring the input dimensionality. The model's expressive power is greatly increased by the design's use of a larger middle layer with two smaller ends. Even with a growing number of channels, this novel topology enables the network to effectively handle the development of parameters while capturing more complicated aspects. Improved Grouped Convolution balances expressive strength and computational economy, boosting the network's ability to learn and extract features, which improves overall performance.

Attention mechanism

Attention mechanisms, modelled after human cognition and efficiently filter pertinent data, improve deep learning approaches. With the size of the convolution kernel proportionate to the channel dimension, this study optimizes cross-channel interactions to enhance the extraction of features in image processing tasks. This is achieved through the use of Efficient Channel Attention (ECA).

$$C = \varphi(k).$$

The exponential operation of two is typically used to set the channel dimension C. The linear function(k) $= 2^{(\gamma * k - b)}$ is extended to a nonlinear function to present a potential solution.

$$C = \varphi(k) = 2^{(\gamma * k - b)}. \tag{5}$$

The parameter γ in Equation (5) is a scaling or growth factor that controls the rate at which the function$\varphi(k)$rises or falls in relation to the variable k. It controls the exponential increase or decay of the output C by directly affecting the exponent in the expression $2^{(\gamma * k - b)}$. The kernelk size can adaptively determine the convolution size k size given the dimension of the channel C.

$$k = \varphi(C) = OddRound\left(\frac{\log_2 C + b}{\gamma}\right). \tag{6}$$

In this case, b is a bias term that is utilized to modify the value, γ is a scaling factor, and C is a positive integer input parameter. In order to maintain symmetrical and effective group partitioning in convolutional processes, the function $OddRound(\cdot)$ rounds the calculated result to the next odd integer, and the logarithmic term $\log_2 C$ ensures scale adaptability.

3.4 Feature selection

The goal of Improved Binary COA (IBCOA) feature selection in cognitive state prediction is to find the most pertinent features for accurate categorization. By decreasing the number of features, IBCOA improves classification performance while streamlining the model, accelerating training, and enhancing interpretability. IBCOA contributes to more accurate and effective cognitive state predictions by concentrating on the most informative features.

3.4.1 Proposed improved binary COA (IBCOA) for feature selection

Enhancing classifier accuracy often involves focusing on the most relevant features. Feature selection (FS) aims to reduce high-dimensional data by retaining only essential attributes, discarding irrelevant ones. FS operates in a binary manner, marking important features as 1 and unimportant ones as 0. This study introduces the Improved Binary Crayfish Optimization Algorithm (IBCOA) for effective FS. The IBCOA framework includes key phases: population initialization, binary conversion, position updating using the Crayfish Optimization Algorithm (COA), and improved exploration and exploitation strategies. Each phase is designed to refine feature selection and boost classification performance.

3.4.1.1. Binary encoding of crayfish position

An exceptional global exploration capability in the form of a V-shaped transfer function, expressed as

$$\nu(k) = \alpha \times \frac{\arctan(x) \times \dfrac{\pi}{\sqrt{1+x^2}}}{\pi}. \tag{7}$$

It is ensured that the transfer function result is within $[0, 1]$ if $\alpha < 0.64$, where x is the constant spot value acquired for the Crayfish position. Next, it is defined that an upgrade rule for the IBCOA's position is

$$p_i^{bin} = \begin{cases} 1 & \text{if } rand < \nu(p_i) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Which $\nu(p_i)$ represents a transfer function that converts the real-valued p_i to a probability, and $rand \in [0,1]$ is a uniformly distributed random number. The sigmoid function is specifically employed:

$$\nu(p_i) = \frac{1}{1+e^{-p_i}}. \tag{9}$$

A smooth transition from continuous to binary space is provided by this function, where extreme values drive the decision toward a more deterministic result, while values closer to zero offer selection probabilities that are close to 0.5, enabling exploration. In the binary feature space, this encoding technique effectively strikes a balance between exploration and exploitation, improving IBCOA's search capabilities.

3.4.1.2. Initial population generation

The first stage is to create the initial individuals of N positions for the IBCOA, representing possible solutions for selecting likelihood in a dimensional space within the range $[0,1]$. The number of people is calculated using

$$N = Round\left(10 + 2 \times \sqrt{dim()}\right). \tag{10}$$

In this case, N represents the entire population size, which is dynamically calculated according to the feature space's dimensionality. The number of features or decision variables in the dataset is indicated by the variable dim. The final value is rounded to the closest integer using the $Round(\cdot)$ function after the formula scales the population size with problem complexity using a square-root-based heuristic.

3.4.1.3. Position update in IBCOA

Position enhancing uses the COA formulas described in the third section. After that, the fitness function analyzes the binary-transformed vector to find the errors in classification while preserving the initial structure of the vector for future enhancements.

3.4.1.4. Fitness function evaluation

A fitness function is formulated to balance minimal feature selection with maximal classifier accuracy, as mathematically expressed below.

$$f = \omega_1 \times (1 - \text{accuracy}) + \omega_2 \times \left|\frac{|\ d^*\ |}{|\ D\ |}\right|. \tag{11}$$

In this case, the consequences of the size of the characteristics selected and the efficacy of classification are addressed by $\omega_2 = 1 - \omega_1 \cdot \omega_1$ is a randomized digit among $[0,1] \cdot \omega_2 = 0.01$ and $\omega_1 = 0.99$, using thorough tests from earlier research, is used. Since the goal is to increase accuracy rather than reduce the size of the selected attributes, $\omega_2 < \omega_1$. D stands for the attributes. The above equation defines a fitness function for feature selection optimization. In this case, the goal to be decreased throughout the optimization process is the fitness value f. By balancing the number of features chosen and classification accuracy, the weighting factors ω_1 and ω_2 enable the model to give priority to either limiting the feature subset size or lowering prediction error. While d^* is the number of features chosen during the optimization phase, classification accuracy (accuracy) is a measure of how well the model works with the chosen features. $|\ D\ |$ is a representation of the dataset's total number of features. In order to properly optimize the feature set, the norm operator ($\|\cdot\|$) makes sure that the fitness value stays non-negative.

