جـامعة بيروت العربية

BEIRUT ARAB UNIVERSITY

EngageSense AI: A Rule Based System for Real Time Student Engagement Detection Using Head Pose and Eye Gaze

Rein Ghattas

202301814

Hadi Mostafa

202303776

Research

Methodology

1. Introduction

With the rapid growth of online learning, maintaining student engagement has become a major challenge. Traditional indicators such as eye contact, posture, and attentiveness are harder to observe in virtual classrooms [5]. Existing AI systems mainly rely on emotion-recognition models, which use datasets like DAiSEE and AffectNet.[15],[19] However, emotional states do not always reflect actual attentional focus [9], and emotional labeling is subjective and computationally expensive.

Recent advances in computer vision—such as MediaPipe FaceMesh and Dlib facial landmark models—enable accurate extraction of behavioral cues like eye gaze and head pose [6],[7]. These physical indicators correlate strongly with real attentional engagement [10], making them more suitable for real-time educational environments.

This project introduces EngageSense AI, a lightweight and interpretable system that detects student engagement using purely behavioral cues.

2. Problem Statement

Current engagement detection systems suffer from several limitations:

- Emotion-based models rely on large datasets and are computationally heavy [13].
- Deep-learning approaches are black-box systems with low interpretability [16].
- Many systems cannot run in real time on CPU-only devices.
- Emotional expressions often fail to reflect true behavioral engagement [9].

As a result, teachers and educational platforms lack a fast, reliable, and transparent method for detecting real-time student attention.

This research addresses the need for an interpretable, rule-based behavioral model that works in real time without the need for training or GPU resources.

3. Aim of the Work

The aim of this research is to design and evaluate **EngageSense AI**, a rule-based behavioral engagement detection system that operates in real time using eye gaze and head pose.

4. Objectives

- Extract real-time gaze and head-pose features using MediaPipe and Dlib.
- Develop a transparent, rule-based inference model with no training required.

- Achieve real-time performance (≥ 25 FPS) on CPU-only hardware.
- Construct a manually labeled behavioral dataset (~1,000 frames).
- Evaluate system accuracy using precision, recall, and F1-score.
- Compare predictions with human-labeled ground truth.

5. Methodology

5.1 Literature Review

A review of behavioral and emotional engagement detection studies was conducted, focusing on:

- Limitations of emotion-based datasets (DAiSEE, AffectNet) [15],[19]
- Behavioral indicators such as gaze direction and head orientation [10]
- Real-time computer vision pipelines (MediaPipe, Dlib) [6],[7]
- Gaps in transparency and computational efficiency [14]

5.2 System Design

EngageSense AI consists of four modules:

1. **Frame Capture** – Using OpenCV for real-time video input.
2. **Feature Extraction** – MediaPipe FaceMesh for eye landmarks [7], Dlib for head pose.
3. **Rule-Based Scoring Model** – Weighted combination of gaze (0.6) and head pose (0.4).
4. **Output Module** – Displays engagement state in real time.

5.3 Dataset Creation

- Recorded 720p webcam sessions from 10 participants.
- Labeled ~1,000 frames manually into Engaged, Partially Engaged, and Not Engaged.
- Compared system predictions to ground truth to compute accuracy.

5.4 Evaluation

Engagement classification was evaluated using:

- Accuracy
- Precision
- Recall

- F1-Score
- Confusion matrix

Ethical measures were followed by discarding all raw video frames and storing only numerical behavioral features.

6. Expected Outcomes

The system is expected to:

- Provide accurate real-time engagement classification (~82% accuracy).
- Operate at 25 FPS on CPU without GPU or training.
- Demonstrate that behavioral cues alone are strong indicators of attention.
- Offer transparent decision-making compared to deep-learning models.
- Enable scalable and ethical deployment in online classrooms.

7. Significance of the Study

This research contributes:

- A lightweight and interpretable alternative to emotion-based deep learning.
- A practical engagement monitoring tool for educators.
- A full methodology pipeline: behavioral modeling, dataset creation, rule-based design, and evaluation.
- A foundation for future hybrid behavioral-emotional models and multi-student tracking.

References

[1] S. Zhang, A. Abedi and S. S. Khan, "Supervised Contrastive Learning for Ordinal Engagement Measurement," *arXiv preprint arXiv:2505.20676,* 2025.

[2] X. S. a. H. L. Y. Zhao, "Hybrid CNN-LSTM for Engagement Recognition," *IEEE Transactions on Multimedia,* 2022.

[3] Y. Z. a. J. L. X. Li, "Deep Learning for Gaze Estimation: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2020.

[4] C.-H. Wu, S.-Y. Liu, X. Huang, X. Wang, R. Zhang, L. Minciullo, K. K. Wong, K. Kwan and K.-T. Cheng, "CMOSE: Comprehensive Multi-Modality Online Student Engagement Dataset with High-Quality Labels," in *IEEE/CVF CVPR Workshops (CVPRW)*, 2024.

[5] M. Trowler, "Student Engagement Literature Review," Higher Education Academy, 2010.

[6] P. R. a. L. M. T. Baltrušaitis, "OpenFace 2.0: Facial Behavior Analysis Toolkit," in *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2018.

[7] G. Research, "MediaPipe FaceMesh Documentation," 2022. [Online]. Available: https://google.github.io/mediapipe/.

[8] S. M. a. P. K. R. Suresh, "Multimodal Fusion for Student Engagement Analytics," *Computers & Education: Artificial Intelligence,* 2023.

[9] C. Pekrun, "Emotions and Learning," UNESCO (Educational Practices Series), 2014.

[10] Y. L. a. P. W. M. Zhang, "Eye Gaze-Based Attention Recognition in Online Learning Environments," *Computers & Education,* 2021.

[11] M. B. a. J. M. L. Whitehill, "Automatic Estimation of Engagement from Facial Expressions," *IEEE Transactions on Affective Computing,* 2014.

[12] P. S. R. &. S. A. Kaur, Prediction and Localization of Student Engagement in the Wild., International Conference on Multimodal Interaction (ICMI), 2018.

[13] P. C. B. a. A. H. P. J. A. Fredricks, "School Engagement: Potential of the Concept, State of the Evidence," *Review of Educational Research,* vol. 74, no. 1, p. 109, 2004.

[14] R. Gupta, M. Sharma and D. Chauhan, A multimodal facial cues based engagement detection, vol. 5, Frontiers in Computer Science, 2023.

[15] A. Gupta, A. D’Cunha, K. Awasthi and V. N. Balasubramanian, "DAiSEE: Towards User Engagement Recognition in the Wild," arXiv, 2016.

[16] G. T. a. H. Dey, "Deep Learning for Real-Time Emotion Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020.

[17] K. Cao, M. Long, J. Wang and M. (. p. Jordan, "Rank-consistent Ordinal Regression for Deep Learning," *AAAI Conference on Artificial Intelligence,* 2021.

[18] W. B. I. A. T. D. D. C. a. R. P. B. Woolf, "Affect-Aware Tutors: Recognizing and Responding to Student Affect," *International Journal of Learning Technology,* vol. 4, no. 3, 2009.

[19] D. C. a. M. H. M. A. Mollahosseini, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," in *EEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[20] D. D. a. J. W. A. Gupta, "DAiSEE: Dataset for Affective Student Engagement in the Wild," in *Proceedings of the ACM on Multimedia*, 2016.