Seminars in Medical Writing and Education. 2025; 4:846
doi: 10.56294/mw2025846

ORIGINAL

AI-Based Pattern Recognition Model for Enhancing Student Engagement and Supporting Academic Planning

Modelo de reconocimiento de patrones basado en IA para mejorar la participación estudiantil y apoyar la planificación académica

Huynh Gia Bao[1], Nguyen Van Dai[2]

[1]Department of Chemistry Education, School of Education, Can Tho University. Viet Nam.
[2]Faculty of Chemistry, Hanoi Pedagogical University 2. Viet Nam.

Cite as: Gia Bao H, Van Dai N. AI-Based Pattern Recognition Model for Enhancing Student Engagement and Supporting Academic Planning. Seminars in Medical Writing and Education. 2025; 4:846. https://doi.org/10.56294/mw2025846

Submitted: 23-03-2025 **Revised:** 09-07-2025 **Accepted:** 20-10-2025 **Published:** 21-10-2025

Editor: PhD. Prof. Estela Morales Peralta

Corresponding Author: Huynh Gia Bao

ABSTRACT

With the increase in online education, maintaining student engagement and real-time monitoring has become a significant challenge. Manual monitoring is labour-intensive and ineffective in dynamic virtual learning environments. The purpose of this research is to develop an artificial intelligence (AI)-based pattern recognition model to improve student engagement tracking and academic planning utilizing the PPDAC (Problem, Plan, Data, Analysis, Conclusion) framework. The problem is ineffective engagement monitoring in online situations. The plan entails incorporating AI for emotion and behavior analysis. The data consists of facial expressions and student activity. The analysis uses deep learning (DL) to interpret engagement, and the conclusion supports adaptive, tailored training. Data are collected from Student Engagement and Emotion Recognition Dataset. Use Gaussian blur and Histogram Equalization (HE) to smooth facial images, improve image contrast, and help in low-light situations. Texture-based features are used to recognize expressions using Local Binary Patterns (LBP). This research proposed a novel 3D Convo Flying fox Neural Network (3D-CFNN) model for student engagement monitoring and academic planning. A hybrid proposed model combining a 3D Convolutional Neural Network (3D CNN) for emotion recognition and Flying Fox Optimization (FFO) for parameter tuning was employed. The proposed 3D-CFNN method achieved higher accuracy (0,995), precision (0,988), recall (0,980), and F1-score (0,994) in student identification and emotional state classification, outperforming conventional methods. The proposed AI-based pattern recognition model enables automated, real-time engagement tracking and supports personalized academic planning, leading to improved learning outcomes in virtual environments.

Keywords: Student Engagement; Virtual Environments; Pattern Recognition; Academic Planning; 3D Convo Flying Fox Neural Network (3D-CFNN).

RESUMEN

Con el aumento de la educación en línea, mantener la participación estudiantil y el monitoreo en tiempo real se ha convertido en un desafío significativo. El monitoreo manual es laborioso e ineficaz en entornos dinámicos de aprendizaje virtual. El propósito de esta investigación es desarrollar un modelo de reconocimiento de patrones basado en inteligencia artificial (IA) para mejorar el seguimiento de la participación estudiantil y la planificación académica utilizando el marco PPDAC (Problema, Plan, Datos, Análisis, Conclusión). El problema es la ineficacia del monitoreo de la participación en situaciones en línea. El plan implica incorporar IA para el análisis de emociones y comportamiento. Los datos consisten en expresiones faciales y actividad estudiantil.

El análisis utiliza aprendizaje profundo (DL) para interpretar la participación, y la conclusión respalda el entrenamiento adaptativo y personalizado. Los datos se recopilan del conjunto de datos de participación estudiantil y reconocimiento de emociones. Utilice el desenfoque gaussiano y la ecualización de histograma (HE) para suavizar las imágenes faciales, mejorar el contraste de la imagen y ayudar en situaciones de poca luz. Las características basadas en texturas se utilizan para reconocer expresiones utilizando patrones binarios locales (LBP). Esta investigación propuso un novedoso modelo de Red Neuronal Convolucional 3D (3D-CFNN) para el monitoreo de la participación estudiantil y la planificación académica. Se empleó un modelo híbrido que combina una Red Neuronal Convolucional 3D (3D CNN) para el reconocimiento de emociones y la Optimización Flying Fox (FFO) para el ajuste de parámetros. El método 3D-CFNN propuesto logró mayor precisión (0,995), exactitud (0,988), recuperación (0,980) y puntuación F1 (0,994) en la identificación de estudiantes y la clasificación del estado emocional, superando a los métodos convencionales. El modelo de reconocimiento de patrones basado en IA propuesto permite el seguimiento automatizado de la participación en tiempo real y apoya la planificación académica personalizada, lo que resulta en mejores resultados de aprendizaje en entornos virtuales.

Palabras clave: Participación Estudiantil; Entornos Virtuales; Reconocimiento de Patrones; Planificación Académica; Red Neuronal 3D Convo Flying Fox (3D-CFNN).

INTRODUCTION

Various sectors have transformed as a result of artificial intelligence (AI), which systematizes human work. Pattern recognition is a fundamental factor of AI, allowing robots to evaluate data, spot trends, and formulate well-informed predictions. To improve learning experiences, institutions all around the world are progressively adding digital technologies to courses.(1) Information Technology (IT) in 12-year compulsory schooling clearly places emphasis on supporting students' development of thinking skills in computational and design thinking, while helping to integrate knowledge through meaningful experiential learning and interdisciplinary contexts. Due to the variety and fast-paced evolution of technology materials, curriculum design is becoming increasingly difficult for pedagogical researchers, educators and educational professionals from various backgrounds.(2) AI technology is receiving increasing interest in education given its AI capabilities, which consist of learning, adapting, self-correcting and handling complex problems. AI has the potential to change education in both distributed learning environments and traditional classrooms. Recent studies indicate the significance of AI for flexible language learning, as well as developing students' digital learning wellbeing and self-efficacy.(3)

AI technology in education promotes inclusive education through all learning styles and all levels of ability, creating equal opportunities for all students to be successful. AI supports the lifelong learning journey by recommending personalized feedback for the ongoing development of skills, from early childhood education to graduate school. The possibilities of AI advances indicate more personalization, which increases motivation, engagement, and academic outcomes. The aim of this trend is to create flexible and efficient learning environments to maximize each learner's potential and democratize learning.(4) AI is divided into three types: super intelligent, broad, and narrow. AI uses machine learning (ML) algorithms to pretend human intelligence. The most prevalent application of narrow AI is in education, where it powers technologies like adaptive learning, virtual reality (VR), Augmented Reality (AR), and facial recognition. Applications span from virtual and individualized instruction to automated grading. Classrooms are being transformed by these technologies, which have interactive and intelligent systems. AI in education must support academic abilities while also encouraging creativity and teamwork to remain relevant.(5) The main participants in educational organizations are the students; therefore, the performance of the institution is reflected in their achievement. An important factor in assessing an institution's reputation is the quality of its graduates and postgraduates. In the present competitive academic world, schools work hard to preserve both reputation and quality. Nonetheless, the status of the institution is frequently valued more highly than the actual ability of education provided.(6)

