# AI Based Student Engagement Analysis


Project Team

| | |
|---|---|
| Abdurrahman Khan | i221148 |
| Bilal Javed | i220788 |
| Sohaib Sattar | i220962 |


Session 2021-2025


Supervised by

Dr. Qaisar Shafi

Co-Supervised by

Sir Aadil ur Rehman

**Department of Computer Science**

**National University of Computer and Emerging Sciences**
**Islamabad, Pakistan**


**April, 2025**

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

This document presents a comprehensive project proposal for developing an AI-based student engagement analysis system. The system aims to provide educators with evidence-based insights into student behavior during classroom sessions, enabling more effective teaching strategies and improved learning outcomes.

## 1.1 Motivation

Student engagement plays a crucial role in the learning process and significantly impacts academic performance. Traditional methods of assessing student engagement rely heavily on subjective teacher observations and some feedback forms, which can be inconsistent and may miss subtle behavioral cues. With the advancement of artificial intelligence and computer vision technologies, there is an opportunity to develop automated systems that can objectively analyze student engagement in real-time.

The motivation behind this project is to bridge the gap between technology and education, providing teachers with data-driven insights that can help them adapt their teaching methods to better engage their students. By analyzing both facial expressions and behavioral patterns, our system will offer a more comprehensive understanding of student engagement levels.

## 1.2 Problem Statement

Existing approaches to student engagement analysis remain limited in scope and practicality. Most rely on facial-only datasets, ignoring critical cues such as posture, gestures, and actions. Engagement is often reduced to a simple binary (engaged vs. disengaged), overlooking the nuanced spectrum of student states. Current systems also separate emotional and behavioral analysis, provide only group-level insights, and rarely operate in real time.

The lack of multimodal datasets and tools for individual-level analysis leaves teachers without actionable feedback during live classrooms. This research addresses the gap by developing a multimodal framework that fuses emotion recognition with action detection for real-time, student-specific engagement monitoring.

## 1.3 Proposed Solution/Method

Our proposed solution is a comprehensive AI-based student engagement analysis system that combines computer vision, deep learning, and video processing to provide detailed insights into individual student engagement levels. The system analyzes pre-recorded classroom videos to detect and classify both emotional states and physical actions of students.

The core innovation lies in our multimodal approach that simultaneously processes facial expressions for emotion recognition and full-body analysis for action detection. The system identifies multiple engagement states beyond simple binary classification, including emotions such as happiness, sadness, confusion, boredom, neutrality, and physical indicators like yawning. For behavioral analysis, the system detects actions including hand raising, note-taking, mobile phone usage, looking away, and attentive posture.

The system provides individual-level analysis for each student in the classroom frame, enabling teachers to identify specific students who may need additional attention or support. Real-time visualization displays engagement labels and generates comprehensive reports that summarize engagement patterns over entire lecture sessions. This allows educators to make timely interventions and adjust their teaching strategies based on concrete data.

The solution includes a user-friendly web-based dashboard where teachers can upload video recordings, view real-time analysis results, and access detailed engagement reports. The system generates both immediate feedback during video processing and comprehensive summaries that highlight engagement trends, identify periods of low engagement.

By focusing on physical classroom pre-recorded videos as input, the system ensures practical applicability while maintaining privacy and ethical considerations. The comprehensive analysis of both emotions and actions provides a holistic view of student engagement that goes beyond traditional observation methods.

# Chapter 2

# Preliminary Literature Review

A literature review is a survey of scholarly sources on a specific topic. It provides a critical overview of current knowledge, allowing you to identify relevant theories, methods, and gaps in existing research. This chapter examines recent developments in student engagement analysis using artificial intelligence and computer vision technologies.

## 2.1 Overview of Student Engagement Analysis

Student engagement analysis has evolved significantly with the advancement of machine learning and computer vision technologies. Traditional methods relied heavily on subjective teacher observations and self-reported surveys, which often lacked objectivity and real-time capabilities. Recent research has focused on developing automated systems that can analyze student behavior and emotional states to provide objective engagement assessments.

The field has progressed from simple binary classification approaches to more sophisticated multimodal systems that combine facial expression analysis, behavioral recognition, and temporal modeling. However, significant challenges remain in developing comprehensive solutions that can handle real classroom environments effectively.

