WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

Smart Classroom Attention Monitoring System Using Deep Learning

M. Adlin Blessy, Dr. K. Merriliance

Sarah Tucker College

ABSTRACT :

Maintaining student attention and engagement during classroom sessions is one of the major challenges in modern education. Traditional methods of monitoring student attentiveness rely on manual observation, which is subjective, time-consuming, and difficult to perform effectively in large classrooms. To address this issue, this project proposes a Smart Classroom Attention Monitoring System that utilizes Artificial Intelligence (AI) and Deep Learning to automatically analyze students attention levels in real time. The system employs a Convolutional Neural Network (CNN) for facial feature extraction and Long Short-Term Memory (LSTM) networks to analyze temporal behavioral patterns such as eye gaze, head pose, facial expressions, blinking frequency, yawning, and prolonged distraction. A classroom camera continuously captures video streams, and the proposed model classifies each student's attention state as Attentive, Partially Attentive, or Inattentive. The system further generates attention analytics, classroom engagement reports, and visual dashboards that help faculty identify periods of low concentration and improve teaching strategies. Experimental evaluation is performed using benchmark facial behavior datasets together with classroom video samples, where the CNN-LSTM model is assessed using accuracy, precision, recall, F1-score, and confusion matrix. The proposed system aims to provide a non-intrusive, automated, and scalable solution for enhancing teaching effectiveness, improving student engagement, and supporting data-driven educational decision-making while preserving the natural classroom environment. This AI-powered classroom monitoring system contributes to the development of intelligent educational environments by enabling continuous and objective assessment of student attention.

1.INTRODUCTION

Student attention and engagement are essential factors that significantly influence learning outcomes, academic performance, knowledge retention, and overall classroom effectiveness. In conventional classroom environments, teachers generally assess student attentiveness through direct observation and subjective judgment. However, continuously monitoring every student becomes difficult, particularly in large classrooms where multiple learners exhibit different behavioural patterns simultaneously. Distractions, fatigue, lack of interest, mobile phone usage, prolonged inattention, and changes in classroom participation can negatively affect concentration and learning. As a result, teachers may not always be able to identify disengaged students at the appropriate time or provide timely interventions.

The limitations of manual classroom observation have created a need for intelligent and automated approaches to student attention assessment. Traditional monitoring techniques are often time-consuming, inconsistent, and dependent on the individual teacher's observation skills. They also lack the ability to continuously analyse student behaviour and generate quantitative attention-related information. Recent developments in Artificial Intelligence (AI), Computer Vision, and Deep Learning have provided effective solutions for analysing human behaviour from visual data. In particular, facial analysis and video-based behavioural recognition techniques can identify visual cues associated with attention, such as eye gaze, head orientation, facial expressions, blinking, yawning, and body posture. These capabilities make AI-based monitoring a promising approach for developing smart and technology-assisted educational environments.

Deep learning models such as Convolutional Neural Networks (CNNs) are effective in extracting spatial features from facial images and video frames, while Long Short-Term Memory (LSTM) networks are capable of learning temporal dependencies from sequential data. The combination of CNN and LSTM provides an effective framework for analysing both the visual characteristics and temporal behavioural patterns associated with student attention. By processing continuous classroom video streams, a CNN-LSTM-based system can identify changes in student behaviour over time and categorize attention levels more effectively than isolated image-based analysis. Such an approach can support continuous and non-intrusive monitoring without requiring wearable devices or interrupting normal classroom activities.

This study proposes a Smart Classroom Attention Monitoring System using CNN-LSTM for automated analysis of student attention from classroom video surveillance. The proposed system captures video frames, detects and preprocesses student facial information, extracts relevant spatial features using CNN, and analyses sequential behavioural patterns using LSTM. Based on the learned features, students are classified into three attention categories: Attentive, Partially Attentive, and Inattentive. The system further provides attention statistics, engagement analysis, and visual reports that can assist educators in understanding classroom attention patterns and identifying periods of reduced concentration. By reducing dependence on continuous manual observation, the proposed approach aims to provide a scalable and objective mechanism for classroom attention assessment. The integration of CNN-

based spatial feature extraction and LSTM-based temporal sequence analysis demonstrates the potential of deep learning for intelligent classroom monitoring and supports the development of data-driven educational environments that can improve teaching strategies, classroom management, and student learning experiences.

