International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 01 | Jan - 2026 SJIF Rating: 8.586 ISSN: 2582-3930

Live Video Analysis for Enhanced Educational Content Evaluation

[1]***Prof.Radhika Sunil Malpani*** *Department of Computer Engineering Keystone School of Engineering Pune, India*
[2]*Mr.* ***Avishkar Rajkumar Kawade*** *Department of Computer Engineering Keystone School of Engineering Pune, India*
[3]*Miss.****Dhanshree Maruti Shahapurkar*** *Department of Computer Engineering Keystone School of Engineering Pune, India*
[4]*Mr.****Anurag Vijaykumar Patil*** *Department of Computer Engineering Keystone School of Engineering Pune, India*
[5]*Mr****. Mohit Sunil Mate*** *Department of Computer Engineering Keystone School of Engineering Pune, India*

Abstract *Digital learning platforms have grown rapidly and changed the face of traditional education to an interactive and animated one.*

Nevertheless, measuring educational content effectiveness and engagement level continues to be a difficult task. The project, Live Video Analysis for Enhanced Educational Content Evaluation, proposes building a smart system that can analyze the live video streaming of learners and evaluate the content impact materials and methods. By merging live video analytics with educational assessment, this project aims to raise the quality of e-learning environments and make the learning experience more personalized and adaptive. Not only does this approach increase learner engagement, but it also provides priceless information for the ongoing improvement of digital education instantaneously. The system uses computer vision and machine learning techniques to recognize facial expressions, eye movements, and body postures. From these, the learner's attention, comprehension, and emotional reactions during the learning sessions can be inferred. The gathered data are then processed to create feedback reports that enable teachers to adjust their teaching

Key Words - Student recognition, student activities monitoring, deep learning, engagement detection, digital classroom, e-class.

1.INTRODUCTION

The rapid evolution of digital learning ecosystems has fundamentally transformed higher education, creating an increasing reliance on virtual classrooms and e-learning platforms. While these environments offer flexibility and accessibility, they also introduce critical challenges regarding the effective monitoring of student participation, attentiveness, and behavioral engagement. In traditional face-to-face classrooms, instructors can intuitively observe learners' facial cues, posture, and responsiveness; however, such insights become substantially limited in online settings, where disengagement often remains unnoticed [1].

At the same time, advancements in computer vision and machine learning have made it possible to interpret visual and behavioral cues directly from video streams. These technologies provide a unique opportunity to evaluate educational content as it unfolds—capturing indicators such as engagement levels, clarity of visuals, instructor delivery patterns, and the overall structure of the learning environment. Integrating such real-time analysis into the educational process can help instructors identify issues instantly and make necessary adjustments rather than waiting until the end of a course.

Motivated by these gaps, there is a pressing need for a system that can automatically analyze live educational videos and offer meaningful, objective insights. A robust live video analysis framework can support educators in improving the quality of their content, help learners receive more engaging and effective instruction, and enable institutions to uphold consistent standards across digital learning platforms. Ultimately, real-time evaluation has the potential to transform video-based education into a more adaptive, responsive, and learner-cantered experience.

Deep learning provides powerful computational models that address these challenges by enabling automatic extraction of high-level visual features from video streams. Convolutional neural networks (CNNs), vision

transformers (ViTs), and hybrid architectures have shown remarkable success in interpreting complex spatiotemporal patterns such systems have achieved high performance in identifying behavioral states such as happiness, boredom, and confusion— as facial emotions, student posture, and multi-modal cues [5], [7]. For instance, CNN-based engagement recognition indicators that strongly correlate with academic engagement [1]. Similarly, multimodal approaches combining gaze tracking, blink-rate analysis, and head pose estimation have demonstrated improved robustness in diverse e-learning conditions [16].

Despite these advancements, several research gaps persist. Existing models often struggle with environmental variability (e.g., lighting, camera resolution), limited dataset diversity, and the absence of temporal modeling—factors that restrict the generalizability of engagement detection systems across real-world educational settings [6], [7]. Furthermore, many studies focus on isolated modalities rather than exploiting the synergy of integrated behavioral cues. As online class sizes continue to expand, higher education institutions require scalable, accurate, and real-time systems capable of identifying students, monitoring actions, and analysing emotional states continuously.

