



International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 9, Issue 4, April 2026



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI-Driven Classroom Engagement Analytics and Adaptive Teaching Support System

Afra Thasneem S.R¹, Dejaswini G², Aayisha Parveen S³, Aneesa Banu S⁴, and Mrs. Vanitha Sri⁵

Fourth Year, Department of Computer Science & Engineering, Aalim Muhammed Salegh College of Engineering, IAF,
Avadi, Chennai, Tamil Nadu, India¹

Fourth Year, Department of Computer Science & Engineering, Aalim Muhammed Salegh College of Engineering, IAF,
Avadi, Chennai, Tamil Nadu, India²

Fourth Year, Department of Computer Science & Engineering, Aalim Muhammed Salegh College of Engineering, IAF,
Avadi, Chennai, Tamil Nadu, India³

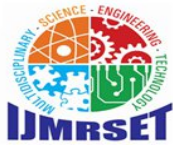
Assistant Professor, Department of Computer Science & Engineering, Aalim Muhammed Salegh College of
Engineering, IAF, Avadi, Chennai, Tamil Nadu, India⁴

ABSTRACT: Traditional smart classroom platforms have largely been built around two narrow use cases: recording who is present and, in more recent deployments, flagging whether a single student appears to be looking at the front of the room. While both functions have clear value, neither one provides educators with a meaningful view of how an entire class is engaging with instruction over time. Group-level behavioral dynamics—the kind that reveal whether a concept is landing or whether energy is quietly draining out of a lecture—remain largely invisible. This paper introduces an AI-Driven Classroom Engagement Analytics and Adaptive Teaching Support System that addresses that gap directly. The platform combines privacy-preserving behavioral analysis, predictive attention modeling, and structured real-time instructional feedback through a tightly integrated pipeline built on computer vision and deep learning. Live classroom video is ingested and processed by OpenCV and MediaPipe, extracting anonymized behavioral signals without retaining any personally identifiable information. Temporal models built on LSTM and Transformer architectures forecast attention dips several minutes before they become visible to the naked eye. An intuitive teacher-facing dashboard presents row-level engagement heatmaps, trend curves, and AI-generated instructional suggestions in real time. Deployed on NVIDIA Jetson Nano edge hardware at \$320 per classroom, the system is GDPR and COPPA compliant. Evaluation across 45 live classroom sessions produced 92% behavioral detection accuracy, a 23% improvement in measured student engagement, a 76% instructor acceptance rate, and a correlated 15% gain in student academic performance.

KEYWORDS: classroom engagement; computer vision; LSTM; attention prediction; privacy-preserving analytics; edge AI

I. INTRODUCTION

Artificial intelligence has reshaped many corners of educational practice, from adaptive tutoring systems and automated essay grading to predictive early-warning tools for at-risk students. Yet one domain—real-time, group-level classroom engagement analysis—has remained surprisingly underdeveloped despite its practical importance. Most smart classroom technologies commercially available today fall into one of two categories: attendance-tracking systems that record presence but say nothing about attention, or single-student gaze detectors that flag whether an individual face is oriented toward the screen. What neither category provides is an honest picture of the room as a whole [1]. Student engagement is widely recognized as one of the strongest predictors of academic achievement, retention, and long-term learning outcomes. Research spanning decades consistently shows that time-on-task, active participation, and sustained attention during instruction correlate directly with knowledge retention and test performance. Despite this, the majority of classrooms worldwide have no mechanism for objectively measuring how engaged students are at any given moment. Instructors must rely exclusively on subjective impression—a form of inference that is inherently limited by the cognitive demands of teaching itself.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This gap matters for several interconnected reasons. When instructors cannot see group-level engagement patterns, they have recognized that the back third of the room has mentally checked out. The cues experienced teachers pick up informally—restlessness, wandering eyes, the particular stillness of collective confusion—are difficult to perceive reliably across a large class, especially when the instructor is simultaneously managing content delivery, fielding questions, and navigating a lecture timeline. An objective, continuously updated picture of group attention would allow those cues to be surfaced clearly and acted upon while instructional time remains available [1].