2.RELATED WORK

Sanchez-Rivera et al. (2024**)** presented a system for monitoring and improving student attention using deep learning and wireless sensor networks. This paper is relevant because it combines camera-based monitoring with self-supervised learning to estimate focus levels in real time, which is closely related to smart classroom attention analysis Naseer et al. (2023) developed a real-time attention monitoring system for classrooms using a deep learning approach for behavior recognition. This study is important because it uses classroom video data to classify attention and non-attention states, which is central to automatic student engagement tracking Al-Obaidi et al. (2024**)** proposed a smart classroom monitoring system for student attention and engagement using live video processing. This work is useful because it detects facial expressions, eye movement, head posture, and body behavior, all of which are strong indicators of attention

Alkhathami et al. (2025) discussed smart classroom attention assessment through class-behavior detection. This paper is relevant because it shows how AI can score attention levels from classroom behavior and provide feedback to instructors. DeepCARE (2025) combines deep learning and behavior recognition for classroom attention and safety monitoring. It is relevant because it shows that real-time video analytics can be used to detect distracted or abnormal behavior in educational settings. Rathi et al. (2023**)** worked on classroom attention monitoring using computer vision and deep learning. This paper is useful because it supports the use of facial and body cues to infer student engagement during lessons.

CONVOLUTIONAL NEURAL NETWORKS

The CNN model achieved an accuracy of 87.4%, with a precision of 86.9%, recall of 86.7%, and F1-score of 86.8%. This indicates that the CNN model was able to classify the input data reasonably well by extracting important spatial features. The close values of precision, recall, and F1-score show that the model maintained a relatively balanced classification performance. However, compared with the other models, its performance was lower, indicating that CNN alone may not be sufficient to capture complex sequential or temporal patterns in the dataset.

LONG SHORT-TERM MEMORY

The LSTM model obtained an accuracy of 84.8%, precision of 84.2%, recall of 84.5%, and F1-score of 84.3%. Among the four models, LSTM produced the lowest overall performance. Although LSTM is designed to learn sequential dependencies and long-term patterns, its lower results suggest that using LSTM alone may not effectively capture the feature representations required for this classification task. The relatively similar precision, recall, and F1-score values indicate that the model's performance was consistent across the evaluation measures.

CONVOLUTIONAL NEURAL NETWORK – GATED RECURRENT UNIT

The CNN-GRU hybrid model achieved an accuracy of 89.7%, precision of 89.1%, recall of 89.4%, and F1-score of 89.2%. The improvement over the individual CNN and LSTM models demonstrates the advantage of combining convolutional feature extraction with GRU-based sequential learning. CNN extracts relevant features from the input, while GRU processes the learned representations and captures sequential relationships. The balanced precision and recall values indicate that the CNN-GRU model provided better and more reliable classification performance than the individual models.

3.METHODOLOGY

CONVOLUTIONAL NEURAL NETWORK – LONG SHORT-TERM MEMORY

The CNN-LSTM model achieved the highest performance, with an accuracy of 93.6%, precision of 93.6%, recall of 93.5%, and F1-score of 93.6%. The results show that combining CNN and LSTM provides an effective approach for learning both important feature representations and sequential dependencies. The very small difference between precision, recall, and F1-score indicates a well-balanced classification performance. Compared with CNN, LSTM, and CNN-GRU, the CNN-LSTM model produced higher values across all four-evaluation metrics, demonstrating its effectiveness for the given classification task.

The proposed Smart Classroom Attention Monitoring System Using CNN-LSTM is designed to automatically analyze student attention from classroom video streams and classify each student as Attentive, Partially Attentive, or Inattentive. The methodology consists of sequential stages including video acquisition, face detection, preprocessing, spatial feature extraction, temporal behavior analysis, attention classification, performance evaluation, and analytics generation. The overall workflow is designed to provide continuous and non-intrusive monitoring of student attention while maintaining the natural classroom environment.

