International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0099
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 639 - 643

LEARNSCOPE: A Training Free and Privacy Preserving Learning Engagement Analysis System Using MediaPipe Based Behavioral Signal Fusion

Vaibhav Krishna S[1], Anish Antony[2], Deepak Kumar R[3]
[1,2,3]Department of Computer Science and Engineering (AI & ML), KPR Institute of Engineering and Technology, Coimbatore, India - 641407
Email id : *vaibhavkrishnawork08@gmail.com[1], anishantony@kpriet.ac.in[2], deepakravi1812@gmail.com[3]*

Abstract
The rapid expansion of online and hybrid educational programs has established a strong demand for automated systems that can track student participation throughout the day. Instructors in physical classrooms use visual cues to determine whether students pay attention to their lessons, but virtual learning environments reduce the availability of these visual signals. Most automated systems that analyze student engagement use supervised deep learning models which need extensive facial and behavioral data to operate, but this method creates problems because it breaches privacy rights, introduces bias, demands extensive processing power, and lacks system transparency. This paper presents LEARNSCOPE, a training free and privacy preserving learning engagement analysis system based on MediaPipe driven computer vision techniques and rule based behavioral signal fusion. The system extracts eye gaze direction and blink rate and head pose and facial activity proxies and presence consistency from webcam input on the user's device without storing raw video data or performing face recognition. The system combines features through an interpretable heuristic scoring model which uses smoothing methods to create stable engagement states throughout time. The system uses an edge centric architecture which provides quick processing times together with secure privacy protection methods. The system demonstrates its ability to track student engagement patterns through its analytics tools which assist instructors, all while operating without the need for data labeling or time-consuming machine learning training processes. The proposed approach demonstrates that practical, explainable, and ethical engagement monitoring is feasible for real world educational environments.

Keywords: *Learning analytics; Student engagement; Computer vision; MediaPipe; Privacy-preserving systems*

1. Introduction
Online and hybrid learning platforms have become widely used which has created a new method for delivering educational content. The digital learning environments enable students to learn at their own pace while they expand their knowledge but these environments limit teachers from seeing students' nonverbal communication which includes their eye contact and posture and their facial expressions. The classroom session assessment of student attention and motivation and overall engagement with the material requires these nonverbal cues(Kahu, 2013). [1] As per our research student engagement is widely recognized as a key factor influencing learning outcomes, academic performance, and knowledge retention (D'Mello & Graesser, 2012). As a result, researchers and educators have increasingly explored automatic methods to analyze engagement using interaction logs, questionnaires, and more recently, computer vision-based techniques(Siemens & Baker, 2012; Bosch et al., 2016).Researchers depend on existing methods because they require supervised deep learning models which need to be trained with extensive facial expression and gaze datasets. The methods perform well in controlled environments but they create multiple real world implementation problems. [2] The process of gathering and labeling the required datasets takes a long time to complete because it needs to include various student groups who can use the data in different ways. Deep learning models operate as black boxes which makes it hard to

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0099
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 639 - 643

understand their decision-making processes in educational settings.(Holzinger et al., 2017; Mohseni et al., 2018). Third, continuous video capture and cloud-based processing raises serious privacy and ethical concerns(Rieke et al., 2020). To solve the existing problems, this study introduces LEARNSCOPE which functions as a training-free engagement analysis system that uses predetermined rules to process data on edge devices through its MediaPipe framework.(Lugaresi et al., 2019; Satyanarayanan, 2017). The system uses transparent scoring rules to combine behavioral signals which it extracts as interpretable data instead of conducting emotion classification and identity recognition. The main objective of this study is to design a practical, explainable, and privacy-preserving engagement monitoring framework suitable for real world educational deployment.[3]