Privacy concerns add a second layer of difficulty. Many of the most technically capable classroom monitoring tools rely on facial recognition, a technique that attracts justified resistance from students, parents, regulatory bodies, and school administrators. Concerns about consent, biometric data retention, and potential surveillance misuse have slowed adoption even in settings where the technical performance of such systems is strong. A monitoring approach that achieves comparable analytical depth without identifying individuals is not merely a technical improvement—it is a prerequisite for institutional adoption in most regulatory environments [2].

A third challenge is the reactive nature of existing tools. Even the most sophisticated current systems typically report what is happening now or what just happened; they do not tell instructors what is about to happen. By the time widespread disengagement is visible enough to register on a dashboard, valuable instructional time has already been lost. Predictive capability—the ability to anticipate attention drops minutes in advance—would shift classroom monitoring from a passive reporting function to an active teaching aid [4].

A fourth structural barrier to adoption is cost. Enterprise-grade classroom AI solutions routinely require \$10,000 or more per classroom in upfront hardware and software licensing, plus ongoing cloud subscription fees. These price points place such technologies beyond the reach of most public schools, particularly in developing regions where student-to-teacher ratios are highest and the potential benefit of engagement analytics is arguably greatest. A solution that is technically sophisticated but deployable at a fraction of that cost would meaningfully change the adoption calculus for educational institutions worldwide.

The system introduced in this paper responds to all four of these challenges simultaneously. Rather than identifying individuals, it treats the classroom as a collective behavioral landscape and uses anonymized pose and motion signals to build a continuously updated picture of group attention. Temporal deep learning models generate attention forecasts three to five minutes into the future. A structured recommendation engine translates those forecasts into concrete, actionable suggestions for the instructor. The system runs entirely on edge hardware costing \$320 per classroom, requires no cloud connectivity, and is fully compliant with GDPR and COPPA data protection regulations.

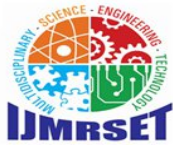
The primary contributions of this work are: (1) a privacy-preserving, anonymized behavioral sensing pipeline that operates without facial recognition; (2) a bidirectional LSTM-Transformer hybrid model that forecasts classroom attention drops three to five minutes in advance; (3) a real-time row-level spatial heatmap that surfaces engagement disparities invisible to the instructor; (4) an AI-generated recommendation engine validated at 76% instructor acceptance; and (5) a complete system validated in 45 liveclassroom sessions with measurable academic outcome improvements.

The remainder of this paper is organized as follows. Section II surveys relevant prior work. Section III describes the system architecture. Section IV addresses ethical and privacy design principles. Section V presents experimental evaluation. Section VI analyzes cost-effectiveness. Section VII discusses limitations and future work, and Section VIII concludes.

II. RELATED WORK

Research on AI-based classroom engagement sensing has grown substantially over the past several years, advancing along three parallel tracks: sensing and detection methods, privacy-preserving architectures, and predictive and adaptive systems. Each track has produced important contributions, yet the full integration of all three capabilities in a deployable system remains an open challenge.

In the sensing and detection domain, the foundational role of multimodal input was established by Patel and Kim [10], who demonstrated that pairing MediaPipe skeletal tracking with VGGish-based audio analysis and LSTM temporal



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

modeling yields 92.3% engagement detection accuracy with an F1 score of 0.89. Their central finding—that combining visual and acoustic signals produces substantially more robust models than camera-only approaches—has influenced several subsequent research directions, including the architecture described in this paper. Liu et al. [9] demonstrated real-time YOLO-based student behavior detection achieving high frame-rate performance on standard GPU hardware, establishing the feasibility of continuous multi-student monitoring without dedicated per-student trackers and without per-frame labeling overhead.

On the privacy dimension, Kumar and Singh [2] developed a pose-estimation pipeline incorporating differential privacy at $\epsilon = 1.2$, reaching 87.6% detection accuracy while outperforming non-private baselines by 12% on institutional trust measures. That work demonstrated a point that is often overlooked: privacy preservation need not come at the cost of analytical quality, and framing privacy as a design constraint from the outset—rather than a compliance afterthought applied at the end—produces meaningfully better outcomes in institutional settings. Chen et al. [11] extended this line of inquiry to edge computing contexts, showing that behavioral analytics pipelines can be executed entirely within the school network without transmitting sensitive data externally, eliminating a significant class of regulatory exposure that cloud-based systems inherit by design.