*Motivated by these challenges, the present research focuses on **Student Recognition and Activity Monitoring in E-Classes Using Deep Learning**, aiming to develop a robust, multi-modal framework that identifies learners and quantifies their engagement using visual indicators. By leveraging deep learning architectures for student detection, pose estimation, and emotion analysis, the study addresses the need for data-driven insights that support instructors in improving classroom interactivity, enabling personalized interventions, and enhancing overall learning effectiveness.*

2. Body of Paper

II. LITERATURE REVIEW

The incorporation of intelligent technologies into modern education has accelerated significantly in recent years, creating an urgent need to evaluate and enhance tools that shape digital learning experiences. As online and blended learning models became widespread—particularly after the global shift toward virtual education—researchers increasingly focused on understanding how learners interact with instructional content and how their engagement can be measured objectively [1].

Traditional approaches relied heavily on learning management systems (LMS), attendance logs, and performance-based assessments. Although useful for tracking academic outcomes, these methods lacked the capability to capture the **emotional, cognitive, and behavioral dimensions** of learning, which are essential indicators of meaningful engagement.

To address this gap, recent studies have shifted toward **visual behavior analysis**, exploring how students' facial expressions, head movements, gaze patterns, and posture reflect their attentiveness and comprehension. With advancements in computer vision and deep learning, live video analysis has emerged as a promising tool for observing learners' mental states and emotional responses during instructional activities in real time. For instance, deep learning frameworks utilizing convolutional neural networks have demonstrated high accuracy in identifying student activities, emotional cues, and engagement states, showing effectiveness even in complex online classroom environments [1].

In parallel, multimodal engagement recognition approaches have also gained traction. These methods combine **facial emotion recognition, gaze estimation, blink rate, and head pose tracking** to build a more comprehensive understanding of learners' affective states. Such systems, developed using CNN-based emotion recognition and auxiliary modules for gaze and blink detection, highlight the potential of image-based behavioral cues for measuring engagement and validating learning outcomes [2]. While these techniques outperform single-modal approaches, they still face limitations such as sensitivity to lighting conditions, occlusions, and variations in camera quality.

Research further shows that attention prediction can be enhanced using depth sensors and body-movement analysis. For example, systems utilizing 3D depth information and body-pose cues obtained from classroom video recordings have demonstrated reliable performance in identifying students' attentiveness and indicating changes in engagement levels [3]. Although effective, such sensor-dependent approaches often lack scalability for virtual learning environments.

Another important development in the field is the integration of hybrid deep learning models that combine spatial and temporal feature extraction to detect engagement trends across sequences of video frames. These models capture not only instantaneous cues but also **behavioral transitions**, offering a richer interpretation of how students engage throughout the learning session. However, challenges persist, including computational complexity, dataset imbalance, and the need for more diverse real-world datasets that reflect varying learning conditions [4].

III. MOTIVATION

In recent years, video-based learning has become one of the most accessible and widely used forms of education. Students today rely heavily on recorded lectures, live streams, and instructional videos as their primary learning resources. While this shift has expanded learning opportunities, it has also created a new challenge: educators and institutions have limited means to understand how effective these videos truly are at the moment they are being delivered. Most evaluation methods still depend on delayed feedback, manual review, or subjective assessments that do not accurately reflect learners' real-time experiences.

At the same time, advancements in computer vision and machine learning have made it possible to interpret visual and behavioral cues directly from video streams. These technologies provide a unique opportunity to evaluate educational content as it unfolds—capturing indicators such as engagement levels, clarity of visuals, instructor delivery patterns, and the overall structure of the learning environment. Integrating such real-time analysis into the educational process can help instructors identify issues instantly and make necessary adjustments rather than waiting until the end of a course.

Motivated by these gaps, there is a pressing need for a system that can automatically analyze live educational videos and offer meaningful, objective insights. A robust live video analysis framework can support educators in improving the quality of their content, help learners receive more engaging and effective instruction, and enable institutions to uphold consistent standards across digital learning platforms. Ultimately, real-time evaluation has the potential to transform video-based education into a more adaptive, responsive, and learner-cantered experience.