## 2.2 Literature Review

### 2.2.1 Facial Expression-Based Approaches

[2] developed a CNN-based model for engagement detection using static facial expressions. Their approach utilizes the FER-2013 dataset with 48×48 grayscale images and applies data augmentation techniques to improve model robustness. The study demon-

strates promising results for emotion recognition but is limited to facial-only analysis and static images, missing important behavioral cues and temporal dynamics.

### 2.2.2 Multimodal Behavioral and Emotional Analysis

[5] proposed a comprehensive approach combining behavioral and emotional features using deep learning models. Their methodology employs ResNet-50 for feature extraction combined with Temporal Convolutional Networks (TCN) for temporal analysis of video clips.

### 2.2.3 Hybrid Network Approaches

[1] developed an improved approach using ResNet and TCN hybrid networks for engagement detection. Their multimodal methodology combines behavioral and emotional analysis with a novel engagement metric that provides more nuanced assessment than binary classification.

### 2.2.4 Real-time Detection Systems

[9] created a facial video-based engagement classification system for real classroom environments. [7] proposed a YOLOv4-based system for binary engagement detection with feedback mechanisms. [8] also demonstrated a practical classroom CV system for binary engagement detection using YOLOv4 with GAN-based augmentation.

### 2.2.5 Advanced Multimodal Systems

[6] introduced the DIPSER dataset for multimodal engagement recognition, combining visual and IMU data. [3] conducted a systematic literature review (SLR) of 113 CV-based engagement studies. [4] presented a meta-analysis of 71 studies, identifying internal and external factors affecting student engagement.

### 2.2.6 Gesture and Gaze-Based Multimodal Approaches

[11] investigated gaze tracking and group interaction detection using video data and process mining. Their approach demonstrated how non-verbal cues such as gaze behaviours can provide insights into collaborative learning interactions, though it relies on rule-based logic and does not capture emotions or gestures.

[10] proposed a vision-based engagement assessment model using deep learning for gesture recognition. Their system incorporates temporal segmentation, feature fusion, and self-attention mechanisms to recognize classroom gestures in real time. While effective in gesture analysis, the method does not account for emotions or group-level engagement.

### 2.2.7 Tabular Analysis of Research Papers

Table 2.1: Comparative analysis of research studies on student engagement detection

| **Paper** | **Methodology** | **Contribution** | **Limitations** |
|---|---|---|---|
| Detecting Student Engagement with CNN and FER [2] | CNN on FER-2013 dataset (48×48 grayscale images with augmentation) | CNN model for engagement detection using static facial expressions | Facial-only analysis, static images, no posture/action data, not real-time |
| Measuring Student Engagement through Behavioral and Emotional Features [5] | ResNet-50 for features + TCN for temporal analysis on video clips | ResNet + TCN hybrid for engagement detection from classroom videos | Basic emotions only, misses confusion/curiosity, ignores gestures and posture |
| Improving Student Engagement Detection with ResNet + TCN Hybrid [1] | Transfer learning (ResNet50/VGG16/InceptionV3) with behavioral gating + emotional fusion | Multimodal approach combining behavior and emotions with new engagement metric | Oversimplified cues, ignores fine actions like note-taking or posture changes |
| Multimodal Engagement Analysis from Facial Videos [9] | ResNet-50 Attention-Net + Affect-Net, multiple classifiers, real classroom data | Facial video-based engagement classification in real classrooms with multi-state labels | Mostly facial expression-based, tested in small controlled lab settings |
| Student Engagement Detection Using YOLOv4 [7] | YOLOv4 with GAN-augmented dataset (1,276 images), binary classification + feedback | YOLOv4 system for binary engagement with weighted scoring and feedback | Binary classification, lacks multimodal feedback, heavy computational needs |

