

A Fusion of Transfer Learning with Osprey Optimization Algorithm Based Student Engagement Detection for Improving Performance of Virtual Classroom Environments

Hassan Gise

dr.hassan7778@hotmail.com

King Saud University

Ali Muhammad Aljodea

King Saud University

Research Article

Keywords: Fusion of Transfer Learning, Student Engagement Detection, Virtual Classrooms, Osprey Optimization Algorithm, Image Pre-processing

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11000799/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A Fusion of Transfer Learning with Osprey Optimization Algorithm Based Student Engagement Detection for Improving Performance of Virtual Classroom Environments

Author-1: Hassan Gise*

Department of Curriculum and Instruction,

College of Education, King Saud University, Riyadh - 11495, Saudi Arabia.

dr.hassan7778@hotmail.com

Author-2: Ali Muhammad Aljodea

Department of Curriculum and Instruction, College of Education,

King Saud University, Riyadh – 11495, Saudi Arabia.

aaljodea@ksu.edu.sa,

*Corresponding author: Hassan Gise dr.hassan7778@hotmail.com

Abstract

Education is the main factor for the social change and economic evolution of any nation. The method of education is becoming much easier every day. Presently, a virtual classroom is an online learning model which is performed through the internet and delivers interaction for distance learners like traditional classrooms. This mode of education is a high trend in the future days and will be substituting the conventional method of education as internet-dependent learning is developing quickly in almost every college and school. When compared to conventional classrooms, virtual classrooms are discovered to be more susceptible. With the developments in Artificial Intelligence (AI) and the rising amount of online educational data, the Deep Learning (DL) and Machine learning (ML) approach plays a significant part in forecasting the performance of students. Recent growths have aided teachers in defining the weaknesses and strengths of student attainment. This study develops a Fusion of Transfer Learning to Student Engagement Detection for Virtual Classrooms Using the Osprey Optimization Algorithm (FTLSED-VCOOA). The main intention of the FTLSED-VCOOA technique is to detect students' performance in virtual classrooms using advanced deep-learning techniques. At first, the adaptive bilateral filtering (ABF) method is employed in the image preprocessing stage to eliminate the noise from input data. Moreover, the feature extraction process has been executed by fusion models such as Squeezenet, Shallow Capsule

Networks (Shallow CapsNets), and MobileNetV2. Besides, the convolutional neural network and long short-term memory (CNN-LSTM) technique have been deployed for the classification process. Eventually, the Osprey Optimization Algorithm (OOA)-based hyperparameter range method is achieved for optimizing the classification outcomes of the CNN-LSTM system. The simulation analysis of the FTLSED-VCOOA system is verified on a benchmark image database, and the results are measured under many measures. The experimental outcome emphasized the development of the FTLSED-VCOOA approach in student engagement detection process.

Keywords: Fusion of Transfer Learning; Student Engagement Detection; Virtual Classrooms; Osprey Optimization Algorithm; Image Pre-processing

1. Introduction

Education is deliberated as the major factor for social change and economic growth of every country [1]. The way of education is getting easier every day. The novel education methods utilize several technologies to provide education. Presently, the most trending domain of education is the virtual classroom. The word virtual describes simulation of a real thing. A virtual classroom is an online learning approach that is performed through the internet and gives communication for distance learners just like a face-to-face classroom [2]. It aims at technology to give education to students who are not physically accessible on site to get their education. The class can be accompanied by learners from each part of the world from some type of gadget such as a desktop, mobile phone, or laptop [3]. This education field is going to a higher trend in the future days and soon will be substituting the conventional way of education as internet-dependent learning is increasing quickly in almost every college and school.

A virtual learning environment (VLE) is a software method intended to support learning and teaching [4]. VLE typically operates the World Wide Web, thus accessing either on or off-campus, offered the consumers are registered and admittance an Internet. It overwhelms the restriction of conventional face-to-face teaching and guarantees that learning is either limited to environmental position or time. VLEs becomes more embedded and popular in multiple higher education institutes throughout the sphere [5]. It is due to their adaptability and also gives a broad range of devices or abilities, comprising communications, assessment, content delivery, evaluation, and more. There are multiple commercial VLE systems or products that are accessible, like COSE, Blackboard, WebCT, and Learnwise [6]. The benefit of choosing an optimum VLE method is to assist faculty members of module contents and also improve the

learning knowledge of students over the assessment, communication tools, and evaluation. Thus, a choice and evaluation of the VLE method is a vital strategic decision for a better education institution [7].

In comparison with conventional, virtual classrooms are initiated to be more adaptable [8]. It allows the learner to communicate with the tutor from each part of the world for a specific course at the proper time. Primarily, learner can utilize their free time for learning as they need not be present in the classroom physically [9]. To identify performance of students in virtual classrooms, Artificial Intelligence (AI) is a vital device to explain certain student behaviour patterns in the virtual classroom. Predictive models intend to decipher the essential relation of data that can substantially help institutions and teachers face educational problems [10]. Particularly, Machine Learning (ML) models permit us to recognize these relations and establish the grades of students depend on them. This method, knowing that certain representative data of student in the course, we can measure their performance.

This study develops a Fusion of Transfer Learning to Student Engagement Detection for Virtual Classrooms Using the Osprey Optimization Algorithm (FTLSED-VCOOA). At first, the adaptive bilateral filtering (ABF) method is employed in the image preprocessing stage to eliminate the noise from input data. Moreover, the extraction of feature procedure is executed by fusion models such as Squeezenet, Shallow CapsNets, and MobileNetV2. Besides, the convolutional neural network and long short-term memory (CNN-LSTM) technique have been deployed for the classification process. Eventually, the osprey optimization algorithm (OOA)-based hyperparameter range procedure is achieved for optimizing the classification outcomes of the CNN-LSTM system. The simulation analysis of the FTLSED-VCOOA system is verified on a benchmark image database, and outcomes are measured under numerous measures.

2. Literature of Works

In [11], a novel flipped classroom model in universities and colleges that depends on the Virtual Reality (VR) model is intended to integrate the model of Contrastive Language-Image Pre-Training (CLIP). Over cross-modal data fusion, this method integrates student operation behaviour with teaching content and develops impacts around the intellectual feedback mechanism. Pabba and Kumar [12] projected a real-world vision-based multi-cues method. This model attained three types of scores: Primarily, a distress score depends on the negative and positive academic conditions. Then, head movement score is dependent upon head displacement distance. The FER method is projected for academic affective state

classification. Guo et al. [13] employ the presage-process-product (3P) method of student learning and involvement for inspecting the dynamic among scrutinizing, institutional and personal elements, their effect on student interpersonal communication, and its succeeding involvement in HIPs. Emphasizing the outcomes of the substantial role of student's social and emotional skills (SES) in improving their involvement in HIP.

Mazumder et al. [14] developed a mechanism utilizing ensemble learning and computer vision. This method integrates engagement classification and emotion detection methods depending on ResNet50, DenseNet121, and VGG16, a custom CNN trained on a database of student facial expressions. A head pose estimation module evaluates gaze direction and head orientation, incorporating these outputs with the engagement and emotion methods. This holistic method measures engagement by deliberating head pose and emotional states. Furthermore, Gradient-weighted Class Activation Mapping (Grad-CAM) is leveraged to deliver visual explanations for the predictions of the model. In [15], an AI-supported teaching method was advanced, employing AI devices to evaluate the emotional response of students and allowing educators to modify teaching approaches depending on emotional data. This paper utilized a three-step approach: analytical, experimental, and theoretical stages. The theoretical stage involved a wide-ranging literature survey aimed at the part of EI in structural education. In the analytical stage, a review is carried out to assess the ability of students to overcome learning problems. Olaniyan et al. [16] advance a Multi-Modal Emotion Detection Method for a Virtual Classroom utilizing a non-intrusive, contactless method. This method utilizes remote Photoplethysmography (rPPG) to take physiological aspects through real-world video streaming from webcams. The incorporated method employs DL techniques for FER and speech recognition. The outcomes contribute to online education by offering a reliable and non-intrusive model for monitoring students' engagement levels and emotional states.

In [17], the pre-processed data and essential aspects are removed utilizing Term Frequency & Inverse Gravity Moment and Word2Vec. The primary innovation is to employ the DL model on Bi-LSTM-CNN with a gated mechanism and for discovering the academic performance of students. Maddu and Murugappan [18] developed an innovative engagement forecast model of students for online learners depending on facial emotion. The pre-processing stage is the primary level in process of Face recognition. Succeeding the pre-processing level, the feature extractor proceeds with the removal of ResNet features, Shape Local Binary Texture (SLBT), Global Binary Pattern (GBP), and Improved Active Appearance Model (AAM). The retrieved

features are eventually exposed to recognizing emotion through Hybrid Classification method that integrates method comprising CNN and an Improved Deep Belief Network (IDBN).

3. Materials and Methods

In this study, we have developed a new FTLSED-VCOOA algorithm. The main aim of the FTLSED-VCOOA system is to detect students' performance in virtual classrooms using advanced DL techniques. It includes four kinds of phases like image pre-processing, fusion of feature extractor, classification process, and parameter selection. Fig. 1 depicts the complete procedure of FTLSED-VCOOA methodology.