3.4.1.5. Optimizing exploitation through local search process

This section outlines the local search (LS) principles incorporated into the IBCOA to improve its exploitation capability and overall efficiency. The goal is to generate new individuals by leveraging optimal positions while maintaining the original structure of the algorithm.

$$p^{L+1} = P^L + \beta p^L. \tag{12}$$

Wherein N $(0.0, 0.4)$ is the standard deviation of a random factor represented by β and L is the number of local search iterations. A small mutation causes the resulting solution to differ slightly from the best PG available. When looking

for local search methods with a fixed size threshold ($LS_{\max}$), the best solution is first included in an empty set. Next, a mutated solution is created on the current PG. This solution is then converted to binary, and its fitness is evaluated if it performs better than the previous ones. The computational complexity of the algorithm is represented as

$$O_{time}(IBCOA) = O_{time}(N) + O_{time}\left(FEdim_{\max} + O_{time}\left(FE_{\max}()_{time}\left(FEdim\times_{\max_{\max}} = \mid O_{time}\left(FEdim\times_{\max_{\max}}\right)\right)\right)\right). \tag{13}$$

3.5 Fuzzy classification system

Takagi-Sugeno and Mamdani fuzzy inference systems (FIS) are combined to provide a hybrid fuzzy-based classification method in the proposed cognitive state detection framework. The necessity for clear, understandable, and effective decision-making which is particularly important in educational institutions where feedback is closely related to student behaviour and performance, motivates this design choice. While the Takagi-Sugeno model produces smoother outputs appropriate for incorporation with deep learning networks like IMTCNN and ABiLSTM, the Mamdani method offers a simple rule-based interpretation. Even though solely data-driven deep learning techniques, like Transformers or reinforcement learning models, have shown remarkable effectiveness in challenges involving the prediction of emotions and engagement, they frequently function as complex models with little interpretability. In contrast, fuzzy systems preserve model transparency while managing ambiguous or imprecise inputs that are typical of cognitive and affective states. This combination makes the technology more reliable and applicable in real-time learning environments. Nevertheless, we recognize the progress in fuzzy inference systems, especially the rise of Fuzzy Fuzzy Inference Systems (FFIS), which combine hierarchical rule evolution and self-adaptive learning processes to increase the learning capacity of conventional FIS algorithms.

Takagi-Sugeno (TS) and Mamdani fuzzy systems are integrated into this method to improve the classification of cognitive states in e-learning. The Mamdani fuzzy system can capture broad patterns in learner behaviour because it offers interpretability through human-readable rules. On the other hand, the Takagi-Sugeno system precisely models intricate connections among inputs and outputs through mathematical functions.

The ensemble combines the accuracy and flexibility of the TS model with the interpretability of Mamdani fuzzy logic by combining these two systems. The ensemble method more precisely classifies cognitive states like engagement, confusion, and frustration by processing learner input data, including behavioural and physiological signals. This method offers a robust and adaptable solution that will eventually enhance learner outcomes and engagement in online learning environments by personalizing educational opportunities based on real-time cognitive state monitoring. Intrinsic fuzzy logic systems, Takagi-Sugeno fuzzy systems, and Mamdani fuzzy systems are the most common fuzzy algorithms for real-world engineering. As a result, the investigation's basic unit is a Mamdani-type fuzzy system. This approach implements a fuzzy max-min synthesis process as follows.

$$\mu_B(y) = \underset{x\subset X}{V}[\mu_A(x) \wedge \mu_R(x,y)]. \tag{14}$$

The set of fuzzy values A, B, and R are represented by $\mu_A(x)$, $\mu_B(y)$ and $\mu_R(x,y)$ with $\wedge$ representing the minimum and $\vee$ the maximum. Fuzzy-NMS selects the sum as a result of the synthesis process. Likewise, defuzzification uses a centroid approach, which provides smoother outcome inference than alternatives. In this case, even if the input variable alterations slightly, the result will vary significantly. This procedure can be represented as follows:

$$\nu_0 = \frac{\int_\nu \nu\mu_\nu(\nu)d\nu}{\int_\nu \nu\mu_\nu(\nu)d\nu}. \tag{15}$$

While ν_0 is the center of gravity for the area bounded by the membership function curve and the abscissa ν is the membership function curve and ν is a fuzzy input variable. In the proposed fuzzy-based classification method, the efficacy and interpretability of the fuzzy inference system are greatly influenced by the selection of membership functions. We chose the triangle membership function for this study because of a number of significant benefits it provides for the e-learning environment's cognitive state detection. The triangular membership function's interpretability, computational efficiency, and simplicity make it a popular choice for fuzzy logic systems. It can give a simple representation of membership degrees because a peak and two linear slopes characterize it. This structure allows for rapid computations without compromising accuracy, which makes it especially helpful when working with real-time input in dynamic contexts. The triangular membership function is favored for its sharp peak, enabling it to define clear membership boundaries-an essential feature for distinguishing varying learner cognitive states. It integrates effectively with fuzzy logic rules applied to deep learning models like ABiLSTM and IMTCNN. While Gaussian and trapezoidal functions are also common in fuzzy systems, the triangular function offers an optimal balance between interpretability and

computational efficiency. This makes it ideal for real-time learning environments that demand fast and transparent decision-making. A fuzzy statistical method is used to determine input and output membership functions. The sample value u is adjusted, and the fuzzy set $A*A*$ is updated. After n evaluations, the membership frequency PA of u with respect to A is given by $PA = \tau A/n$, where τA is the number of times u belongs to $A*A*$. The triangular membership function is then defined using the parameters a, b, and c.

$$f(x,a,b,c) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ \dfrac{c-b}{c-x} & b \le x \le c \\ 0 & x > c \end{cases} \tag{16}$$

The "foot" of the triangle is determined by a and c, the "peak" by b, the input to the membership function is represented by x, and the value of the membership is defined by $f(x,a,b,c)$. Multiple dimensions fuzzy rules are then used to create a fuzzy rule library. Depending on past knowledge, a fuzzy rule table is also made to ensure that fuzzy rules are full. The fuzzy system is subsequently utilized to forecast bounding boxes by counting the data's inherent mathematical properties as previous knowledge for categorization.