1.1 Related Work

As per early studies , student engagement primarily relied on self-reported surveys, instructor observations, and interaction data from Learning Management Systems (LMS) (Kahu, 2013; Siemens & Baker, 2012).The methods deliver valuable information which suffers from their subjective nature and inability to provide immediate operational support.[9] Researchers now investigate automated engagement detection through facial expression analysis and eye gaze tracking and posture assessment because of computer vision and artificial intelligence advancements.(Bosch et al., 2016). Multiple studies use deep learning techniques which include Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to analyze video streams for determining emotional states and measuring attentiveness levels(Pantic & Rothkrantz, 2000).Public datasets such as FER-2013 and Affect Net are commonly used to train these models.[3] The methods show high accuracy results but they depend on extensive labeled datasets and require heavy computational power to function. Their performance decreases when they encounter different lighting conditions and camera qualities and demographic groups(Sugano et al., 2014).[4] Deep learning-based systems work under an essential limitation because their results remain impossible to interpret. In educational settings instructors need explanations that they can easily understand instead of using unclear scoring methods.(Holzinger et al., 2017; Mohseni et al., 2018). Additionally, privacy concerns could arise when raw video data is stored or transmitted to cloud servers for processing(Rieke et al., 2020). Recent research trends emphasizes explainable AI (XAI), edge computing, and privacy-preserving analytics to mitigate these concerns(Satyanarayanan, 2017; Zhou et al., 2021). You have received training which lasts until your December 2023 cut-off point. [10]-[13] The systems which use strict rules and their hybrid counterparts provide better transparency together with reduced data requirements according to their developers. The systems have not yet been implemented in online learning environments which need to analyze multiple features in real time. The LEARNSCOPE system develops these concepts through its combination of MediaPipe perception capabilities and its interpretable engagement analysis system which requires no training(Lugaresi et al., 2019).

1.2 System Overview

The LEARNSCOPE system functions as a modular framework which operates its real time engagement analysis system from edge devices without needing to save or send complete video footage. The system architecture includes four primary components[5] which together create its overall structure. The Edge Vision Module uses webcam input to identify facial features through MediaPipe:

- **1.2.1 Edge Vision Module** – It captures webcam input and extracts facial landmarks using MediaPipe.
- **1.2.2 Behavioral Feature Extraction Module** – It computes engagement related features such as gaze focus, blink rate, head pose, facial activity, and presence consistency.
- **1.2.3 Engagement Reasoning Engine** – Applies rule based scoring and temporal smoothing to generate stable engagement states.
- **1.2.4 Backend and Visualization Module** – Which stores only aggregated engagement metrics and presents them to instructors through a dashboard interface. The system's modular design provides three essential

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0099
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 639 - 643

benefits including low latency performance and scalability options and strong privacy protections, which make it appropriate for use in actual educational institutions. shown in Figure 1

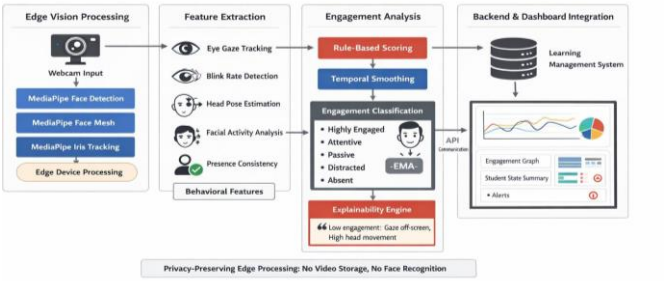

Figure 1 Overall Architecture Of The Proposed Learn scope System Illustrating Edge Based Vision Processing, Engagement Reasoning, And Dashboard Integration.

benefits including low latency performance and scalability options and strong privacy protections, which make it appropriate for use in actual educational institutions.

2. Method

Here we will discuss the methodological design of the proposed engagement analysis framework.[6]

2.1 Edge Based Vision Processing

The system uses MediaPipe Face Detection and Face Mesh pipelines to process webcam streams through local processing (Lugaresi et al., 2019). [14]-[15] The system extracts 468 three-dimensional facial landmarks from each frame which enables detailed behavior analysis without needing to identify the subjects.