Table 2.1: Comparative analysis of research studies on student engagement detection

| **Paper** | **Methodology** | **Contribution** | **Limitations** |
|---|---|---|---|
| Student Engagement Detection in Classrooms with CV [8] | YOLOv4-based empirical CV system, with GAN data augmentation | Practical classroom demonstration of YOLOv4 for real-time CV monitoring | Binary classification only, limited dataset, no multimodal fusion |
| DIPSER Dataset: Multimodal Engagement Recognition [6] | Visual + IMU fusion (YOLO, MIVIOLC, DeepFace) on Raspberry Pi clusters | Comprehensive multimodal dataset and system for engagement recognition | Ignores posture, gaze, and head movement, not tested in real classrooms |
| SLR of Computer Vision in Student Engagement [3] | Meta-analysis of 113 CV-based engagement studies across modalities | Comprehensive synthesis of CV methods, datasets, and engagement trends | No experimental model; highlights lack of public datasets |
| Dynamic Interaction Between Student Behaviour and Environment [4] | Meta-analysis of 148 effects from 71 studies (93,189 participants) | Identifies emotional, behavioral, and cognitive factors influencing engagement | Heterogeneity, publication bias, limited to higher education |
| Detecting Non-verbal Speech and Gaze Behaviours [11] | Uses video data for gaze tracking and group interaction detection with rule-based logic; process mining to study collaboration patterns | Introduced seven group interaction types, showed how interaction loops improve group learning, and connected findings with collaborative learning theory | Focuses only on gaze, no emotion detection, no student action analysis, rule-based methods may not generalize, limited classroom applicability |
| Vision-Based Gesture Recognition for Engagement Assessment [10] | Uses hand joint tracking with deep learning (temporal segmentation, convolution feature fusion, self-attention), tested on gesture datasets | Provides gesture-based engagement analysis, robust across classroom conditions, supports real-time deployment without special hardware | Focuses only on gestures, no emotion analysis, not fully multimodal, limited to individual actions without group-level context |

## 2.3 Research Gap Analysis

The reviewed literature highlights significant progress in the development of automated systems for student engagement detection. However, several critical gaps remain that must be addressed to enable practical, scalable, and secure deployment in real educational environments:

- **Limited Multimodal Integration:** Most studies rely heavily on either facial expressions [2, 9] or behavioral cues [5], with only a few attempting multimodal fusion [6]. The lack of holistic models limits accuracy, particularly in complex classroom scenarios.

- **Oversimplified Classification:** Several works reduce engagement to binary categories (engaged/disengaged) [7, 8], which fails to capture nuanced states such as confusion, curiosity, and frustration. More fine-grained and context-aware classification systems are needed.

- **Dataset Limitations:** Many models depend on small, controlled datasets (e.g., FER-2013 or GAN-augmented datasets). Systematic reviews [3] emphasize the scarcity of large, diverse, and publicly available engagement datasets, which hinders model generalization across cultures, age groups, and environments.

- **Deployment and Real-world Validation:** While several methods demonstrate promising results in controlled settings, very few are validated in large-scale, real classroom deployments [9]. Real-world testing is critical to assess robustness against environmental noise, occlusion, and classroom dynamics.

- **Neglected Factors:** Meta-analysis studies [4] show that emotional, behavioral, and environmental factors interact in complex ways to influence engagement. Yet most computer vision approaches ignore environmental context (e.g., classroom atmosphere, peer interaction).

- **Security and Ethical Concerns:** Current works overlook data security, privacy, and ethical aspects. No existing system ensures multi-factor authentication for teacher dashboards, encryption of student data, or compliance with privacy regulations.

Addressing these gaps requires multimodal, ethically responsible, and deployable solutions that integrate emotion, behavior, and contextual signals while ensuring security and privacy.

## 2.4 Summary and Research Direction

The literature review demonstrates an evolutionary trajectory in student engagement research — from facial-expression-only methods [2] to multimodal approaches integrating behavior and temporal dynamics [1, 5], and more recently to datasets and frameworks aimed at real-time recognition [6]. Despite this progress, challenges remain regarding multimodal integration, real-world deployment, and the ethical use of engagement analytics.

Our research will build on these findings by:

- Developing a **multimodal engagement detection framework** that combines facial expressions, body posture, and behavioral cues for holistic analysis.
- Introducing a **fine-grained engagement taxonomy** that goes beyond binary classification, capturing states such as attention, confusion, boredom, and curiosity.
- Designing the system for **real-time classroom deployment**, validated with real-world datasets rather than controlled lab conditions.
- Embedding **security mechanisms** such as encrypted data pipelines, access control, and secure visualization dashboards to address privacy and ethical concerns.
- Contributing to the research community by generating or enhancing a **publicly accessible dataset** that incorporates multimodal engagement features in natural classroom settings.