One benefit of using online learning is that it takes into account each student's unique characteristics. Learning through presentations, e-books, animated movies, video lectures, and online articles is among the categories of learning resources mentioned. Teachers arrange these educational materials in the online Learning Management System (LMS) portal so that students access them more easily.(7) The lack of face-to-face interactions in virtual classrooms presents special difficulties, particularly when it comes to sustaining the interest of students and focus in real time. To guarantee the best possible learning experiences and offer insightful information about student participation during online sessions, an efficient attention monitoring system is necessary. Online learning still faces challenges such as inadequate visual cues, poor internet access, device constraints, and distractions at home.(8) According to recent studies, using laptops and cellphones during class causes students to become less focused. The issue of permitting or restricting technology use in the classroom is starting to gain significant traction. The issue is how to use technology to enhance learning while shielding students from needless distractions. From the perspective of the instructor, technology-related diversions among students are equally significant.[9]

Digital platforms are transforming education, reducing student anxiety and promoting active engagement. The acceptance of ML and AI in teaching has accelerated, offering advanced answers like smart tutoring systems. These technologies enhance education quality, enable personalized evaluation, and support the evolution of smart classrooms and virtual learning environments.[10] When compared to standard lectures, active learning strategies that prioritize teacher mentoring and cooperation have been demonstrated to enhance relationships, critical thinking, academic accomplishment, and student retention. However, many online courses during the transition to remote learning lacked interactive components, which decreased student-faculty connections, engagement, and feeling of community.[11] While there are many advantages to distance learning, there are also significant drawbacks, such as the inability to predict the emotions of the student, which are critical for academic performance. Because of the physical distance and lack of useful emotional monitoring tools, instructors find it difficult to detect students' emotions or problems in online learning environments. Numerous solutions have been developed to solve these problems, which involve emotion detection algorithms to assist teachers in better understanding and providing remote support to students.[12]

Strong emotional indicators, such as facial expressions, can provide a wealth of multifaceted information about a person's innermost thoughts and feelings. Emotions are closely linked to cognition and have a big impact on learning in a lot of different areas, such as language, skills, and ethics. Emotions in educational environments influence student engagement and results, as well as the classroom atmosphere. Therefore, to improve the teaching-learning process, educators must comprehend the learning context and analyze students' facial expressions.[13] Traditional classroom instruction has a number of drawbacks, including rigid physical space requirements, inflexibility for both staff and students, accessibility problems, obstacles for people with impairments, high expenses, and transportation-related environmental effects. On the contrary, online education provides more accessibility, lower expenses, less environmental impact, and flexibility in terms of time and place. Thus, online education can overcome or greatly minimize many of the drawbacks of in-person learning.[14] Smart classrooms enhance instruction by assessing the environment and modifying materials or approaches as necessary. Determine students' emotional states using information and communication technology to create more interesting learning opportunities. Teachers evaluate both the emotional and cognitive components of students' learning with the aid of excitement recognition software.[15]

Aim of the Research

The purpose of this research is to develop a 3D Convo Flying Fox Neural Network (3D CFNN) module to improve personalized academic planning and real-time engagement monitoring in virtual learning spaces. The developed approach includes a Flying Fox optimization (FFO) module that will regulate parameters for emotion detection using a 3D Convolutional Neural Network (CNN), while keeping reliability and flexibility across different learning situations. The method is designed to extract local binary pattern (LBP) features while utilizing complex image pre-processing, such as Gaussian Blur and Histogram Equalization, to provide a better consequence of emotional recognition over several lighting scenarios and video quality. Automated, robust, and context specific engagement analysis creates enhancements for suggested context-adaptive pedagogical interventions and improved online learning outcomes.

Key contribution

To create a 3D Convo Flying Fox Neural Network (3D-CFNN) method that accurately monitors and measures student engagement in a virtual learning environment in real-time.

To develop a hybrid model that combines Flying Fox Optimization (FFO) for adaptive parameters with 3D CNN for emotion detection to enhance monitoring.

To implement Local Binary Pattern (LBP) feature extraction to enhance emotion detection in a variable lighting and video quality environment.

The rest of the research follows the phases. Related works on emotional detection were showcased in Phase 2. Phase 3 addresses the methodological framework of the (AI)-based pattern recognition method to improve student engagement tracking, which included feature extraction, data preparation, and the FFO technique. Phase 4 discusses the experiment results and discussion around improvements in accuracy of real-time engagement detection of students and personalized academic planning. Phase 5 concludes the research, emphasizing the framework's potential to provide accurate, scalable, and context-aware monitoring of student engagement across a variety of virtual learning environments.

Related works

To examine children's computational thinking (CT), a supervised deep neural network (DNN) was constructed using images taken from multimodal video data. A high degree of agreement was shown by confusion matrices

https://doi.org/10.56294/mw2025846 ISSN: 3008-8127

comparing AI predictions with human evaluations, suggesting that the AI successfully supported the in-depth examination of students' multimodal interactions. Biases in data categorization continued, as did the need for broader validation across a variety of educational contexts.[16]

Three hybrid deep learning (DL) models were built with EfficientNetB7 with Temporal Convolutional Network (TCN), Long Short-Term Memory network (LSTM), and Bidirectional LSTM (BiLSTM) to categorize student assignation in online educational videos to maximize accuracy by capturing spatial and temporal characteristics.[17] The accuracy of the models was 94,47 %, which was higher than the other existing approaches. The models were also able to accurately identify patterns of engagement. Generalizability was limited.

A quasi-experimental design was used to investigate a combined concept mapping and image recognition (CM-IR) approach to assist students with engagement in scientific inquiry.[18] In studying the CM-IR approach, the researchers were attempting to foster motivation and enhance achievement by scaffolding students to acquire and uniquely organize knowledge. The results showed that CM-IR had a high mental load with increased learning achievement, attitudes, and intrinsic motivation. Limitations of the research involve a small sample size and a single school situation that might limit generalizability of the results.

Using neurocognitive data from functional near-infrared spectroscopy (fNIRS),[19] a method was developed to predict student outcomes assuming the rate of learning within an adaptive synthetic learning environment. The analysis indicated that the approach achieved 85 % predictive accuracy. In addition, this method performed better than conventional data-driven methods that utilize a minimal amount of relevant learning session data. However, this method analyzed focused on only one specific part of lesson content. The method requires further testing to validate for larger populations.

A real-time artificial intelligence (AI)-based framework[20] was created to research students' engagement in the classroom by examining behaviors. Robust engagement evaluations were measured, which allowed for the real-time use of teaching and learning interventions. The trials illustrated the capacity to quantitatively compare the various teaching methods, and showed greater flexibility and accuracy. Limitations of the framework include a need for future trials in a more diverse range of educational contexts, and issues related to the potential privacy of students.

The best algorithm to pull data from a virtual learning environment (VLE) to forecast student contribution in courses was recognized.[21] Accuracy, precision, recall, and area under the receiver operating characteristic (AUC) curve scores were employed to evaluate and test five different classification algorithms using the data through cross-validation. The categorical boosting (CATBoost) model was the best model of all the models with respect to both accuracy and recall. This demonstrates that it is able to predict engagement in a useful way. Limitations include being reliant on VLE data and needing to validate results in other educational contexts.

A system in real time that tracked student group participation by facial features was created.[22] The methodology included estimating group involvement on a frame-wise basis, face identification, and a CNN-based facial emotional recognition model. Performance showed a strong correlation to students' self-reports of participation, demonstrating a testing and training accuracy of 76,90 % and 78,70 %, respectively. However, the dependence on the visibility and quality of students' faces in the video feed indicated limitations.

An AI-powered automated system that uses voice and face recognition to track teachers' and students' speech, facial expressions, and interactions in real time.[23] The results demonstrated that this one-on-one online learning environment attracted the interest of young children. The concentration on a single learning format and the short sample size were limitations that could compromise generalizability.