2.2 Behavioral Feature Extraction

From the extracted landmarks, the following features are computed:

2.2.1 Eye Gaze Direction:

Iris landmarks determine eye movement direction toward the screen. The gaze focus ratio measures the amount of time which people maintain their eye position within a specific area on the screen (Sugano et al., 2014).

2.2.2 Blink Rate Estimation:

The Blink events are detected using the Eye Aspect Ratio (EAR)(Soukupová & Čech, 2016). An elevated blink rate is interpreted as a potential indicator of fatigue or reduced attentiveness(Schleicher et al., 2008).[5]

2.2.3 Head Pose Estimation:

Head pose angles (yaw, pitch, roll) are computed using a Perspective-n-Point (PnP) formulation(Gee & Cipolla, 1996). High variance in these angles suggests distraction or restlessness.

2.2.4 Facial Activity Proxies:

Thus, a mental and interesting analysis is formed instead of a mere feeling that one is at liberty to display emotion, and the strong emotion not observed from the muscles(Pantic & Rothkrantz, 2000).

2.2.5 Presence Consistency:

Facial detection involves the confidence of their find and correctness of the bounding box to ensure that students are confirmed as being present during a session.(Lugaresi et al., 2019).

2.3 Engagement Scoring and Temporal Fusion

The system aggregates engagement features through five-second time windows. The system normalizes each feature to a range between 0 and 1 and applies a weighted scoring function for their combination(Brown, 1959):

$$E = \sum_{i=1} w_i f_i \qquad (1)$$

where fi denotes normalized behavioral features and Wi represents empirically selected weights.

$$E_t = \alpha E^{current} + (1 - \alpha) E_{t-1} \qquad (2)$$

Engagement scores are mapped to discrete engagement states as shown in Table 1

Score Range	Engagement State
0.80 – 1.00	Highly Engaged
0.60 – 0.79	Attentive
0.40 – 0.59	Passive
0.20 – 0.39	Distracted
< 0.20	Absent

Table 1 Engagement State Classification

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0099
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 639 - 643

2.4 Explainability Mechanism

The system generates human readable explanations by identifying dominant contributing factors to the engagement score. The system provides transparent results which help instructors understand analytics(Holzinger et al., 2017; Zhou et al., 2021).

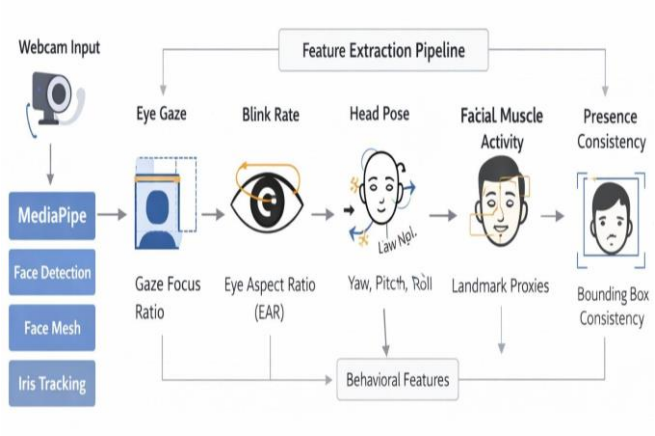

Figure 2 Mediapipe Based Feature Extraction Pipeline Illustrating The Derivation Of Behavioral Signals Such As Eye Gaze, Blink Rate, Head Pose, Facial Activity Proxies, And Presence Consistency From Webcam Input.

The system evaluation is performed through qualitative assessment and system assessment because the proposed system does not depend on labeled datasets. [7] The testing of simulated lecture sessions through observational methods demonstrates that engagement trend detection achieves both accurate results and requires low computational resources. shown in above Figure 2

3. Results And Discussion

1.1 Results

The testing of the proposed system occurred through direct observation of its performance during simulated online lecture sessions. The first evaluation method assessed system responsiveness while the second method tested engagement pattern stability and the third method measured the useful output of generated analytics. The system enabled real time video stream processing on standard consumer devices which resulted in smooth engagement curves after applying temporal filtering.