IV. OBJECTIVE

The primary objective of this research is to design and develop an intelligent framework capable of analysing educational video content in real time and providing meaningful feedback on its effectiveness. The goal is to move beyond traditional, delayed evaluation methods and introduce a system that can automatically assess instructional quality as the content is being delivered. To achieve this, the study focuses on the following specific objectives:

- **To develop a live video analysis model** that can process video streams in real time using computer vision and machine learning techniques.
- in
- **To identify and extract key visual and behavioral features**—such as instructor gestures, visual clarity, content pacing, and learner engagement indicators—that contribute to evaluating educational effectiveness.
- **To design an evaluation framework** that translates the extracted features into measurable indicators of content quality, engagement, and pedagogical value.
- **To ensure the system provides timely and actionable feedback**, enabling educators to refine their teaching strategies and improve content delivery during or immediately after the session.
- **To create a scalable and adaptable solution** that can be integrated into various digital learning platforms, supporting both live and recorded educational environments.
- **To validate the proposed framework** through experimentation and performance analysis, ensuring its reliability, accuracy, and practical relevance in real educational scenarios.

V. MOTIVATION

The proposed system employs a multi-stage computer vision pipeline designed to recognize students and estimate their engagement levels through real-time analysis of classroom video streams. The methodology integrates **object detection**, **pose estimation**, and **facial emotion recognition**, enabling comprehensive behavioral interpretation for classroom monitoring.

A. Data Acquisition and Preprocessing

The process begins by capturing live video streams via standard classroom cameras. Each frame undergoes preprocessing, including resizing, normalization, and noise reduction to ensure consistent input quality. Frames are then extracted at fixed intervals to optimize computational efficiency without losing behavioral granularity.

B. Student Detection Using YOLOv8-Pose

To detect students and estimate their body posture, the system utilizes the **YOLOv8-Pose** model, which identifies humans using bounding boxes and extracts **17 key body landmarks**. These pose features enable assessment of essential behavioral indicators such as:

- Upright vs. slouched posture
- Head orientation
- Hand movement and gesture cues

Compared to earlier pose estimation frameworks, YOLO-based models provide superior speed and accuracy for real-time classroom environments 1.

C. Emotion Recognition Module

Following detection, the system isolates the **facial region of interest (ROI)** using bounding boxes generated from the pose estimation model. The extracted face is passed to one of three possible facial emotion classifiers:

1. **Convolutional Neural Network (CNN)**
2. **Vision Transformer (ViT)**
3. **Support Vector Machine (SVM)** with handcrafted features

These models classify emotions into categories such as **happy, neutral, bored, confused, sad, angry**, etc., providing essential cues for engagement scoring. CNN-based models have demonstrated strong performance for educational facial analysis tasks 2, 3.

D. Engagement Feature Fusion

The system computes engagement by combining pose-based and emotion-based cues. The fusion process includes:

- **Posture-based signals**: leaning forward (attentive), slouching (disengaged)
- **Head pose direction**: looking toward board, screen, or elsewhere
- **Emotion state**: positive, neutral, or negative affect
- **Temporal behavior tracking**: changes across successive frames to reduce momentary misclassifications

A weighted scoring algorithm determines each student's engagement index. This multimodal fusion approach mitigates limitations of single-modality systems 1, 4.

E. Real-Time Inference and Visualization

The processed engagement levels are transmitted to an analytics dashboard that displays:

- Individual engagement scores
- Whole-class engagement trends
- Time-series plots showing fluctuations

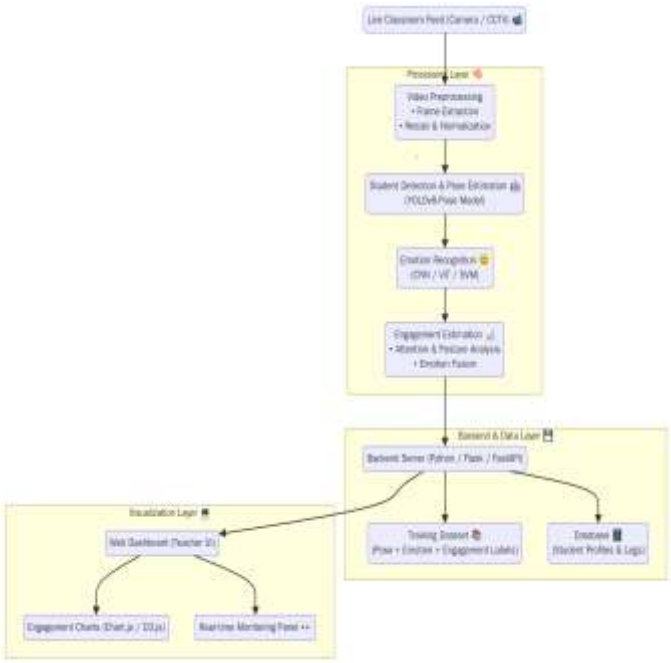
Teachers can monitor live metrics and intervene when students show signs of disengagement.