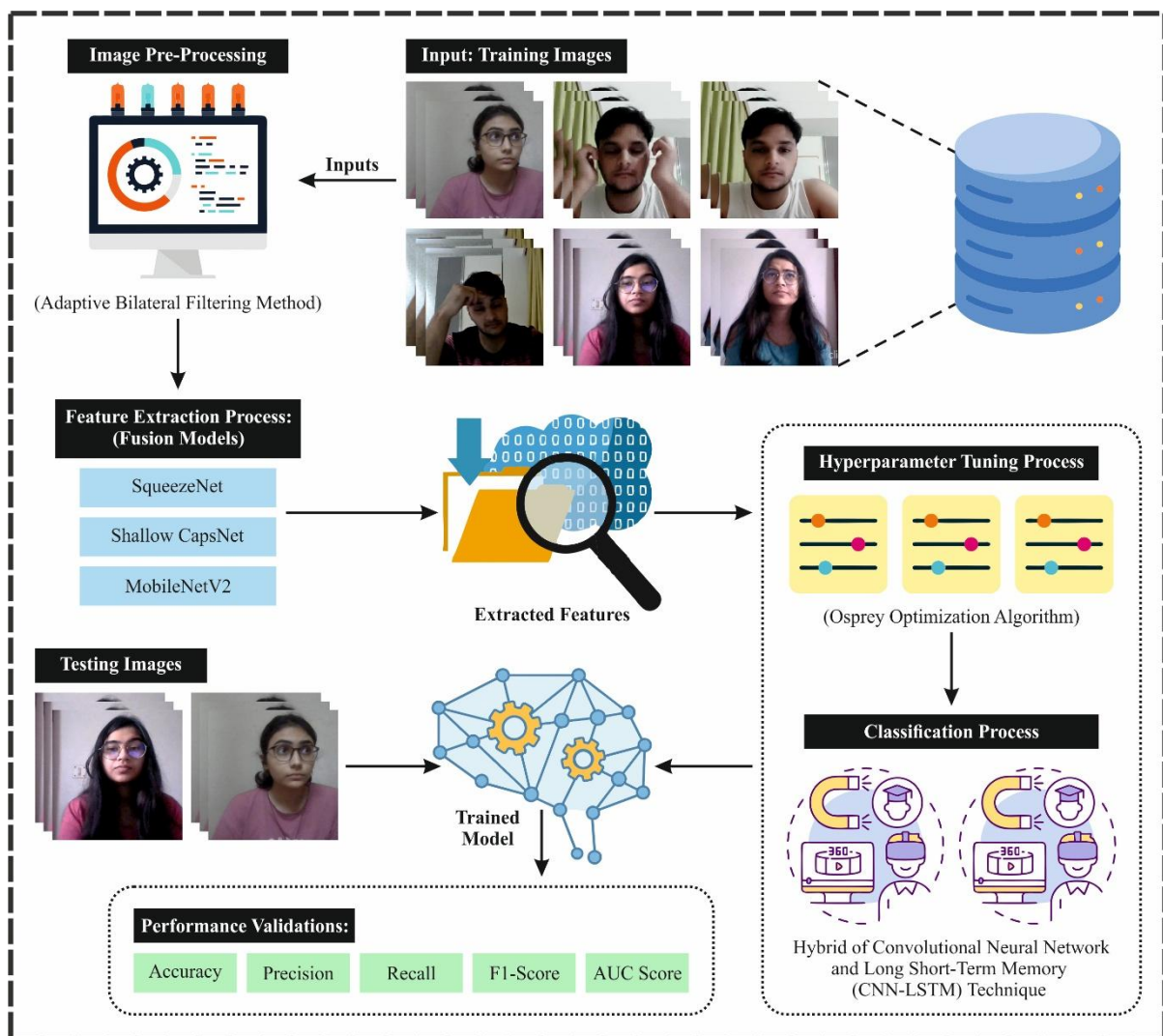


Fig. 1. Overall process of FTLSED-VCOOA methodology

3.1. Image Preprocessing

At first, the ABF method is employed in the image preprocessing stage to eliminate the noise from input data. ABF is a great image-processing model that upholds edges while decreasing noise, making it appropriate for student engagement recognition in virtual classrooms [19]. This filtering system alters intensity and spatial parameters dynamically, certifying that vital facial features, like expressions and eye movements, stay together while eliminating irrelevant noise. By improving the video frame quality, ABF enhances the accuracy of DL systems in identifying engagement-related cues. It certifies that variants in background distractions, lighting, and slight alterations do not affect the process of recognition. When combined with student engagement analysis methods, it advances real-time performance by improving input data before the extraction of features. This technique mains to more reliable and effectual engagement detection, eventually enhancing the virtual learning experiences.

3.2. Fusion of Feature Extraction

Moreover, the feature extraction process is implemented by fusion models such as Squeezenet, Shallow CapsNets, and MobileNetV2.

3.2.1. Squeezenet

The SqueezeNet structure for image classification includes numerous important modules [20]. The input is getting ahead of a primary convolution layer by an 7x7 filter and a stride of 2, accompanied by the activation function of ReLU, efficiently decreasing the spatial size of the input. The SqueezeNet's fundamental building blocks are the *fire* units, comprising expand and squeeze layers. The squeeze layer uses 1x1 convolutions to decrease input channel sizes, enhancing feature compression. The expanded layer comprises parallel routes using 1x1 and 3x3 convolutions, improving spatial coverage and channel count. The outputs from these routes are connected through the channel size. Down-sampling performed by utilizing max – pooling layers with a stride of 2, tactically situated on reduce feature mapping sizes whereas maintaining important information. The system determines a global average pooling layer, making a fixed-size feature vector. This can be accomplished by the layer of fully connected (FC) with softmax for last classification.

3.2.2. Shallow CapsNets

Shallow CapsNets are an initial intended DNN presented in that comprises capsules in their structure [21]. The capsule is a crucial element in CapsNets. It summarizes each significant information regarding the state features, which are identifying in a vector form (instead of a scalar that a neuron outputs). In nature, in vector area, all vectors such as the output vector of the capsule contains an orientation and the length. The vector length specifies the network probability of identifying a specific feature. In the meantime, the representation contains the parameterized states of that attribute, like size, color, and saturation.

Determine that CapsNets consist of the group of neural networks identified as autoencoders (AEs). This class includes dual portions: initially, an encoder mainly designed for compressing the input into a considerable representation; next, a decoder rebuilds the input from this model.

CapsNets encoder has three layers such as DigitCaps, the PrimaryCaps, and convolutional. The convolutional layer has the responsibility of identifying elementary characteristics. Its basic function is to alter the output from the convolutional layer into a formal expression of features in the capsule form. Every capsule in the PrimaryCaps layer characterizes a collection of small capsules organized in a spatial formation, making an eight-dimensional vector. These capsules seizure significant spatial relations and feature stimulations, which are then applied by the high-level capsules in the layer of DigitCaps for additional handling and classification. The dynamic routing model occurs amongst the lower-level capsules of PrimaryCaps layer and the higher-level capsules of DigitCaps layer.

3.2.3. MobileNetV2

MobileNet-V2 is a CNN structure tailored for the effective classification of an image, mainly appropriate for mobile and edge devices [22]. MobileNet-V2 relies on the unique MobileNet structure with numerous main improvements, which enhance efficiency and performance. The primary innovation of MobileNet-V2 is the usage of linear bottlenecks and reversed residuals. This model includes utilizing a thin layer of bottleneck and expanding it to the high-dimension area, whereas depth-wise convolution are utilized. The projection and expansion processes are implemented utilizing 1x1 convolutions, which aid in protecting the expressiveness of the model while preserving efficacy by decreasing the parameter counts and computational cost. As its predecessor, MobileNet-V2 uses depth-wise separable convolutions. This kind of convolution factorizes a standard convolutional into the depth-wise and a point-wise

convolution (1x1 convolution), considerably decreasing the computational complexity and the parameter counts while preserving higher performance. The structure is mainly renowned for its effective layer model that improves either accuracy or speed, making it appropriate for utilization in resource-restricted surroundings like mobile devices.

3.3. Classification using CNN-LSTM Technique

Besides, the hybrid of CNN-LSTM technique have been deployed for the classification process. The CNN method is common in the area [23]. It uses shared weights and local connections to get efficient models straight from raw data by consecutively utilizing pooling and convolutional layers. This model automatically removes local features from data and makes dense, whole feature vectors. In this paper, CNN has been to remove features from data types.

The objective of LSTM is to solve the issue of vanishing gradients present in conventional recurrent neural networks (RNNs) in training. LSTM presents a special networking architecture, which contains input, output, and forget gates in memory cells, successfully resolving the problem of the vanishing gradient. These gate mechanisms permit the network to discriminatingly forget or remember information, allowing it to learn long-range dependencies in sequences. This LSTM characteristic makes it implement well in processing sequential data like time-series data.

