EngageSense AI

A Rule-Based System for Real-Time Student Engagement Detection Using Head Pose & Eye Gaze

Research Methodolgy COMP500

- Hadi Mostafa 202303776
- Rein Ghattas 202301814

جامعة بيروت العربية
BEIRUT ARAB UNIVERSITY

Outline

01

Introduction

- Motivation
- Problem Statement
- Objectives

02

Literature Review

- Emotion-Based Approaches
- Behavioral Approaches
- Research Gap

03

System Design

- System Architecture
- Feature Extraction
- Engagement Scoring Model

04

Methodology

- Dataset & Labeling
- Ethical Handling
- Evaluation Approach

05

Results & Discussion

- Accuracy & Performance
- Strengths & Limitations

06

Conclusion & Future Work

- Conclusion
- Future Work

01

INTRODUCTION

1. Motivation
2. Problem Statement
3. Objectives

➢ INTRODUCTION

1. Motivation

- Online classes make it hard to monitor engagement
- Teachers cannot track: attention, gaze, posture
- Existing AI relies on **emotions**, which do **not** reflect real attention
- Need a **lightweight, interpretable, real-time** system

➢ INTRODUCTION

2. Problem Statement

Existing systems struggle with:

- Black-box deep learning (no interpretability)
- High computational cost
- Dependency on massive emotion datasets
- Slow, not real-time
- Emotions ≠ attention

Our solution: Rule-based + behavioral = fast + transparent

➢ INTRODUCTION

3. Objectives

- Detect engagement using **gaze + head pose**
- Build a fully **interpretable rule-based model**
- Achieve **≥ 25 FPS** real-time
- Validate with a manually labeled dataset
- Keep processing lightweight (CPU only)

02

Literature Review

- Emotion-Based Approaches
- Behavioral Approaches
- Research Gap

➢ Literature Review

1. Emotional-Based Approaches

Study	Method used	Dataset	Limitations
Whitehill et al., 2014	Facial expression analysis	Custom	A student can look neutral or bored but still be paying attention. Expressions do not equal engagement.
Gupta et al., 2016	CNN emotion recognition	DAiSEE	Their CNN model is too simple to detect DAiSEE's engagement levels
Mollahosseini et al., 2017	Deep CNN for facial emotions	AffectNet	Needs a strong GPU. Too slow for real-time classroom use.
Suresh et al., 2023	Multimodal (face + audio)	Private dataset	Very complex and heavy model → not suitable for laptops or normal classroom devices.

➢ Literature Review

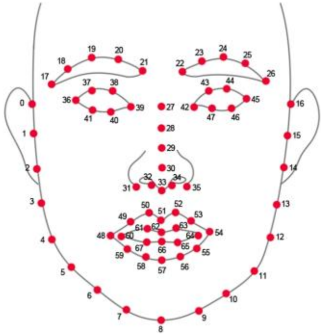
2. Behavioral Approaches

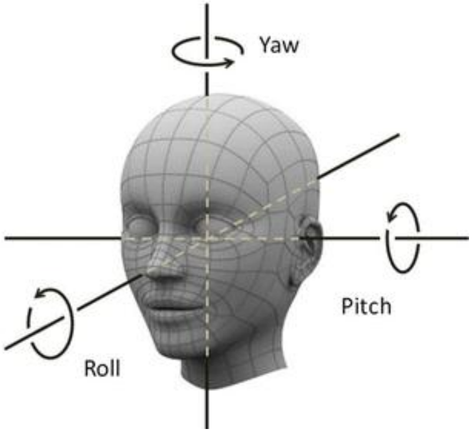
- Behavior-based engagement uses:

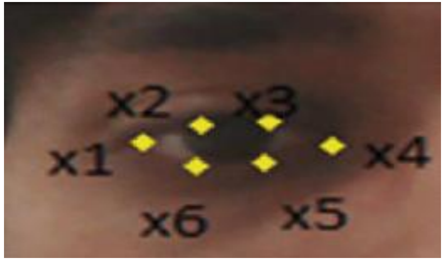
- Eye gaze
- Head pose
- Facial landmarks

- Tools used:

- MediaPipe FaceMesh
- Dlib 68-point landmarks
- OpenFace tools







➢ Literature Review

3. Research Gap

No system offers:

- Real-time performance
- Full transparency
- No training required
- Purely behavioral rules

EngageSense AI fills this gap.