3.6 Enhanced mother optimization algorithm

It has been recognized that families play a special educational role in society, with mothers playing a particularly significant role in raising children. The mother uses the life lessons and advice she has received to train and improve her children. Children and their mothers have three different kinds of relationships: instruction, upbringing, and learning. Thus, the suggested MOA design predates the care and habits teaching mathematical framework.

3.6.1 Mathematical model of MOA

This population-based meta-heuristic algorithm solves optimization problems through an iterative approach. The algorithm's population consists of potential solutions. The population was started using Eq. (17) at the start of the optimization process, and it was modelled using the matrix in Eq. (18). Each member calculated the values of the decision parameters based on their location inside the issue's search scope.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_j \\ \vdots \\ X_N \end{bmatrix}_{N\times d} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,i} & \cdots & x_{1,d} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{j,1} & \cdots & x_{j,i} & \cdots & x_{j,d} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,i} & \cdots & x_{N,d} \end{bmatrix}_{N\times d}. \tag{17}$$

$$x_{j,1} = lb_i + rand(0,1) \times (ub_i - lb_i),\ j = 1,2,\cdots,N,\ i = 1,2,\cdots,d. \tag{18}$$

The variables $x_{N,d}$ represent the individual matrix of the proposed MOA, the number of individuals, the decision variables' volume, and the candidate's j^{th} solution. The value $x_{j,1}$ demonstrates the i^{th} variable for which the probability function generates a chaotic amount among $[0,1]$ and $ub_i lb_i$, in turn, indicates the upper and lower bounds of the i^{th}.

Every member of the collection can contribute a method to the problem that is being improved in MOA. The values that every individual suggests for the decision factors are added up to determine the cost function of the problem. Eq. (19) provides a mathematical representation of the functions of cost values in a vector's properties.

$$V = \begin{bmatrix} V_1 \\ \vdots \\ F_j \\ \vdots \\ F_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} V(X_1) \\ \vdots \\ F(X_j) \\ \vdots \\ F(X_N) \end{bmatrix}. \tag{19}$$

The value for the j^{th} relevant solution is represented by V and V_1 makes up the vector of cost function values. The cost function's values evaluate the calibre of the solutions that each individual's members provide. Furthermore, by

analyzing the worst and best values of the cost function, it is possible to determine the population is the best and the poorest. While the positions of individuals in the community are enhanced with each iteration, the best individual must be updated as well. As a result, the best member of the community can solve the problem in the last iteration of the method. The number of individuals in the MOA design has been enhanced in three areas, considering the mathematical concept of mother-child growth through interaction. Moreover, it is covered in detail in the sections that follow.

1^{st} stage: Exploration

The children's education with the suggested method served as the inspiration for this phase. This research aims to significantly change people's whereabouts to enhance exploration and global search capabilities. This behaviour has been used to replicate this phase because the mother is thought to be the best member and provides the children with excellent instruction. Every person has been given an alternate position by applying Equation (20) in this step.

$$x_{j,i}^{P1} = x_{j,i} + rand(0,1) \times (D_i - rand(2) \times x_{j,i}), \tag{20}$$

$$X_j = \begin{cases} X_j^{P1} & V_j^{P1} \leq F_j \\ X_j & \text{else} \end{cases} \tag{21}$$

The location update and selection methods of a metaheuristic algorithm are described by formulas (20) and (21) respectively. By perturbing the existing solution $x_{j,i}$ with a guiding vector Di and random components to induce exploration, formula (20) produces a new candidate solution X_j^{P1}. In order to ensure performance-driven evolution, formula (21) then uses a greedy selection rule, where the new candidate X_j^{P1} replaces the existing solution X_j only if it achieves a superior or equivalent fitness value $V_j^{P1} \leq F_j$.

2^{nd} stage: Exploration

A mother's most important role in raising her children is to advise them, not enable them to misbehave. The second stage of the individuals' algorithmic update employed this advisory method. This phase modifies the individual's position to a significant extent, facilitating MOA in exploration and global search. Any population member's spot that is surpassed by others with a higher cost function value is regarded by the MOA framework as an abnormal behaviour that needs to be avoided. Every individual's bad behaviour (BBi) was ascertained through a comparative analysis of the cost function values. Employing a uniform distribution, a member is chosen randomly from the set of undesirable behaviours (BBi) for every Xi.

3^{rd} stage: Exploration and upbringing

Mothers utilize a variety of ways to encourage their kids to progress in their education. However, by modifying members' locations, parenting helps users increase their capacity to exploit and conduct local searches. Employing Eq. (22), a unique location has been made for each person based on the way children's personalities develop to replicate this stage. As Eq. (23) demonstrated, the cost function assumes the previous position if its value increases at the new location.

$$x_{j,i}^{P3} = x_{j,1} + (1 - 2 \times rand(0,1)) \times \frac{ub_i - lb_i}{t} \tag{22}$$

$$X_j = \begin{cases} x_j^{P3} & V_j^{P3} \leq V_j \\ X_j & \text{else} \end{cases} \tag{23}$$

The new spot for the jth member of the population is represented by $x_{j,i}^{P3}j$, while the Ith dimension is defined by $x_{j,i}^{P3}$. The cost function's value is represented by V_j^{P3} the rand function, generating a random number ranging from $[0, 1]$. Finally, the value of the iteration counter is represented by t.

3.6.1.1. Enhanced MOA

The MO strategy may be improved by updating it, adapting it to a broader range of problem domains, making it more scalable and effective, incorporating novel concepts and techniques, and subjecting it to a thorough empirical analysis. This might lead to the development of a more flexible optimization algorithm that can handle a wider variety of real-world scenarios. Adjustments incorporating efficiency, scalability, and adaptability to specific challenges can address

issues like handling massive problems, optimizing computing resources, and premature convergence. Adding novel concepts and methods, like fusing ideas from different algorithms or applying state-of-the-art optimization techniques, can further enhance the algorithm's adaptability and performance.