An improved CNN was created to precisely identify students' emotions to expand results in learning, motivation, and attention in classrooms.[24] The model's design, which included convolutional, batch normalization, and dropout layers, allowed it to classify seven emotions. Strong real-time applicability was indicated by the results, which exhibited 95 % test accuracy with great precision and recall across all categories. Reliance on a single dataset and the requirement for validation in a variety of authentic classroom settings are among the limitations.

A system for recognizing emotional expressions in an AI-based classroom, by using a Control-Value Theory framework.[25] It tried to identify the capacity of the system to recognize reactions accurately, and then how closely it approximated self-report metrics. There was a strong agreement in the results reported in the systems and the self-reporting of the students who disclosed that they preferred active learning approaches over passive. However, the precision was not as good as Control-Value Theory (CVT), and it was not sensitive enough to volitional elements.

A real-time video facial processing pipeline[26] was suggested to predict students' levels of engagement, individual affect, and group affect. The method included face detection, tracking, and clustering, then a robust optimization scheme and a pre-trained neural network (NN) optimized for facial expression detection. The experimental result showed that produced lesson summaries from emotion and engagement clips were definitely better than existing individual models available. Nonetheless, the dependency on lighting and facial visibility posed challenges for trustworthiness with respect to reliability.

https://doi.org/10.56294/mw2025846 ISSN: 3008-8127

An online learning environment that is scalable, tracks students' progress and accepts instructional modifications in real-time based on facial expression detection was proposed.[27] Utilized a ResNet-50-based model that was enhanced with a convolutional attention mechanism and modified residual down-sampling module to help reduce noise and increase relationship with features. The proposed model yielded better results than the base ResNet-50 and other state-of-the-art methods with 87,62 % accuracy. One limitation is that the system relies on camera placement and image quality to operate effectively.

An Image Emotion Recognition-Based Online Learning Behavior Analysis Model was suggested[28] to identify and react to shifts in students' emotions. The model used an attention mechanism-based classification strategy for emotion recognition and an enhanced Local Binary Pattern (LBP) and wavelet transform for key frame extraction. The outcomes validated the algorithm's capacity to analyze online learning practices effectively. Nevertheless, drawbacks include the possibility of errors in dimly lit environments, limited flexibility in recognizing a range of facial expressions, and dependence on camera quality for precise identification.

Real-time tracking of students' academic performance was made possible by the Academic Progress Monitoring Framework, which combined AI and the Internet of Everything (IOE).[29] It generated individualized learning paths, properly assessed progress, and identified learning gaps. The outcomes demonstrated increased student engagement, more instructional flexibility, and more successful individualized learning opportunities. For educators, the system facilitated scalable, data-driven decision-making. However, the intricacy of large-scale AI-IoE integration, data protection, and ethical issues continues to be a major obstacle.

A facial cues-based multimodal engagement detection framework was created to monitor student engagement during e-learning by analyzing facial reactions, head movements, and eye blinks using DL models, such as Residual network with 50 layers (ResNet-50) and Visual Geometry Group network with 19 layers (VGG-19).[30] The combined data generated an Engagement Index (EI) to organize engagement states. Results showed the system achieved 92,58 % accuracy, outperforming existing approaches. However, its performance depends on video quality and was affected by lighting, camera angles, and network issues.

Research gap

With the rise of online education, maintaining student engagement in real time is challenging, as manual monitoring is labor-intensive and ineffective. Existing methods, such as supervised DNN for multimodal Computed Tomography (CT) analysis,[16] hybrid EfficientNet with Temporal Convolutional Networks (EfficientNet+TCN)/ Long Short-Term Memory (LSTM) networks, and Bidirectional LSTM (BiLSTM),[17] Functional Near-Infrared Spectroscopy (fNIRS)-based neurocognitive predictors,[19] real-time behavior-based AI frameworks,[20] and facial-cues multimodal engagement detection using VGG-19/ResNet-50,[30] have advanced engagement detection but remain limited by bias, poor generalizability, sensitivity to lighting and occlusion, high computational cost, and static optimization. To address these gaps, this study proposes a 3D Convo Flying Fox Neural Network (3D-CFNN) combining 3D CNN for spatiotemporal emotion recognition with Flying Fox Optimization (FFO) for adaptive parameter tuning. Preprocessing uses Gaussian blur, histogram equalization, and Local Binary Pattern (LBP) for robust feature extraction. The 3D-CFNN enables real-time, scalable student engagement tracking and adaptive academic planning, improving accuracy and personalization in virtual learning environments.

METHOD

The AI-Based student engagement monitoring framework enhances academic planning and real-time engagement in online education by applying DL and optimization. It gathers multimodal data, such as student activity and facial expressions, then enhances images using Gaussian Blur and HE. LBP is used to extract features. A hybrid 3D-CFNN combines FFO for adaptive tuning with 3D CNN for spatiotemporal emotion identification. Figure 1 illustrates the entire workflow of the proposed method.

Data collection

Student Engagement and Emotion Recognition Dataset is obtained from Kaggle (https://www.kaggle.com/datasets/zyan1999/student-engagement-and-emotion-recognition-dataset/data). This dataset includes documentation of students' emotional states and levels of involvement in virtual learning settings. In addition to ambient context like illumination, it contains student identities, session information, timestamps, and activity metrics like mouse and keyboard usage, chat interactions, and screen focus. With an emotion state label of bored, delight, neutral, surprise, or upset as the target variable, academic performance scores are recorded for every entry. By making it possible to examine student behavior, engagement trends, and emotional patterns in virtual classrooms, the dataset aids educational analytics research. It works well for tasks that require analysis and prediction to improve learning outcomes and flexible teaching methods in online learning.

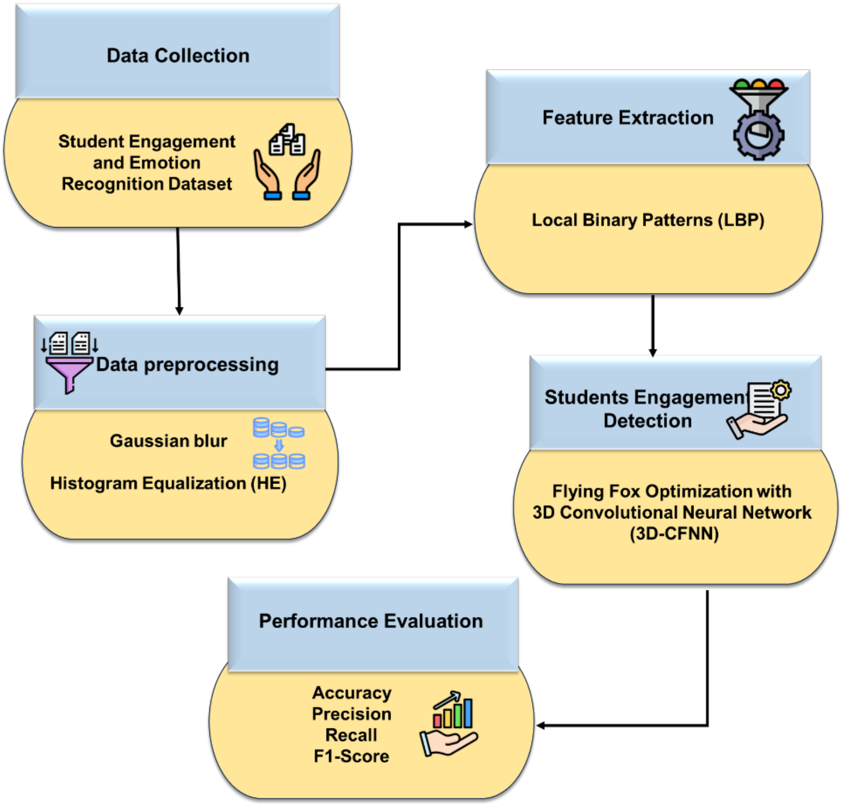

Figure 1. Proposed flow of 3D-CFNN method

Data preprocessing

Gaussian blur and histogram equalization are image preprocessing techniques that also improve the facial image quality and accuracy of emotion recognition to enhance real-time student engagement processes in online learning environments.