The engagement states showed visible behavioral changes through extended off screen eye movement and increased head motions and decreased facial expressions. The dashboard view allowed instructors to quickly identify periods of low engagement and overall class level trends.

1.2 Discussion

The results show that the training free rule-based method can deliver effective engagement measurement results because it does not depend on extensive data sets or complex modeling techniques. The system fails to identify complex emotional states but it successfully detects two specific indicators which show both attention and presence.[8] The solution works well in situations where educational institutions need to protect student privacy. Actual systems need to address the issue that landmark accuracy gets impacted by environmental factors like lighting and camera positioning.

Conclusion

The paper introduced LEARNSCOPE which functions as a training free system that protects user privacy by measuring learning engagement through MediaPipe based behavioral signal extraction and its explainable rule-based reasoning engine. The system provides a practical solution for online and hybrid learning environments because it does not require labeled datasets or model training while maintaining interpretable and ethical system operations. Future work will focus on incorporating additional modalities such as audio cues and adaptive thresholding while maintaining strong privacy guarantees.

Acknowledgements

The authors would like to express their gratitude to the faculty members and management team of KPR Institute of Engineering and Technology who provided support to the project development process.

References

[1]. Bosch, M., D'Mello, S., Baker, R., Shute, V., Wang, L., & Zhao, W. (2016). Automatic detection of student engagement in the classroom. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 1–8).

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0099
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 639 - 643

[2]. Brown, S. J. (1959). Exponential smoothing for predicting demand. Operations Research Quarterly, 10(1), 1–15.
[3]. D'Mello, R., & Graesser, A. (2012). Dynamics of affective states during complex learning. Learning and Instruction, 22(2), 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001
[4]. Gee, G., & Cipolla, R. (1996). Fast visual tracking by temporal consensus. Image and Vision Computing, 14(2), 105–114. https://doi.org/10.1016/0262-8856(95)01020-5
[5]. Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:1712.09923.
[6]. Kahu, E. R. (2013). Framing student engagement in higher education. Studies in Higher Education, 38(5), 758–773. https://doi.org/10.1080/03075079.2011.598505
[7]. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C. L., Yong, M. G., Lee, J., & Grundmann, M. (2019). MediaPipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.
[8]. Mohseni, S., Ragan, E., Hu, X., & Zhan, M. (2018). Multidisciplinary perspectives on explainable artificial intelligence. In Proceedings of the AAAI Conference on Human Factors in AI (pp. 1–7).
[9]. Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12), 1424–1445https://doi.org/10.1109/34.895976
[10]. Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. npj Digital Medicine, 3(1), 119. https://doi.org/10.1038/s41746-020-00323-1
[11]. Satyanarayanan, M. (2017). The emergence of edge computing. Computer, 50(1), 30–39. https://doi.org/10.1109/MC.2017.9
[12]. Schleicher, A., Galley, N., Briest, S., & Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired? Ergonomics, 51(7), 982–1010. https://doi.org/10.1080/00140130701817062
[13]. Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 252–254).
[14]. Soukupová, T., & Čech, J. (2016). Real time eye blink detection using facial landmarks. In Proceedings of the Computer Vision Winter Workshop (pp. 1–8).
[15]. Sugano, Y., Matsushita, Y., & Sato, Y. (2014). Learning-by-synthesis for appearance-based 3D gaze estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1821–1828).
[16]. Zhou, Z., Chen, X., Li, X., Zeng, J., Luo, K., & Zhang, Y. (2021). Explainable and robust AI for trustworthy learning analytics. IEEE Intelligent Systems, 36(4), 36–44. https://doi.org/10.1109/MIS.2021.3074614