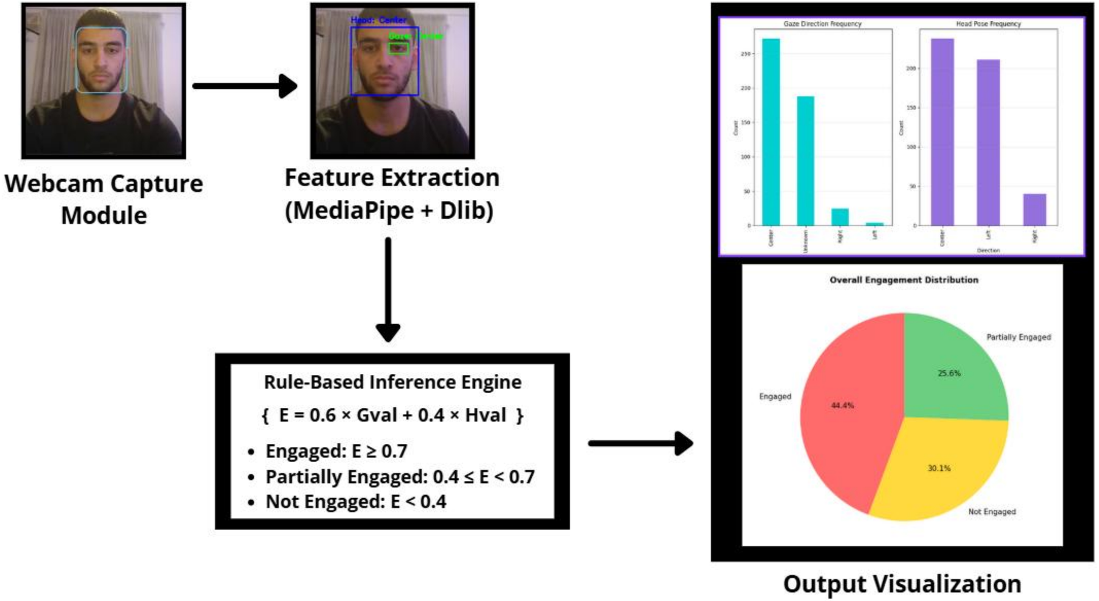
03

System Design

- System Architecture
- Feature Extraction
- Engagement Scoring Model

➢ System Design

1. System Architecture



➢ System Design

2. Feature Extraction & Tracking Techniques

MediaPipe FaceMesh (Eyes + Facial Landmarks)
Extracts:
– Eye region landmarks
– Iris center points
– Eyelid contours
– Facial geometry for stable gaze direction estimation

Dlib (Head Pose Estimation)

Uses 3D facial model + 2D landmarks to compute:

– **Yaw** (left ↔ right rotation)
– **Pitch** (up ↔ down rotation)
– **Roll** (tilt angle)

➢ System Design

3. Engagement Scoring

$$\boldsymbol{E = 0.6G + 0.4H}$$

$$\boldsymbol{H}\ (\boldsymbol{Head\ Pose}) \in \{1, 0.5, 0\} = \{Center, Slightly\ Right|Left, Fully\ Right|Left\}$$

$$\boldsymbol{G}\ (\boldsymbol{Gaze}) \in \{1, 0.5, 0\} = \{Center, Slightly\ Right|Left, Fully\ Right|Left\}$$

- **Classification:**

Engaged ≥ 0.7

Partially Engaged 0.4–0.69

Not Engaged < 0.4

Why 0.6 for gaze?
Gaze strongly correlates with attention.