Two modifications can be made to the MO algorithm to increase its effectiveness: self-adaptive individuals and Lévy flight. By incorporating long-distance hops into the search space, Lévy flight enables the procedure to break out of local optima and explore new regions. This haphazard, far-reaching movement finds a balance between exploitation and exploration, potentially producing better answers. While:

$$x_{j,i}^{P1} = x_{j,i} + Lf(\delta) \times (D_i - Lf(\delta) \times x_{j,i}) \,. \tag{24}$$

$$Lf(\delta) \cong \frac{1}{\delta^{1+\xi}}. \tag{25}$$

$$\delta_i = \frac{A}{|B|^{\frac{1}{\xi_i}}}. \tag{26}$$

$$\sigma^2 = \left\{ \frac{\Gamma(1+\xi_i)}{\xi_i \Gamma\left((1+\xi_i)/2\right)} \frac{\sin(\pi\xi_i/2)}{2^{(1+\xi_i)}/2} \right\}. \tag{27}$$

Formulas (24) and (25) are adopted from Mantegna's algorithm for generating Lévy flight steps, where the step length δ_i is determined using a scale parameter σ derived from a symmetric Lévy distribution. Equation (26) calculates δ_i using normally distributed random values, while Equation (27) provides the corresponding scale parameter σ^2, derived from the Lévy distribution's statistical properties. The variable ξ in this study has been set to 1.5, which denotes a random value between 0 and 2. The symbol for the Gamma function is $\Gamma(\cdot)$. Step size is represented by the variable w, while the variance (σ^2) and mean value of 0 are indicated by the variables B and A.

$$size(x) = 10 \times Dim. \tag{28}$$

While Dim denotes the issue in dimension. The process to determine the revised number of individuals is as follows:

$$Size(x)_{new} = round(size(x) + \beta \times Size(x)). \tag{29}$$

3.7 Attention-based bi-directional long-short term memory network for cognitive state prediction

The advanced model, known as the Attention-Based Bi-Directional Long-Short Term Memory (Bi-LSTM) Network, is intended to forecast cognitive states, particularly attention and inattention. By combining the benefits of attention mechanisms and Bi-LSTM networks, this method improves the precision and comprehensibility of cognitive state predictions. By processing sequences in both forward and backward directions, the Bi-LSTM network expands upon the conventional LSTM. The model's ability to capture dependencies from past and future contexts makes it easier to understand temporal patterns in cognitive states easier due to its bidirectional approach. By improving the model's capacity to concentrate on essential portions of the input data, the attention mechanism enables the model to assess the relative significance of various time steps in the sequence. This aids the model in focusing on pertinent features while decreasing the weight of irrelevant data. The Luong or multiplicative attention, was selected as the attention mechanism for this study. Because it operates more quickly than additive attention, this mechanism was chosen. Prioritizing the attention layer over the flattened layer, the attention width was established at 20 inputs from the past. This layer was regularized using $L1$ and $L2$ as well.

4 Result and discussion

This section summarizes the findings from an investigation of the proposed system based on the optimized learning framework. The models' effectiveness will be ascertained by analyzing the outcomes using these metrics. Table 3 shows the parameters and their values. The learning rate was set to 0.001, with 200 training epochs and a batch size of 128. To prevent overfitting, $L1$ regularization was applied with a value of 0.01, and dropout rates ranged between 0.2 and 0.5. The model was optimized using the IMOA optimizer.

4.1 Experimental setup

It runs Windows 10 and has an Intel $i5$ 2.60 GHz processor with 32 GB of RAM. TensorFlow, KERAS, and Python are used for the evaluations, which are conducted against the backdrop of the Anaconda3 environment.

4.2 Dataset description

Real-world affective faces (RAF-DB) (dataset 1): Researchers can benefit from the extensive collection of over 30,000 facial images annotated by 40 human coders skilled in basic or complex expression recognition found in the RAFDB (Ryerson Audio-Visual Database of Emotional Speech and audio). We utilized $3,068$ images for testing and $12,271$ for training from this subset of the RAF-DB.
Extended Cohn-Kanade dataset (CK+) (dataset 2): FER techniques, assuming in static or dynamic (video) settings, are frequently tested and validated by researchers using the Extended Cohn-Kanade (CK+) (Gupta, 2018) image database. The CK+ dataset includes 594 video clips with 123 people between 18 and 50. The 640×490 or 640×480-pixel resolution photos in the CK+ database were captured from frontal views. Although there are colour images in the CK+ database, they were transformed to grayscale specifically for this research.
AffectNet Dataset (dataset 3): This facial expression dataset was gathered by looking up 1,250 emotional keywords on search engines like Google, Bing, and Yahoo. Every image was labelled and allocated to a single annotator. A train set of 283 901 visuals and a set for validation of 3,500 visuals make up 287,401 visuals linked to 7 classes.
Facial Expression Recognition 2013 (FER-2013) (dataset 4): FER-2013 (Facial Expression Recognition 2013) is a well-known dataset that was featured in the 2013 ICML Kaggle facial expression recognition challenge and has since gained widespread usage. Sample images from the proposed dataset are shown in Figure 3.

4.3 Evaluation metrics

The metrics used to evaluate the simulation are listed below:

$$\text{Accuracy} = \frac{Tn + Tp}{Tn + Fn + Tp + Fp}. \tag{30}$$

$$\text{Precision} = \frac{Tp}{Fp + Tp}. \tag{31}$$

$$\text{Recall} = \frac{Tp}{Fn + Tp}. \tag{32}$$

$$F1 - \text{Measure} = 2 \times \frac{\text{recision} \times \text{Recall}}{\text{recision} + \text{Recall}} \tag{33}$$

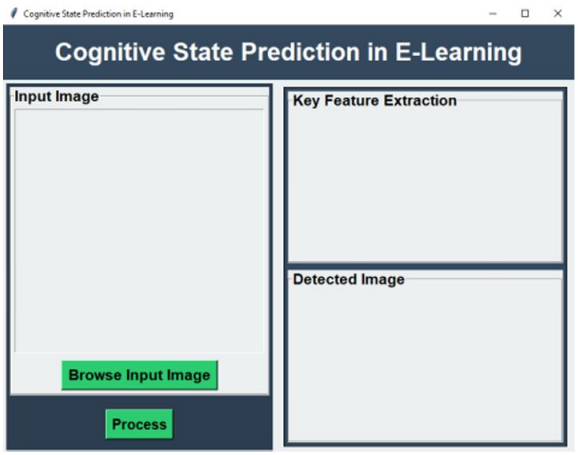


Figure 2: Initial GUI page

Figure 2 depicts the key points that illustrate various emotions. The initial GUI page is shown in Figure 5.