In doing so, this research seeks to bridge the gap between theoretical models and practical applications, advancing both the technical state-of-the-art and the responsible, secure deployment of AI-driven engagement monitoring in education.

# Chapter 3

# Methodology

This chapter outlines the comprehensive methodology for developing the AI-based student engagement analysis system. Our approach combines computer vision, deep learning, and software engineering principles to create a robust and scalable solution.

## 3.1 System Architecture

The proposed system follows a modular architecture consisting of four main components: data preprocessing, AI model pipeline, backend API services, and frontend visualization dashboard. The architecture is designed to handle video input processing, data storage, and user interaction seamlessly.

The data preprocessing module handles video input, frame extraction, and data augmentation. The AI model pipeline consists of two parallel networks: one for emotion recognition and another for action detection. The backend services manage API endpoints, database operations, and report generation, while the frontend provides an intuitive interface for teachers to interact with the system.

## 3.2 Data Collection and Preprocessing

The data collection phase involves creating a custom dataset designed to capture the complexity of real classroom environments. The dataset includes recordings of students in various classroom scenarios, with each participant recorded individually to ensure high-quality annotations. The preprocessing pipeline handles video input processing, frame extraction, and data augmentation to prepare the data for model training. The dataset captures six primary emotions and five key behavioral actions relevant to student engagement analysis.

## 3.3 Model Development

The model development process involves creating two specialized neural networks for comprehensive engagement analysis. The emotion recognition network utilizes convolutional neural networks specifically designed for facial expression analysis, while the action detection network employs object detection and pose estimation techniques to identify student behaviors. Both networks are integrated through a fusion mechanism that combines their outputs to provide comprehensive engagement scores using weighted scoring and temporal smoothing techniques.

## 3.4 System Implementation

The system implementation encompasses both backend and frontend development to create a complete solution. The backend system is implemented using Python frameworks to provide RESTful API services, handle video processing, and manage data storage using both relational and NoSQL databases. The frontend interface is built using modern web technologies to provide an intuitive dashboard for educators, featuring video upload capabilities, real-time processing status, engagement visualization, and comprehensive report generation tools.

## 3.5 Evaluation Methodology

The system evaluation follows a multi-phase approach to ensure comprehensive assessment of performance, usability, and practical applicability. Technical evaluation focuses on model performance using standard metrics, real-time processing capabilities, and computational efficiency. User evaluation involves educators testing the system with real classroom recordings to assess practical utility, accuracy of engagement predictions, and effectiveness in providing actionable insights for improving classroom instruction.

# Chapter 4

# Dataset Description and Tools

This chapter provides detailed information about the dataset creation process, tools and technologies used in the development of the AI-based student engagement analysis system.

## 4.1 Dataset Specification

### 4.1.1 Objective and Scope

The primary objective of our dataset is to capture comprehensive facial emotions and classroom actions that accurately represent student engagement states in real educational environments. Unlike existing datasets that focus solely on facial expressions, our dataset incorporates both emotional and behavioral dimensions of engagement.

The dataset is specifically designed to address the limitations of current engagement analysis systems by providing multimodal annotations that include both fine-grained emotional states and observable classroom behaviors.

### 4.1.2 Participants and Data Collection

The dataset includes recordings of more than 10 student participants in controlled classroom-like environments. Each participant was recorded individually to ensure high-quality video data and precise annotations. The recording sessions were designed to capture natural variations in engagement levels and behavioral patterns.

Participants were asked to simulate various engagement scenarios including attentive listening, note-taking, confusion, boredom, and distraction. This approach ensures comprehensive coverage of engagement states commonly observed in real classroom settings.

