EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

# LeSA: Applying a Computer Vision-based AI Approach To Monitoring Sustained Attention In Online Learning

doi: 10.52987/edc.2025.004

**Pichaya Promla**
King Mongkut's University of Technology North Bangkok,
Department of Computer Education
Bangkok, Thailand
pichaya.promla@gmail.com

**Somkid Saelee**
King Mongkut's University of Technology North Bangkok,
Department of Computer Education
Bangkok, Thailand
somkid.l@fte.kmutnb.ac.th

## ABSTRACT

As online learning continues to expand, its flexibility in time and location supports individual learning differences and lifelong learning, contributing to sustainable social and economic development. However, online learners require greater self-discipline and responsibility due to limited interaction and reduced instructor oversight, making it difficult to monitor engagement and adapt teaching strategies. One of the most significant challenges in this environment is sustaining student attention, which is essential for effective learning. A lack of sustained attention can lead to decreased motivation and higher dropout rates, impacting overall learning outcomes.

This study explores the use of computer vision-based AI to analyze behavioral patterns that indicate sustained attention in online learning. By tracking key behavioral indicators, including gaze direction, facial expressions, blinking patterns, and eye closure frequency, this approach provides valuable insights into student engagement. The Learner Sustained Attention (LeSA) model was developed and evaluated using four machine learning algorithms to determine the most effective method for identifying attention patterns.

Findings indicate that integrating computer vision-based AI into online education provides a scalable and data-driven solution for real-time attention monitoring. The Random Forest model demonstrated the highest accuracy (84.07%), effectively identifying behavioral patterns associated with sustained attention. These insights can support instructors in adapting teaching methods, enhancing student engagement, and reducing dropout rates. This study contributes to the ongoing innovation in educational technology, reinforcing the role of AI-driven behavioral tracking in improving learning experiences and instructional effectiveness in online education

**KEYWORDS**: Online Learning, Computer Vision, Sustained Attention, Artificial Intelligence, Educational Innovation, Behavioral Patterns, Engagement Tracking, Learning Analytics

EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

# 15 INTRODUCTION

The digital transformation has significantly reshaped education, making learning more flexible, accessible, and adaptable to individual needs (Wichuda, 1999). Online learning has emerged as a key driver in promoting lifelong learning and sustainable development (Torres, 2011; Vieira, 2019), enabling learners to engage with educational content regardless of time and location. However, despite these advantages, maintaining student engagement and attention remains a critical challenge, particularly due to the lack of real-time observation and interaction between instructors and learners (Guo et al., 2015; Rothkrantz, 2017). Addressing this challenge requires innovative strategies that ensure students remain focused and engaged, ultimately enhancing their learning experiences and outcomes. Previous research has explored three primary approaches to tackling this issue (Hakimi et al., 2024):

**The Instructional Strategies Approach** focuses on improving engagement and motivation by implementing student-centered learning methods such as project-based learning, interactive activities, and collaborative tasks (Achuthan et al., 2024; Balalle, 2024; Jiang & Cheong, 2024;Marland & Store, 1993; Samandar, 2024). These approaches foster active participation and reduce dropout rates by making online learning more interactive and personalized.

**The Content Delivery Approach** ensures that learning materials are structured, engaging, and aligned with students' learning styles. Poor content organization can lead to cognitive overload, frustration, and disengagement. Effective delivery includes clear learning objectives, modular content, multimedia integration, gamification, and interactive learning tools, all of which enhance student engagement and knowledge retention (Afzal et al., 2024; Khasawneh et al., 2023; Muhammad, 2024; Sadanala et al., n.d.).

**The Monitoring Learners' Behavior Approach** leverages data-driven interventions and AI-powered student monitoring to analyze learning behaviors and apply timely interventions. AI models can track facial expressions (Cha & Kim, 2015; Dewan et al., 2018; D'Mello et al., 2012; Hutt et al., 2017; Murali et al., 2021; Sun et al., 2019; Zeng et al., 2019), gaze behavior (Bedenlier et al., 2021; Hutt et al., 2017; Murali et al., 2021; Yang et al., 2015), head posture (Ahuja et al., 2021; Murali et al., 2021; Raca & Dillenbourg, 2015), and physiological responses such as brain waves (Hassib et al., 2017; Kosmyna & Maes, 2019) and galvanic skin response (GSR) (Al Machot et al., 2018; Di Lascio et al., 2018). These insights provide real-time feedback, enabling instructors to better understand student engagement levels and implement adaptive interventions.