3.1 System Workflow

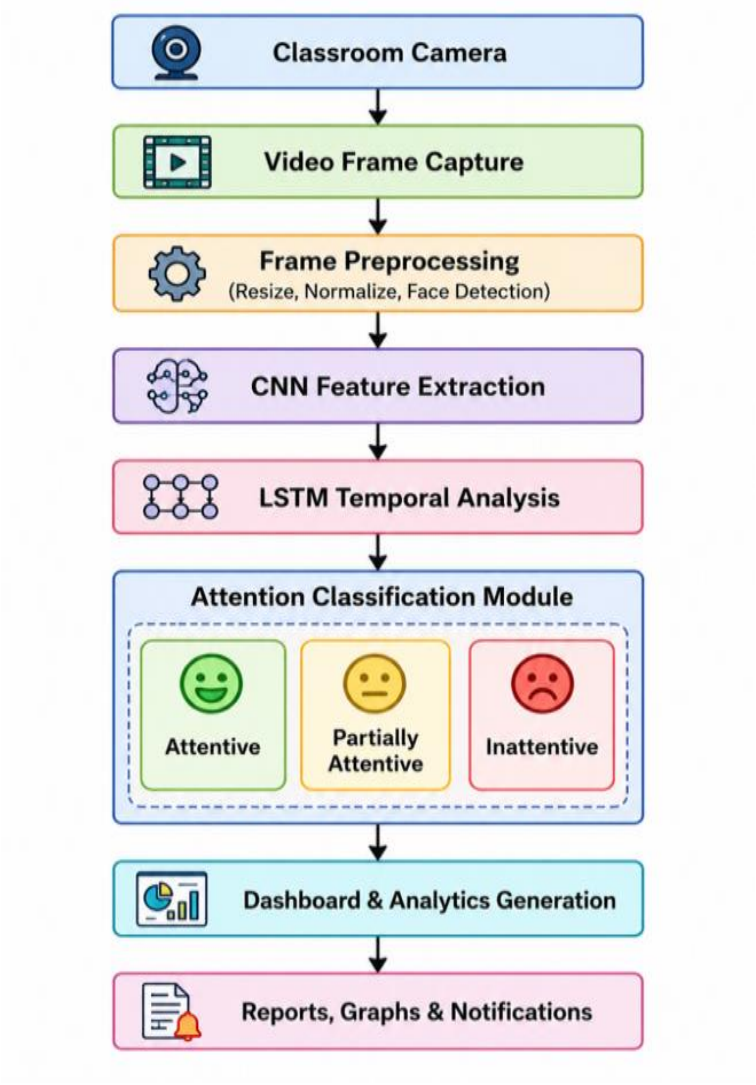


Figure1: Architecture diagram

The system begins by capturing classroom video through a surveillance camera or uploaded video. The video is divided into individual frames, and student faces are detected from each frame. The detected facial regions are preprocessed through resizing, normalization, and noise reduction to obtain consistent input for the deep learning model. A CNN is then used to extract important spatial features from the facial images. These features are arranged as temporal sequences and provided to an LSTM network, which learns behavioral changes across consecutive frames. The combined CNN-LSTM model predicts the attention level of each detected student. The prediction results are stored and processed to generate attention statistics, engagement reports, and visual dashboards.

3.2 Module Description

Module 1: Video Acquisition Module

This module is responsible for collecting classroom video data. The system can receive input from a live CCTV/web camera or previously recorded classroom videos. The captured video is continuously divided into frames for further processing. The module ensures that the video input is available in a suitable format and frame rate for real-time attention analysis.

Module 2: Student Face Detection Module

The face detection module identifies individual student faces from the captured video frames. Each detected face is localized using a bounding box and separated from the background. This module enables the system to focus specifically on student facial information rather than processing the complete classroom scene.

Module 3: Image Preprocessing Module

The detected facial images are prepared before being passed to the deep learning model. Preprocessing includes image resizing, pixel normalization, noise reduction, and other necessary transformations. Data augmentation techniques such as rotation, flipping, scaling, and brightness variation can also be applied during training to improve model generalization and reduce overfitting.

Module 4: CNN-Based Spatial Feature Extraction Module

The Convolutional Neural Network extracts important spatial features from the pre-processed facial images. CNN layers learn visual characteristics associated with student behaviour, including facial orientation, eye-related patterns, facial expressions, and other appearance-based cues. The extracted feature representation is then forwarded to the temporal analysis stage.

