International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. 11 Special Issue 03 | 2026

Cursor Cam : ML Based User Attention and Engagement Tracker

Tulasi Raorane[1], Pavan Rane[2], Kajal Kupale[3], Ruturaj Thakur[4], Suyog Sawant[5]
[1,2,3,4,5]dept. CSE (AI and ML) S.S.P.M's College of Engineering Kankavli, Maharashtra, India

DOI: 10.5281/zenodo.20631423

ABSTRACT

This paper presents CursorCam, a machine learning-based system designed to accurately assess the user attention and engagement through a multimodal approach. Traditional engagement measurement techniques rely primarily on passive metrics such as click rates and time spent, which often fail to capture the true cognitive and emotional state of users. To address this limitation, the proposed system integrates both visual and behavioral data streams to provide a more comprehensive and dynamic understanding of user interaction. The visual component utilizes webcam-based facial analysis to extract key features such as eye gaze direction, blink rate, head pose, and facial expressions, enabling real-time inference of user focus and emotional state. Simultaneously, the behavioral component captures mouse cursor dynamics, including movement patterns, velocity, click frequency, and idle time, which reflect the user's interaction behavior and engagement level. These heterogeneous data sources are combined and processed using a Random Forest model to classify and predict user engagement effectively.Experimental results are demonstrate that the proposed sys- tem provides improved accuracy and reliability compared to traditional single-modality approaches. The fusion of visual and behavioral cues enables a more nuanced interpretation of user attention, making CursorCam suitable for applications in e-learning, human-computer interaction, usability testing, and adaptive user interfaces. Overall, the system contributes to the development of intelligent and responsive digital environments by offering deeper insights into user engagement.

Index Terms—User Attention Detection, Cursor Movement Tracking, Eye Gaze Detection, Facial Expression Analysis, Ran- dom Forest, Machine Learning,

1.INTRODUCTION

In the modern digital ecosystem, understanding user at- tention and engagement has become increasingly important across domains such as e-learning, digital marketing, hu- man–computer interaction (HCI), and web usability research. Accurate measurement of user engagement enables the design of adaptive systems, improves content delivery, and enhances overall user experience. Traditional analytics tools primarily rely on passive interaction metrics such as page views, clickc ounts, session duration, and scroll depth to evaluate user behavior. While these metrics is provide quantitative insights, they fail to capture the underlying cognitive and emotional states that truly reflect user engagement. Consequently, such approaches often lead to incomplete or misleading interpreta- tions, highlighting the need for more intelligent and context- aware engagement analysis systems. Recent advancements in machine learning, computer vision, and affective computing have enabled the real-time analysis of human behavioral and visual cues. Techniques such as eye gaze tracking, facial expression recognition, and emo- tion detection provide valuable insights into user attention, cognitive load, and affective states. Eye gaze indicates the region of interest on a screen, while facial expressions and blink patterns reveal levels of concentration, fatigue, or frustra- tion. In parallel, behavioral signals derived from user interac- tions—particularly mouse cursor movements—have emerged as strong indicators of user intent and engagement. Features such as cursor velocity, trajectory smoothness, click frequency, scrolling behavior, and idle time exhibit meaningful correla- tions with visual attention and cognitive processing patterns. Despite these advancements, most existing systems rely on a single modality, either visual or behavioral, which limits their robustness and accuracy. Visual-only systems are often affected by environmental factors such as lighting conditions and camera quality, whereas behavioral-only systems may lack contextual understanding of the user's emotional and cognitive state. Therefore, there is a very clear need for a multimodal ap- proach that integrates complementary data sources to achieve a more comprehensive and reliable understanding of user engagement.To address this challenge, this paper proposes CursorCam, a multimodal machine learning-based system for real-time user attention and engagement tracking. The system integrates two primary data streams: visual data captured through a webcam and behavioral data derived from cursor interactions.The visual module extracts non-verbal cues such as gaze direction, blink rate, head pose, and facial expressions using computer vision techniques. Simultaneously, the behavioral module records cursor dynamics including movement velocity, trajectory smoothness, click distribution, scrolling behavior, and idle time. These features are preprocessed, fused, and analyzed using a Random Forest model to estimate user engagement levels accurately.The core contribution of this work lies in the development of a real-time multimodal fusion framework that combines visual and behavioral features to improve engagement prediction performance. The use of the Random Forest algorithm ensures robustness, handles high-dimensional data efficiently, and cap- tures complex nonlinear relationships between features. In ad- dition to engagement prediction, the system provides a graph- ical representation of user engagement over time, enabling intuitive visualization of attention patterns. Furthermore, the performance of the model is evaluated using accuracy metrics, demonstrating the effectiveness and reliability of the proposed approach.CursorCam has the significant applications in adaptive learning systems, usability testing, intelligent tutoring plat- forms, and personalized user interfaces. By providing both quantitative metrics (accuracy) and qualitative insights (en- gagement graphs), the system enables a deeper understanding of user behavior and supports data-driven decision-making in digital environments.