Gaussian blur is used as a smoothing filter to improve the quality of a facial image when used in student engagement recognition. This method is used to reduce noise and improve the accuracy of feature extraction. This processing method applies a Gaussian kernel for a convolution of the image by weighting pixel values based on a 2D Gaussian distribution. The Gaussian kernel uses the Gaussian function in equation (1) to fix the distance from the kernel center, which regulates the weights assigned to each pixel in the image.

$$H(y,z) = \frac{1}{2\pi\sigma^2} exp\left(\frac{-y^2+z^2}{2\sigma^2}\right) \quad (1)$$

Where y, z are the coordinates relative to the kernel center and σ is the standard deviation controlling the kernel's spread. H acts as a low-pass filter. The kernel transfers over the image during convolution, multiplying overlapping pixel values by the related kernel weights and adding up the results to control the new value for each pixel. By dropping high-frequency noise, this technique smooths the image while maintaining vital structural elements like edges and face shapes. In the proposed AI-based pattern recognition model, enhancing facial image contrast is vital for accurate emotion recognition and engagement monitoring. Histogram equalization is employed as an image enhancement technique to reallocate pixel intensities, refining brightness and contrast without losing significant details.

Histogram Equalization aims to spread the intensity values so that all gray levels have approximately equal representation, which helps highlight facial features even in poorly lit conditions. Given a digital image H with gray levels ranging from [0,H-1], the Probability Density Function (PDF) for each gray level s_i is calculated in equation (2).

$$PDF\ (s_i) = \frac{p_i}{P}, i = 0,1,2, \dots, H-1 \quad (2)$$

Where p_i refers to the number of pixels at gray level s_i, and P is the total number of pixels in the image. Next, the Cumulative Distribution Function (CDF) is computed using the PDF values in equation (3).

https://doi.org/10.56294/mw2025846 ISSN: 3008-8127

$$CDF\ (s_i) = \sum_{k=0}^{i} PDF(r_k) \quad (3)$$

The equalized pixel intensity T_i is then derived by scaling the PDF by the maximum gray level in equation (4).

$$T_i = CDF\ (s_i) \times (H-1) \quad (4)$$

After rounding, each pixel value in the image is replaced by its corresponding T_i, resulting in an image with enhanced contrast. By applying HE, the proposed model ensures improved image quality, aiding the extraction of texture-based features like Local Binary Patterns (LBP) and enhancing the performance of the 3D-CFNN for student engagement detection.

Feature Extraction

To extract unique face expression patterns from pre-processed images, the suggested AI-based pattern recognition model uses Local Binary Patterns (LBP), a reliable texture-based feature extraction technique. Because LBP encodes local spatial features in grayscale images and is invariant to monotonic illumination variations, it is especially useful for emotion recognition. This is crucial for online learning scenarios where lighting conditions change. For a grayscale image, the basic LBP operator is defined across a 3 × 3 neighborhood. The local texture is then represented by converting the binary pattern into a decimal number calculated in equation (5).

$$t(h_q - h_d) = \begin{cases} 1, h_q - h_d \geq 0 \\ 0, h_q - h_d < 0 \end{cases} q = 0,1,2, \ldots, Q-1 \quad (5)$$

As $t(h_q$-$h_d)$ threshold, the gray value of the central pixel (h_d) is compared to the gray values of each bordering pixel (h_q). It attains a binary value of 1 if the value of the surrounding pixel is greater than or equal to the value of the central pixel, and 0 otherwise.

Where h_q refers to the gray value of the q^{th} neighboring pixel. h_d denotes the gray value of the central pixel, and Q is the number of neighboring pixels. The LBP code for the central pixel is computed in equation (6).

$$LBP_{Q,S} = \sum_{q=0}^{Q-1} t\ (h_q - h_d) \times 2^q \quad (6)$$

Where S,Q is the radius of the neighborhood. Multi-scale texture patterns are captured by using a circular neighborhood with radius S, as opposed to the fixed 3×3 neighborhood used by the original LBP. By rotating the binary pattern in a cyclical fashion and choosing the lowest value as the final LBP code, rotation invariance is accomplished. The suggested 3D-CFNN enhances the accuracy of real-time student involvement recognition in virtual learning settings by using LBP and its variations on Gaussian-blurred and histogram-equalized images to extract stable and discriminative emotion variables.

Student engagement detection using 3D Convolutional Flying Neural Network (3D-CFNN)

To provide real-time engagement monitoring of students in online learning, this research presents a hybrid 3D-CFNN approach that combines a 3D-CNN with FFO to adjust deep learning parameters on the fly, and to recognize students across a broad range of activity and light conditions while improving recognition accuracy. The 3D-CNN extracts spatiotemporal features from facial expressions and behavioral inputs, FFO provides adaptive parameter adjustment in dynamic scenarios of interaction. Gaussian blur and HE enhance image clarity and robust feature extraction, and LBP is utilized for identifying emotion from textural features. Consequently, the hybrid method enables automated and user-independent personalized academic planning, and low-latency engagement monitoring.

3D Convolutional Neural Network (3D-CNN)

A 2D Convolutional Neural Network (2D-CNN) extracts information only from the spatial dimension by applying convolution operations to a two-dimensional feature map to classify the emotional state of the students. The output value w_{jk}^{yz} at position (y,z) in the kth feature map of the jth layer is formalized as follows in equation (7).

$$w_{jk}^{yz} = \left(\phi\ c_{jk} + \sum_n \sum_{q=0}^{Q_j-1} \sum_{r=0}^{R_j-1} x_{jkn}^{qr} \times w_{(j-1)n}^{(y+q)(z+r)}\right) \quad (7)$$

ϕ is the activation function, c_{jk} is the bias for the feature map, and x_{jkn}^{qr} is the weight parameter at kernel position (q,r) connected to the n^{th} feature map in the previous layer. Q_j and R_j are the height and width of the kernel.

A 3D Convolutional Neural Network (3D-CNN) applies convolution kernels across three-dimensional data volumes, concurrently extracting characteristics from temporal and spatial dimensions. This is especially crucial for facial engagement recognition when expressions change over time and convey essential information. Position (y,z,a) in the k^{th} feature map of the j^{th} layer has the following value in equation (8).

$$w_{jk}^{yza} = \left(\phi\, c_{jk} + \sum_n \sum_{q=0}^{Q_j-1} \sum_{r=0}^{R_j-1} \sum_{s=0}^{S_j-1} x_{jkn}^{qrs} \times w_{(j-1)n}^{(y+q)(z+r)(a+r)}\right) \qquad (8)$$

Where x_{jkn}^{qrs} is the weight parameter at position (q,r,s) of the kernel associated with the n^{th} feature map in the preceding layer, and S_j is the kernel's depth. Both the temporal dynamics and the spatial facial structure of expressions from online learning sessions are captured by the 3D-CNN component of the proposed 3D-CFNN. Improved accuracy in real-time student engagement monitoring results from the model's ability to identify invisible, time-dependent engagement indicators that 2D-CNNs overlook.

Flying Fox Optimization (FFO): Flying Fox Optimization (FFO) is leveraged to optimize hyperparameters of the proposed 3D-CFNN model to enhance the accuracy of emotion recognition and engagement classification in e-learning contexts. FFO is designed for high-dimensional, non-linear search spaces, providing adaptability for more complex optimization problems.