The primary units of LSTM contain input, output, and forget gates. c_t and c_{t-1} characterize the present state and preceding time step of an LSTM component, correspondingly, h_{t-1} embodies the LSTM unit's output at the preceding time step, h_t means the output of the present component, c'_t refers to the present unit state at the recent time step, and x_t denote input at present step. The forget gate defines which information must be leftover from the memory cell, and the forget ratio f_t is computed as demonstrated:

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

Here, b_f and W_f represent the bias term and weight matrix of the forget gate individually, and σ denotes the sigmoid activation function.

The input gate comprises dual portions: the *tanh* and *sigmoid* layer. The previous defines which novel data must be kept in the memory cell, and the final incorporates the present time

step input unit state c'_t by the preceding unit state c_{t-1} to create a novel unit state c_t . The computation equation is as demonstrated:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$

$$c'_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot c'_t \quad (4)$$

Whereas b_i, b_c, W_i , and W_c represent bias terms and weighted matrix associated with the input gate, correspondingly, $\tanh()$ denotes an activation function of hyperbolic tangent. The output gate of the LSTM controls which information must be output from the memory cell to the following layer of the network.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

Here, b_o and W_o refers to bias term and weight matrix associated with the output gate, correspondingly.

The architecture of the CNN-LSTM-based hybrid analyst is presented in Fig. 2. The hybrid predictor comprises dual key portions: the CNN, which is mainly responsible for removing influencing dynamic features from date types, and the LSTM, which is mostly responsible for load forecasting. Its input contains the output from the CNN and feature quantities Z^P and Z^T gained from encoder load time-series data.

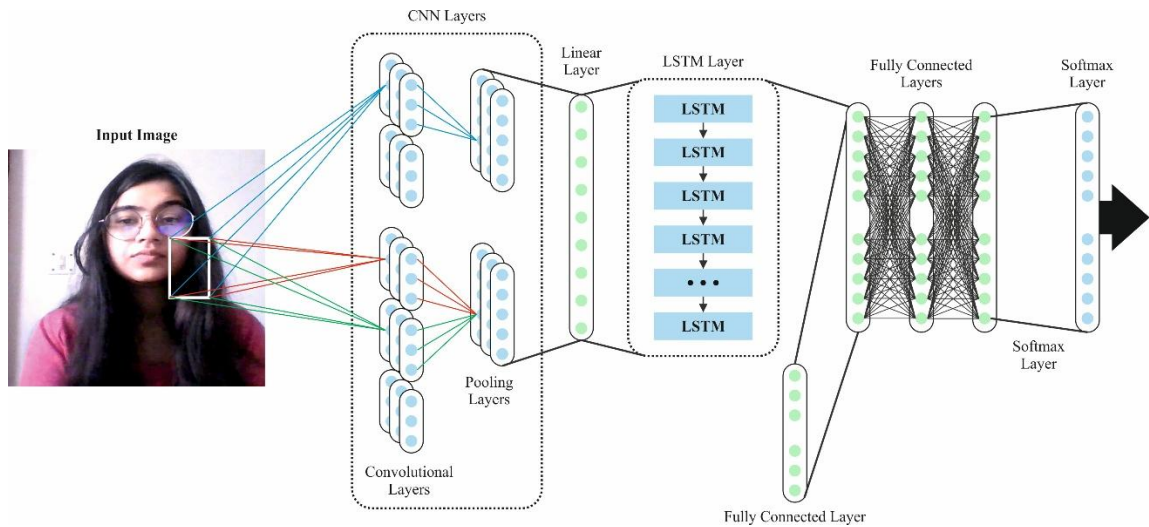


Fig. 2. CNN-LSTM Architecture

3.4. Parameter Optimizer using OOA

Finally, the OOA-based hyperparameter selection process was executed to optimize the classification outcomes of the CNN-LSTM system. The OOA is a new and effective meta-heuristic optimizer model stimulated by the individual foraging behavior of ospreys [24]. Naturally, ospreys depend on their acute sight to find prey and use accurate tactics to seize it. This model follows this procedure, utilizing such a mechanism to look for optimum solutions inside composite solution areas. In the wild, the osprey's foraging process shows either cooperation or individual intelligence. The basic notion of the OOA is for modeling osprey behavior inside the solution area, incorporating global and local search tactics to recognize the optimum solution. This work offers the basic equations of OOA and discusses its application in the optimizer procedure.

During this OOA, every solution in the search area is characterized by the location of the osprey with location upgrades and velocity changes subject to mathematic equations. Let the present location of the i th osprey in the solution area be signified as x_i and its velocity as v_i .

The osprey behavior is defined by the iterative upgrades of its velocity and position. The location-updated equation for the osprey is as demonstrated:

$$X(t + 1) = D' e^{bl} \cos(2\pi l) + X^*(t) \quad (7)$$

$$x_i^{t+1} = x_i^t + v_i^t \quad (8)$$

Here, x_i^t characterizes the location of the i th osprey at time step t , and v_i^t signifies its velocity. This equation considers the osprey's direction of movement and step size according to its present state, enabling exploration inside the solution area. The velocity updated equation is the main module of the method, balancing global and local search. This upgrade reflects the individual finest location, the global optimal location, and the inertia feature, guaranteeing that the model successfully implements either local refinement or global exploration. The velocity updated equation is normally stated as shown:

$$v_i^{t+1} = wv_i^t + c_1r_1(x_{i,best} - x_i^t) + c_2r_2(x_{globalbest} - x_i^t) \quad (9)$$

Now, w denotes inertia weight that controls the influence of the osprey's preceding velocity c_2 and c_1 signifies acceleration constants for social learning and individual, correspondingly,

however r_1 and r_2 are randomly generated values inside the range $[0,1]$. $x_{i,best}$ characterizes the optimal location determined by the i th osprey in its historic search.

Ospreys enhance their searching efficacy over cooperation and information exchange, enabling the identification of potential areas inside the searching region. This procedure is simplified by the distribution of individual and the global best solutions. Throughout all iterations, ospreys relate the fitness values of their present solutions to upgrade either their global best or their individual best. When an osprey finds out the superior solution, it upgrades its individual best. Over cautious parameter tuning, the OOA offers an efficient solution approach through different optimization settings, establishing important probable for real-world applications.

The OOA system originates a fitness function (FF) for attaining an enhanced performance of classification. For signifying the enhanced performance of the candidate solution, the positive number is defined. The classification error rate reduction is measured as FF. Eq. (10) is the mathematical model of FF.

$$\begin{aligned} fitness(x_i) &= ClassifierErrorRate(x_i) \\ &= \frac{no. of misclassified samples}{Total no. of samples} * 100 \end{aligned} \quad (10)$$

4. Experimental Analysis

The experimental outcome of the FTLSED-VCOOA technique is shown below the student–engagement dataset [25]. The dataset consists of 15000 images six class labels. The complete details of this dataset is illustrated in Table 1. Fig. 3 demonstrates the sample images.