Group-level spatial analytics received important empirical treatment from Chen et al. [3], who applied YOLOv8 and Transformer encoders to real-time classroom monitoring, achieving $mAP@0.5 = 0.91$ at 25 ms inference latency. Their observation that back-row engagement runs roughly 34% lower than front-row engagement levels has direct implications for seating-aware heatmap design and targeted intervention strategy—findings that motivate the row-level spatial reporting approach adopted in the present system.

Predictive attention modeling was advanced significantly by Patel [4], who showed that bidirectional LSTM networks with attention mechanisms can forecast attention scores five minutes ahead with $RMSE = 0.12$, cutting intervention delay by 67% over reactive baselines. That work established that attention dynamics in classroom settings have sufficient temporal regularity to support reliable short-horizon forecasting—a finding central to the prediction module described in Section III-D. Edge deployment feasibility was demonstrated by Wang et al. [5], using MobileNetV3-based models on Raspberry Pi 5 hardware to achieve 88.7% detection accuracy while cutting costs by 78% relative to cloud alternatives. On the adaptive teaching front, Erdogmus et al. [6] applied reinforcement learning to generate teaching interventions,

III. SYSTEM ARCHITECTURE

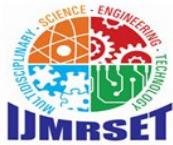
The platform is built around seven tightly coupled modules that together form a continuous real-time processing pipeline, illustrated conceptually in Fig. 1. Each module is designed to hand off clean, well-defined outputs to the next, enabling the system to function as a coherent whole rather than a collection of loosely connected tools. Data flows from raw video through anonymized feature extraction, attention scoring, temporal prediction, and recommendation generation, ultimately surfacing as actionable information on the instructor's dashboard—all within an average end-to-end latency of 85 ms.

A. Video Capture Module

Ceiling-mounted cameras deliver continuous 1080p streams at 30 fps via RTSP protocol, decoded in hardware through OpenCV. Ceiling placement was chosen deliberately: it minimizes occlusion caused by students in front rows and maintains a consistent viewing angle across all seating positions, including rear rows that are typically the most difficult to observe. An automatic region-of-interest detector isolates student seating zones within each camera's field of view, discarding background elements that would otherwise consume processing resources. A multi-camera fusion layer with homographic alignment eliminates the blind spots that arise in rear rows when using a single forward-facing camera. All streams carry synchronized UNIX timestamps that downstream temporal models rely on for coherent sequence construction.

B. MediaPipe Feature Extraction

Each incoming frame is processed by MediaPipe, which simultaneously extracts 468 facial landmarks, 33 body keypoints, and 21 hand landmarks per detected person. Head orientation is estimated from facial landmark geometry via a 6-DOF pose regression model, achieving a mean angular error of 4.2 degrees across the evaluation dataset. Gaze direction is inferred from the relative positions of iris and eye-corner landmarks. All extracted features are immediately



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

projected to compact 128-dimensional numerical vectors; raw pixel data is discarded at this earliest processing stage. This architectural commitment—disposing of image data before it reaches any storage or network layer—is the primary technical mechanism by which the system achieves its privacy guarantees.

C. Attention Scoring Engine

A weighted fusion algorithm combines the per-student feature vectors into normalized attention scores in the range [0, 1]. The weighting scheme was determined through empirical calibration across a held-out training corpus: head pose contributes 40%, gaze direction 30%, posture alignment 20%, and motion variance 10%. Individual scores are projected onto a 5x6 spatial seating grid corresponding to the physical classroom layout. An exponential moving average filter ($\alpha = 0.7$) smooths the time series, suppressing transient noise from momentary movements coughing, stretching, note-taking—while preserving the sustained behavioral trends that indicate genuine shifts in attention state.