The approach is inspired by the flying foxes' behavior for survival by performing better global exploration alongside better local exploitation, preventing any unwanted premature convergence. The learning ability of FFO along with the dynamically adaptable balance helps prevent overfitting to improve generalization and helps to push the model toward potentially better solutions during training, thus improving performance from training to testing and increasing reliability, as calculated in equation (9).

$$y_{j,k}^{u+1} = y_{j,k}^{u} + \alpha.rand.(cool_{j,k} - y_{j,k}^{u}) \qquad (9)$$

Where $y_{j,k}^{u}$ Current position, $y_{j,k}^{u+1}$ updated position of the FFO then α is the attraction constant and stochastic exploration is guaranteed by rand~V (0,1) refers to attraction constant. $cool_{j,k}$ denotes best known current position. The candidate solution investigates nearby areas to prevent convergence stagnation for close to optimal solution in Equations (10-11).

$$cx_{jk}^{u+1} = y_{jk}^{u} + rand1_k.\left(cool_{jk} - y_{ik}^{u}\right) + rand2_k.\times (y_{s1,k}^{u} - y_{s2,k}^{u}) \qquad (10)$$

$$y_{j,k}^{u+1} \begin{cases} cx_{jk}^{u+1}, \; if\ k = m\ or\ rand\ 3_k \geq qb \\ y_{j,k}^{u+1}, \quad otherwise \end{cases} \qquad (11)$$

Where $y_{s1,k}^{u}$, and $y_{s2,k}^{u}$ are parameters from two random individuals, and qb refers to the probability constant controlling element update. m denotes a random parameter index to ensure diversity. If a flying fox is too far from the "coolest" location and cannot return, it "dies" and is replaced using the average from a survival list (SL) in equation (12).

$$y_{j,k}^{u+1} = \frac{\sum_{l=1}^{c} SL_{l,k}^{u}}{c} \qquad (12)$$

Where c is between 2 and the SL size, and $SL_{l,k}^{u}$ is the k^{th} parameter of the l^{th} best solution. Under suffocation, when multiple individuals have the same values, the suffocation probability in equation (13).

$$qE = \frac{cd-1}{Population\ size} \qquad (13)$$

where cd indicates the number of clusters or categories taken into account during parameter tuning, and the qE population size indicates the total number of potential solutions in the optimization process. To ensure that the distribution of people is normalized over these intervals, the number of intervals between clusters is subtracted by one. By minimizing premature convergence and preserving a balance between exploration and exploitation during Flying Fox Optimization, this normalization enhances the precision and consistency of engagement recognition in dynamic online learning settings.

Model parameters, including convolution kernel sizes, learning rates, and filter counts, are adaptively adjusted by integrating FFO into the 3D-CFNN training procedure. In the varied lighting, facial expression, and student behavior variables found in online learning environments, this guarantees quicker convergence, less overfitting, and enhanced identification accuracy. Algorithm 1 shows the proposed 3D-CFNN method.

Algorithm 1: 3D-CFNN method

```
Step 1: Preprocess video clip frames
def Preprocess_Frame(frame):
    blurred = gaussian_blur(frame, ksize=(5,5), sigma=1,0)
    equalized = histogram_equalization(blurred)
    gray = to_grayscale(equalized)
    lbp_map = local_binary_pattern(gray, P=8, R=1)
    return stack_channels([gray, lbp_map])
def Preprocess_Clip(frames):
    processed_frames = [Preprocess_Frame(f) for f in frames]
    return normalize(np.stack(processed_frames, axis=0))
Step 2: Define the 3D-CFNN model
def ThreeD_CNN_Model(data, params):
    model = build_3dcnn(in_channels=params['in_channels'],
                    num_classes=params['num_classes'],
                    base_filters=params['filters'],
                    dropout=params['dropout'])
    output = model.forward(data)
    return output
Step 3: Define the fitness function for optimization
def fitness_func(params):
    outputs = ThreeD_CNN_Model(train_data, params)
    return evaluate(outputs, val_labels)
Step 4: Optimize 3D-CFNN hyperparameters using FFO
def FFO_Optimizer(fitness_func, bounds, pop_size, max_iter):
    population = initialize_population(bounds, pop_size)
    for iter in range(max_iter):
        exploration_rate = 1 - (iter / max_iter)
        for i, agent in enumerate(population):
            fitness = fitness_func(agent)
            agent = update_position_flyingfox(agent, population, exploration_rate)
            agent = clip(agent, bounds)
            update_best(agent, fitness)
            population[i] = agent
    return best_solution
Step 5: Academic planning based on engagement predictions
def Academic_Planning(engagement_predictions):
    if np.mean(engagement_predictions[-10:]) < 0,6:
        return {'action': 'intervene', 'strategy': 'short_quiz'}
    else:
        return {'action': 'advance', 'strategy': 'next_module'}
Load and preprocess test clip
processed_clip = Preprocess_Clip(test_frames)
Optimize 3D-CFNN using FFO
best_params = FFO_Optimizer(fitness_func,
                    bounds={'in_channels': (2, 2),
                        'num_classes': (5, 5),
                        'filters': (16, 64),
                        'dropout': (0,0, 0,6)},
                    pop_size=10,
                    max_iter=20)
Predict engagement using best parameters
engagement_predictions = ThreeD_CNN_Model(processed_clip, best_params)
Generate academic action plan
plan = Academic_Planning(engagement_predictions)
```

RESULTS AND DISCUSSION

An experimental setup is used to develop and estimate the proposed 3D-CFNN-based student engagement monitoring and academic planning model. This section outlines the system specifications, software environment, and programming tools employed during the implementation and evaluation of the proposed AI-based pattern recognition framework. Table 1 represents the experimental setup for the 3D-CFNN method.

Table 1. Experimental setup for AI-based student engagement monitoring system

Component	Specification
Processor (CPU)	Intel Core i7-12900K
GPU	NVIDIA RTX 3090
RAM	64 GB DDR4
Operating System	Windows 10
Programming Language	Python 3.10
Libraries & Tools	OpenCV 4,7, Scikit-learn, NumPy, Pandas

Overall performance outcome

Figure 2 presents the multi-class Precision-Recall curves applied to the 3D CFNN model for student emotion recognition. This figure demonstrates classification performance across five emotions - boredom, joy, neutral surprise, upset, and the model has developed some ability to discriminate across emotional states.