Building upon these approaches, this study introduces LeSA (Learner Sustained Attention), an AI-powered solution utilizing computer vision techniques to analyze students' sustained attention in online learning. By extracting behavioral patterns such as gaze tracking, facial expressions, and blinking patterns, the proposed model offers a practical tool for monitoring engagement and enhancing instructional strategies.

**Research Objectives**

1. To investigate online learners' behavioral patterns using computer vision techniques and their impact on sustained attention levels.
2. To analyze the relationship between these behavioral features and students' engagement in online learning environments.
3. To evaluate the effectiveness of AI-powered attention monitoring tools in supporting instructors to enhance student engagement and improve instructional strategies in online education.

EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

## 16 RESEARCH METHODOLOGY

This research follows a research and development approach using mixed methods design with the following steps:

### 16.1 Data Collection

A dataset was collected from 25 student participants, consisting of 20 males and 5 females, all Asian teenagers aged between 15–20 years old, with 2 participants wearing glasses. Prior to data collection, all participants were informed about research ethics and provided consent for data collection.
Each participant recorded two 20-minutes video sessions using their own devices to reflect real-world diversity in online environments. The sessions included:

- **Active online activities** that participants found engaging and enjoyable, where they were naturally. Such as using social media, playing online games, and streaming movies.
- **Passive online activities** that were perceived as less engaging and more likely to cause distraction or reduced attention, such as attending online lectures, reading e-books, and listening to monotonous podcasts.

Each video was required to capture upper-body movements, facial expressions, and gaze direction to facilitate behavioral analysis.

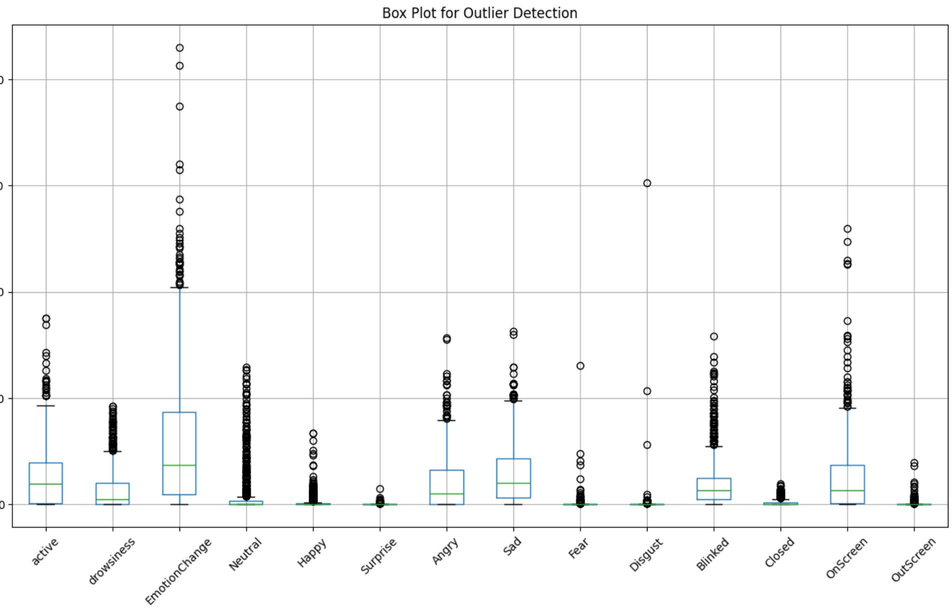
Figure 1: Example of the dataset

After three days, participants reviewed their own videos (without being informed of the session type) and self-assessed their sustained attention levels for each minute, categorizing them into high, moderate, low, or no attention. Meanwhile, three online learning experts independently assessed the same videos, using a randomized order to prevent assessment bias. The final sustained attention level for each video segment was determined using the following rules:

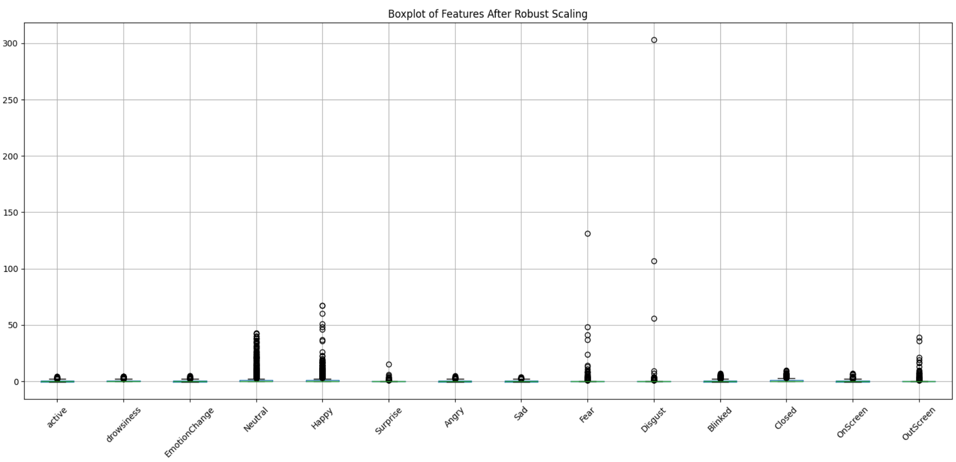
1. If all assessments matched (from three experts and from participants), that level was assigned.
2. If there was disagreement, only the mode value from the three expert assessments was used.
3. If no mode value was found, the participant's self-assessment was included to determine the mode value.
4. If no mode value still could not be determined, the data was discarded.

EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

Meanwhile, a computer vision-based AI model was used to extract 14 behavioral features from the record videos: blink rate per minute, frequency of eye closure , frequency of looking at the screen, frequency of looking away from the screen, frequency of active facial expressions, frequency of drowsy facial expressions, frequency of angry facial expressions, frequency of sad facial expressions, frequency of disgusted facial expressions, frequency of happy facial expressions, frequency of surprise facial expressions, frequency of neutral facial expressions, frequency of fear facial expressions, and frequency of emotional facial expression changes.
These 14 behavioral features were then combined with the categorized sustained attention levels to analyze behavioral patterns affecting sustained attention in online learning environment. Dataset contains 1000 records. After removing records with missing values and duplicates, robust scaling was applied to handle inconsistencies in the data, ensuring that outliers remained within an acceptable range without distorting the analysis. The final dataset contains 764 records.




(a) Dataset before applied robust scaling




(b) Dataset after applied robust scaling

Figure 2: Dataset before and after handling outliers with robust scaling.

EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

## 16.2 Feature Selection and Feature Engineering

To identify behavioral features relevant to sustained attention levels, correlation analysis was conducted. Features which have a correlation coefficient of 0.3 or higher were selected for further analysis. Based on this criterion, six behavioral features were chosen for training the AI model:

1. Frequency of active facial expressions.
2. Frequency of emotional facial expression changes.
3. Blink rate per minute.
4. Frequency of looking at the screen.
5. Frequency of looking away from the screen.
6. Frequency of eye closure.

The selected featured exhibited linear relationships with sustained attention levels, making them suitable for training a supervised AI model for effective prediction. To address potential training inefficiencies caused by limited dataset and class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to resample the data. The dataset was then 976 records, with 337 records in each category. The dataset was split to an 80% training set (780 records) and a 20% test set (196 records) for model evaluation.

## 16.3 Model Training

Six machine learning algorithms were selected to train AI models, categorized based on their computation complexity and ability to handle different data relationships:
**Lightweight models:** Suitable for linear relationships data and requiring minimal computation resources including Support Vector Machine (SVM), and Logistic Regression.
**Medium-weight models:** Capable of handling non-linear relationships with moderate computational resources: Random Forest
**Heavyweight models:** Designed for high accuracy and complex relationships, requiring higher computational resources: Gradient Boost, XGBoost, and LightBGM.
The AI models were trained in two rounds: First round, AI models were trained using default parameters, with 10-fold, 20-fold, and 30-fold cross validation to ensure robust performance.

Table 19 Default Parameters Used in First Round Training.

| AI Model | parameter |
|---|---|
| SVM | probability=True, random_state=42, Kernel= RBF |
| Logistic Regression | max_iter=1000, random_state=42, Solver= lbfgs |
| Random Forrest | random_state=42, n_estimators=100, max_depth=None |
| Gradient Boost | random_state=42, learning_rate=0.1, n_estimators=100, max_depth=3 |
| XGBoost | use_label_encoder=False, eval_metric=mlogloss, random_state=42, n_estimators=100 |
| LightBGM | random_state=42, num_leaves=no define, learning_rate = 0.1 |

After evaluating the first round of training, 20-fold cross validation demonstrates the best balance between bias and variance. Consequently, for the second round all AI models were fine-tuned using RandomizationSearchCV, retrained with optimized hyperparameters, and their performances were compared.

