AI Based Cognitive Load Estimation via Natural Activity Monitoring

Final Year Project Report Session

2022-2026

A literature review submitted in partial fulfillment of the
BS in Computer Science
University of the Punjab

Department of Computer Science

University of the Punjab Lahore
Pakistan

Project Detail

Type (Nature of project)	[ ] **D**evelopment		[ ] **R&D**	
Area of specialization				
Project Group Members				
Sr.#	Reg. #	Student Name	Email ID	*Signature
(i)				
(ii)				
(iii)				

*The candidates confirm that the work submitted is their own and appropriate credit has been given where reference has been made to work of others ,

Plagiarism Free Certificate

This is to certify that, I am ___________________ S/D/o___________________________, group leader of FYP under registration no /__________________/ at Computer Science Department, University of the Punjab, Lahore, Pakistan. I declare that my FYP proposal is checked by my supervisor, and the similarity index is ___% that is less than 20%, an acceptable limit by HEC. Report is attached herewith as Appendix A.

Date: _____________ Name of Group Leader: _________________________ Signature:

Name of Supervisor: ________________________ Co-Supervisor (if any): _____________________

Designation: __________________________ Designation: _______________________

Signature: __________________________ Signature: _______________________

Abstract

This project aims at the design and implementation of an artificial intelligence (AI) based cognitive load estimation system to monitor the natural user activity to assess the level of mental engagement during study/work sessions. In contrast to the usage of traditional digital wellbeing applications that serve generic usage statistics, this system uses real-time behavioral signals from webcam, keyboard and mouse interaction to determine if a user is **overloaded, optimally engaged or disengaged**.

Using combinations of facial expressions, such as frowning, yawning or raised eyebrows, and eye gaze tracking and blink rate, the system reads signs of fatigue, confusion or distraction. Keyboard typing patterns, frequency of pauses, corrections of errors and behaviors of movement of mouse are also recorded to get more subtle signs of cognitive effort. These features are then run through machine learning models to classify the user's cognitive load into three categories: Low, Medium and High.

The solution is intended to give individual feedback as well as aggregated information. For individuals, real-time popups suggest taking breaks in case of cognitive overload or increasing focus on the periods of disengagement. A web-based application will be constantly pulling activities in the background, whereas a dashboard will give **detailed logs, visualizations and engagement metrics**. This project is a practical meeting of the fields of computer vision, machine learning and human computer interaction that contributes to the goal of smarter productivity tools and more personalized learning environments.

Executive Summary

The project, AI-Based Cognitive Load Estimation via Natural Activity Monitoring, is aimed at creating a system that detects whether users are mentally overloaded, engaged or disengaged during digital activities. Traditional digital wellbeing tools only measure screen time, but this project aims at real-time behavioral monitoring for a better and more accurate assessment of mental effort.

The system combines several different modalities: computer vision to analyze the **facial cues, gaze and blinks** from the webcam; keyboard and mouse interaction patterns for pauses, corrections and hesitation; and for collaborative situations. Extracted features will be fed to machine learning models like Random Forests for simpler classification or CNN+LSTM model architectures for temporal signals.

Some of the key sub-tasks are activity monitoring, feature extraction, cognitive load classification and feedback provision through web-based app and dashboard. The success of the project will be tested with an accuracy test on known data sets such as **DAiSEE** as well as a demonstration of real-time feedback and visualization. Evidence of achievement will include experimental results, system performance control and a working prototype that tracks and reports user load well.

TABLE OF CONTENTS

List of Tables

CHAPTER 1: INTRODUCTION

The project, AI-Based Cognitive Load Estimation via Natural Activity Monitoring, is aimed at creating a system that detects whether users are mentally overloaded, engaged, or disengaged during digital activities. Traditional digital wellbeing tools only measure screen time, but this project aims at real-time behavioral monitoring for a better and more accurate assessment of **mental effort**.

The system combines several different modalities: computer vision to analyze facial cues, gaze, and blinks from the webcam; keyboard and mouse interaction patterns for **pauses, corrections, and hesitation**. Extracted features will be fed to machine learning models like Random Forests for simpler classification or CNN+LSTM model architectures for temporal signals.