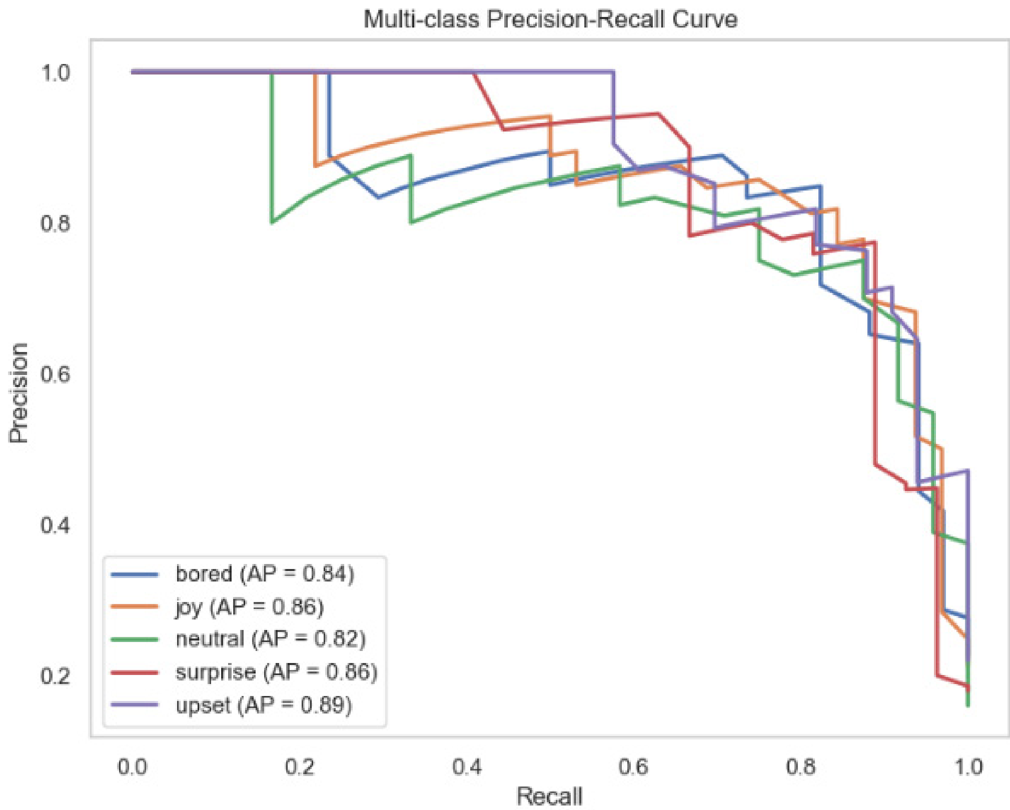

Figure 2. Multi-class Precision-Recall curves

As shown in figure 3, the multi-class ROC curves of the proposed 3D CFNN method for student emotion recognition demonstrate comparative classification performance for five emotional states: bored, joy, neutral, surprise, and upset. The multi-class ROC curves provide the ability to showcase the method's strong discriminative ability across all classes.

The confusion matrix for the proposed 3D CFNN model for student emotion recognition is given in figure 4. The matrix shows correct and incorrect predictions for five emotional states: bored joy, neutral, surprise, and upset. Thus describing both the model's accuracy classification and the errors it made.

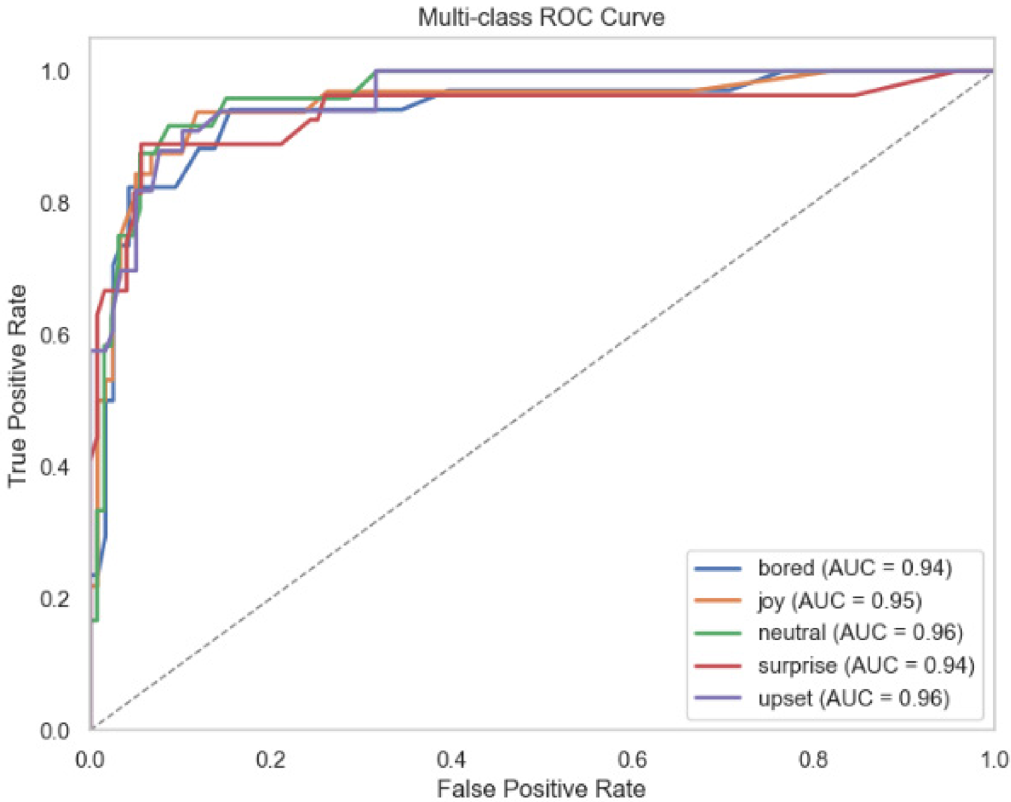

Figure 3. Multi-class ROC curves

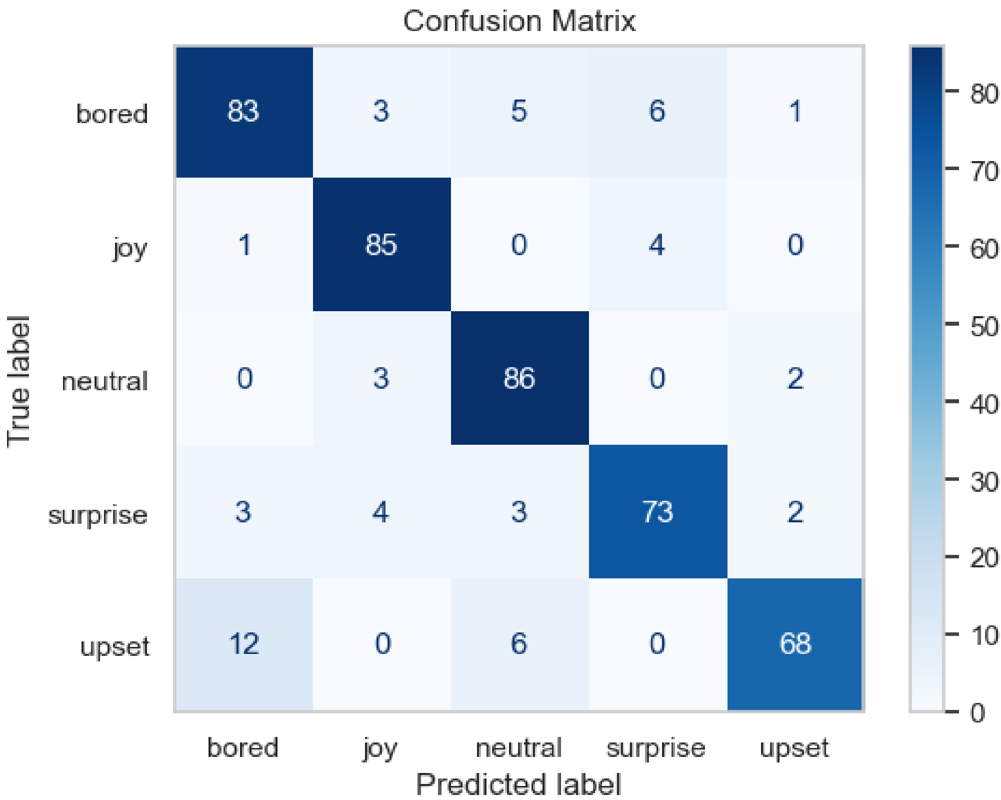

Figure 4. Confusion matrix

Comparison phase

The proposed 3D-CFNN method is compared to the existing CNN method[31] with the three-evaluation metrics precision, recall, and F1 score, and deep neural network (DNN) (combining dual-modality spatiotemporal feature learning) method[31] in terms of accuracy in student engagement monitoring.