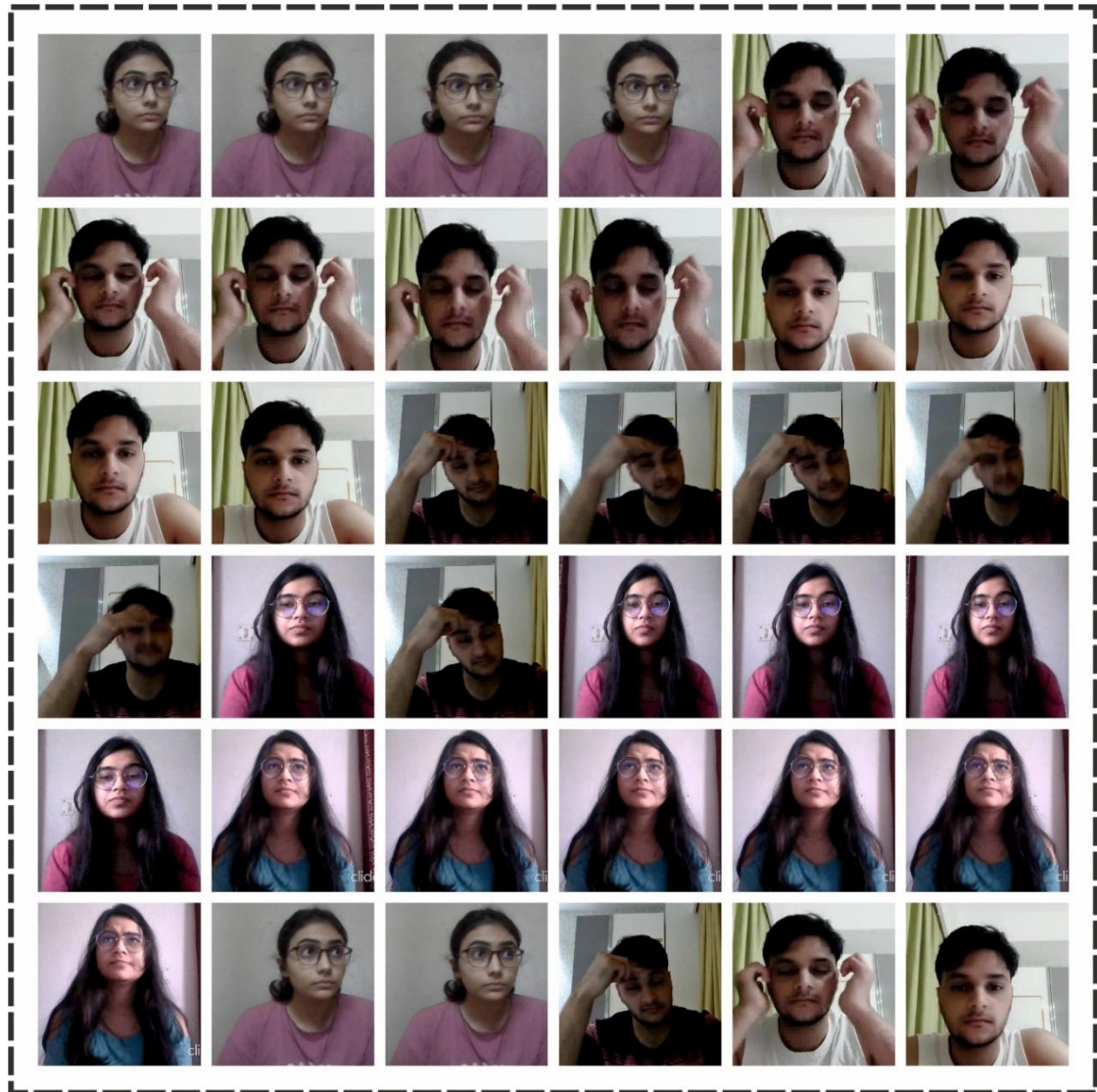


Fig. 3. Sample Images

Table 1 Details of dataset

Class Labels	No. of Images
Confused	250
Engaged	250
Frustrated	250
Bored	250
Drowsy	250
Looking Away	250
Total Images	1500

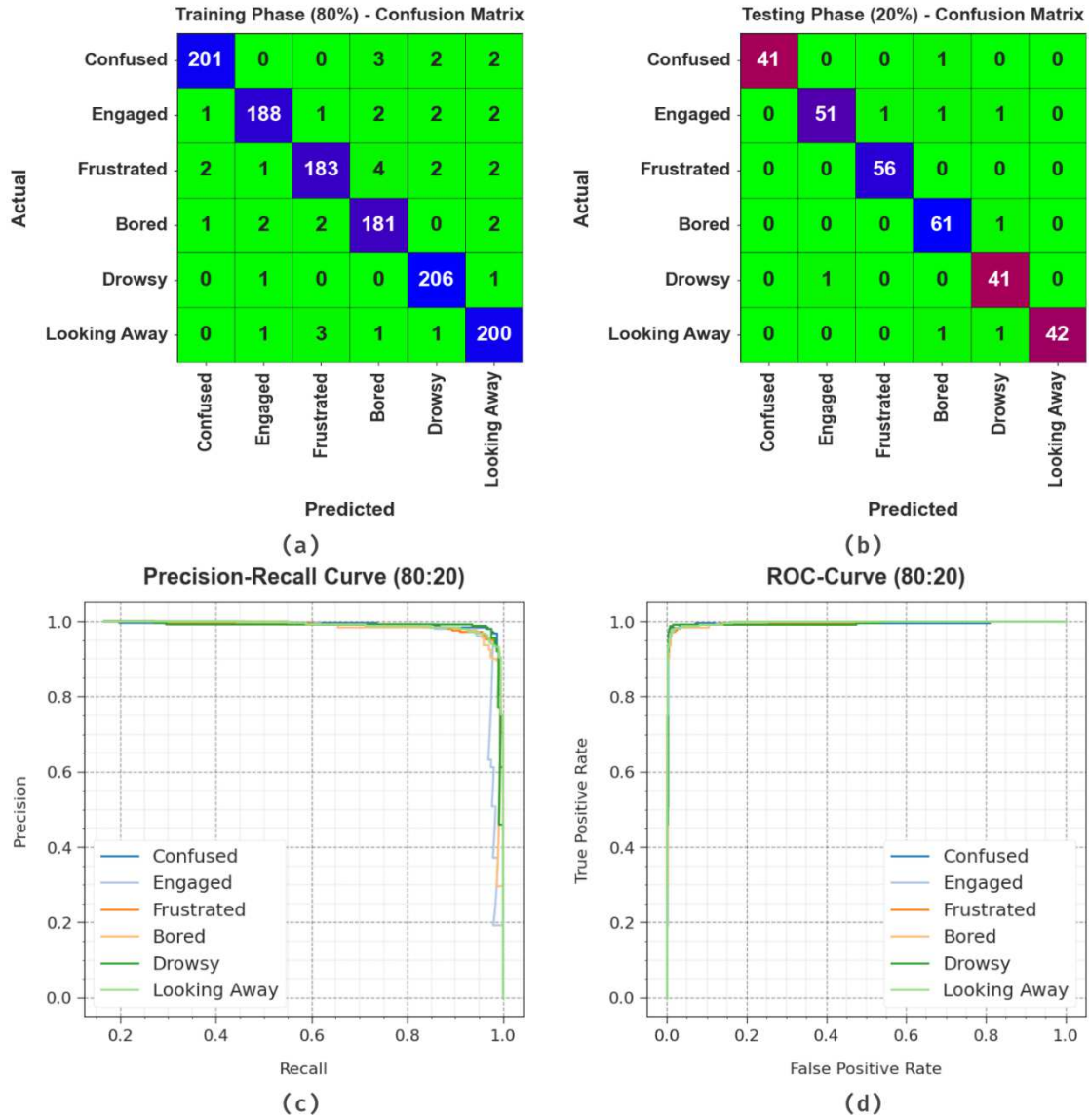


Fig. 4. 80% and 20% of (a-b) confusion matrix and (c-d) PR and ROC curves

Figure. 4 displays the classifier performances of the FTLSED-VCOOA method below 80%TRPHA and 20%TSPHA. Figures. 4a-4b signifies the confusion matrix using specific identification and classification of distinct classes. Figure. 4c reveals the PR investigation, notifying superior outcomes through all class labels. At last, Figures. 4d represents the ROC examination, illustrating skillful outcomes through high ROC values for discrete class labels.

Table 2 and Figure. 5 exemplify the student engagement detection of the FTLSED-VCOOA technique under 80:20 of TRPHA/TSPHA. The performances indicated that the FTLSED-VCOOA system has properly classified all the different classes. According to 80% TRPHA, the proposed FTLSED-VCOOA method attains an average $accu_y$ of 98.86%, $prec_n$ of 96.58%, $reca_l$ of 96.55%, $F_{Measure}$ of 96.55%, and MCC of 95.88%. Additionally, using

20%TSPHA, the proposed FTLSED-VCOOA model accomplishes an average $accu_y$ of 99.11%, $prec_n$ of 97.47%, $reca_l$ of 97.25%, $F_{Measure}$ of 97.33%, and MCC of 96.82%.

Table 2 Student engagement detection of FTLSED-VCOOA model below 80%TRPHA and 20%TSPHA

Classes	$Accu_y$	$Prec_n$	$Reca_l$	$F_{Measure}$	MCC
TRPHA (80%)					
Confused	99.08	98.05	96.63	97.34	96.79
Engaged	98.92	97.41	95.92	96.66	96.02
Frustrated	98.58	96.83	94.33	95.56	94.73
Bored	98.58	94.76	96.28	95.51	94.68
Drowsy	99.25	96.71	99.04	97.86	97.42
Looking Away	98.75	95.69	97.09	96.39	95.63
Average	98.86	96.58	96.55	96.55	95.88
TSPHA (20%)					
Confused	99.67	100.00	97.62	98.80	98.61
Engaged	98.67	98.08	94.44	96.23	95.44
Frustrated	99.67	98.25	100.00	99.12	98.92
Bored	98.67	95.31	98.39	96.83	96.00
Drowsy	98.67	93.18	97.62	95.35	94.61
Looking Away	99.33	100.00	95.45	97.67	97.32
Average	99.11	97.47	97.25	97.33	96.82