2.THEORICAL FRAMEWORK

A. Engagement as a Latent Cognitive Construct

User engagement is considered a latent variable, meaning it cannot be directly measured but can be inferred through observable signals. In digital environments, engagement man- ifests through: Visual attention (eye gaze direction, fixation stability), Facial expressions (indicating emotional states), Behavioral interaction patterns (cursor movement dynamics)

B. Multimodal Data Fusion Theory

The concept of multimodal data fusion is based on the principle that combining multiple sources of information leads to more reliable and robust predictions compared to single- modality systems. Each modality captures a different aspect of user behavior:

- Visual data provides insights into attention and emotional state
- Behavioral data reflects interaction patterns and user intent

Fusion can be categorized into:

- Feature-level fusion: combining features before model training
- Decision-level fusion: combining outputs of multiple models

In the proposed system, feature-level fusion is used to integrate gaze, facial, and cursor features into a unified representation, improving the accuracy of engagement prediction.

C. Visual Attention Theory

Visual attention theory explains how users allocate cognitive resources to specific areas of a screen. Eye gaze tracking is a direct indicator of where attention is focused. Key concepts include:

- Fixation: maintaining gaze on a specific point, indicating information processing
- Saccades: rapid eye movements between fixations
- Gaze duration: longer duration indicates deeper engage- ment

These metrics are crucial for estimating user focus and cog- nitive load.

D. Behavioral Interaction Theory

Behavioral interaction theory suggests that user intent and engagement can be inferred from interaction patterns. Cursor movement has been shown to correlate with eye movement and attention. Important behavioral indicators include:

- Movement speed and acceleration → indicates urgency or hesitation
- Trajectory smoothness → reflects confidence or confusion
- Click patterns and scrolling behavior → indicate interac- tion intensity
- Idle time → may represent disengagement or deep con- centration

These features provide indirect but valuable insights into user engagement.

3.METHODOLOGY

The methodological framework of CursorCam is designed to study user engagement within a naturalistic digital interaction environment. The research is grounded in human–computer interaction principles, where engagement is observed as a measurable outcome of real-time user interaction with web- based content.

Participants are interact with digital material under a normal browsing conditions to ensure ecological validity. Engagement is treated as categorical outcome (e.g., high, medium, low), enabling supervised learning. The experimental context en- sures that both visual and behavioral signals reflect authentic attentional states rather than artificial responses.

A. Context and Participants

The study was conducted in a controlled digital interaction environment designed to simulate real-world web usage con- ditions. Participants interacted with web-based content such as reading material, informational pages,

and interactive elements while their visual and behavioral data were recorded in real time.
- A functional webcam for visual data capture
- A mouse input device for cursor tracking
- A stable internet browser environment

B. Data Collection

The CursorCam system collected two primary data streams simultaneously: visual data and behavioral data.

1) Visual Data Collection (Webcam Based)

The webcam captured real-time facial video frames dur- ing user interaction. Computer vision techniques applied to extract:

- Eye gaze direction
- Fixation duration
- Blink rate
- Head pose orientation
- Facial expression indicators

2) Behavioral Data Collection (Cursor Tracking)

Mouse cursor activity was continuously recorded using browser-based event listeners. The following features were captured:

- Cursor position coordinates (x, y)
- Movement velocity
- Acceleration
- Movement smoothness
- Click frequency and location
- Idle time duration

C. *Data Analysis*

The analytical framework employs a supervised machine learning approach using the Random Forest algorithm.

1) Feature Engineering : Extracted raw signals are trans- formed into statistical and temporal features to represent engagement-relevant patterns. Feature-level fusion is ap- plied by combining visual and behavioral attributes into a single feature vector.

2) Random Forest Classification : Random Forest, an en- semble learning method based on multiple decision trees, is used to model the non-linear relationships between multimodal features and engagement levels. The model improves prediction robustness by averaging multiple tree outputs, reducing variance and overfitting. The prediction function can be expressed as: The pre- diction model is defined as:

The proposed system, CursorCam, is designed to analyze user attention and engagement by combining visual data from a webcam and behavioral data from mouse cursor interactions. The system follows a multi-stage processing pipeline consist- ing of data acquisition, feature extraction, machine learning analysis, and engagement prediction. This structured workflow allows the system to capture both visual and interaction-based indicators of user attention in real time.