Some of the key sub-tasks are activity monitoring, feature extraction, cognitive load classification, and feedback provision through a web dashboard. The success of the project will be tested with an accuracy test on known datasets such as DAiSEE and CLT as well as a demonstration of real-time feedback and visualization. Evidence of achievement will include experimental results, system performance control, and a working prototype that tracks and reports **user load effectively**.

1.1 Purpose of this Document

This literature review document serves as a comprehensive examination of existing research in the domain of cognitive load estimation and natural activity monitoring. The primary purpose is to establish a theoretical foundation for the proposed AI-Based Cognitive Load Estimation system by analyzing current methodologies, identifying research gaps, and understanding the state-of-the-art approaches in this field.

Unlike traditional digital wellbeing tools that provide generic usage statistics, there is limited research on systems that **combine multimodal behavioral signals** (facial expressions, gaze tracking, keyboard/mouse interactions) for real-time cognitive load assessment. This document explores how existing studies have addressed cognitive load measurement, what techniques have been employed, and where opportunities exist for innovation.

The review will examine various aspects including physiological signal processing, computer vision techniques for behavioral analysis, machine learning models for load classification, and real-time feedback mechanisms. By synthesizing findings from multiple research domains, including educational technology, human-computer interaction, and cognitive psychology, this document aims to justify the proposed approach and demonstrate its novelty in creating a **practical, non-intrusive system** for cognitive load estimation.

1.2 Intended Audience

This literature review is intended for multiple stakeholders involved in or interested in cognitive load estimation research and its applications:

Academic Supervisors and Faculty: This document provides a comprehensive overview of the research landscape, helping evaluate the theoretical soundness and novelty of the proposed project approach.

Students and Researchers: Those working on similar projects in cognitive science, educational technology, or human-computer interaction will find valuable insights into current methodologies, datasets, and evaluation metrics used in cognitive load research.

Educators and Learning Designers: Understanding how cognitive load can be measured automatically helps in designing better learning environments and interventions for student engagement and wellbeing.

Developers and Product Managers: Those interested in building productivity tools or digital wellbeing applications will gain insights into technical approaches for implementing cognitive load monitoring features.

Healthcare and Workplace Wellness Professionals: Understanding non-intrusive methods for monitoring mental fatigue and engagement can inform strategies for preventing burnout and improving workplace wellbeing.

1.3 Definitions, Acronyms, and Abbreviations

1.3.1 Definitions

Cognitive Load: The amount of mental effort being used in working memory during learning or task performance. It represents the total amount of mental activity imposed on working memory at an instance in time.

Natural Activity Monitoring: The process of observing and analyzing user behaviors through naturally occurring interactions with digital devices (webcam, keyboard, mouse) without requiring additional sensors or specialized equipment.

Multimodal Fusion: The technique of combining information from multiple input sources or modalities (visual, behavioral, temporal) to make more accurate predictions or classifications.

Facial Action Units: Anatomically based system for describing all visually discernible facial movements, used to detect micro-expressions that indicate cognitive states.

Gaze Tracking: The process of measuring either the point of gaze or the motion of an eye relative to the head, used to infer attention and focus levels.

Temporal Patterns: Time-based sequences of behaviors or signals that reveal trends and changes in cognitive state over duration of activity.

1.3.2 Acronyms and Abbreviations

AI: Artificial Intelligence
CNN: Convolutional Neural Network
LSTM: Long Short-Term Memory
ML: Machine Learning
HCI: Human-Computer Interaction
DAiSEE: Dataset for Affective States in E-Environments
CLT: Cognitive Load Theory
SVM: Support Vector Machine
RF: Random Forest
EEG: Electroencephalography
ECG: Electrocardiography
API: Application Programming Interface
UI: User Interface
FYP: Final Year Project

CHAPTER 2: PROJECT VISION

This chapter entails our project vision. Any good project is only as good as its vision. We aim to outline a comprehensive vision where we define the problem domain, goals, and scope of the project. Our plan is to be ambitious while setting realistic goals that we are capable of achieving within the constraints of available resources and time.