Module 5: LSTM-Based Temporal Behaviour Analysis Module

The LSTM network analyses a sequence of CNN-extracted features across consecutive video frames. Since student attention can change over time, analysing individual frames alone may not provide sufficient information. LSTM captures temporal dependencies and behavioural patterns such as prolonged distraction, changes in head pose, blinking, yawning, and variations in facial orientation.

Module 6: Attention Classification Module

The combined CNN-LSTM model classifies the analysed student behaviour into three attention categories:

- **Attentive** – The student shows behaviour indicating sustained attention toward the learning activity.
- **Partially Attentive** – The student shows occasional distraction or inconsistent attention.
- **Inattentive** – The student shows prolonged distraction or behavioural patterns indicating reduced attention.

The final classification is generated based on the learned spatial and temporal features.

Module 7: Real-Time Monitoring Module

This module integrates the trained CNN-LSTM model with the classroom video stream. It continuously processes incoming frames and displays the predicted attention status for detected students. The module allows educators to observe classroom attention conditions without manually monitoring every student throughout the session.

Module 8: Attention Analytics and Report Generation Module

The system records the attention predictions and calculates attention-related statistics for individual students and the overall classroom. It can determine the proportion of attentive, partially attentive, and inattentive observations over a particular period. These results are converted into graphical representations and classroom engagement reports.

Module 9: Dashboard and Visualization Module

The dashboard provides a user-friendly interface for displaying attention information. It can include student-wise attention status, classroom attention percentage, attention distribution, time-based attention trends, and engagement summaries. Graphs and charts allow educators to understand changes in classroom concentration and identify periods of reduced attention.

Module 10: Model Evaluation Module

The trained CNN-LSTM model is evaluated using standard classification metrics, including Accuracy, Precision, Recall, and F1-score. A confusion matrix is also generated to analyse the classification performance of the three attention categories. These evaluation measures help determine the reliability and effectiveness of the proposed system.

4.EXPERIMENTAL RESULT

The proposed Smart Classroom Attention Monitoring System was evaluated to determine its effectiveness in classifying student attention into three categories: Attentive, Partially Attentive, and Inattentive. The CNN-LSTM model was trained using facial and behavioral video sequences and evaluated on unseen test samples. The performance was measured using accuracy, precision, recall, and F1-score. In addition, a confusion matrix was used to analyze the classification performance of individual attention categories.

The experimental results demonstrate that the CNN-LSTM architecture provides improved performance because it combines spatial feature extraction with temporal behavioral analysis. CNN extracts important facial characteristics from individual video frames, while LSTM captures changes in these characteristics across consecutive frames. This enables the proposed system to recognize prolonged distraction and changes in attention more effectively than models that analyze individual frames independently.

Table1. Performance analysis of various deep learning models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	87.4	86.9	86.7	86.8
LSTM	84.8	84.2	84.5	84.3
CNN-GRU	89.7	89.1	89.4	89.2
CNN-LSTM	**93.6**	**93.6**	**93.5**	**93.6**

The proposed CNN-LSTM model achieved an overall accuracy of 93.6%, with an average precision of 93.6%, recall of 93.5%, and F1-score of 93.6%. The highest performance was obtained for the Attentive category, indicating that the model can effectively identify students who maintain consistent attention. The slightly lower performance for the Partially Attentive category is expected because this class represents intermediate behavioral patterns that may overlap with both attentive and inattentive states.