Accuracy: reflects how effectively the engagement monitoring model correctly identifies both engaged and disengaged students in online learning sessions, offering an overall measure of the system's ability to classify participation states accurately across diverse conditions.

Precision: shows how many students are truly engaged, ensuring reliable recognition of genuine participation and minimizing false positives that could mislead adaptive academic planning strategies in virtual classroom environments.

Recall: measures the model's ability to detect all genuinely engaged students, reducing missed detections and supporting more complete engagement tracking, which is vital for providing timely academic interventions in online knowledge contexts.

F1-score: combines recall and precision into a stable performance metric, showing the 3D-CFNN model's capability to identify engagement accurately while maintaining low rates of both missed detections and incorrect positive classifications.

Table 2 shows the outcome of the student monitoring 3D-CFNN method achieved in the highest precision (0,988), recall (0,980), and F1-score (0,994) compared to the existing methods involving CNN.[31] Table 2 and figure 5 show the outcome of the proposed 3D-CFNN method.

Table 2. Comparison with existing method

Method	Precision	Recall	F1-score
CNN[31]	0,983	0,972	0,991
3D-CFNN [Proposed]	0,988	0,980	0,994

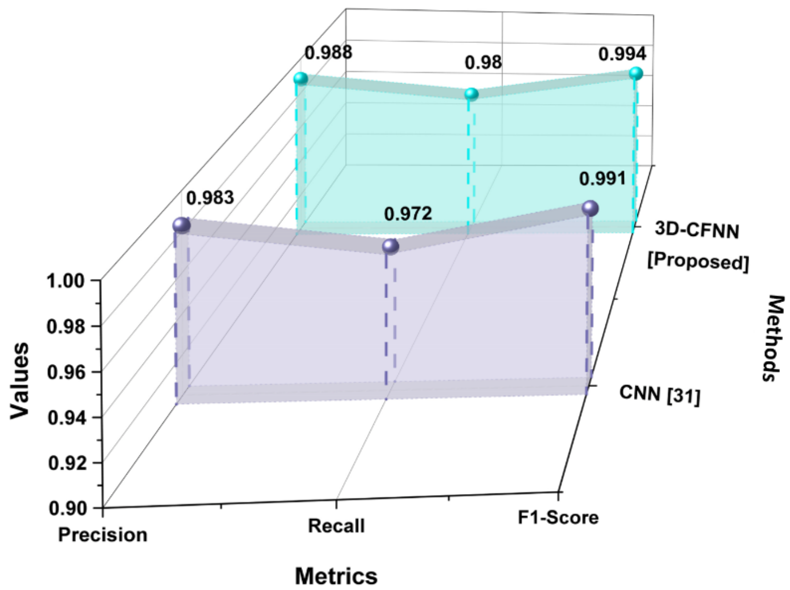

Figure 5. Representation of metrics

Table 3 and figure 6 show the outcome of the student monitoring method achieved in accuracy (0,995) compared to the existing methods, which involve CNN[31] and DNN.[32]

Table 3. Comparison of Accuracy

Method	Accuracy
DNN[32]	0,839
CNN[31]	0,991
3D-CFNN [Proposed]	0,995

In dynamic virtual learning environments, existing DL techniques for identifying student engagement have significant advantages but also significant drawbacks. While the CNN[31] method demonstrates outperformance on various recognition metrics, it is solely independent of spatial features, which limits flexibility and adaptability for detecting and measuring slight temporal changes in student engagement. Because of this dependence on

https://doi.org/10.56294/mw2025846 ISSN: 3008-8127

spatial features, it decreases the ability to fully observe fast changes in facial expressions and behavioral changes, making the CNN method less applicable in dynamic online learning environments.

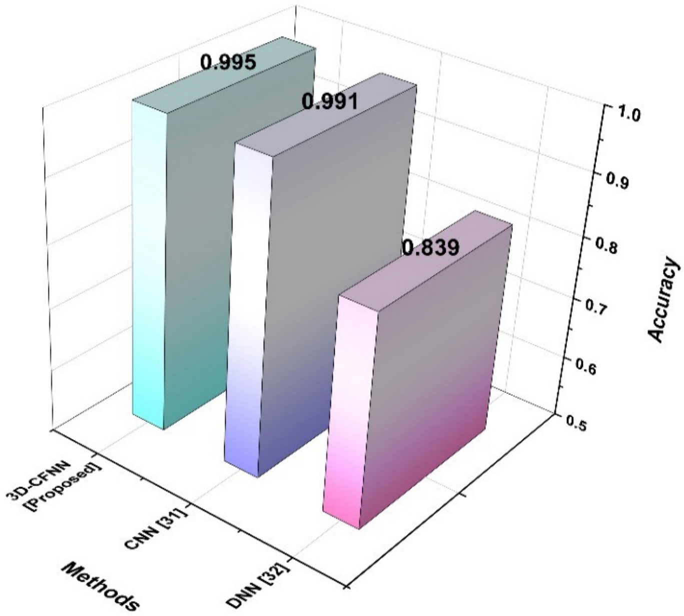

Figure 6. Representation of accuracy

The DNN[32] method displays the ability for multimodal spatiotemporal feature integration and provides a better representation of student behavior for exploratory pedagogical practices. However, without a proper optimization method and a lack of flexibility under differing environmental conditions involving low light and different angles of the face, the DNN method faced limitations in robust use in practical student learning designs. The proposed 3D-CFNN offers a solution to these shortcomings through enhanced spatiotemporal modelling and well-tuned parameters that are more suitable for detecting engagement occurring in real time in a virtual space. To address the issues of the DNN model, a combination of 3D CNN-based feature extraction and the FFO method, resulting in better performance, improved generalization, and increased durability in student learning contexts.

CONCLUSIONS

In online learning environments, this research provides a strong AI-based framework for adaptive academic planning and real-time student involvement monitoring by utilizing the proposed 3D-CFNN method. Student identification and emotional state classification accuracy are increased by the use of 3D CNN for emotion detection and Flying Fox Optimization for parameter tuning. Results from experiments employing pre-processed face and behavioral data show better performance with higher accuracy (0,995), precision (0,988), recall (0,980), and F1-score (0,994) than traditional approaches, allowing for more individualized and successful learning interventions in online learning environments. Extreme shadowing or low-quality video feeds may cause the model to perform less accurately, and extensive datasets may need a lot of processing power during initial training. Future research will investigate how to expand accessibility and scalability by incorporating multimodal sensing data, implementing adaptive learning for ongoing model updates, and optimizing the architecture for deployment on low-power edge devices.

BIBLIOGRAPHIC REFERENCES

1. Kazimova, D., Serikbayeva, N., Samashova, G., Zatyneyko, A., and Sarsenbayeva, B. Evaluating the impact of pattern recognition on AI skills development in universities. In Frontiers in Education. Frontiers Media SA. 2025. https://doi.org/10.3389/feduc.2025.1592209.

2. Hsu, T.C., Abelson, H., Lao, N., Tseng, Y.H. and Lin, Y.T. Behavioral-pattern exploration and development of an instructional tool for young children to learn AI. Computers and Education: Artificial Intelligence. 2021. https://doi.org/10.1016/j.caeai.2021.100012.

3. Chen, X., Xie, H., Qin, S.J., Wang, F.L. and Hou, Y. Artificial Intelligence-Supported Student Engagement Research: Text Mining and Systematic Analysis. European Journal of Education. 2025. https://doi.org/10.1111/ejed.70008.