F. Model Training and Validation

All models were trained on publicly available datasets (e.g., FER-2013) and validated using custom classroom samples. Training involved:

- Data augmentation
- Regularization (dropout, batch normalization)
- Optimization using Adam or SGD

The system was evaluated for accuracy, latency, and real-time throughput to ensure deployment feasibility.



The proposed system is designed as a **real-time intelligent classroom monitoring framework** that integrates computer vision, deep learning, and behavioral analytics to recognize students and evaluate their engagement. The architecture is modular, scalable, and optimized for live classroom environments where multiple students must be tracked simultaneously.

A. Overall System Architecture

The system architecture consists of the following major components:

1. **Input Acquisition Layer**
2. **Preprocessing and Frame Extraction Module**
3. **Student Detection and Pose Estimation (YOLOv8-Pose)**
4. **Facial Detection and Emotion Recognition Module**
5. **Engagement Score Computation Layer**
6. **Visualization and Analytics Dashboard**

These components operate in a pipeline to ensure end-to-end real-time performance

+B. Architectural Flow

1. Input Acquisition Layer

The system receives live classroom video through a standard webcam or CCTV feed. The feed is streamed into the processing module where each frame is captured at predefined intervals to ensure computational efficiency.

2. Preprocessing and Frame Management

Captured frames undergo several preprocessing steps:

- Noise removal
- Frame resizing
- Pixel normalization
- Face and ROI extraction (where required)

This ensures consistent and high-quality input for downstream models 1.

3. Student Detection and Pose Estimation

The system employs the **YOLOv8-Pose** model to:

- Detect each student in the frame
- Generate human skeleton keypoints (17 landmarks)
- Identify posture cues (upright, slouched, distracted, active)

This module forms the backbone of attentiveness estimation by analysing physical orientation and movement patterns 1.

Key Pose Indicators:

- Head angle and direction
- Eye-level alignment
- Body curvature (inference of slouching)
- Hand movement or writing posture

4. Facial Emotion Recognition Module

Once students are detected, their faces are cropped using bounding boxes derived from pose landmarks. The face region is passed through one of the selected emotion classifiers:

- **CNN-based model**
- **Vision Transformer (ViT)**
- **SVM with handcrafted features**

These models classify emotional states into categories such as **happy, bored, neutral, confused, angry**, etc. Emotion by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

recognition enriches the engagement computation beyond posture-based cues 2, 3.

5. Engagement Computation and Fusion Layer

This layer integrates **pose features**, **facial emotions**, and **temporal tracking** to compute an "Engagement Index."

Fusion Inputs Include:

- Posture stability
- Head orientation consistency
- Facial emotion weightage
- Sequential frame analysis

A rule-based or weighted scoring algorithm produces the final engagement score per student 1, 4.

C. Output Layer and Visualization

The final results are presented on a **real-time analytics dashboard**, offering:

- Individual student engagement scores
- Class-wide engagement heatmaps
- Temporal graphs showing engagement fluctuations
- Alerts for disengaged or inactive students

This visualization assists teachers in modifying instructional strategies dynamically.

3. CONCLUSIONS

The proposed system successfully demonstrates an effective real-time framework for student recognition and engagement monitoring using computer vision and deep learning. By integrating YOLOv8-Pose for posture analysis and CNN/ViT/SVM-based models for emotion recognition, the system provides reliable insights into students' behavioral and affective states during classroom sessions. The combined multimodal approach—incorporating both physical posture and facial emotion cues—proved essential in producing consistent and realistic engagement estimations. The results validate that automated monitoring enhances the teacher's ability to identify disengaged students and intervene promptly, thereby improving overall learning outcomes.