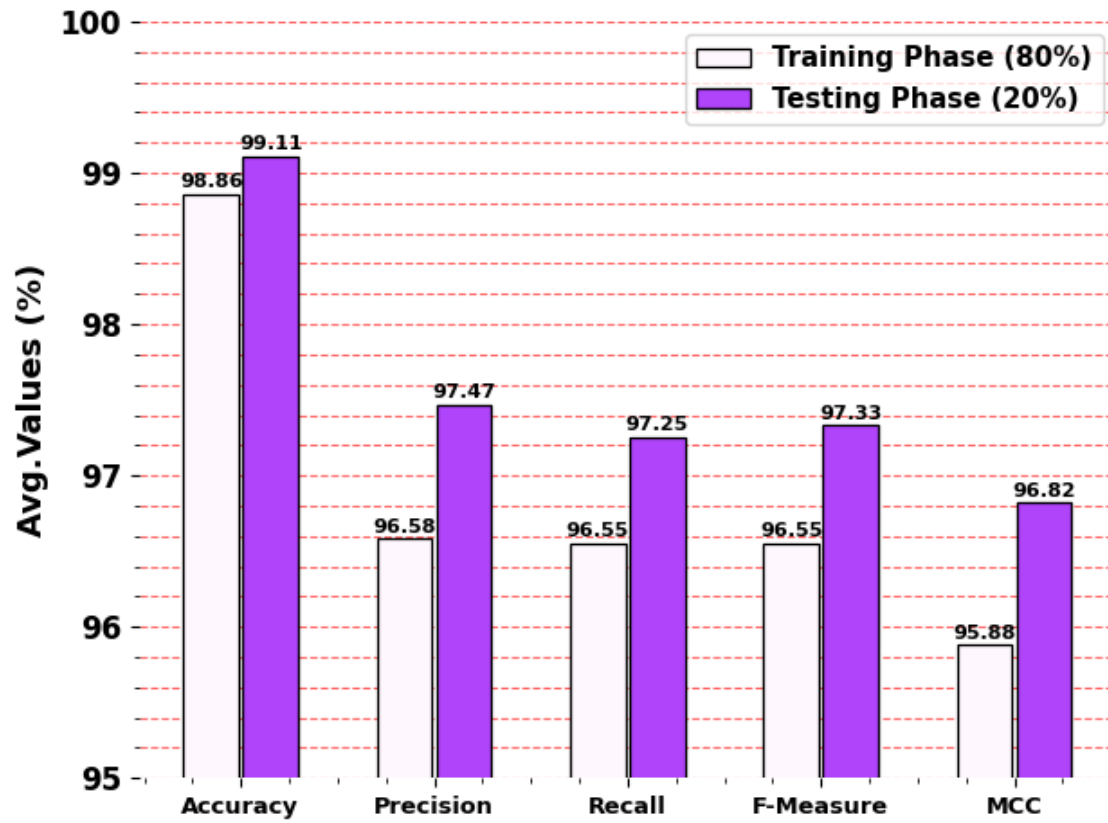


Fig. 5. Average of FTLSED-VCOOA technique below 80%TRPHA and 20%TSPHA

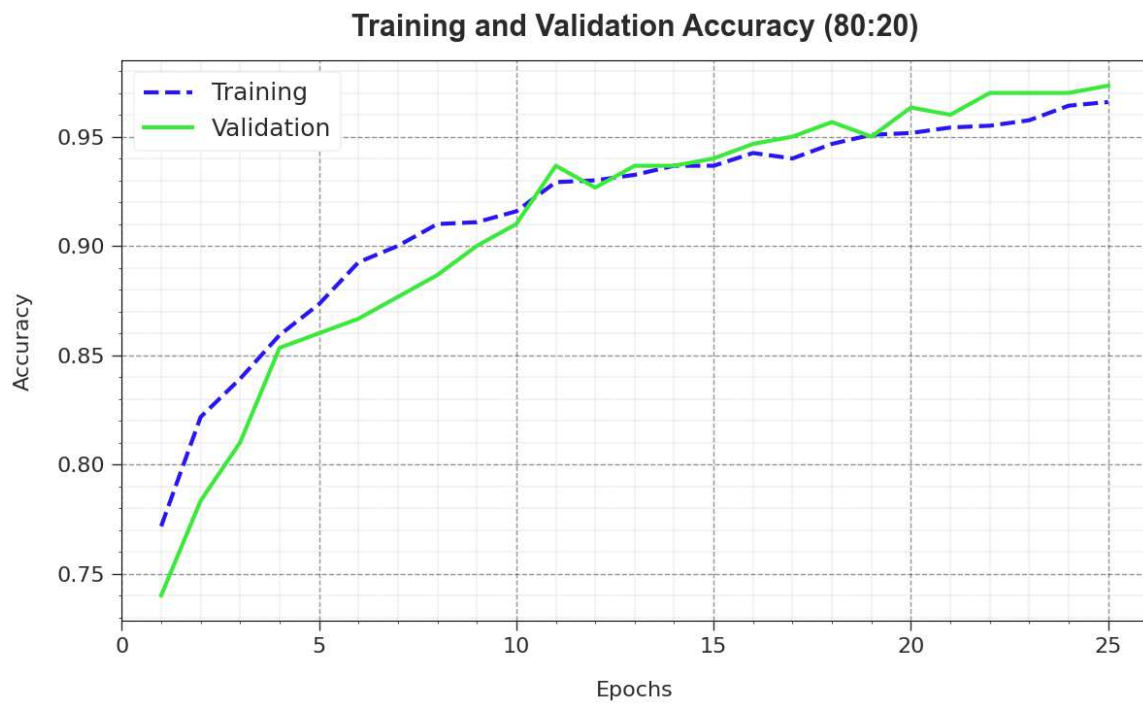


Fig. 6. $Accu_y$ curve of FTLSED-VCOOA method below 80%TRPHA and 20%TSPHA

In Fig. 6, the training (TRAN) $accu_y$ and validation (VALN) $accu_y$ performances of the FTLSED-VCOOA system under 80%TRPHA and 20%TSPHA is depicted. The values of $accu_y$ are calculated through the time period of 0-25 epochs. The figure underscored that the values of TRAN and VALN $accu_y$ express a cumulative tendency, indicating the proficiency of the FTLSED-VCOOA algorithm with enhanced outcomes through numerous repetitions. Furthermore, the TRAN and VALN $accu_y$ values remain close through the epochs, notifying decreased overfitting and presenting higher performance of the FTLSED-VCOOA approach, which guarantees reliable calculation on hidden samples.

In Fig. 7, the TRA loss (TRALOSS) and VAL loss (VALLOSS) graph of the FTLSED-VCOOA method under 80%TRPHA and 20%TSPHA is exemplified. The loss values are computed through a time period of 0-25 epochs. It is depicted that the values of TRALOSS and VALLOSS signify a diminishing tendency, which indicates the competency of the FTLSED-VCOOA system in corresponding an equilibrium between data fitting and generalization. The consecutive dilution in values of loss as well as securities is the maximum outcome of the FTLSED-VCOOA approach and tune the calculation solutions after a while.