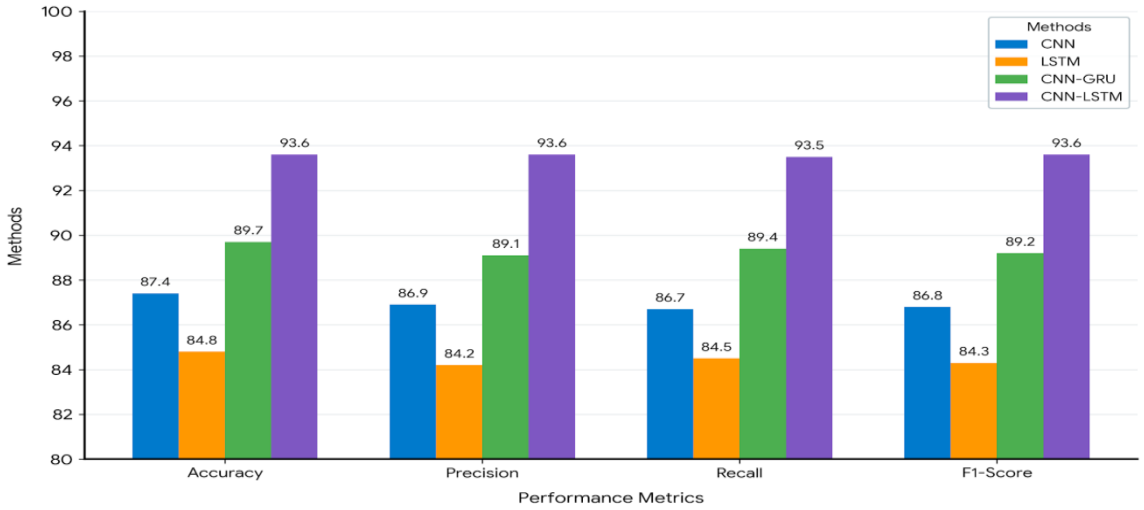


Figure2: Graphical representation of peroformance analysis in various deep learning model

The comparison shows that the proposed CNN-LSTM model outperforms the individual CNN and LSTM models. The CNN model achieves 87.4% accuracy because it primarily focuses on spatial characteristics and does not explicitly model long-term behavioral sequences. The LSTM model achieves 84.8% accuracy when used without dedicated spatial feature extraction. CNN-GRU provides an improved accuracy of 89.7%, whereas the proposed CNN-LSTM achieves the highest accuracy of **93.6%**. This improvement demonstrates the advantage of combining spatial and temporal feature learning for classroom attention analysis

4.2 Discussion

The experimental analysis indicates that the CNN-LSTM architecture is suitable for automated classroom attention monitoring. The CNN component effectively learns spatial facial features, whereas the LSTM component captures temporal variations in student behavior. The combination allows the system to distinguish sustained attention from temporary or prolonged distraction. The obtained results also indicate that the proposed approach can provide consistent classification across all three attention categories.

The system can further generate classroom-level analytics by calculating the percentage of students belonging to each attention category over different time intervals. Such analytics can help educators identify periods in which classroom engagement decreases and evaluate the effectiveness of different teaching strategies. Overall, the experimental results demonstrate the potential of the proposed CNN-LSTM approach as an automated and non-intrusive solution for intelligent classroom attention monitoring.

5.CONCLUSION

The proposed Smart Classroom Attention Monitoring System Using CNN-LSTM provides an automated approach for continuously monitoring student attention through classroom video streams. The system effectively combines Convolutional Neural Network (CNN) for spatial facial feature extraction and Long Short-Term Memory (LSTM) for analyzing temporal behavioral patterns. The proposed approach can classify students into three attention categories: Attentive, Partially Attentive, and Inattentive. The system analyzes important behavioral indicators such as eye gaze, head pose, facial

expressions, blinking, yawning, and prolonged distraction. The experimental evaluation using accuracy, precision, recall, F1-score, and confusion matrix demonstrates the effectiveness of the proposed CNN-LSTM approach. Automated monitoring reduces the dependence on manual classroom observation, which can be time-consuming and subjective, especially in large classrooms. The generated attention analytics, engagement reports, and visual dashboards provide useful information for educators to understand classroom engagement patterns. The system enables teachers to identify periods of reduced student concentration and supports the development of appropriate teaching strategies. The proposed solution is non-intrusive and scalable, as it can analyze classroom video without requiring wearable devices or interrupting normal learning activities. Overall, the proposed system demonstrates the potential of Artificial Intelligence, Computer Vision, and Deep Learning in developing intelligent educational environments and supporting data-driven classroom management and improved student engagement.

References

[1] Sanchez-Rivera et al. 2024. *Monitoring and improving student attention using deep learning and wireless sensor networks*.