### 4.1.3 Annotation Categories

#### 4.1.3.1 Emotional States

The dataset includes annotations for six primary emotional states that are directly relevant to student engagement:

- **Happy:** Positive emotional state indicating satisfaction or enjoyment with the learning content
- **Sad:** Negative emotional state that may indicate frustration or disappointment
- **Confused:** Cognitive state indicating difficulty understanding the material
- **Bored:** Disengagement state characterized by lack of interest
- **Neutral:** Baseline emotional state with no strong emotional indicators
- **Yawning:** Physical indicator of fatigue or disinterest

#### 4.1.3.2 Behavioral Actions

Five key behavioral actions are annotated to capture physical indicators of engagement:

- **Raising Hand:** Active participation behavior indicating desire to contribute
- **Writing Notes:** Active learning behavior showing engagement with content
- **Using Mobile Phone:** Distraction behavior indicating disengagement
- **Looking Away:** Attention diversion behavior suggesting loss of focus
- **Attentive Posture:** Physical positioning indicating focus and engagement

### 4.1.4 Data Format and Organization

The dataset follows standard computer vision dataset formats to ensure compatibility with existing machine learning frameworks. Video files are stored in MP4 format with consistent resolution and frame rates. Individual frames are extracted and stored as high-quality JPEG/PNG images.

Annotations are maintained in structured formats using both CSV and JSON files. The CSV format provides tabular organization suitable for statistical analysis, while JSON format offers hierarchical structure for complex annotation relationships.

The dataset is partitioned into training, validation, and test sets using stratified sampling to ensure balanced representation of all classes across splits. This partitioning strategy ensures robust model training and unbiased evaluation.

## 4.2 Tools and Technologies

### 4.2.1 Programming Languages and Frameworks

The development stack is built using Python as the primary programming language due to its extensive machine learning and computer vision libraries. TensorFlow and PyTorch frameworks are utilized for deep learning model development, providing flexibility and performance optimization.

OpenCV library is employed for computer vision operations including image preprocessing, face detection, and video processing. The combination of these tools provides a comprehensive foundation for developing robust AI models.

### 4.2.2 Backend and API Development

Flask and FastAPI frameworks are used for building RESTful API services that handle video upload, processing requests, and result delivery. These frameworks provide lightweight and scalable solutions for web service development.

### 4.2.3 Database Systems

A hybrid database approach is implemented using both relational and non-relational database systems. MySQL is used for storing structured data including user information, session metadata, engagement scores, and system configuration parameters.

MongoDB handles unstructured data storage including raw video files, processing logs, detailed annotation data, and temporary files generated during analysis. This dual-database approach optimizes storage efficiency and query performance.

### 4.2.4 Frontend Development

React.js framework provides the foundation for building responsive and interactive user interfaces. The component-based architecture ensures maintainable and scalable frontend development.

Chart.js library enables dynamic data visualization including real-time engagement plots, statistical summaries, and comparative analysis charts. The visualization components are designed to present complex engagement data in intuitive and actionable formats.

### 4.2.5 Development and Collaboration Tools

Visual Studio Code serves as the primary integrated development environment, providing comprehensive support for Python, JavaScript, and web development. The development workflow incorporates modern software engineering practices including version control and collaborative development.

Git and GitHub provide version control and collaborative development capabilities, enabling efficient team coordination and code management. The repository structure supports modular development and maintains clear separation between different system components.

Jupyter Notebook and Google Colab platforms are utilized for experimental development, model prototyping, and data analysis tasks. These tools provide interactive development environments that facilitate rapid iteration and hypothesis testing.

### 4.2.6 Deployment and Testing

Docker containerization ensures consistent deployment across different environments and simplifies system deployment and scaling. Container orchestration enables efficient resource management and system monitoring.

Cloud deployment platforms including AWS and Azure provide scalable infrastructure for production deployment. These platforms offer comprehensive services for compute, storage, and networking requirements.

Streamlit framework is employed for rapid prototyping and demonstration purposes, enabling quick model testing and user interface development during the iterative development process.

# Chapter 5

# Project Timeline and Work Division

This chapter outlines the comprehensive project timeline and work distribution strategy for the AI-based student engagement analysis system development.

## 5.1 Project Timeline

The project is structured into four distinct iterations, each spanning approximately two months, with specific deliverables and milestones to ensure systematic progress and quality outcomes.

### 5.1.1 Iteration 1: Data Preparation and Model Foundation (September - October)

The first iteration focuses on establishing the foundational components of the system, with primary emphasis on data collection, preprocessing, and initial model development.