2.1 Problem Domain Overview

The digital age has transformed how people work, learn, and interact. With remote work and online education becoming increasingly prevalent, individuals spend extended hours engaged with digital devices. However, current digital wellbeing tools provide limited insight into users' mental states, focusing primarily on screen time metrics rather than the quality of engagement or cognitive effort involved.

Cognitive overload, a state where mental demands exceed working memory capacity, can lead to decreased productivity, learning difficulties, and mental fatigue. Conversely, disengagement indicates insufficient challenge or attention, which also hampers effective learning and work performance. Traditional methods of assessing cognitive load, such as self-reporting or physiological sensors (EEG, ECG), are either subjective or intrusive, making them impractical for everyday use.

Our project operates in the intersection of computer vision, behavioral analysis, and machine learning. By leveraging naturally available signals from webcams, keyboards, and mouse interactions, we aim to create a non-intrusive system that continuously monitors and classifies cognitive load states. This approach differs from existing research by focusing on practical implementation using multimodal fusion of accessible input sources, rather than relying on specialized hardware or controlled laboratory settings.

2.2 Problem Statement

Current digital wellbeing and productivity tools lack the capability to accurately assess users' cognitive load in real-time using non-intrusive methods. While physiological monitoring provides accurate measurements, it requires specialized equipment that is impractical for everyday use. Existing behavioral monitoring systems focus on single modalities and fail to provide actionable feedback that could help users optimize their work or study patterns.

There is a need for an intelligent system that can automatically detect cognitive overload, optimal engagement, and disengagement by analyzing natural user activities during digital work sessions and provide timely interventions to enhance **productivity** and **prevent mental fatigue.**

2.3 Problem Elaboration

The problem can be broken down into several interconnected challenges:

1. Detection Accuracy: Cognitive load is a complex internal state that manifests through subtle behavioral cues. Accurately detecting it from external observations requires sophisticated pattern recognition that can distinguish between different cognitive states while accounting for individual variations in behavior.

2. Real-time Processing: For the system to be practical, it must process multimodal inputs and provide feedback in real-time without causing significant computational overhead or system lag that would interfere with the user's primary tasks.

3. Privacy and Intrusiveness: Webcam-based monitoring raises valid privacy concerns. The system must be designed with privacy-first principles, processing data locally when possible and providing users with complete control over monitoring features.

4. Multimodal Integration: Different modalities (facial expressions, gaze patterns, typing behavior, mouse movements) provide complementary information. Effectively fusing these signals while handling missing or noisy data from any single modality presents a technical challenge.

5. Actionable Feedback: Beyond detection, the system must translate cognitive load assessments into meaningful, timely interventions, such as suggesting breaks during overload or recommending more challenging tasks during disengagement, without being disruptive.

6. Generalization: The system should work across different users, tasks, and contexts without requiring extensive per-user calibration, while still being able to adapt to individual behavioral patterns over time.

2.4 Goals and Objectives

The overarching goal of this project is to design, develop, and validate an AI-based system for real-time cognitive load estimation using natural activity monitoring. The specific objectives are:

1. Data Collection and Preprocessing: To collect and preprocess multimodal data including facial cues, gaze patterns, typing behavior, and mouse activity from both public datasets (DAiSEE, CLT) and pilot studies for training and validation purposes.

2. Feature Engineering: To identify and extract salient features from visual and behavioral signals that correlate strongly with low, medium, and high cognitive load states, including temporal patterns that emerge over time.

3. Model Development: To build and evaluate multiple machine learning models (Random Forest, CNN+LSTM, multimodal fusion architectures) for cognitive load classification, comparing their performance in terms of accuracy, real-time processing capability, and robustness to missing data.

4. System Implementation: To develop a functional application (desktop or browser-based) for real-time activity capture and monitoring that operates efficiently in the background without disrupting user workflow.

5. Dashboard Development: To create a web-based dashboard that presents cognitive load trends, engagement metrics, and historical data in an intuitive, actionable format for individual users and educators.

6. Validation and Testing: To validate the system using benchmark datasets and real-time trials, demonstrating its accuracy, usability, and practical value in real-world educational or workplace scenarios.