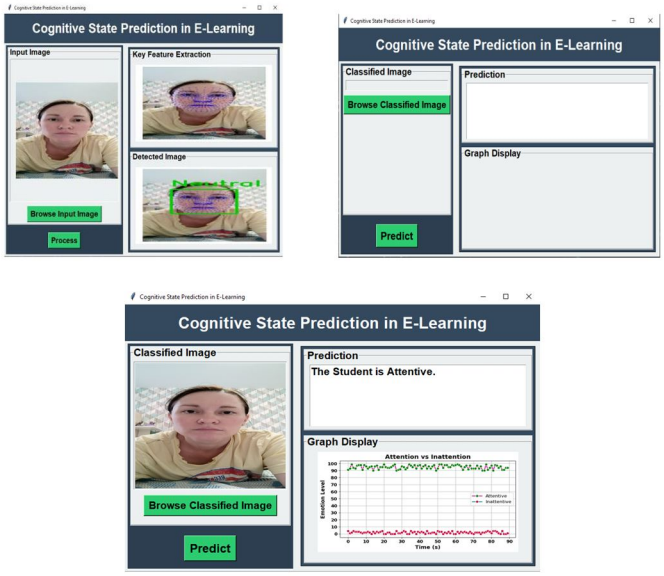


Figure 3: Sample GUI for attentive learner

This GUI is designed to browse input images and show the extracted key features and detect facial emotion based on extracted features. And the cognitive state of the student whether it's attentive or not will show and their attentive or non-attentive level is displayed as a graph. A sample proposed attentive learner GUI is illustrated in Figure 3.

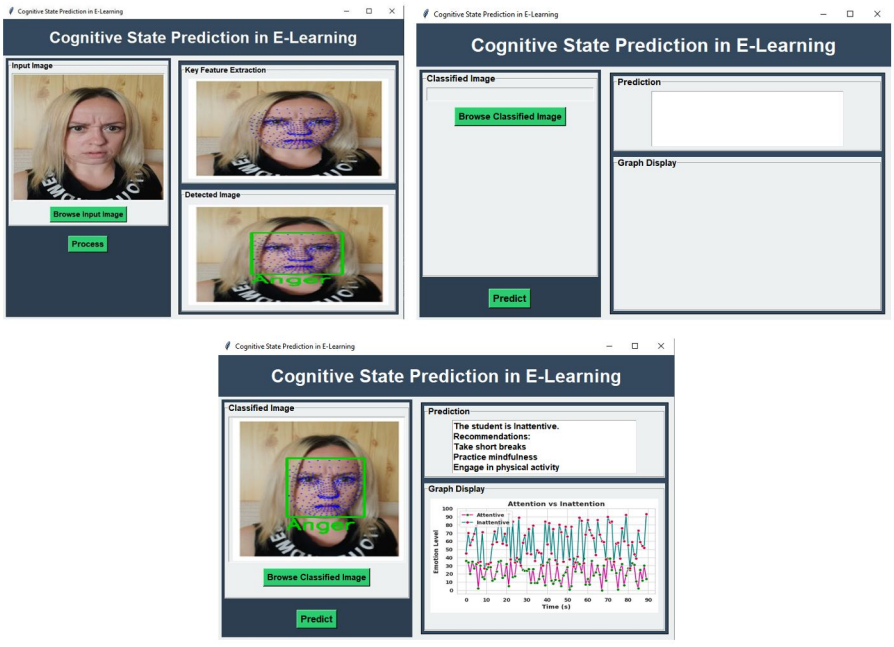


Figure 4: Sample GUI for inattentive learner

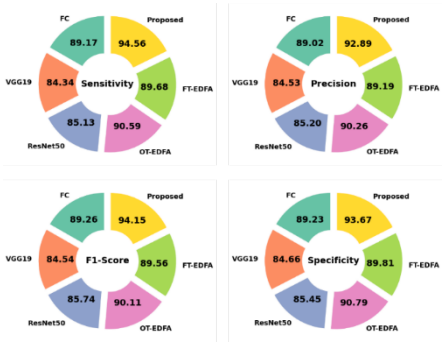


Figure 5: Graphical performance differentiation of proposed dataset 1

The suggested approach achieves the highest sensitivity (94.56%), precision (92.89%), F1-Score (94.15%), and specificity (93.67%), surpassing the performance of current methods like VGG19, ResNet50, and OT-EDFA in all metrics.

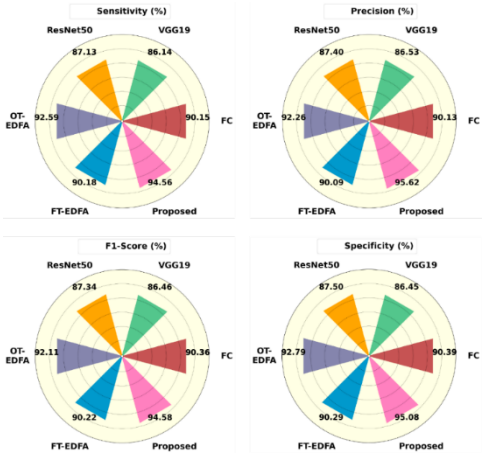


Figure 6: Graphical performance differentiation of proposed dataset 2.

With a sensitivity of 94.56%, precision of 95.62%, F1-Score of 94.58%, and specificity of 95.08%, the suggested method accomplishes the best results across the board. These outcomes demonstrate the proposed method's superior capacity to precisely identify and categorize emotions in the CK+ dataset.