#### 5.1.1.1 Key Deliverables

- Complete dataset creation with comprehensive video recordings of student participants
- Systematic annotation process implementation for both emotional states and behavioral actions
- Data preprocessing pipeline development including frame extraction and augmentation
- Initial model architecture design for both emotion recognition and action detection

- Baseline model training with preliminary performance evaluation

#### 5.1.1.2 Technical Milestones

Data collection protocols will be established and implemented, ensuring consistent recording quality and comprehensive coverage of engagement scenarios. The annotation framework will be designed to maintain consistency across different annotators while capturing the nuanced aspects of student engagement.

Initial model architectures will be implemented using transfer learning approaches to leverage existing computer vision models. Preliminary training will establish baseline performance metrics and identify areas requiring optimization.

### 5.1.2 Iteration 2: Backend Infrastructure and API Development (November - December)

The second iteration concentrates on building robust backend infrastructure and developing comprehensive API services to support the AI pipeline and user interactions.

#### 5.1.2.1 Key Deliverables

- Backend API development using Flask/FastAPI frameworks
- Database schema design and implementation for both structured and unstructured data
- Video upload and processing pipeline integration
- Model inference optimization for real-time performance
- API endpoint testing and documentation

#### 5.1.2.2 Technical Milestones

RESTful API services will be developed to handle video upload, processing queue management, and result delivery. The database architecture will be optimized to support efficient storage and retrieval of video data, annotations, and analysis results.

Integration between the AI models and backend services will be implemented with focus on performance optimization and error handling. Comprehensive API testing will ensure reliability and scalability of the backend infrastructure.

### 5.1.3 Iteration 3: Frontend Development and System Integration (January - February)

The third iteration focuses on developing the user interface and achieving complete system integration with emphasis on user experience and functionality.

#### 5.1.3.1 Key Deliverables

- Teacher portal development with authentication and user management
- Dashboard creation for engagement visualization and analysis
- Report generation module with comprehensive analytics
- Real-time visualization components for engagement monitoring
- Complete system integration testing

#### 5.1.3.2 Technical Milestones

The frontend interface will be developed using React.js with focus on responsive design and intuitive user experience. Dashboard components will provide comprehensive visualization of engagement data using Chart.js for dynamic charts and graphs.

Integration testing will ensure seamless communication between frontend, backend, and AI components. User interface testing will validate usability and functionality from the educator's perspective.

### 5.1.4 Iteration 4: Testing, Deployment, and Refinement (March - April)

The final iteration concentrates on comprehensive system testing, deployment preparation, and performance optimization based on real-world usage scenarios.

#### 5.1.4.1 Key Deliverables

- Comprehensive system testing with recorded lecture videos
- Performance optimization and scalability improvements
- Cloud deployment implementation using Docker containers

- User acceptance testing with educator feedback
- Final system documentation and user guides

#### 5.1.4.2 Technical Milestones

Extensive testing will be conducted using real classroom recordings to validate system accuracy and practical utility. Performance optimization will focus on processing speed, resource utilization, and system responsiveness.

Cloud deployment will be implemented using containerization technologies to ensure scalable and reliable system access. Final documentation will include technical specifications, user manuals, and deployment guides.

## 5.2 Work Division and Team Responsibilities

The project team consists of three members, each with specialized roles and responsibilities aligned with their expertise and the project requirements.

### 5.2.1 AI and Computer Vision Development

Bilal will handle most of the model development work, focusing on implementing and training both the emotion recognition and action detection networks using convolutional neural networks and deep learning techniques. He will be responsible for model architecture design, transfer learning implementation, performance optimization, and integration with the backend API services. Abdurrahman and Sohaib will contribute to the AI development by collecting and annotating the custom dataset, ensuring data quality and consistency across the six emotional states and five behavioral actions. They will assist with data preprocessing, augmentation strategies, and validation processes to support robust model training.

### 5.2.2 Frontend Development and User Interface

Abdurrahman will lead the frontend development using React.js to create a responsive and intuitive user interface for educators. His work includes implementing the dashboard for engagement visualization, integrating Chart.js for dynamic data representation, developing the video upload interface with real-time processing status monitoring, and creating comprehensive report generation capabilities. He will focus on user experience design, ensuring the interface remains accessible to educators with varying technical expertise, and implementing responsive design principles for multiple device compatibility.