7. Feedback Mechanism: To implement and evaluate an intelligent feedback system that provides timely notifications and suggestions based on detected cognitive states, helping users optimize their productivity and wellbeing.

2.5 Project Scope

The scope of this project encompasses the design, development, and evaluation of a cognitive load estimation system based on natural activity monitoring. The project will leverage facial behavior analysis, gaze tracking, and keyboard/mouse interaction analysis to classify users into three cognitive load categories: Low, Medium, and High.

2.5.1 Project Goals

- To construct a multimodal system combining visual signals (from webcam) and behavioral signals (from keyboard/mouse) for cognitive load estimation
- To implement machine learning models capable of real-time classification with acceptable latency

- To provide feedback and visualization through user-friendly interfaces accessible via web browsers

2.5.2 Deliverables

- A monitoring application (browser-based or desktop, depending on technical feasibility) for continuous activity capture from webcam, keyboard, and mouse
- A web dashboard for presenting results, historical logs, and engagement trends with interactive visualizations
- Trained AI models capable of classifying cognitive load levels with documented performance metrics
- Comprehensive project documentation including literature review, methodology, results analysis, and user manuals
- Source code repository with clear documentation for future development and research

2.5.3 Features and Functions

- Real-time monitoring of user activities through webcam (facial expressions, gaze, blinks) and input devices (keyboard, mouse)
- Three-level classification of cognitive load: Low (disengaged), Medium (optimally engaged), and High (overloaded)
- Intelligent feedback notifications (e.g., "Consider taking a break" during overload periods or "Try a more challenging task" during disengagement)
- Session logging with timestamped cognitive load assessments
- Visual analytics dashboard showing cognitive load trends over time, daily/weekly patterns, and productivity insights
- Privacy controls allowing users to pause monitoring, delete data, and configure feedback preferences

2.5.4 Constraints

Timeline Constraints: The project must align with the official FYP timeline, with key milestones including proposal defense (September 2025), mid-term report (November 2025), and final documentation (May 2026).

Computational Resources: Deep learning models, particularly CNN+LSTM architectures, require significant computational resources. Training will depend on available GPU access, which may limit model complexity or training duration.

Data Availability: The project relies on public datasets (DAiSEE, CLT) for initial training and validation. Additional pilot data collection will be limited by participant availability and consent procedures.

Technical Limitations: Browser-based implementations may face limitations in accessing system-level input monitoring. If browser APIs prove insufficient, a desktop application may be necessary, which would increase development complexity.

Privacy and Ethics: All monitoring features must comply with privacy regulations and ethical guidelines. User consent, data security, and transparency in data processing are mandatory requirements that may constrain certain features.

Scope Boundaries: The system will focus on cognitive load estimation during digital work/study sessions. It will not diagnose medical conditions, replace professional psychological assessment, or monitor activities outside of user-initiated sessions.

2.6 Sustainable Development Goal (SDG)

This project aligns with several United Nations Sustainable Development Goals, contributing to global efforts for creating equitable, inclusive, and sustainable societies.

2.6.1 Quality Education (SDG 4)

By providing educators with objective insights into student engagement and cognitive load, this system supports SDG 4's target of ensuring inclusive and **equitable quality education**. The system can help identify students who may be struggling (high cognitive load) or disengaged (low cognitive load), enabling timely interventions. In remote and hybrid learning environments, where direct observation is limited, automated cognitive load monitoring can help maintain educational quality and personalize learning experiences. This contributes to creating more **effective learning environments** that adapt to individual student needs, ultimately improving learning outcomes and reducing dropout rates.

2.6.2 Good Health and Well-being (SDG 3)

Mental health and cognitive wellbeing are essential components of SDG 3. By detecting cognitive overload and prompting users to take breaks, the system actively contributes to preventing mental fatigue, burnout, and associated health issues. In an era where digital overwork is increasingly common, tools that promote **healthy work patterns** and prevent cognitive **exhaustion** align directly with global health objectives. The system's feedback mechanism encourages sustainable work practices, helping users maintain a healthy balance between productivity and rest.