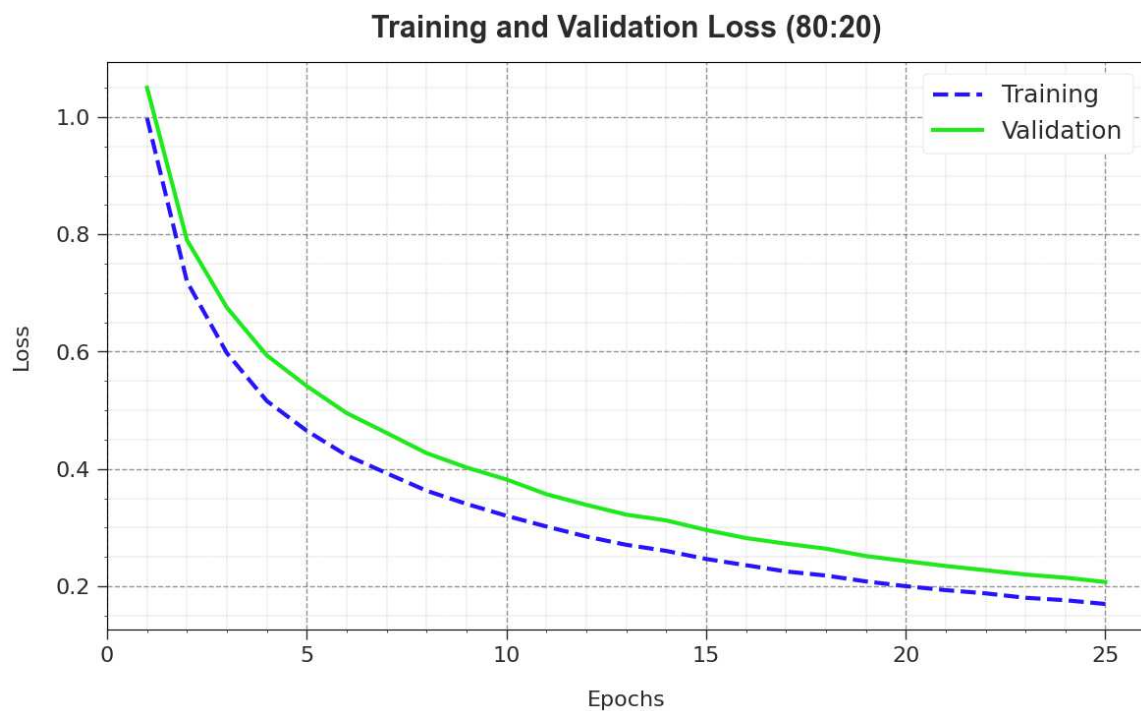


Fig. 7. Loss curve of FTLSED-VCOOA technique below 80%TRPHA and 20%TSPHA

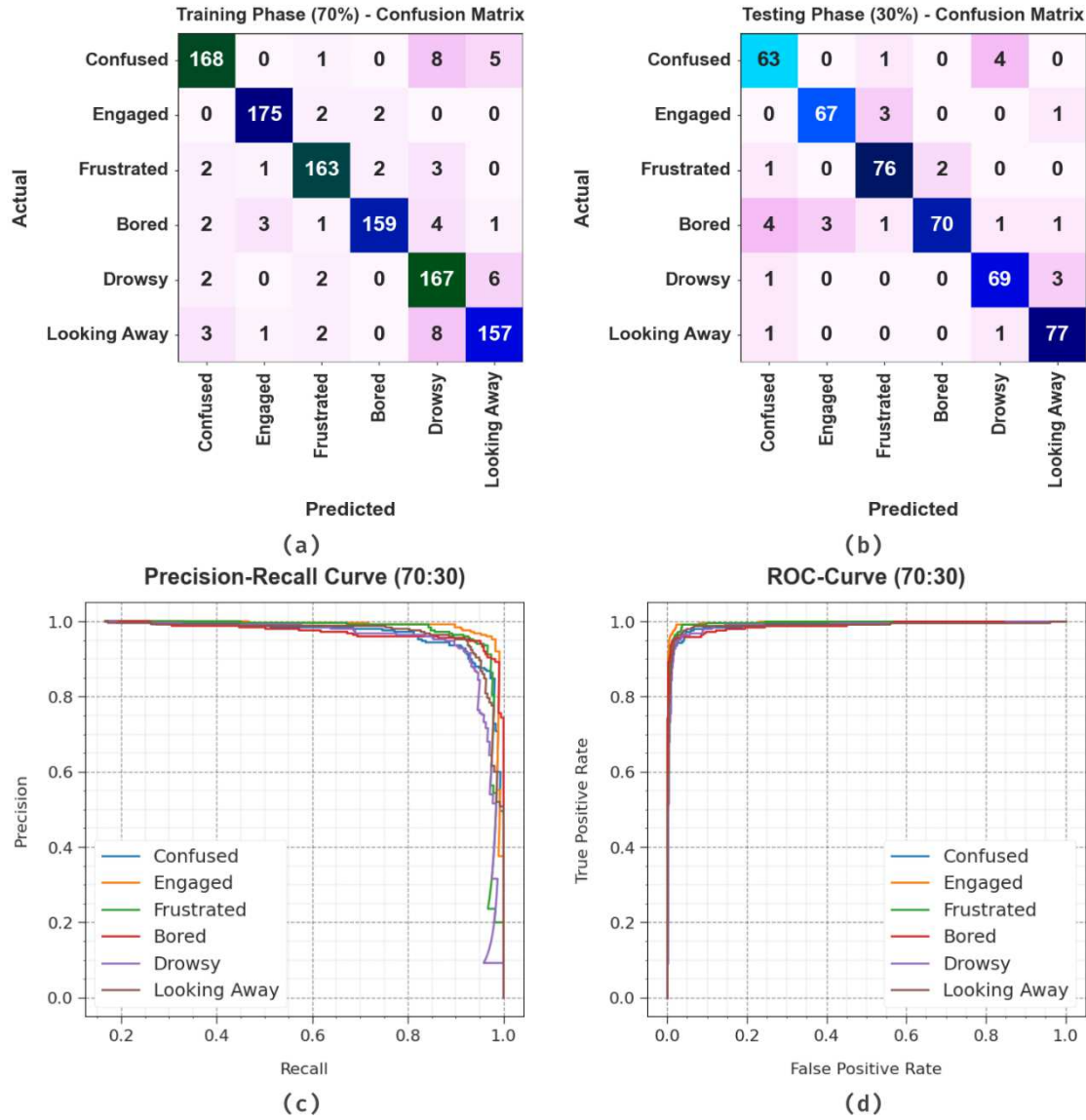


Fig. 8. 70% and 30% of (a-b) confusion matrix and (c-d) PR and ROC curves

Fig. 8 shows the classifier performances of the FTLSED-VCOOA method below 70%TRPHA and 30%TSPHA. Figs. 8a-8b displays the confusion matrix using specific identification and classification of all different classes. Fig. 8c showcases the PR investigation, notifying increased outcomes through all classes. Finally, Fig. 8d signifies the ROC outcome, representing skillful performances through high ROC values for discrete classes.

Table 3 and Fig. 9 reveal the student engagement detection of the FTLSED-VCOOA technique under 70:30 of TRPHA/TSPHA. The performance implied that the FTLSED-VCOOA system has suitably classified all the distinct classes. Using 70% TRPHA, the proposed FTLSED-VCOOA approach attains an average $accu_y$ of 98.06%, $prec_n$ of 94.30%, $reca_l$ of 94.18%, $F_{Measure}$ of 94.21%, and MCC of 93.07%. In addition, using 30%TSPHA, the proposed

FTLSED-VCOOA methodology obtains an average $accu_y$ of 97.93%, $prec_n$ of 93.78%, $reca_l$ of 93.78%, $F_{Measure}$ of 93.72%, and MCC of 92.52%.

Table 3 Student engagement detection of FTLSED-VCOOA model under 70%TRPHA and 30%TSPHA

Classes	$Accu_y$	$Prec_n$	$Reca_l$	$F_{Measure}$	MCC
TRPHA (70%)					
Confused	97.81	94.92	92.31	93.59	92.29
Engaged	99.14	97.22	97.77	97.49	96.98
Frustrated	98.48	95.32	95.32	95.32	94.41
Bored	98.57	97.55	93.53	95.50	94.68
Drowsy	96.86	87.89	94.35	91.01	89.19
Looking Away	97.52	92.90	91.81	92.35	90.88
Average	98.06	94.30	94.18	94.21	93.07
TSPHA (30%)					
Confused	97.33	90.00	92.65	91.30	89.74
Engaged	98.44	95.71	94.37	95.04	94.12
Frustrated	98.22	93.83	96.20	95.00	93.93
Bored	97.33	97.22	87.50	92.11	90.69
Drowsy	97.78	92.00	94.52	93.24	91.93
Looking Away	98.44	93.90	97.47	95.65	94.73
Average	97.93	93.78	93.78	93.72	92.52

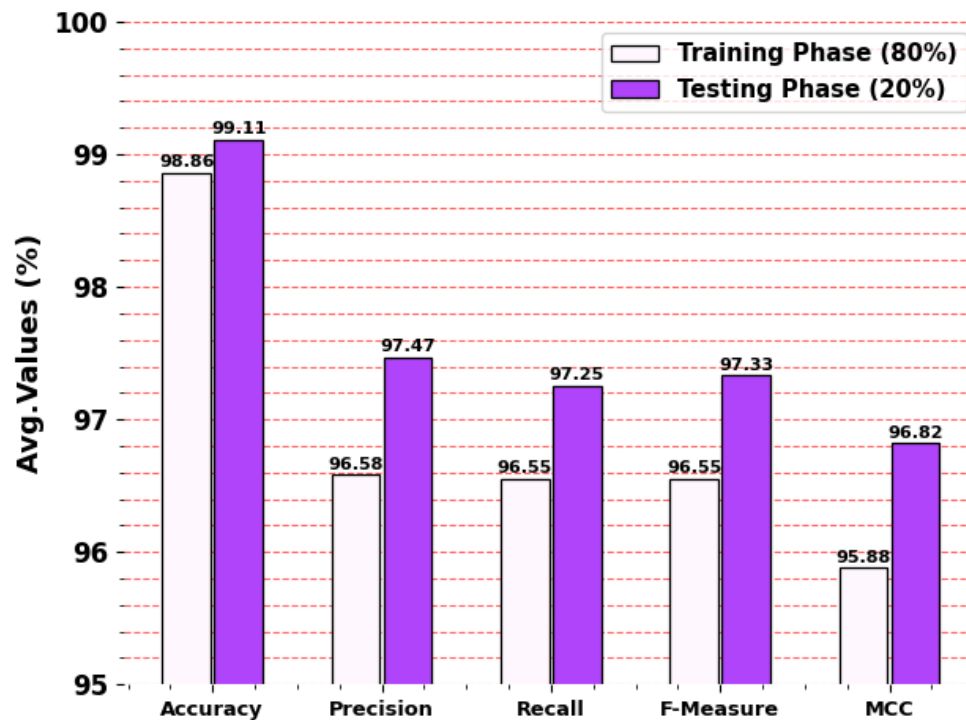


Fig. 9. Average of FTLSED-VCOOA model below 70%TRPHA and 30%TSPHA



Fig. 10. $Accu_y$ curve of FTLSED-VCOOA model below 70%TRPHA and 30%TSPHA

In Fig. 10, the TRAN $accu_y$ and VALN $accu_y$ performances of the FTLSED-VCOOA method below 70%TRPHA and 30%TSPHA is exposed. The values of $accu_y$ are computed across a time period of 0-25 epochs. The figure underscored that the values of TRAN and VALN $accu_y$ present an increasing trend, indicating the capacity of the FTLSED-VCOOA technique through maximum performance across multiple repetitions. Moreover, the TRAN and VALN $accu_y$ values remain close across the epochs, notifying diminished overfitting and expressing the higher outcome of the FTLSED-VCOOA system, which guarantees reliable calculation on hidden samples.