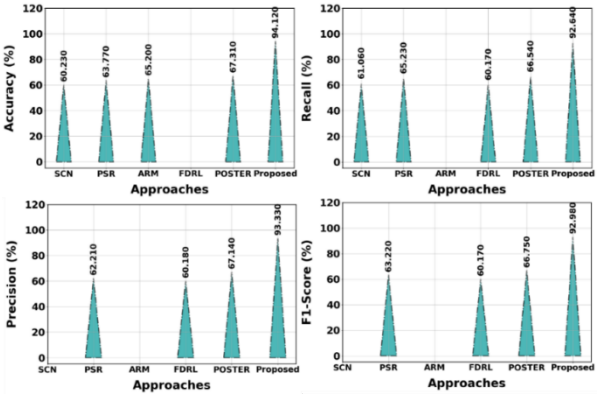


Figure 7: Graphical performance differentiation of proposed dataset 3

This figure 7 uses Dataset 3 (AffectNet) to compare the performance of different approaches (SCN, PSR, ARM, FDRL, POSTER) with the suggested method. The proposed approach achieves the highest accuracy of 94.12%, recall of 92.64%, precision of 93.33%, and F1-score of 92.98%, indicating a significant improvement over current methods.

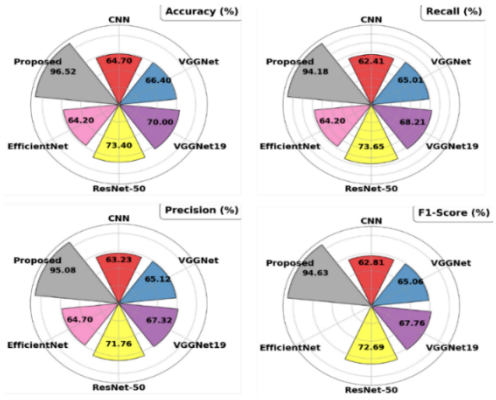


Figure 8: Graphical performance differentiation of proposed dataset 4

Figure 8 contrast the suggested method's performance on Dataset 4 (FER-2013) with other existing approaches. The suggested approach achieves the best accuracy of 96.52%, recall of 94.18%, precision of 95.08%, and F1-score of 94.63%, considerably surpassing all other methods.

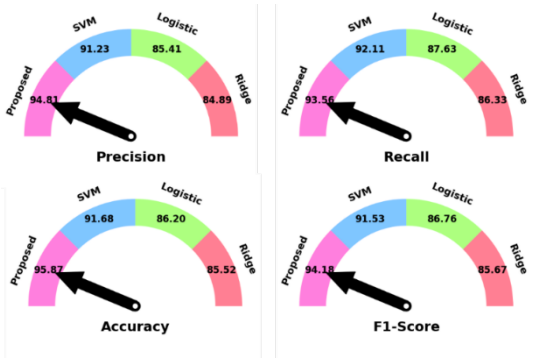


Figure 9: Graphical performance differentiation of dataset 4

The proposed approach outperforms the other methods in terms of precision, recall, accuracy, and F1-Score, as shown by the figure 9 that compares their respective performances in predicting attention and inattention.

The training and testing procedure for the proposed dataset was carried out over 200 epochs at a learning rate of 0.001. Evaluation of training and testing accuracy and loss is shown in figure 10.

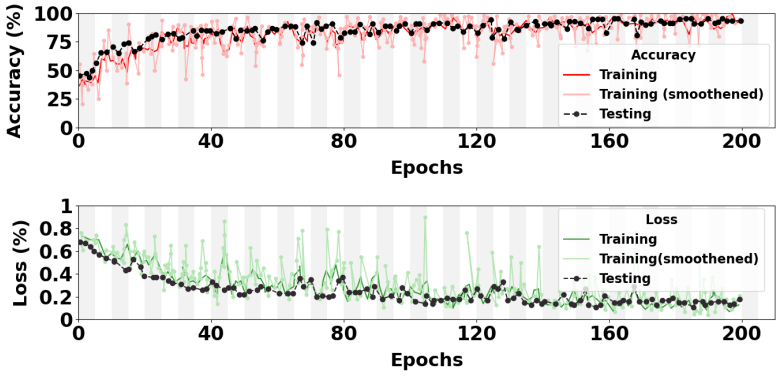
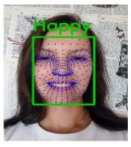
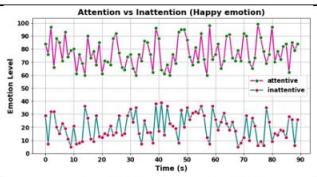
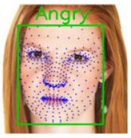
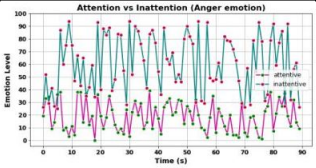
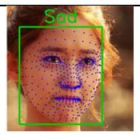
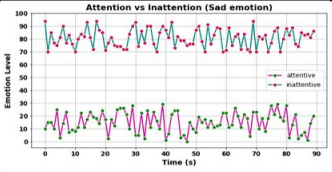
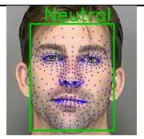
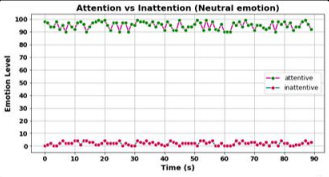
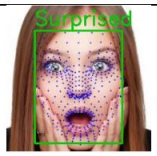
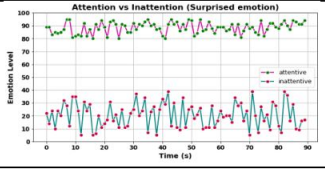


Figure 10: Training and Testing Evaluation

4.3.1 Temporal facial emotion recognition for cognitive states

This section explores real-time facial emotion recognition to assess students' cognitive states. Capturing temporal changes in emotion and attention is crucial, as emotional reactions are often brief and vary quickly. To address this, we incorporate an ABiLSTM network, which captures sequential patterns and contextual shifts in learners' behavior. Unlike static models, ABiLSTM considers both past and future temporal dependencies, enabling more accurate detection of attention and mood changes. The model's attention mechanism focuses on key time frames, improving the relevance of feedback. As a result, when students show signs of disengagement, timely interventions, such as motivational cues or content adjustments, are delivered. This approach enhances proactive teacher responses, offering a more personalized, engaging online learning experience. Table 2 displays the corresponding temporal graphs for those emotions as cognitive states. The graphs show that students are attentive when feeling neutral, happy, or surprised. On the other hand, learners exhibit deficient levels of attentiveness and high levels of inattentive cognitive state when experiencing fear, sadness, or anger.