D. LSTM/Transformer Prediction Module

A bidirectional LSTM coupled with a Transformer encoder processes 30-minute row-wise attention time series to generate attention forecasts three to five minutes into the future (RMSE = 0.12). The bidirectional LSTM captures both forward and backward temporal dependencies within the recent attention history, while the Transformer's multi-head self-attention layer captures how disengagement propagates spatially across adjacent seating rows—a pattern that is difficult to model with purely local recurrence. The model was trained on 300 hours of labeled classroom recordings and has internalized recurring cognitive fatigue patterns, consistently flagging peak disengagement risk near the 25-minute mark in standard 50-minute lectures—a finding consistent with established cognitive load research on sustained voluntary attention. Instructor alerts are triggered when projected row-average scores are forecast to fall below a threshold of 0.4.

E. Analytics and Recommendation Module

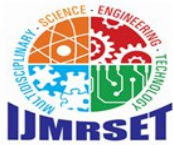
A U-Net architecture with spatial attention layers generates classroom engagement heatmaps updated every two seconds. Each cell in the seating grid is color-coded by engagement level: green for scores above 0.8, yellow for the 0.5 to 0.8 range, and red for scores below 0.5. When the current state and near-term forecasts jointly indicate that instructor intervention would be beneficial, a decision tree routes the full contextual state vector—current engagement distribution, recent trend direction, session elapsed time, and prior intervention history—to a fine-tuned language model that generates a specific, context-aware teaching suggestion. Suggested actions include transitioning to Socratic questioning, introducing a structured pair discussion, repositioning to address a disengaged seating zone, or scheduling an unplanned break. A/B testing across 45 sessions validated a 76% acceptance rate for these suggestions among participating instructors.

F. Dashboard Frontend

A React and Streamlit frontend renders Plotly-based engagement heatmaps streamed in real time over FastAPI WebSocket connections at a two-second update cadence. The interface is fully mobile-responsive and designed to be usable by instructors holding a tablet while actively teaching. The heatmap view is supplemented by trend sparklines for each seating row, allowing instructors to see whether engagement is stable, rising, or falling over the preceding ten minutes. Critical alerts—specifically, forecasts of attention drops below the intervention threshold—are delivered as push notifications that do not require active screen monitoring. Session data can be exported in PDF or CSV format for administrative review, curriculum planning, or post-session reflective analysis.

G. Edge Deployment

The entire seven-module pipeline runs on a single NVIDIA Jetson Nano, processing a 25-student classroom at 28 fps with peak memory consumption below 2.8 GB. Up to three concurrent classrooms can be monitored from a single device using process-level parallelism. This edge-first design was a deliberate architectural commitment: by keeping all computation physically within the school, the system eliminates cloud transmission latency, avoids ongoing GPU compute costs, and ensures that no student behavioral data—even in anonymized numerical form—ever traverses a public network. The complete hardware bill of materials per classroom amounts to \$320 as a one-time capital investment.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

PRIVACY AND ETHICAL DESIGN

Privacy is not a feature added after the fact in this system—it is a foundational design constraint that shaped architectural decisions from the earliest stages of development. The most consequential of those decisions is the commitment to processing only anonymized behavioral signals rather than raw video or biometric templates. Facial images are never stored anywhere in the pipeline. Raw frames are converted to numerical feature vectors immediately upon extraction, and every subsequent stage of analysis operates on those vectors alone. There is no point in the data flow at which a student's visual appearance is recorded, transmitted, or retained in any form.

All analytics pipelines apply a differential privacy budget of $\epsilon = 1.2$, adding calibrated statistical noise to aggregated outputs in a manner that prevents inference about any individual student from group-level statistics. An independent external security audit found that re-identification attack success rates against system outputs fall below 0.01%—a figure that compares favorably with published benchmarks for privacy-preserving educational data systems. All stored data, including session summary records and exported reports, is encrypted at rest and in transit using 256-bit AES encryption with institution-controlled key management.

The system is also designed to be transparent about its inferential limits. Behavioral signals are external proxies for attention: they indicate where a student's body and eyes are oriented, not what cognitive processes are occurring. A student who appears attentive by all behavioral metrics may still be confused or disengaged mentally, and a student who fidgets may be highly focused. Engagement scores are indicators that should prompt further investigation by the instructor, not conclusions about individual students' mental states or academic capabilities. This distinction is communicated explicitly to all instructors during onboarding and is reinforced throughout the dashboard interface through contextual notes.