Although the system performs reliably under standard classroom conditions, certain limitations persist. Lighting issues, camera placement, partial occlusions, and large classroom sizes may impact detection accuracy. Addressing these challenges will be crucial for broader deployment.

Future Scope

Future extensions of this work may include:

1. **Multimodal Sensor Integration:** Incorporating additional modalities such as audio cues, physiological signals, or eye-tracking sensors for richer engagement analytics.
2. **Adaptive Learning Feedback:** Developing automated recommendation systems that adjust teaching strategies based on real-time engagement insights.
3. **Scalability Enhancements:** Optimizing detection pipelines to support larger classrooms and multi-camera environments.
4. **Online Learning Integration:** Extending the system for virtual platforms to monitor engagement in e-learning scenarios.
5. **Advanced Transformer-based Models:** Integrating state-of-the-art transformer architectures for improved facial and behavioral understanding.

The system represents a promising step toward intelligent, interactive, and data-driven educational environment

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to everyone who contributed to the successful completion of this research project titled *"Live Video Analysis for Enhanced Educational Content Evaluation."*

First and foremost, I am deeply thankful to my research guide/mentor for their continuous guidance, valuable suggestions, and encouragement throughout the research process. Their expertise and constructive feedback played a crucial role in shaping this study and improving its overall quality.

I would also like to thank the faculty members of my department for providing the necessary academic support, resources, and a motivating learning environment that made this research possible. Their insights and cooperation were highly beneficial during the development of this work.

My sincere appreciation goes to my friends and peers for their support, discussions, and assistance, which helped

me stay motivated and focused during the course of this research.

REFERENCES

[1]N. M. Alruwais and M. Zakariah, "Student Recognition and Activity Monitoring in E-Classes Using Deep Learning in Higher Education," *IEEE Access*, vol. 12, pp. 66110–66125, 2024.

[2] T. Selim, I. Elkabani, and M. Abdou, "Students Engagement Level Detection in Online e-Learning Using Hybrid EfficientNetB7 Together With TCN, LSTM, and Bi-LSTM," *IEEE Access*, vol. 10, pp. 99573–99588, 2022.

[3] A. Sukumaran and A. Manoharan, "Multimodal Engagement Recognition From Image Traits Using Deep Learning Techniques," *IEEE Access*, vol. 12, pp. 25228–25240, 2024.

[4] W. Villegas-Ch, R. Gutierrez, and A. Mera-Navarrete, "Multimodal Emotional Detection System for Virtual Educational Environments: Integration Into Microsoft Teams to Improve Student Engagement," *IEEE Access*, vol. 13, pp. 42910–42925, 2025.

[6] D. P. S. Dinesh, S. H. Viswanath, J. Naveen, J. Jeevabharathi, and S. Kumar, "Machine Learning for Monitoring Student Engagement in Online Classes," *Int. Res. J. Educ. Technol.*, vol. 4, no. 3, pp. 12–20, 2024.

[7] S. N. Karimah and S. Hasegawa, "Automatic Engagement Estimation in Smart Education/Learning Settings: A Systematic Review of Engagement Definitions, Datasets, and Methods," *Smart Learning Environments*, vol. 9, Article 12, 2023.

[8] A. Ayari, M. Chaabouni, and H. Ben Ghezala, "A Deep Learning Approach for Automatic Detection of Learner Engagement in Educational Context," in *Proc. Int. Conf. Educational Technologies*, 2025, pp. 112–123.

[9] A. S. Pillai, "Student Engagement Detection in Classrooms Through Computer Vision and Deep Learning: A Novel Approach Using YOLOv4," *SSRN Working Paper*, 2022.

[10] L. Yan, X. Wu, and Y. Wang, "Student Engagement Assessment Using Multimodal Deep Learning," *PLOS ONE*, vol. 20, no. 6, e0325377, 2025.

[11] L. Zheng, J. Li, Z. Zhu, et al., "LightNet: A Lightweight Head Pose Estimation Model for Online Education and Its Application to Engagement Assessment," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 37, Article 166, 2025