Table 2 Fine-Tuned Parameter Used in Second Round Training.

| AI Model | parameter |
|---|---|

EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

| | |
|---|---|
| SVM | kernel=rbf, gamma=1, C=10 |
| Logistic Regression | solver=lbfgs, penalty=l2, max_iter= 100, C=0.1 |
| Random Forrest | n_estimators= 200, min_samples_split= 2, min_samples_leaf=1, max_depth=None |
| Gradient Boost | n_estimators=200, max_depth=7, learning_rate=0.2 |
| XGBoost | subsample=0.8, n_estimators=100, max_depth=7, learning_rate=0.2 |
| LightBGM | num_leaves=50, n_estimators=300, max_depth= -1, learning_rate= 0.2 |

### 16.4 Model Evaluation

To evaluate the performance of the AI models, a confusion matrix was used to measure accuracy metrics:
**Accuracy** measures the overall correctness of a model by calculating the percentage of correctly predicted instances out of all predictions. Higher accuracy generally indicates better performance; however, it can be misleading in imbalanced datasets where one class dominates
**Precision** quantifies how many of the predicted positive cases are actually correct, with higher precision meaning fewer false positives, making the model more reliable when false positives need to be minimized.
**Recall (Sensitivity or True Positive Rate)** assesses how well the model identifies actual positive cases, measuring the proportion of true positives correctly predicted. A high recall indicates that the model effectively detects most positive cases but may increase false positives.
**F1-score** is the harmonic means of precision and recall, balancing both metrics to provide a more comprehensive evaluation when there is an imbalance between false positives and false negatives. A high F1-score suggests that the model maintains an optimal trade-off between precision and recall.
**Receiver Operating Characteristic - Area Under the Curve (ROC-AUC)** evaluates a model's ability to distinguish between different classes by plotting the True Positive Rate (Recall) against the False Positive Rate at various classification thresholds. The AUC score quantifies the model's performance, where a value of 1.0 indicates perfect classification, 0.5 suggests random guessing, and below 0.5 means the model performs worse than random chance. A higher AUC indicates better class separation, making ROC-AUC particularly useful for evaluating probabilistic models and assessing their discriminatory power across different classification thresholds.
Additionally, given the real-world application in online learning environments, **computation time (latency)** was measured to evaluate how long each model took to predict new data. All models were compared across these metrics to determine the most situation model for monitoring sustained attention in online leaning.

## 17 RESULTS

This section presents research results according to research objectives.

### 17.1 Investigating Online Learners' Behavioral Patterns Using Computer Vision Techniques and Their Impact on Sustained Attention Levels.

To investigate the behavioral patterns of online learners and their impact on sustained attention levels, 14 behavioral features were extracted using computer vision-based methods. Each feature was analyzed to determine its relationship with sustained attention.
**Behavioral Feature Analysis**: An initial correlation analysis was conducted to assess the relationship between each behavioral feature and sustained attention level. The results

EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

indicated that all 14 features influenced sustained attention to some degree, however, only six features exhibited a statistically significant correlation (≥ 0.3) with sustained attention level. These six key behavioral features were selected for further study:

1. Frequency of active facial expressions: Facial movement, such as raising eyebrows, smiling, or reacting expressively, was positively correlated with sustained attention. Learners who frequently display active facial expression tended to remain more engaged in the learning sessions.

2. Frequency of emotional facial expression changes: Frequent shifts between different emotional expressions indicate fluctuating attention levels. Learners who rapidly changed facial expressions showed lower sustained attention, possibly due to distractions or shifting emotional states.

3. Blink rate per minute: Engaged learners blinked between 15-30 times per minute. A higher blink rate was associated with lower sustained attention, as increased blinking may indicate fatigue, disengagement, or cognitive overload.

4. Frequency of looking at the screen: Direct eye contact with the screen was a strong indicator of sustained attention. Learners who consistently maintained screen focus exhibited higher attention levels.

5. Frequency of looking away from the screen: In contrast to looking at the screen, frequent eye movement away from the screen was negatively correlated with sustained attention. A higher off-screen gaze frequency suggested distraction or reduced cognitive engagement.

6. Frequency of eye closure: Prolonged eye closures were linked to fatigue, reduced alertness, and lower sustained attention. Learners who frequently closed their eyes during the session were more likely to experience attention lapses.