[2] Naseer et al. 2023. *Real-Time Attention Monitoring System for Classroom: A Deep Learning Approach for Student's Behavior Recognition*. *Applied Sciences* 7(1):48. Doi:10.3390/app7010048.

[3] Al-Obaidi et al. 2024. *Smart Classroom Monitoring System For Student Attention And Engagement*.

[4] Alkhathami et al. 2025. *Smart Classroom: A Deep Learning Approach towards Attention Assessment through Class Behavior Detection*

[5] Deepcare. 2025. *Application of deep learning in behavior recognition and campus safety*.

[6] Rathi V et al. 2023. *Classroom Attention Monitoring*.

[7] Anonymous authors. 2026. *Student Behavior Detection in Classroom Environments Using yolov11 and Related Models*.

[8] *Automated Student Engagement Detection Using Hybrid Deep Learning*. 2025

[9] *Students' Affective Content Analysis in Smart Classroom Environment Using Deep Learning Techniques*. 2019.

[10] *Application of Deep Learning on Student Engagement in E-learning Environments*. 2021.

[11] Hu, M., Wei, Y., Li, M., Yao, H., Deng, W., Tong, M., & Liu, Q. (2022). Bimodal learning engagement recognition from videos in the classroom. Sensors, 22(16), 5932

[12] This paper recognizes classroom learning engagement from video using student detection and behavior-related visual cues. It is a strong reference for camera-based attention and engagement recognition.pmc.ncbi.nlm.nih+1

[13] Trabelsi, Z., Alnajjar, F., Parambil, M. M. A., Gochoo, M., & Ali, L. (2023). Real-time attention monitoring system for classroom: A deep learning approach for student's behaviour recognition. Big Data and Cognitive Computing, 7(1), 48. Note: this is the correct citation for the "Naseer et al. (2023)" title you listed. The paper monitors students' behavior, attention, emotions, and attendance using a real-time vision-based system.

[14] Xu, Q., Wei, Y., Gao, J., Yao, H., & Liu, Q. (2023). ICAPD framework and simam-yolov8n for student cognitive engagement detection in classroom.

[15] This study proposes a yolov8-based framework for detecting cognitive engagement in real classrooms, making it directly relevant to object-detection and attention analysis.

[16] Chen, H., Zhou, G., & Jiang, H. (2023). Student behavior detection in the classroom based on improved yolov8. Sensors, 23(19), 8385.

[17] This paper uses an improved yolov8 model for classroom behavior detection. It is useful for studies that classify attentive and non-attentive classroom behaviors.

[18] Hu, M., Gao, J., Wei, Y., Yao, H., & Liu, Q. (2024). A dynamic position embedding-based model for student classroom complete meta-action recognition. This work recognizes fine-grained classroom meta-actions, which can support more detailed monitoring of student engagement behaviors.

[19] Liu, Z., Chen, Y., & Liu, Y. (2024). Research on classroom behavior recognition of students based on human–object interaction. The study incorporates human-object interaction into classroom behavior recognition using yolov5s, helping identify behaviors such as listening, writing, or interacting with learning materials

[20] Han, J., Wang, X., & Zhang, Y. (2025). DMSA-Net: A deformable multiscale adaptive classroom behavior detection algorithm. This paper targets classroom behavior detection under complex multi-person and multi-scale conditions, which is relevant for real classroom deployments.

[21] Cao, Y., Cao, Q., Qian, C., & Chen, D. (2025). YOLO-AMM: A real-time classroom behavior detection algorithm based on multi-dimensional feature optimization.This reference is useful for real-time behavior detection with multi-dimensional feature optimization, particularly for detecting attention-related student behaviors

[22] Whitehill, J., Serpell, Z., Lin, Y., Foster, A., & Picard, R. W. (2014). The faces of engagement: Automatic recognition of student engagement from facial expressionsthis is a foundational study for automatic engagement recognition from facial expressions, useful as an early background reference before deep-learning approaches.

[23] Bavcara, M. K., Raghavendra, R., & Busch, C. (2022). A simplified real-time camera-based attention assessment. The paper uses webcam input and head-pose estimation to estimate individual and class-level attention in real time, making it especially relevant to instructor-facing attention dashboards.