In Fig. 11, the TRALOSS and VALLOSS graph of the FTLSED-VCOOA method below 70%TRPHA and 30%TSPHA is shown. The loss values are computed through a time period of 0-25 epochs. It is represented that the values of TRALOSS and VALLOSS represent a diminishing tendency, indicating the competency of the FTLSED-VCOOA approach in corresponding an equilibrium between data fitting and generalization. The succeeding dilution in values of loss as well securities the superior outcome of the FTLSED-VCOOA technique and tune the calculation results gradually.

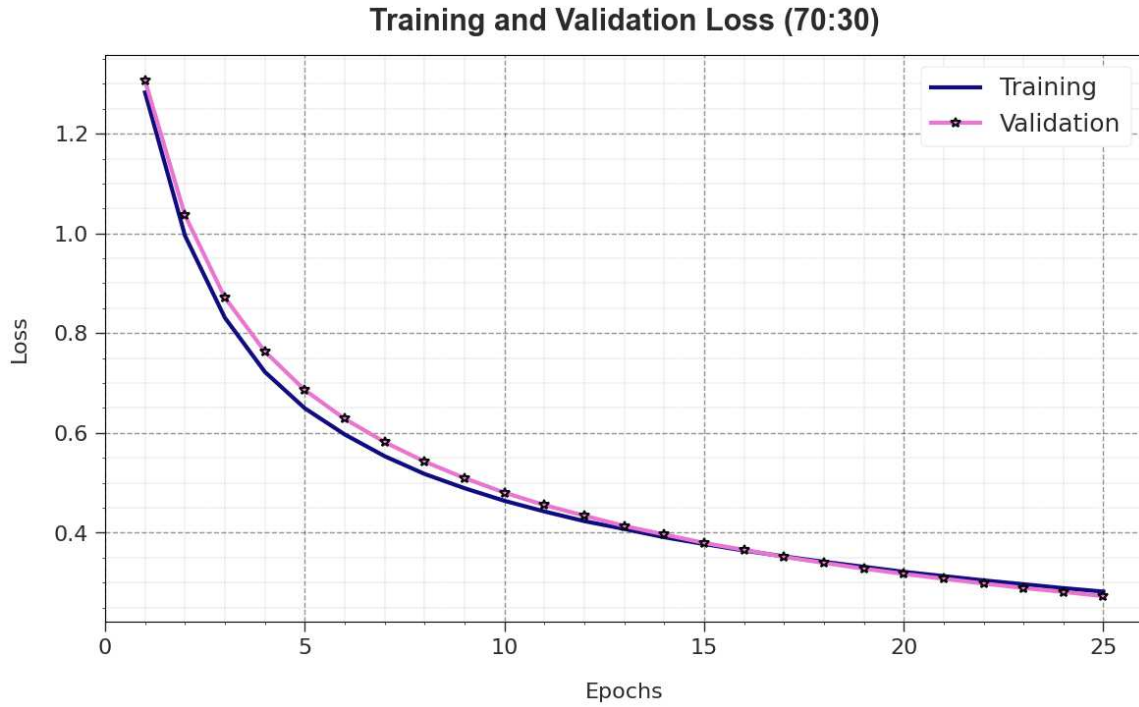


Fig. 11. Loss curve of FTLSED-VCOOA technique below 70%TRPHA and 30%TSPHA

Table 4 and Fig. 12 study the comparison outcome of the FTLSED-VCOOA model through the existing techniques [26-28]. The performances underscored that the VGG16, Inception V3, ResNet50, Fast-R-CNN, Retinanet, and YOLOv5s techniques have reported poorer performance. Afterward, SVM system has gained closer solutions through $accu_y$, $prec_n$, $reca_l$, and $F_{Measure}$ of 98.99%, 96.06%, 92.16%, and 91.82%, correspondingly. In addition, the FTLSED-VCOOA approach indicated maximum performance with increased $accu_y$, $prec_n$, $reca_l$, and $F_{Measure}$ of 99.11%, 97.47%, 97.25%, and 97.33%, correspondingly.

Table 4 Comparative analysis of FTLSED-VCOOA system with existing techniques

Model	$Accu_y$	$Prec_n$	$Reca_l$	$F_{Measure}$
VGG16 Model	93.50	91.67	92.45	96.81
Inception V3	93.88	92.89	91.93	94.38
ResNet50	90.22	93.40	90.44	96.38
Fast-R-CNN	95.31	95.82	90.21	91.37
Retinanet	93.91	92.55	93.09	92.77
YOLOv5s	94.10	91.61	92.50	96.93
SVM Classifier	98.99	96.06	92.16	91.82
FTLSED-VCOOA	99.11	97.47	97.25	97.33

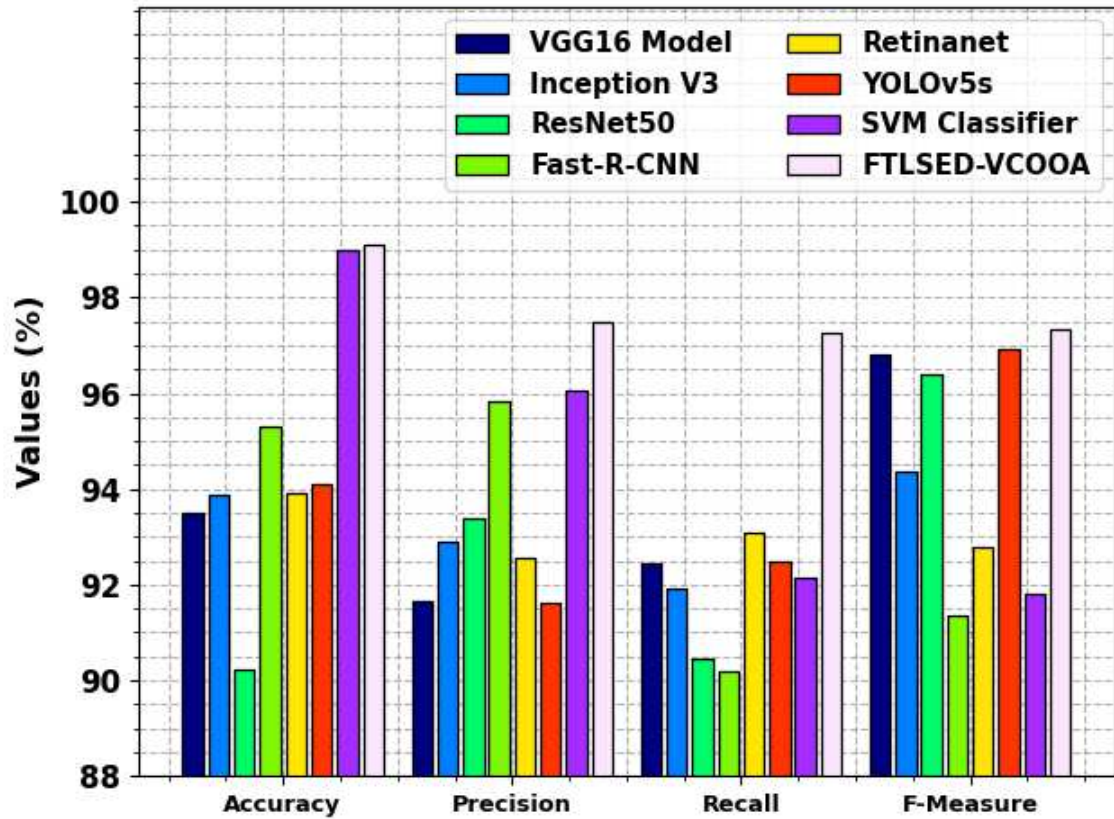


Fig. 12. Comparative analysis of FTLSED-VCOOA system with existing algorithms

In Table 5 and Fig. 13, the processing time (PT) outcome of FTLSED-VCOOA model with existing models is shown. Based on PT, the FTLSED-VCOOA method gets lesser PT of 8.63min whereas the VGG16, Inception V3, ResNet50, Fast-R-CNN, Retinanet, YOLOv5s, and SVM techniques obtain greater PT values of 22.18min, 19.03min, 15.34min, 12.64min, 20.73min, 17.86min, and 14.46min, correspondingly.