**Interpretation of Behavioral Patterns**: The findings suggest that learners who exhibited active facial expressions, maintained eye contact with the screen, and had a moderate blink rate were more likely to sustain their attention during online learning. Conversely, those who frequently looked away from the screen, rapidly changed facial expressions, or closed their eyes experienced reduced attention spans.

These results highlight the importance of visual sustained attention cues in online learning environments. The extracted behavioral patterns provide valuable insights for instructors, allowing them to identify disengaged learners in real time and adjust instructional strategies accordingly.

Table 3 Leaners’ Behavioral Pattern Represent Sustained Attention.

| Sustained Attention level | Learners’ Behavioral Pattern |
|---|---|
| Tend to high level | 1. Exhibited active facial expressions.<br>2. Maintained eye contact with the screen.<br>3. Had a moderate blink rate. |
| Tend to low level | 1. Frequently looked away from the screen.<br>2. Rapidly changed facial expressions.<br>3. Closed their eyes |

## 17.2 Analyzing the Relationship Between Behavioral Features and Sustained Attention

To examine the relationship between behavioral features and sustained attention, Pearson correlation analysis was conducted. The correlation coefficient (r-value) was used to measure

EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

the strength and direction of each relationship, with statistical significance assessed using p-values.

Figure 3 presents the correlation heatmap, illustrating the relationships between the 14 behavioral features and sustained attention levels. The intensity of the color represents the strength and direction of correlation, with darker shades indicating stronger correlations.

The analysis revealed that six key features exhibited statistically significant correlations (≥ 0.3) with sustained attention levels. Table 4 presents the correlation results for the selected six behavioral features.