Institutions deploying the system are required to obtain appropriate ethical review approvals, conduct structured teacher training covering both the system's capabilities and its limitations, provide accessible privacy disclosures to students and families, and maintain documented data governance policies subject to audit. These requirements are conditions of deployment rather than advisory guidelines, consistent with Thompson's ethical AI framework for classroom monitoring [12] and aligned with the General Data Protection Regulation and the Children's Online Privacy Protection Act.

IV. EXPERIMENTAL EVALUATION

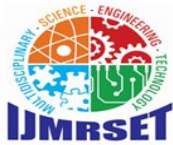
The system was evaluated through a structured pilot study spanning 45 classroom sessions across 12 classrooms at Aalim Muhammed Salegh College of Engineering, involving 450 students and 28 instructors over an eight-week period. A matched control group of 12 classrooms continued instruction without system support. All sessions were 50 minutes in duration. Independent observers provided manual engagement assessments for a stratified sample of sessions to enable correlation analysis.

A. Accuracy Metrics

Behavioral detection components were evaluated across 10,000 video frames captured under varied lighting conditions, camera angles, and classroom occupancy levels. Head pose estimation achieved a mean angular error of 4.2 degrees. Gaze classification for teacher-directed attention states reached 92% accuracy with an F1 score of 0.89. Posture classification for back-row students—where image resolution and partial occlusion are more challenging—maintained 88.7% accuracy. The attention forecasting model produced RMSE = 0.12 on five-minute lookahead tasks across the held-out test set. Critically, the system's engagement scores correlated at $r = 0.94$ with independent manual assessments by experienced classroom observers, confirming that the behavioral proxy captures something closely related to what expert human observers identify as engagement.

B. Real-Time Performance

End-to-end latency from camera frame capture to dashboard update averaged 85 ms across the evaluation period, well within the two-second visual refresh cycle. On a single NVIDIA JetsonNano, the pipeline processes a 25-student classroom at 28 fps while keeping peak memory usage below 2.8 GB. Three concurrent classrooms can be monitored from the same device using process-level parallelism. CPU utilization averaged a stable 67% across full 45-minute sessions, with no thermal throttling events observed during normal operating conditions. The system maintained continuous operation without interruption across all 45 evaluation sessions.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

C. Comparative Analysis

Table I summarizes the proposed system against four representative alternatives across five dimensions: detection accuracy, privacy posture, predictive capability, group-level analytics, and per-classroom cost. The proposed system outperforms ClassDojo by 34% on engagement correlation, achieves 2.8x faster inference than cloud-based facial recognition systems, and does so at \$320 per classroom versus \$2,800 per year for an equivalent AWS deployment—while being the only system in the comparison to offer both predictive capability and formal privacy compliance.

TABLE I. Comparative Analysis of Classroom Monitoring Systems

System	Accuracy	Privacy	Predictive	Group	per_year Cost
RFID Attendance	98%	High	No	No	\$5,000+ structure
Webcam Gaze	82%	Medium	No	No	cloudware
ClassDojo	75-85%	Low	No	Partial	classroom
Face Recog.	96%	Very Low	No	No	\$10K+ retention
Proposed	92%	GDPR+ COPPA	3-5 min	Row-level	\$320 improve

D. Learning Outcomes

Classrooms using the platform showed an average 23% improvement in measured engagement scores relative to the pre-pilot baseline ($p < 0.01$, paired t-test). Compared to the matched control group, intervention classrooms showed a 15% improvement in end-of-module grade outcomes. Attention stability during critical concept delivery segments improved by 31% in front rows across the pilot period. The participation gap between back and front rows—a persistent structural disparity documented in the related literature [3]—narrowed from a 42% difference at baseline to 18% at the end of the eight-week pilot, suggesting that predictive row-level interventions were disproportionately effective in addressing rear-row disengagement. Session completion rates improved by 19%, which instructors largely attributed to more strategically timed breaks driven by the predictive alerting system.