Table 2: Outcome of temporal analysis

Input	Emotions	Emotion detected	Temporal emotion graph	state
		Happy		Attentive
		Angry		Inattentive
		Sad		Inattentive
		Neutral		Attentive
		Surprised		Surprised

In order to illustrate the usefulness of the proposed structure, a GUI for visualization of learners' cognitive states has been created. The classification results based on facial emotion analysis processed by the ABiLSTM model are displayed in the GUI. By using temporal facial expression sequences like happy, neutral, sad, and furious, it shows if a learner is attentive or inattentive. The initial system startup interface, a sample view for an attentive learner, and one for an inattentive learner are the three exemplary GUI screenshots that are given. With the help of these visual outputs, educators can examine emotional patterns and session participation in the past, enabling them to make well-informed decisions about future adaptive content delivery. Despite this, the system GUI effectively converts model predictions into comprehensible feedback for assessment after the session. We simultaneously track the four algorithms hardware resource utilization during the training and testing phases. Upon training the four models, we log the CPU usage, the proportion of CPU occupied by user space, the percentage of CPU consumed by kernel space, the percentage of

physical memory employed, and the total memory count. Table 3 displays the results.

Table 3: Utilization of Hardware Resources during Training

Approach	CPU usage (%)	User space (%)	Running time kernel - Sy (%)	Memory (%)
ResNet-50	25.00	29.50	5.92	5.9
VGG-Net	18.50	22.05	3.11	4.6
CNN	8.67	16.00	4.62	12.5
EfficientNet	16.67	18.97	1.38	4.3
Proposed	15.26	16.00	1.08	3.7

Table 4: Utilization of Hardware Resources during Testing

Approach	CPU usage (%)	User space (%)	Running time kernel - Sy (%)	Memory (%)
ResNet-50	49.78	53.67	15.73	2.9
VGG-Net	21.50	24.67	5.90	3.2
CNN	10.83	15.12	3.95	12.0
EfficientNet	17.5	21.12	2.65	3.1
Proposed	16.62	19.05	1.41	2.04

During testing, ResNet-50 again consumes the most significant resources, mainly CPU and user space, whereas the proposed technique uses the fewest resources. This suggests that the proposed method is more efficient, requiring less computational power during the training and testing phases than standard models such as ResNet-50, VGG-Net, and CNN shown in table 4.

Table 5: Comparison over existing approaches

Approach	Accuracy	Precision	Recall	F1-Score
EEGNet	88.00%	85.00%	83.00%	84.00%
ResNet-50	94.50%	93.00%	92.00%	92.50%
VGG-19	92.80%	91.50%	90.30%	90.90%
CapsNet	93.00%	91.00%	90.00%	90.50%
LSTM	91.50%	89.00%	88.00%	88.50%
Proposed	98.21%	98.14%	98.10%	98.12%

With an accuracy of 98.21%, precision of 98.14%, recall of 98.10%, and F1-score of 98.12%, the proposed approach performs better than any of the other models in the comparison series. EEGNet, a well-known EEG-based model, performs poorly in contrast, whereas models such as ResNet-50 and CapsNet exhibit competitive outcomes with good accuracy and precision which is shown in table 5.

5 Discussion

While e-learning offers convenience, it suffers from high dropout rates largely due to the lack of emotional and cognitive engagement. Existing models typically treat emotional and cognitive feedback separately, leading to limited understanding and ineffective support for learners. To address this, our proposed ensemble learning-based approach integrates both emotional and cognitive cues in real time, offering a more holistic assessment of learner engagement. Unlike traditional approaches that rely heavily on external sensors or computationally expensive methods, our framework uses non-intrusive facial emotion recognition, ensuring scalability, low cost, and real-time responsiveness. The experimental results across FER-2013, CK+, AffectNet, and RAF-DB datasets validate the effectiveness of our model. For example, the proposed model achieved 96.52% accuracy on FER-2013, significantly outperforming ResNet-50, which only achieved 73.4%. Similarly, on AffectNet, our model reached 94.12% accuracy, surpassing POSTER's 67.31%. By accurately detecting subtle emotional cues such as mild confusion or hesitation, the proposed approach addresses the limitation of poor sensitivity to nuanced learner states found in existing systems. Furthermore, through the integration of contextual factors into the model design and training, our method adapts better to diverse learning scenarios compared to static,

one size-fits-all models. Overall, this research closes key gaps by delivering a real-time, scalable, context-aware, and personalized cognitive-emotional monitoring framework for online learning environments. It thus lays the foundation for more responsive, effective, and inclusive e-learning systems. This resulted in improved generalization capabilities for the test datasets. Differentiation of related work with proposed is shown in table 18.

Table 6: Differentiation of proposed with existing prior research

References	Method used	Dataset Used	Accuracy
Gupta et al. [11]	VGG-19, ResNet-50	Wider face, CK+ and FER-2013	92.58%
Rathi et al. [12]	SSRO-based Deep LSTM	Own dataset	96.2%
Kukkar et al. [13]	LSTM, Random Forest (RF), Gradient Boosting (GB)	OULAD and emotion dataset	96%
Ma et al. [14]	Neighborhood-based Collaborative Filtering	Wider face and CK+	93.14%
Benabbes et al. [15]	BiLSTM with FastText,	Emotion dataset	97.08%
Ma et al. [16]	FC-CDF	Real-world dataset	95.00%
Proposed		CK+, FER-2013, AffectNet and RAF-DB	98.21%