Table 5 PT outcome of FTLSED-VCOOA method with existing approaches

Model	Processing Time (min)
VGG16 Model	22.18
Inception V3	19.03
ResNet50	15.34
Fast-R-CNN	12.64
Retinanet	20.73
YOLOv5s	17.86
SVM Classifier	14.46
FTLSED-VCOOA	8.63

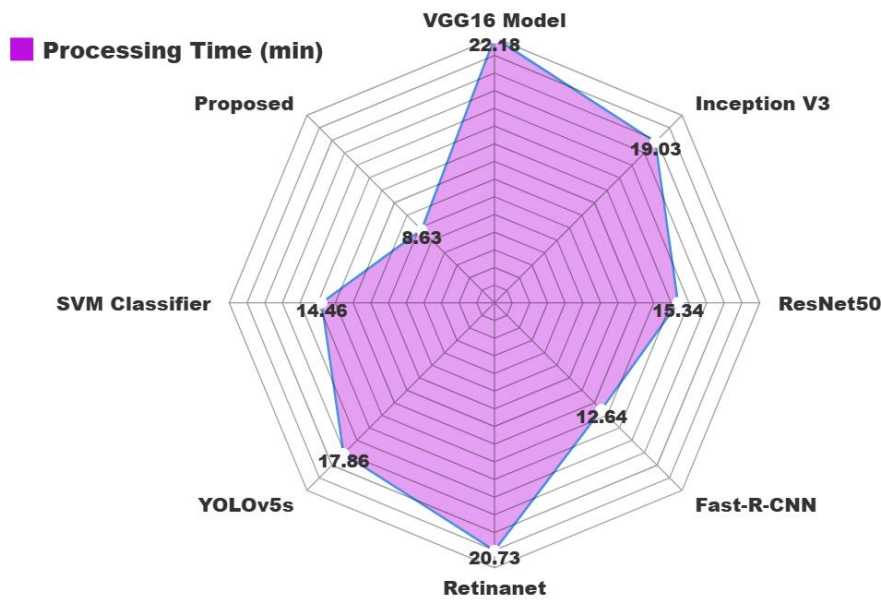


Fig. 13. PT outcome of FTLSED-VCOOA technique with existing models

5. Conclusion

In this paper, we have developed a new FTLSED-VCOOA algorithm. The main aim of the FTLSED-VCOOA system is to detect students' performance in virtual classrooms using advanced DL techniques. At first, the ABF method is employed in the image preprocessing stage to eliminate the noise from an input data. Moreover, the feature extraction process is implemented by fusion models such as Squeezenet, Shallow CapsNets, and MobileNetV2. Besides, the CNN-LSTM technique has been deployed for the classification process. Finally, the OOA-based hyperparameter selection process is achieved to optimize the classification outcomes of the CNN-LSTM system. The simulation analysis of FTLSED-VCOOA system is verified on a benchmark image database and the outcomes are measured under numerous measures. The experimental outcome emphasized the development of FTLSED-VCOOA approach in the student engagement detection process.

Data Availability Statement: The data that support the findings of this study are openly available in Kaggle repository at <https://www.kaggle.com/datasets/joyee19/studentengagement>, reference number [25].

Funding

No funding available in this study.

Declarations

Consent to Publish

This study used a publicly available dataset obtained from the Kaggle repository (Student Engagement Dataset). The authors did not directly recruit participants or collect identifiable personal data. Accordingly, informed consent was not obtained directly by the authors. Any participant consent and ethical requirements related to the original data collection are the responsibility of the dataset creators.

References

- [1] Rivas, A., Gonzalez-Briones, A., Hernandez, G., Prieto, J. and Chamoso, P., 2021. Artificial neural network analysis of the academic performance of students in virtual learning environments. *Neurocomputing*, 423, pp.713-720.
- [2] Waheed, H., Hassan, S.U., Aljohani, N.R., Hardman, J., Alelyani, S. and Nawaz, R., 2020. Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human behavior*, 104, p.106189.
- [3] Chen, Z., Zhang, J., Jiang, X., Hu, Z., Han, X., Xu, M., Savitha, V. and Vivekananda, G.N., 2020. Education 4.0 using artificial intelligence for students performance analysis. *Inteligencia Artificial*, 23(66), pp.124-137.
- [4] Lino, A., Rocha, Á. and Sizo, A., 2019. Virtual teaching and learning environments: automatic evaluation with artificial neural networks. *Cluster Computing*, 22(Suppl 3), pp.7217-7227.
- [5] Bhardwaj, P., Gupta, P.K., Panwar, H., Siddiqui, M.K., Morales-Menendez, R. and Bhaik, A., 2021. Application of deep learning on student engagement in e-learning environments. *Computers & Electrical Engineering*, 93, p.107277.
- [6] Yousafzai, B.K., Khan, S.A., Rahman, T., Khan, I., Ullah, I., Ur Rehman, A., Baz, M., Hamam, H. and Cheikhrouhou, O., 2021. Student-performulator: student academic performance using hybrid deep neural network. *Sustainability*, 13(17), p.9775.
- [7] Chui, K.T., Fung, D.C.L., Lytras, M.D. and Lam, T.M., 2020. Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. *Computers in Human behavior*, 107, p.105584.

- [8] Liu, R., Wang, L., Lei, J., Wang, Q. and Ren, Y., 2020. Effects of an immersive virtual reality-based classroom on students' learning performance in science lessons. *British Journal of Educational Technology*, 51(6), pp.2034-2049.
- [9] Albreiki, B., Zaki, N. and Alashwal, H., 2021. A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences*, 11(9), p.552.
- [10] salh Ali, S., Al'Ameri, J.H. and Abbas, T., 2023. Real Time E-learning Students Monitoring for Optimization Facial Landmark Recognition Based on Hybrid Deep Learning Techniques. *Full Length Article*, 10(2), pp.102-02.
- [11] Dai, W. and Kang, Q., 2025. Improvement of flipped classroom teaching in colleges and universities based on virtual reality assisted by deep learning. *Scientific Reports*, 15(1), p.3204.
- [12] Pabba, C. and Kumar, P., 2024. A vision-based multi-cues approach for individual students' and overall class engagement monitoring in smart classroom environments. *Multimedia Tools and Applications*, 83(17), pp.52621-52652.
- [13] Guo, Y., Xu, B., Lyu, B. and Chen, C., 2025. How social and emotional skills affect Chinese college student engagement in high-impact educational practices: a moderated mediation model. *Studies in Higher Education*, 50(1), pp.27-46.
- [14] Mazumder, D., Chatterjee, A., Chakraborty, A. and Karmakar, R., 2024, June. A Novel Student Engagement Level Detection and Emotion Analysis Using Ensemble Learning. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.
- [15] Zahra, S., Samra, M. and El Gizawi, L., 2025. Working Toward Advanced Architectural Education: Developing an AI-Based Model to Improve Emotional Intelligence in Education. *Buildings*, 15(3), p.356.
- [16] Olaniyan, D., Adebisi, M.O., Adebisi, A.A. and Olaniyan, J., 2024, April. Enhancing Engagement in Virtual Classrooms: A Contactless Multi-Modal Emotion Detection Model. In *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)* (pp. 1-9). IEEE.
- [17] Praveen Kumar, B., Kalpana, A.V. and Nalini, S., 2025. Gated Attention Based Deep Learning Model for Analyzing the Influence of Social Media on Education. *Journal of Experimental & Theoretical Artificial Intelligence*, 37(2), pp.291-305.

- [18] Maddu, R.B.R. and Murugappan, S., 2024. Online learners' engagement detection via facial emotion recognition in online learning context using hybrid classification model. *Social Network Analysis and Mining*, 14(1), p.43.
- [19] Li, H. and Duan, X.L., 2022. SAR ship image speckle noise suppression algorithm based on adaptive bilateral filter. *Wireless Communications and Mobile Computing*, 2022(1), p.9392648.
- [20] Anegundi, R.V., Squeezenet-based Image Classification and Prediction of High Frenal Attachment in the Upper Lip.
- [21] Raji, M., Ahsaei, A.G., Soroush, K. and Ghavami, B., 2025. Progressive Bitwidth Assignment Approaches for Efficient Capsule Networks Quantization. *IEEE Access*.
- [22] Jahan, I., Chowdhury, M.E., Vranic, S., Al Saady, R.M., Kabir, S., Pranto, Z.H., Mim, S.J. and Nobil, S.F., 2025. Deep learning and vision transformers-based framework for breast cancer and subtype identification. *Neural Computing and Applications*, pp.1-20.
- [23] Yang, D., Zhang, Z., Gao, H., Prokhorov, A.V., Chen, Z. and Lv, Y., Variational Autoencoder-Based Learning Intrinsic Periodic-Trend Representations for Short-Term Load Forecasting. *Available at SSRN 5089646*.
- [24] Cao, Y., Wang, W. and He, Y., 2025. Prediction of Heat-Treated Wood Adhesive Strength Using BP Neural Networks Optimized by Four Novel Metaheuristic Algorithms. *Forests*, 16(2), p.291.
- [25] <https://www.kaggle.com/datasets/joyee19/studentengagement>
- [26] Mahmood, N., Bhatti, S.M., Dawood, H., Pradhan, M.R. and Ahmad, H., 2024. Measuring Student Engagement through Behavioral and Emotional Features Using Deep-Learning Models. *Algorithms*, 17(10), p.458.
- [27] Pecuchova, J. and Drlik, M., 2024. Enhancing the Early Student Dropout Prediction Model Through Clustering Analysis of Students' Digital Traces. *IEEE Access*.
- [28] Xu, Q., Wei, Y., Gao, J., Yao, H. and Liu, Q., 2023. ICAPD framework and simAM-YOLOv8n for student cognitive engagement detection in classroom. *IEEE Access*, 11, pp.136063-136076.