4. RESULT

The proposed CursorCam system was implemented and tested to evaluate its ability to detect user attention and engagement using multimodal interaction data. The system integrates webcam-based visual analysis with mouse cursor behavioral tracking to generate real-time predictions of user attentional state.

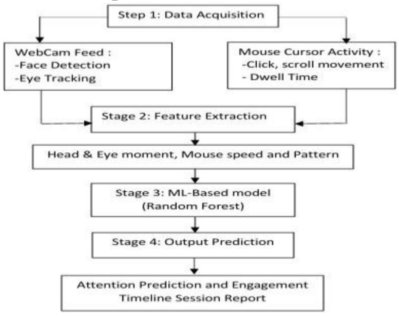

Fig. 1. Implementation

During this experiment, the system continuously captured facial features such as the gaze direction, blink rate, and head pose orientation, along with the behavioral parameters including cursor position, movement velocity, click frequency, and idle time. These features were stored in a structured dataset and used as inputs for a Random Forest classification model.In addition to real-time prediction, the system generates attention analytics by aggregating interaction data over time. The attention trend graph illustrates variations in engagement levels throughout the session. Higher values represent periods of focused interaction, while lower values indicate potential distraction or reduced engagement.The generated dataset contains multiple attributes including timestamp, face detection status, gaze direction, eye state, head pose, cursor coordinates, and click activity, enabling further analysis and potential improvements in model training.Overall, this experimental results demonstrate that com- bining visual attention cues with cursor behavioral patterns provides a reliable method for estimating user engagement. The proposed CursorCam system shows promising potential applications in e-learning environments, usability analysis, and human–computer interaction studies, where monitoring user attention is essential for improving system effectiveness.

A. Engagement Prediction Accuracy

The performance of this model was evaluated using stan- dard classification metrics, with accuracy being the primary measure. The Random Forest model demonstrated strong predictive capability, achieving a high level of accuracy in distinguishing between different engagement states such as engaged, partially engaged, and disengaged users.

B. Visualization of User Engagement

One of key outputs of the CursorCam system is the graphical representation of user engagement over time. The engagement graph provides a visual insight into how user attention fluctuates during interaction with the system. The X-axis represents time (or interaction duration) The Y-axis represents engagement level (ranging from low to hih) The graph typically shows: Stable high values indicating sustained user attention Sudden drops representing distraction or disengagement Fluctuations indicating varying levels of focus and cognitive load

C. Interpretation of Results

The results demonstrate that:Users with consistent gaze focus and smooth cursor movements tend to show higher engagement levelsIrregular cursor patterns combined with diverted gaze often correspond to lower engagementFacial expressions indicating confusion or fatigue corre- late with reduced attentionThe engagement of graph complements the accuracy metric by providing a qualitative understanding of user behavior, while the accuracy score provides a quantitative evaluation of the model's performance.

5. CONCLUSION

The CursorCam system presents a novel approach for ana- lyzing user attention and engagement by combining visual and behavioral data streams. Unlike traditional analytics methods that rely only on simple interaction metrics such as clicks or time spent on a page, this page provides more comprehensive understanding of user behavior by integrating webcam-based facial analysis with mouse cursor activity. By capturing non- verbal cues such as eye movements, head orientation, and facial expressions along with cursor movement patterns, the system is able to infer the attentional state of the user more accurately. The implementation of a Random Forest Machine Learning model enables the system is effectively analyze multiple features extracted from both visual and behavioral data. This approach improves the reliability of engagement prediction by identifying patterns that indicate focused attention, distraction, or interaction intensity. The proposed framework demonstrates how multimodal data fusion can enhance the accuracy of user engagement tracking in digital environments. In the future, the system can be further improved by incorporating more advanced deep learning models, increasing dataset diversity, and enhancing privacy-preserving mecha- nisms for handling webcam data. With these improvements, CursorCam has the potential to become a powerful tool for intelligent user behavior analysis and engagement measurement.

6. REFERENCES

[1] V. Karunakaran and A. Sharma, ”User Engagement Analysis of E- Commerce Websites from the Perspective of Eye Tracking,” in 2022 Fourth International Conference on Emerging Research in Electronics, Computer *Science and Technology (ICERECT)*, 2022, pp. 1–6, doi: 10.1109/ICERECT56837.2022.10060428.

[2] L. Thomas, K. C. Shilpa, M. I. Shilpa, and N. D. Tejaswini, ”User Engagement Prediction via Multi-Modal Emotion Analysis in Static and Dynamic Media,” in *2025 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AM- ATHE)*, 2025, pp. 1–9, doi: 10.1109/AMATHE65477.2025.11081217.