6 Conclusion and future scope

In today's digital age, where online education is increasingly prevalent, ensuring effective e-learning requires continuous feedback and learner support. Monitoring students' cognitive states through computer vision-based algorithms has attracted global research interest, showing promising results with neural networks. This study presents a novel, optimized fuzzy-based classification approach for predicting cognitive states in real-time on e-learning platforms. Using a device camera, facial images are captured and processed through the IMTCNN algorithm to detect the face. Significant facial features are then selected using the Improved Crayfish Optimization Algorithm (ICOA). These features are input into the optimized fuzzy model, which classifies facial emotions by analyzing key facial landmarks. The ABiLSTM network further evaluates these features to determine the learner's cognitive state attentive or inattentive. Experimental results show the proposed method achieved 98.21% accuracy, outperforming existing approaches. Additionally, a user-friendly GUI web application was developed to enable seamless deployment. Future work will focus on integrating EEG signals and additional cues such as audio, eye gaze, and body posture to enhance cognitive state detection. Expanding emotion recognition beyond six basic emotions to include subtle or mixed states (e.g., frustration, confusion) through multimodal fusion and affective modeling can improve system sensitivity. Testing on larger, more diverse datasets will also help improve adaptability across varied learners and environments, enabling more personalized feedback and interventions.

Declaration

Conflict of interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

Funding: No funding was received to assist with the preparation of this manuscript.

Availability of data and material: Data will be available when requested.

Authors' contributions: The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Ethics approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] S. Ahmad, et al., *Multi-clustered mathematical model for student cognitive skills prediction optimization*, IEEE Access, **11** (2023), 65371-65381. https://doi.org/10.1109/ACCESS.2023.3285612

[2] M. Arefi, *Geometric clustering fuzzy regression based on c-means clustering*, Iranian Journal of Fuzzy Systems, **22**(3) (2025), 87-101. https://doi.org/10.22111/ijfs.2025.51386.9078

[3] E. Z. Z. A. I. M. Aymane, D. A. H. B. I. Aziz, H. A. I. D. I. N. E. Abdelfatteh, A. Q. Q. A. L. Abdelhak, *Enabling sustainable learning: A machine learning approach for an eco-friendly multi-factor adaptive E-learning system*, Procedia Computer Science, **236** (2024), 533-540. https://doi.org/10.1016/j.procs.2024.05.063

[4] K. Benabbes, et al., *A new hybrid approach to detect and track learner's engagement in e-learning*, IEEE Access, **11**(8) (2023), 70912-70929. https://doi.org/10.1109/ACCESS.2023.3293827

[5] S. Gupta, et al., *A multimodal facial cues based engagement detection system in e-learning context using deep learning approach*, Multimedia Tools and Applications, **82**(18) (2023), 28589-28615. https://doi.org/10.1007/s11042-023-14392-3

[6] C. Q. Huang, et al., *XKT: Towards explainable knowledge tracing model with cognitive learning theories for questions of multiple knowledge concepts*, IEEE Transactions on Knowledge and Data Engineering, **15**(7) (2024), 7308-7325. https://doi.org/10.1109/TKDE.2024.3418098

[7] M. N. Kouahla, et al., *Emorec: A new approach for detecting and improving the emotional state of learners in an e-learning environment*, Interactive Learning Environments, **31**(10) (2023), 6223-6241. https://doi.org/10.1080/10494820.2022.2029494

[8] R. Kukkar, et al., *Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms*, Education and Information Technologies, **28**(8) (2023), 9655-9684. https://doi.org/10.1007/s10639-022-11573-9

[9] V. T. Lokare, P. M. Jadhav, *An AI-based learning style prediction model for personalized and effective learning*, Thinking Skills and Creativity, **51** (2024), 101421. https://doi.org/10.1016/j.tsc.2023.101421

[10] H. Ma, et al., *Predicting student performance in future exams via neutrosophic cognitive diagnosis in personalized e-learning environment*, IEEE Transactions on Learning Technologies, **16**(5) (2023), 680-693. https://doi.org/10.1109/TLT.2023.3240931

[11] H. Ma, et al., *Predicting examinee performance based on a fuzzy cloud cognitive diagnosis framework in e-learning environment*, Soft Computing, **27**(24) (2023), 18949-18969. https://doi.org/10.1007/s00500-023-08100-4

[12] M. R. Mastani Shabestari, N. Mikaeilvand, *A novel approach to time-fractional equations using fuzzy beta Laplace transform iterative technique and its applications in fluid dynamics*, Iranian Journal of Fuzzy Systems, **22**(3) (2025), 1-19. https://doi.org/10.22111/ijfs.2025.50308.8872

[13] R. Mustapha, G. Soukaina, Q. Mohammed, A. Es-Sâadia, *Towards an adaptive e-learning system based on deep learner profile, machine learning approach, and reinforcement learning*, International Journal of Advanced Computer Science and Applications, **14**(5) (2023), 1-7. https://doi.org/10.1207/s15430421tip4104_2

[14] N. G. Praveena, S. S Nath, *A fuzzy based efficient and blockchain oriented secured routing in vehicular Ad-Hoc networks*, Iranian Journal of Fuzzy Systems, **21**(6) (2024), 15-31. https://doi.org/10.22111/ijfs.2024.48069.8461

[15] H. S. Raghavendra, et al., *Student education analysis of e-learning during COVID-19 using support regression random forest algorithm*, SN Computer Science, **5**(6) (2024), 727. https://doi.org/10.1007/s42979-024-03065-z

[16] S. Rathi, et al., *Affective state prediction of E-learner using SS-ROA based deep LSTM*, Array, **19** (2023), 100315. https://doi.org/10.1016/j.array.2023.100315

[17] G. Shi, et al., *One for all: A unified generative framework for image emotion classification*, IEEE Transactions on Circuits and Systems for Video Technology, **34**(8) (2023), 7057-7068. https://doi.org/10.1109/TCSVT.2023.3341840

[18] R. Wang, L. Chen, A. Ayesh, *Multimodal motivation modelling and computing towards motivationally intelligent E-learning systems*, CCF Transactions on Pervasive Computing and Interactions, **5**(1) (2023), 64-81. https://doi.org/10.1007/s42486-022-00107-4

[19] C. Zuo, et al., *GUGEN: Global user graph enhanced network for next POI recommendation*, IEEE Transactions on Mobile Computing, **8**(9) (2024), 14975-14986. https://doi.org/10.1109/TMC.2024.3455107