2.6.3 Industry, Innovation and Infrastructure (SDG 9)

This project exemplifies innovation in human-computer interaction and AI applications. By developing novel multimodal fusion techniques and demonstrating practical implementation of cognitive load monitoring, the project contributes to advancing technological infrastructure for the **digital age**. The open-source nature of the project (through code repository) enables knowledge sharing and further innovation by researchers and developers globally. This aligns with SDG 9's emphasis on fostering innovation, building resilient infrastructure, and promoting inclusive and sustainable industrialization. The system's potential applications extend beyond education to workplace productivity, healthcare monitoring, and human factors engineering, supporting diverse industries in optimizing human performance while prioritizing **wellbeing**.

CHAPTER 3: LITERATURE REVIEW / RELATED WORK

3.1 Definitions, Acronyms, and Abbreviations

Student Engagement Detection (SED): Automatically identifying whether students are attentive, engaged, or disengaged during e-learning sessions.

Cognitive Load: Mental effort applied to process information during learning tasks, often measured using behavioral, physiological, or interaction signals.

Datasets:

- **DAiSEE:** Dataset for Affective States in E-learning Environments.

Deep Learning Techniques:

- **CNN (Convolutional Neural Networks):** Spatial feature extraction from images or video frames.
- **RNN (Recurrent Neural Networks):** Models sequential/temporal dependencies.
- **LSTM (Long Short-Term Memory):** A type of RNN capable of capturing long-term dependencies.
- **TCN (Temporal Convolutional Network):** Models temporal patterns in sequences.
- **ST-GCN (Spatial-Temporal Graph Convolutional Network):** Captures temporal and spatial relationships in structured data, e.g., facial landmarks.
- **Transformer:** Attention-based architecture suitable for sequence modeling and multimodal fusion.

Other Acronyms:

- **ITL - Human-in-the-Loop:** A methodology where human input is incorporated into machine learning workflows, often for annotation, validation, or improving model performance.
- **RBM - Restricted Boltzmann Machine:** A type of stochastic neural network that learns probability distributions over input data and is used for feature extraction and dimensionality reduction.
- **DAiSEE - Dataset for Affective States in E-learning Environments:** A benchmark dataset for studying student engagement, including labels for engagement levels in video recordings of learners.
- **MMR - Maximal Marginal Relevance:** An algorithm used in information retrieval and text summarization that balances relevance of selected items with diversity, reducing redundancy in results.

3.2 Detailed Literature Review

This section summarizes previous research grouped by methodology, dataset, problem focus, and timeline. All works cited as [*3.3.1.x*] correspond to the Literature Review Summary Table (*3.3.1*) for cross-reference.

3.2.1 Related Research Work 1 - Engagement Detection

3.2.1.1 Summary of the research item

Engagement detection aims to automatically identify student attention and involvement in e-learning or virtual environments. Methods rely on:

- Facial expression analysis (frame-level CNNs, graph-based models)
- Temporal modeling (RNNs, LSTMs, TCNs)
- Graph-based techniques (ST-GCN for landmark features)
- Multimodal fusion (facial + behavioral cues)
- Attention-based transformers
- Human-in-the-loop (HITL) annotation to improve labeling consistency

Representative works:

- CNN-based DAiSEE classification [*3.3.1.9*]
- Hybrid CNN + LSTM EfficientNet models [*3.3.1.10, 3.3.1.7*]
- Graph-based landmark CNNs [*3.3.1.2, 3.3.1.15*]
- Transformer-based models [*3.3.1.14*]
- HITL annotation frameworks [*3.3.1.1*]
- Multimodal fusion combining facial and behavioral cues [*3.3.1.8, 3.3.1.12*]

3.2.1.2 Critical Analysis

- CNN models [*3.3.1.9*] excel at frame-level facial engagement but ignore temporal dependencies.
- Hybrid CNN-RNN / LSTM models [*3.3.1.10, 3.3.1.7*] model temporal changes but are computationally heavy.
- Graph CNN / ST-GCN models *[3.3.1.2, 3.3.1.15*] capture micro-expressions but are sensitive to landmark errors.
- Transformer-based models [*3.3.1.14*] provide high accuracy but need large datasets.
- HITL annotation [*3.3.1.1*] improves labeling but does not enhance model inference.
- Multimodal fusion [*3.3.1.8, 3.3.1.12*] improves accuracy but adds complexity.