[3] Rawat, S. K. Singh, S. Kumar, T. Aggarwal, S. Singh, K. T. Chui, and B. B. Gupta, ”Predictive User Engagement Modeling Using Time Series Analysis for Interactive Media and Gaming,” in *2025 IEEE Gaming, Entertainment, and Media Conference (GEM)*, 2025, pp. 1– 4, doi: 10.1109/GEM66882.2025.11155680.

[4] S. Kamada and T. Ichimura, ”Multimodal Adaptive Structural Deep Belief Network for Emotion Recognition on RAVDESS Dataset,” in *2024 16th IIAI International congress on Advanced Applied*

Informatics (IIAI-AAI), 2024, pp. 292–298.

[5] S. Liu, D. Li, Q. Gao, and Y. Song, ”Facial Emotion Recognition Based on CNN,” in *2020 Chinese Automation Congress (CAC)*, 2020, pp. 398– 403.

[6] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, ”A Review of Emotion Recognition Using Physiological Signals,” *Sensors*, vol. 18, no. 7, p. 2074, 2018.

[7] Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, ”Characterizing User Behavior in Online Social Networks,” in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, 2009, pp. 49– 62.

[8] N. Shafiabady, T. Koo, F. U. Din, K. Sattarshetty, M. Yen, M. Alazab, and E. Alsharaydeh, ”Predicting Postgraduate Student En- gagement Using Artificial Intelligence (AI),” *IEEE Transactions on Artificial Intelligence*, vol. 6, no. 9, pp. 2464–2475, 2025, doi: 10.1109/TAI.2025.3548016.

A. Khan, A. R. Ahmad, N. Jabeur, and M. N. Mahdi, ”An Artificial Intelligence Approach to Monitor Student Performance and Devise Preventive Measures,” *Smart Learning Environments*, vol. 8, no. 1, p. 17, 2021.

[9] K. Benabbes, K. Housni, B. Hmedna, A. Zellou, and A. El Mezouary, ”A New Hybrid Approach to Detect and Track Learner’s Engagement in e-Learning,” *IEEE Access*, vol. 11, pp. 70912–70929, 2023, doi: 10.1109/ACCESS.2023.3293827.

[10] N. Ahmad, Z. Khan, and D. Singh, ”Student Engagement Prediction in MOOCs Using Deep Learning,” in *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2023, pp. 1–6, doi: 10.1109/ESCI56872.2023.10100007.

[11] S. Anand, ”Are Online Classes Wearing Children Out?,” *India Today Insight*, 2020.

[12] M. Dewan, M. Murshed, and F. Lin, ”Engagement Detection in Online Learning: A Review,” *Smart Learning Environments*, vol. 6, no. 1, pp. 1–20, 2019.

[13] V. Subhashini, A. R. Nisha, V. Radhalakshmi, G. Madhumita, K. Selvi, and K. Sudharson, ”Detecting Learning Patterns and Student Engagement in Online Courses Using Deep Learning,” in *2024 International Conference on Science Technology Engineering and Management (ICSTEM)*, 2024, pp. 1–7, doi: 10.1109/ICSTEM61137.2024.10560539.

[14] R. F. Kizilcec and S. Halawa, ”Attrition and Achievement Gaps in Online Learning,” in *Proceedings of the Second ACM Conference on Learning@ Scale*, 2015, pp. 57–66.

[15] J. D. Rose, K. Vijayakumar, and S. Sakthivel, ”Students’ Performance Analysis System Using Cumulative Predictor Algorithm,” *International Journal of Reasoning-based Intelligent Systems*, vol. 11, no. 2, pp. 122– 133, 2019.

[16] X. Tao, A. Shannon-Honson, P. Delaney, L. Li, C. Dann, Y. Li, and H. Xie, ”Data Analytics on Online Student Engagement Data for Academic Performance Modeling,” *IEEE Access*, vol. 10, pp. 103176–103186, 2022, doi: 10.1109/ACCESS.2022.3208953.

[17] Z. Kastrati, F. Dalipi, A. S. Imran, K. P. Nuci, and M. A. Wani, ”Senti- ment Analysis of Students’ Feedback with NLP and Deep Learning: A Systematic Mapping Study,” *Applied Sciences*, vol. 11, no. 9, p. 3986, 2021.

[18] Qarbal, N. Sael, and S. Ouahabi, ”Student’s Engagement Detec- tion Based on Computer Vision: A Systematic Literature Review,” *IEEE Access*, vol. 13, pp. 140519–140545, 2025, doi: 10.1109/ACCESS.2025.3596885.

[19] S. Maloney, M. Axelsen, L. Galligan, J. Turner, P. Redmond, A. Brown, M. Basson, and J. Lawrence, ”Using LMS Log Data to Explore Student Engagement with Coursework Videos,” *Online Learning*, vol. 26, no. 4, pp. 399–423, 2022.