### 5.2.3 Backend Development and System Infrastructure

Sohaib will handle most of the backend development using Python frameworks (Flask or FastAPI) to create robust RESTful API services. His responsibilities include designing and implementing the database architecture using both MySQL for structured data and MongoDB for unstructured data, developing the video processing pipeline with queue management capabilities, implementing user authentication and session management systems, and creating backend services for report generation. He will also focus on system scalability, security implementation including multi-factor authentication and data encryption, and deployment infrastructure setup with performance monitoring capabilities.

## 5.3 Collaboration and Quality Assurance

The team follows collaborative development practices using Git and GitHub for version control and code review processes. Regular team meetings ensure coordination and progress tracking across all development streams.

Quality assurance procedures include code reviews, unit testing, integration testing, and user acceptance testing. Documentation standards are maintained throughout the development process to ensure system maintainability and knowledge transfer.

# Chapter 6

# Conclusions and Future Work

This research successfully developed an AI-based student engagement analysis system that combines computer vision and deep learning techniques to monitor classroom dynamics. The system demonstrates the potential of multimodal analysis by integrating emotion recognition and action detection to provide comprehensive engagement insights for educators.

The implemented solution addresses key challenges in educational technology by offering recorded video processing capabilities, intuitive visualization dashboards, and actionable reporting tools. Through the development of custom datasets and specialized neural networks, the system achieves reliable detection of six emotional states and five behavioral patterns relevant to student engagement assessment.

Future work should focus on expanding the dataset to include more diverse classroom environments and cultural contexts, improving model robustness through advanced deep learning techniques, and incorporating additional behavioral indicators such as eye tracking and speech analysis. Furthermore, longitudinal studies examining the impact of engagement feedback on teaching effectiveness would provide valuable insights into the practical benefits of such systems in educational settings.

# Bibliography


[1] Ali Abedi and Shehroz S Khan. Improving state-of-the-art in detecting student engagement with resnet and tcn hybrid network. In *2021 18th Conference on Robots and Vision (CRV)*, pages 151–157. IEEE, 2021.

[2] Nuha Alruwais and Mohammed Zakariah. Detecting student engagement with convolutional neural networks and facial expression recognition. *Traitement du Signal*, 42(2):229–240, 2025.

[3] Mohamed Garbal, Siham El Janati, and Lahcen El Ghadraoui. Student's engagement detection based on computer vision: systematic literature review. *Education and Information Technologies*, 30(3):2845–2891, 2025.

[4] Xiaojing Li and En Xue. Dynamic interaction between student learning behaviour and environment: Meta-analysis of student engagement in higher education. *Behaviour & Information Technology*, 42(4):471–497, 2023.

[5] Nasir Mahmood, Sohail Masood Bhatti, Hussain Dawood, Manas Ranjan Pradhan, and Haseeb Ahmad. Measuring student engagement through behavioral and emotional features using deep-learning models. *Algorithms*, 17(10):458, 2024.

[6] L. Marquez-Carpintero, S. Suescun-Ferrandiz, C. Lorenzo Álvarez, J. Fernandez-Herrero, D. Viejo, R. Roig-Vila, and M. Cazorla. Dipser: A dataset for in-person student engagement recognition in the wild. *arXiv preprint arXiv:2502.20209*, 2025.

[7] A. S. Pillai. Student engagement detection in classrooms through computer vision and deep learning: A novel approach using yolov4. *Sage Science Review of Educational Technology*, 5(1):87–97, 2023.

[8] Aishwarya S. Pillai and N. Radhika. Student engagement detection in classrooms through computer vision and deep learning. In *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, pages 1–6. IEEE, 2022.

[9] Ömer Sümer, Patricia Goldberg, Sidney D'Mello, Peter Gerjets, Ulrich Trautwein, and Enkelejda Kasneci. Multimodal engagement analysis from facial videos in the classroom. *IEEE Transactions on Affective Computing*, 14(2):1012–1027, 2021.

[10] Xin Zhang and Li Wang. A vision-based gesture recognition and student engagement assessment model for interactive educational environments. *IEEE Transactions on Learning Technologies*, 2025.

[11] Qi Zhou, Wannapon Suraworachet, and Mutlu Cukurova. Detecting non-verbal speech and gaze behaviours with multimodal data and computer vision to interpret effective collaborative learning interactions. *Computers & Education*, 197:104712, 2023.