4. Xu, Z. AI in education: Enhancing learning experiences and student outcomes. Applied and Computational Engineering. 2024. https://doi.org/10.54254/2755-2721/51/20241187.

5. Lin, H. and Chen, Q. Artificial intelligence (AI)-integrated educational applications and college students' creativity and academic emotions: students and teachers' perceptions and attitudes. BMC psychology. 2024. https://doi.org/10.1186/s40359-024-01979-0.

6. Khan, I., Ahmad, A.R., Jabeur, N. and Mahdi, M.N. An artificial intelligence approach to monitor student performance and devise preventive measures. Smart Learning Environments. 2021. https://doi.org/10.1186/s40561-021-00161-y.

7. Pardamean, B., Suparyanto, T., Cenggoro, T.W., Sudigyo, D. and Anugrahana, A. AI-based learning style prediction in online learning for primary education. Ieee Access. 2022. https://doi.org/10.1109/ACCESS.2022.3160177.

8. Hossen, M.K. and Uddin, M.S. Attention monitoring of students during online classes using XGBoost classifier. Computers and Education: Artificial Intelligence. 2023. https://doi.org/10.1016/j.caeai.2023.100191.

9. Su, M.C., Cheng, C.T., Chang, M.C. and Hsieh, Y.Z. A video analytic in-class student concentration monitoring system. IEEE Transactions on Consumer Electronics. 2021. https://doi.org/10.1109/TCE.2021.3126877.

10. Bhardwaj, P., Gupta, P.K., Panwar, H., Siddiqui, M.K., Morales-Menendez, R. and Bhaik, A. Application of deep learning on student engagement in e-learning environments. Computers & Electrical Engineering. 2021. https://doi.org/10.1016/j.compeleceng.2021.107277.

11. Cavinato, A.G., Hunter, R.A., Ott, L.S., and Robinson, J.K. Promoting student interaction, engagement, and success in an online environment. 2021. https://doi.org/10.1007/s00216-021-03178-x.

12. Ozdamli, F., Aljarrah, A., Karagozlu, D. and Ababneh, M. Facial recognition system to detect student emotions and cheating in distance learning. Sustainability. 2022. https://doi.org/10.3390/su142013230.

13. Fakhar, S., Baber, J., Bazai, S.U., Marjan, S., Jasinski, M., Jasinska, E., Chaudhry, M.U., Leonowicz, Z., and Hussain, S. Smart classroom monitoring using novel real-time facial expression recognition system. Applied Sciences. 2022. https://doi.org/10.3390/app122312134.

14. Ahmad, I. A novel deep learning-based online proctoring system using face recognition, eye blinking, and object detection techniques. International Journal of Advanced Computer Science and Applications. 2021. http://dx.doi.org/10.14569/IJACSA.2021.0121094

15. Villegas-Ch, W.E., García-Ortiz, J. and Sánchez-Viteri, S. Identification of emotions from facial gestures in a teaching environment with the use of machine learning techniques. IEEE Access. 2023. https://doi.org/10.1109/ACCESS.2023.3267007.

16. Ocak, C., Kopcha, T.J. and Dey, R. An AI-enhanced pattern recognition approach to temporal and spatial analysis of children's embodied interactions. Computers and Education: Artificial Intelligence. 2023. https://doi.org/10.1016/j.caeai.2023.100146.

17. Selim, T., Elkabani, I. and Abdou, M.A. Students engagement level detection in online e-learning using hybrid efficientnetb7 together with tcn, lstm, and bi-lstm. Ieee Access. 2022. https://doi.org/10.1109/ACCESS.2022.3206779.

18. Huang, H.L., Hwang, G.J. and Chen, P.Y. An integrated concept mapping and image recognition approach to improving students' scientific inquiry course performance. British Journal of Educational Technology. 2022. https://doi.org/10.1111/bjet.13177.

19. Lamb, R., Neumann, K. and Linder, K.A. Real-time prediction of science student learning outcomes using machine learning classification of hemodynamics during virtual reality and online learning sessions. Computers and Education: Artificial Intelligence. 2022. https://doi.org/10.1016/j.caeai.2022.100078.

20. Alkabbany, I., Ali, A.M., Foreman, C., Tretter, T., Hindy, N. and Farag, A. An experimental platform for real-time students engagement measurements from video in STEM classrooms. Sensors. 2023. https://doi.org/10.3390/s23031614.

21. Alruwais, N. and Zakariah, M. Student-engagement detection in classroom using machine learning algorithm. Electronics. 2023. https://doi.org/10.3390/electronics12030731.

22. Pabba, C. and Kumar, P. An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition. Expert Systems. 2022. https://doi.org/10.1111/exsy.12839.

23. Xu, X., Dugdale, D.M., Wei, X. and Mi, W. Leveraging artificial intelligence to predict young learner online learning engagement. American Journal of Distance Education. 2023. https://doi.org/10.1080/08923647.2022.2044663.

24. Salloum, S.A., Alomari, K.M., Alfaisal, A.M., Aljanada, R.A. and Basiouni, A. Emotion recognition for enhanced learning: using AI to detect students' emotions and adjust teaching methods. Smart Learning Environments. 2025. https://doi.org/10.1186/s40561-025-00374-5.

25. Fernández Herrero, J., Gómez Donoso, F. and Roig Vila, R. The first steps for adapting an artificial intelligence emotion expression recognition software for emotional management in the educational context. British Journal of Educational Technology. 2023. https://doi.org/10.1111/bjet.13326.

26. Savchenko, A.V., Savchenko, L.V. and Makarov, I. Classifying emotions and engagement in online learning based on a single facial expression recognition neural network. IEEE Transactions on Affective Computing. 2022. https://doi.org/10.1109/TAFFC.2022.3188390.

27. Aly, M., Ghallab, A. and Fathi, I.S. Enhancing facial expression recognition system in an online learning context using efficient deep learning model. IEEE Access. 2023. https://doi.org/10.1109/ACCESS.2023.3325407.

28. Wang, S. Online Learning Behavior Analysis Based on Image Emotion Recognition. Traitement du Signal. 2021. https://doi.org/10.18280/ts.380333.

29. Marouf, M.A.A., Alchilibi, H., Sakr, D.K., Alkhayyat, A. and Pateriya, N. AI in Real-Time Student Performance Monitoring Using IoE. In Role of Internet of Everything (IOE), VLSI Architecture, and AI in Real-Time Systems. IGI Global Scientific Publishing. 2025. https://doi.org/10.4018/979-8-3693-7367-5.ch019.

30. Gupta, S., Kumar, P. and Tekchandani, R. A multimodal facial cues based engagement detection system in e-learning context using deep learning approach. Multimedia Tools and Applications. 2023. https://doi.org/10.1007/s11042-023-14392-3.

31. Alruwais, N.M. and Zakariah, M. Student recognition and activity monitoring in e-classes using deep learning in higher education. IEEE access. 2024. https://doi.org/10.1109/ACCESS.2024.3354981.

32. Zhu, X. and Chen, Z. Dual-modality spatiotemporal feature learning for spontaneous facial expression recognition in e-learning using hybrid deep neural network. The Visual Computer. 2020. https://doi.org/10.1007/s00371-019-01660-3.

FINANCING
The authors did not receive financing for the development of this research.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION
Conceptualization: Huynh Gia Bao, Nguyen Van Dai.
Data curation: Huynh Gia Bao, Nguyen Van Dai.
Formal analysis: Huynh Gia Bao, Nguyen Van Dai.
Drafting - original draft: Huynh Gia Bao, Nguyen Van Dai.
Writing - proofreading and editing: Huynh Gia Bao, Nguyen Van Dai.