3.2.1.3 Relationship to Proposed Research Work

Our work combines facial features, temporal modeling, and behavioral signals for real-time engagement detection. We benchmark against DAiSEE, leverage HITL strategies for dataset reliability, and focus on practical models balancing **accuracy and efficiency**.

3.2.2 Related Research Work 2 - Cognitive Load / Disengagement Detection

3.2.2.1 Summary of the research item

Cognitive load estimation detects mental effort via eye gaze, blink rates, and interaction patterns, while disengagement detection identifies attention lapses. Methods include:

- TCN-based anomaly detection [*3.3.1.5*]
- Hybrid EfficientNet + LSTM temporal models [*3.3.1.7, 3.3.1.10*]
- Multimodal fusion of facial and behavioral cues [*3.3.1.8, 3.3.1.12*]

3.2.2.2 Representative works

- TCN Autoencoder for Disengagement Detection [*3.3.1.5]*
- Hybrid EfficientNet + Temporal Models [*3.3.1.7*]
- EfficientNet-RNN Hybrid Models [*3.3.1.10*]
- Behavioral + Facial Multimodal Fusion [*3.3.1.8*]
- Facial + Behavioral Fusion for Engagement [*3.3.1.12*]

3.2.2.3 Critical analysis of the research item

- TCN anomaly detection [*3.3.1.5*] works for binary engaged/disengaged detection but not multi-class.
- Hybrid CNN-RNN / LSTM [*3.3.1.10, 3.3.1.7*] robustly models temporal data but is computationally heavy.
- Multimodal fusion [*3.3.1.8, 3.3.1.12*] improves accuracy but increases model complexity.

3.2.2.4 Relationship to the proposed research work

Our research incorporates elements of cognitive load detection by using multimodal behavioral cues such as gaze, blinking, and interaction signals, but extends beyond prior work by focusing on **real-time and scalable models**. Unlike heavy hybrid CNN-LSTM architectures, our approach prioritizes computational efficiency and applicability in real-world e-learning and productivity platforms.

3.2.3 Related Research Work 3 - Video-based Facial Recognition

3.2.3.1 Summary of the research item

Video-based facial recognition methods extend static CNN-based engagement detection to dynamic modeling of facial features across time. Instead of relying on individual frames, these approaches analyze video clips to capture micro-expressions, gaze changes, and temporal facial dynamics. They often integrate spatiotemporal CNNs or CNN-RNN hybrids for sequence-level representation.

Representative works include:

- Large-Scale Engagement Dataset & Baselines [*3.3.1.4*]
- Deep Facial Spatiotemporal Network [*3.3.1.11*]
- Landmark-based Graph CNN [*3.3.1.15*]

3.2.3.2 Critical Analysis

- Video-based methods [*3.3.1.4, 3.3.1.11*] improve robustness by modeling temporal dynamics, but they are computationally expensive for real-time use.
- Landmark-driven video recognition [*3.3.1.15*] captures fine-grained movements but is sensitive to tracking errors in low-quality video.
- Sequence-level recognition enables more naturalistic engagement detection but requires high-quality continuous data and large annotated video datasets.

3.2.3.3 Relationship to the proposed research work

Our work builds upon these advances by combining video-based facial recognition with behavioral signals (keyboard, mouse, and gaze) for multimodal analysis. Unlike prior works that rely solely on video clips, our system integrates video-driven facial dynamics into a real-time multimodal framework designed for practical use in **e-learning and productivity tools**.