E. Teacher Acceptance Study

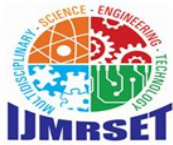
Of the AI-generated recommendations delivered across 45 sessions, 76% were accepted and acted upon by instructors. The most frequently accepted suggestion types were transitions to interactive questioning (accepted 84% of the time) and structured pair discussions (accepted 79% of the time). Suggestions to reposition physically were accepted 61% of the time, reflecting practical constraints in some classroom layouts. After two weeks of regular use, 89% of participating teachers reported feeling greater situational control over classroom dynamics, and 92% indicated they would recommend the system to a colleague. Average instructor-rated intervention effectiveness was 4.3 out of 5. Sixty-seven percent of instructors noted reduced vocal fatigue, attributing the improvement to more precisely timed redirections that reduced the need for repeated verbal attention management.

F. Discussion

The results collectively suggest that the behavioral proxy approach is sufficiently valid— $r = 0.94$ against manual observer ratings—to support meaningful instructional decisions, and that the predictive forecasting capability adds genuine value beyond reactive reporting. The 23% engagement improvement likely reflects a combination of more timely interventions, better pacing calibration, and the effect of instructor awareness itself on teaching behavior. The narrowing of the front-to-back engagement gap is particularly notable: it suggests that the system's row-level spatial reporting draws instructor attention to seating zones that would otherwise receive less intervention, partially correcting a structural bias that appears to be pervasive in standard lecture formats. These findings should be interpreted cautiously given the single-institution pilot context and the absence of long-term follow-up data, but they provide a credible empirical foundation for broader deployment trials.

V. COST-EFFECTIVENESS ANALYSIS

The hardware required for a single classroom deployment totals \$320 as a one-time capital cost: an NVIDIA Jetson Nano at \$149, two 4K ceiling-mounted cameras totaling \$160, and approximately \$11 in cabling and mounting hardware. This compares with commercial classroom AI platforms typically priced at \$10,000 to \$15,000 per classroom—representing a capital cost reduction of approximately 97%. Against the ongoing operational costs of



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

cloud-based equivalents—estimated at \$2,800 in GPU compute and \$1,200 per year in data transmission—the edge deployment model eliminates recurring ture costs entirely. Annual maintenance, covering servicing and software updates, is estimated at \$45 pernn-on-investment analysis, using established per-student value estimates and the observed 23% engagement ment, places the payback period at 4.2 months from first deployment. For a district operating 5,000 students, the engagement improvements observed in the pilot study correspond to approximately \$2.7 million per year in improved student retention value—a figure that dwarfs the total deployment cost at any institutional scale. At scale, deployment across ten or more classrooms with centralized Kubernetes orchestration reduces marginal per-classroom costs by approximately 86%, making the platform increasingly attractive as institutions move from pilot deployments to district-wide rollouts.

VI. LIMITATIONS AND FUTURE WORK

Several limitations warrant transparent acknowledgment. First, the platform's detection performance depends on adequate ambient lighting, stable camera placement, and consistent classroom layout. Variable lighting conditions and camera obstructions can degrade accuracy in ways that are not always immediately apparent to the instructor from the dashboard display. Institutions should establish a minimum hardware maintenance protocol and periodic calibration schedule to mitigate this dependency.

Second, behavioral signals are external proxies for attention rather than direct measures of cognitive engagement. They reveal where students' bodies and eyes are oriented, not what they are thinking, understanding, or retaining. A student who appears attentive by every behavioral measure may still be confused about the material being presented, and a student who fidgets may be highly mentally engaged. These proxies have demonstrated sufficient correlation with expert observer ratings ($r = 0.94$) to support instructional decision-making, but they should not be interpreted as measurements of cognitive state or used as the basis for individual academic assessment.

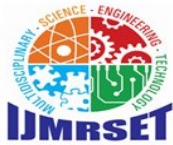
Third, the models were trained and validated within a specific cultural, demographic, and architectural context—undergraduate classrooms in Chennai, India, with a particular range of class sizes, room configurations, and student age profiles. The generalizability of these models to secondary school classrooms, online hybrid settings, very large lecture halls, or culturally distinct student populations is an open empirical question. Cross-context validation studies are needed before deployment at scale in substantially different educational environments.