04

Methodology

- Dataset & Labeling
- Ethical Handling
- Evaluation Approach

➢ Methodology

1. Dataset & Labeling

- **1,000 self-recorded frames**
  Collected from controlled video sessions recorded by a single participant.
- **Manually labeled ground truth**
  Each frame was annotated with one of three engagement states:
  Engaged, Partially Engaged, Not Engaged.
- **Prediction dataset generated automatically**
  The EngageSense AI system processed the same video to produce frame-wise engagement predictions.
- **Frame-by-frame accuracy evaluation**
  Predicted labels were directly compared to the ground-truth labels to compute accuracy, precision, recall, F1-score, and the confusion matrix.

➢ Methodology

2. Ethical Handling

- **No video footage stored** — only numerical engagement features are kept.
- **No identifiable information collected** — no faces, images, or personal data saved.
- **Designed for safe use in educational settings** — fully privacy-preserving and compliant with classroom requirements.

➢ Methodology

3. Evaluation Approach

Metrices

Accuracy	$(Ncorrect/Ntotal)x100$
Precision	$TP/(TP + FP)$
Recall	$TP/(TP + FN)$
F1-score	$2x(\frac{PrecisionxRecall}{Precision + Recall})$

05

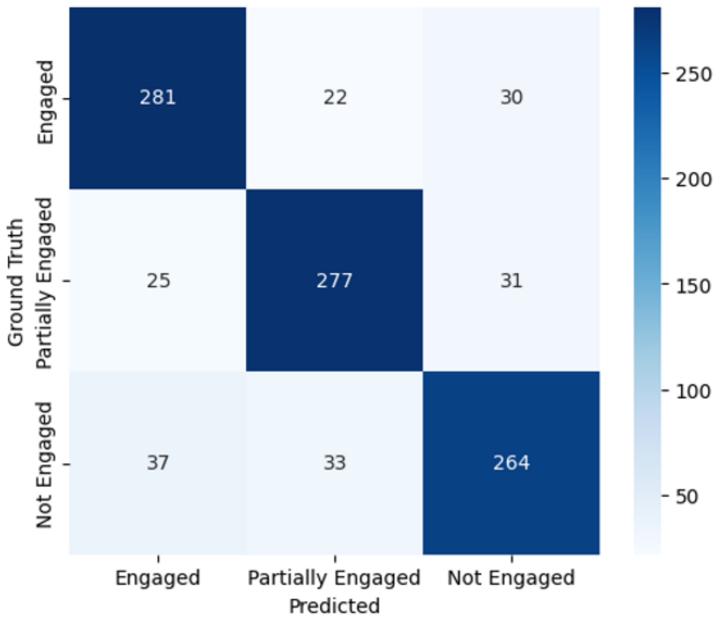
Results & Discussion

- Accuracy & Performance
- Strengths & Limitations

➢ Results & Discussion

1. Accuracy & Performance

- **Accuracy:** 82.2%
- **Metrics:** Precision, Recall, F1-score for all classes
- **Confusion Matrix:** Shows class-wise correctness and misclassification patterns
- **Stable results across Engaged, Partially Engaged, and Not Engaged**



➢ Results & Discussion

2. Strengths & Limitations

Strengths

- Real-time, lightweight performance
- Fully interpretable rule-based design
- No training needed
- Privacy-safe (no video stored)
- Easy to deploy in classrooms

Limitations

- Tested on one participant only
- Limited environmental diversity
- Only behavioral cues (gaze + head pose)
- Needs broader testing for robustness

06

Conclusion & Future Work

- Conclusion
- Future Work

Conclusion & Future Work

Conclusion

- The system accurately detects engagement using gaze and head pose.
- Results closely match the manually labeled ground truth.
- Real-time, and privacy-safe design shows strong potential for educational use.

Future Work

- Test with more participants and real classroom settings.
- Add more behavioral cues (posture, blinking).
- Explore hybrid rule-based + ML approaches.
- Develop a classroom dashboard and LMS integration.

➢ 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.

THANK YOU