3.2.4 By Problem Focus

- **Engagement Detection:** Frame-level or temporal detection [*3.3.1.9, 3.3.1.10, 3.3.1.7*]
- **Disengagement Detection:** Identify low-attention behaviors [*3.3.1.5*]
- **Facial Cues:** Facial expressions or landmarks only [*3.3.1.2, 3.3.1.15*]
- **Multimodal Fusion:** Combine facial, behavioral, and interaction signals [*3.3.1.8, 3.3.1.12*]

3.2.5 By Timeline

- **Early Baselines:** Frame-level CNNs [*3.3.1.9, 3.3.1.13*]
- **Spatiotemporal Models:** CNN-LSTM, TCNs, hybrid EfficientNet [*3.3.1.7, 3.3.1.10*]
- **Graph / Landmark Methods:** ST-GCN, bimodal CNNs [*3.3.1.2, 3.3.1.15*]
- **Modern Approaches:** Transformers, HITL annotation, multimodal fusion [*3.3.1.1, 3.3.1.8, 3.3.1.12, 3.3.1.14*]

3.3 Literature Review Summary Table

Table 3.3.1: *This table contains detailed analysis of the existing research studies for Engagement Detection and Cognitive Load Estimation*

No.	Name	Reference (Heading)	Method	Output	Link
1	Human-in-the-Loop Annotation	*[3.3.1.1]*	HITL labeling strategies	Improved annotation quality	arXiv:2502.07404
2	ST-GCN Facial Landmark Engagement	*[3.3.1.2]*	ST-GCN	Micro-expression detection	IEEE:11029003
3	Affect-driven Ordinal Engagement	*[3.3.1.3]*	Ordinal sequence modeling	Engagement level prediction	arXiv:2403.17175
4	Large-Scale Engagement Dataset & Baselines	*[3.3.1.4]*	CNN / RNN	Benchmark results	arXiv:2302.00431
5	TCN Autoencoder for Disengagement	*[3.3.1.5]*	TCN anomaly detection	Binary disengagement detection	IEEE:9893134
6	Multimedia Tools & Applications Survey	*[3.3.1.6]*	Survey	Overview of engagement methods	Springer
7	Hybrid EfficientNet + Temporal Models	*[3.3.1.7]*	CNN + LSTM	Temporal cognitive load modeling	arXiv:2211.06870
8	Behavioral + Facial Multimodal Fusion	*[3.3.1.8]*	CNN + Behavioral Features	Improved engagement accuracy	MDPI
9	DAiSEE CNN Baseline	*[3.3.1.9]*	CNN frame-level	Engagement classification	Thesai
10	EfficientNet-RNN Hybrid Models	*[3.3.1.10]*	CNN + RNN	Temporal engagement prediction	ResearchGate
11	Deep Facial Spatiotemporal Network	*[3.3.1.11]*	CNN-RNN	Facial temporal modeling	ResearchGate
12	Facial + Behavioral Fusion for Engagement	*[3.3.1.12]*	Multimodal CNN	Higher prediction accuracy	Springer
13	Early CNN Baselines	*[3.3.1.13]*	CNN	Frame-level engagement	IEEE:6786307
14	Transformer-based Engagement Detection	*[3.3.1.14]*	Transformer + Attention	High accuracy engagement	ScienceDirect
15	Landmark-based Graph CNN	*[3.3.1.15]*	Bimodal Graph CNN	Facial expression classification	ResearchGate

3.4 Research Gap and Motivation for Current Work

Research Gap:

Dataset Dependency & Limited Real-World Usability: Most studies rely on benchmark datasets (DAiSEE, CLARE), often collected in controlled environments, limiting real-world application.

Over-Reliance on Physiological or Facial Signals: Methods using only EEG/ECG or facial cues miss broader behavioral patterns.

Lack of Real-Time Multimodal Fusion: Existing fusion models are mostly offline; live webcam, keyboard, and mouse data are rarely combined.

Absence of Feedback & Practical Integration: Engagement detection often ends at classification without feedback or dashboard integration.

Annotation & Generalization Challenges: Labels are noisy, and models trained on one dataset fail to generalize to diverse users and environments.

Motivation for Current Work:

Practical Applicability: Integrate behavioral (typing/mouse) and visual (facial/gaze) signals for natural use without specialized hardware.

Real-Time Engagement Awareness: Provide instant feedback such as "take a break" or "refocus" to enhance productivity and wellbeing.

End-to-End Implementation: Develop a working prototype, including desktop activity monitoring and a web dashboard for visualization.

Contribution to Affective Computing & HCI: Bridge lab research with real-world applications in education and productivity tools.