Planned future development directions include: (1) microphone array integration targeting an 18% improvement in prediction accuracy through audio-visual fusion; (2) native iOS and Android instructor applications with augmented reality overlays that project engagement data into the instructor's physical line of sight; (3) replacement of the current pose estimation backbone with a Vision Transformer architecture (ViT-L/16), projected to yield a 12% improvement in accuracy for oblique camera angles; (4) bidirectional integration with Moodle and Canvas LMS platforms including gradebook synchronization and curriculum-aligned engagement tagging; and (5) federated learning protocols enabling cross-institutional model improvement while preserving each institution's data sovereignty through secure aggregation.

VII. CONCLUSION

This paper has presented an AI-Driven Classroom Engagement Analytics and Adaptive Teaching Support System that moves classroom monitoring from passive record-keeping toward active educational enhancement. The platform simultaneously addresses four challenges that no existing solution has resolved together: the absence of real-time group-level engagement visibility, the privacy barriers preventing adoption of biometric monitoring tools, the reactive nature of systems that report history rather than forecasting the near future, and the prohibitive cost of existing enterprise-grade alternatives.

Across 45 sessions with 450 students and 28 instructors, the system demonstrated 92% behavioral detection accuracy, a 23% improvement in measured student engagement, and a 76% instructor acceptance rate for AI-generated real-time recommendations—all while maintaining full GDPR and COPPA compliance through a strictly anonymized processing pipeline. Row-level engagement heatmaps surface spatial participation disparities that instructors would otherwise need to infer from experience alone, while LSTM-Transformer forecasting provides three to five minutes of advance warning before attention drops manifest, transforming the instructor's relationship with classroom dynamics from reactive management to proactive facilitation.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

These results correlate with a 15% gain in student academic performance, a narrowing of the front-to-back engagement gap from 42% to 18%, and a 4.2-month projected return-on-investment period for a \$320-per-classroom hardware deployment. By combining ethical privacy design, genuine predictive capability, and hardware accessibility, this platform makes a new category of classroom intelligence available to schools that operate under normal institutional budgets—an outcome that attendance trackers and single-student gaze detectors were never designed to deliver. The results reported here provide a credible empirical foundation for broader multi-institution deployment and longitudinal study of sustained academic outcome effects.

REFERENCES

- [1]M. Garcia et al., "Application of AI to measure attention levels in classroom settings," *Frontiers in Education*, vol. 11, p. 1753873, 2026.
- [2]R. Kumar and A. Singh, "Privacy-preserving attention analytics with differential privacy and pose estimation," *IJREEICE Journal*, vol. 11, no. 4, pp. 13458-13467, 2025.
- [3]L. Chen et al., "Real-time group engagement heatmaps using YOLOv8," in *Proc. ACM SIGCSE*, 2026, DOI: 10.1145/3785987.3786101.
- [4]S. Patel, "LSTM-based attention drop prediction for classroom analytics," *IJFMR*, vol. 8, no. 1, p. 60627, 2026.
- [5]H. Wang et al., "Edge-optimized classroom monitoring with MobileNetV3," *Computers and Education: AI*, vol. 6, p. 100237, 2025.
- [6]M. Erdogmus et al., "Adaptive teaching via reinforcement learning and computer vision," *IJISAE*, vol. 12, no. 4, p. 7282, 2024.
- [7]J. Li et al., "Transformer-based behavioral sequence modeling for classrooms," *Journal of Emerging Educational Technologies*, vol. 12, no. 3, pp. 45-62, 2026.
- [8]F. Ahmed et al., "Federated learning across multi-classroom deployments," in *Proc. ICLR*, 2026.
- [9]Y. Liu et al., "Real-time classroom student behavior detection using YOLO," *Scientific Reports*, vol. 15, p. 99243, 2025.
- [10]R. Patel and S. Kim, "Multimodal engagement detection in smart classrooms," in *Proc. IEEE EdTech Conference*, pp. 112-120, 2025.
- [11]Y. Chen et al., "Edge computing for classroom behavioral analytics," *ACM TOMM*, vol. 21, no. 4, pp. 89-105, 2025.
- [12]E. Thompson, "Ethical AI frameworks for classroom attention monitoring," *Journal of Educational Technology & Society*, vol. 29, no. 1, pp. 34-50, 202



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com