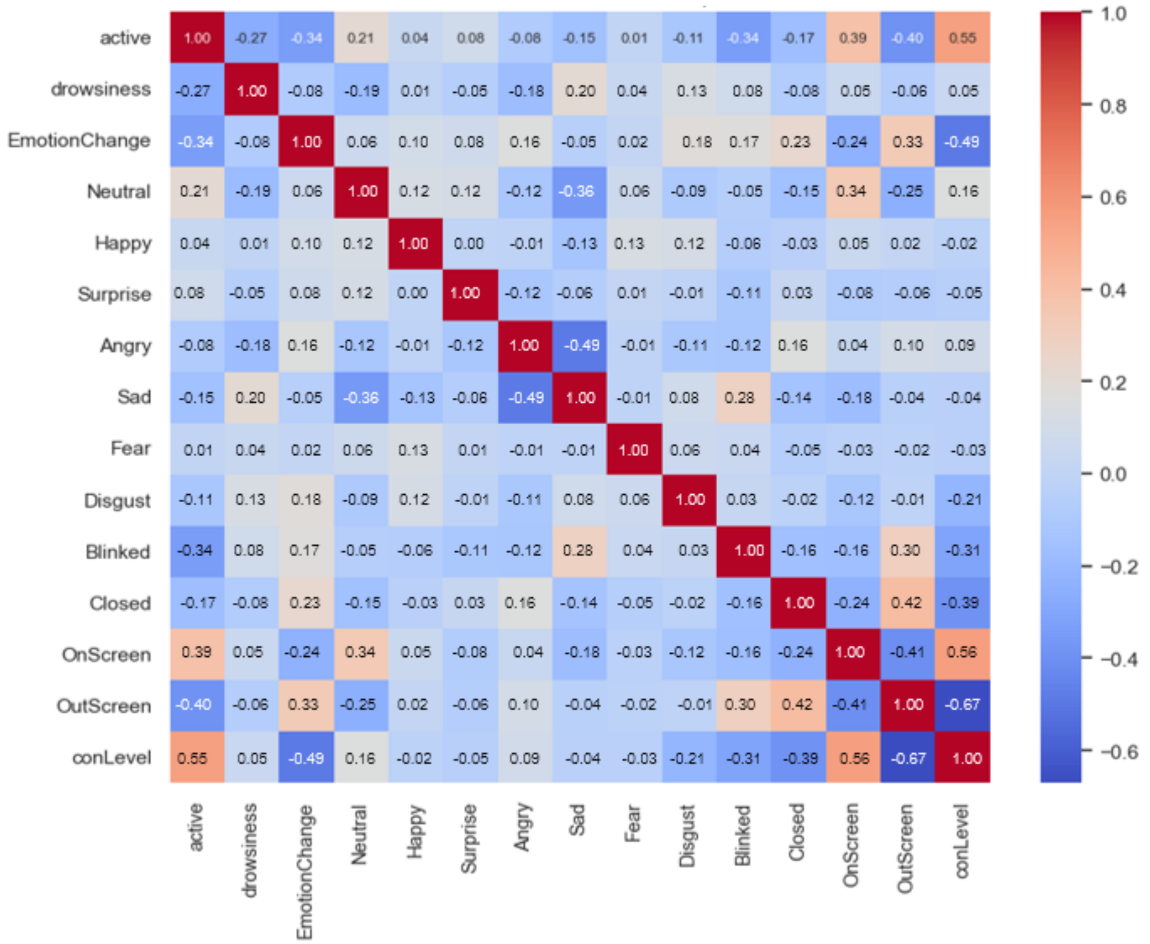



Figure 3: Correlation Heatmap of Behavioral Features and Sustained Attention Levels.

Table 4 The Correlation Results for The Selected Six Behavioral Features.

| **Behavioral Feature** | **Correlation Coefficient (r)** | **Statistical Significance (p-value)** | **Relationship with Sustained Attention** |
|---|---|---|---|
| Frequency of looking at the screen | 0.56 | < 0.05 | Strong Positive |
| Frequency of active facial expressions | 0.55 | < 0.05 | Moderate Positive |
| Blink rate per minute | -0.31 | < 0.01 | Weak Negative |
| Frequency of eye closure | -0.39 | < 0.01 | Moderate Negative |

EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

| | | | |
|---|---|---|---|
| Frequency of emotional facial expression changes | -0.49 | < 0.05 | Moderate Negative |
| Frequency of looking away from the screen | -0.67 | < 0.01 | Strong Negative |

The correlation analysis highlights key behavioral indicators associated with sustained attention in online learning environments.

**Positive Correlations (Indicators of High Sustained Attention)**

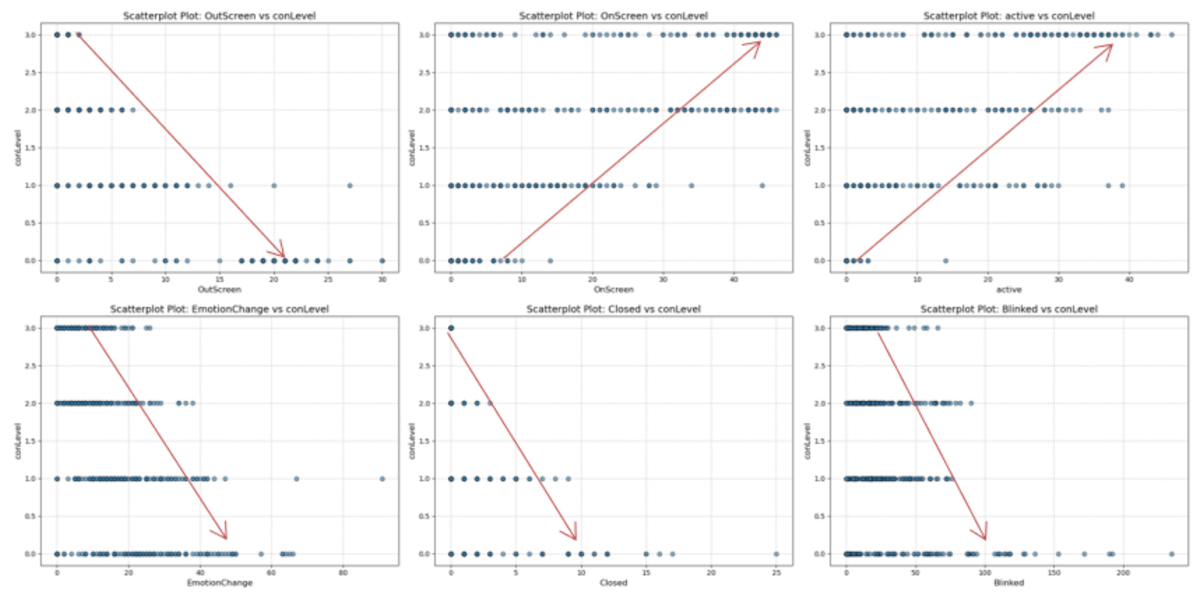
- Looking at the screen ($r = 0.67$, $p < 0.01$): Strongly associated with higher sustained attention, confirming that visual focus is a key determinant of attention retention.
- Active facial expressions ($r = 0.55$, $p < 0.05$): Moderately associated with higher sustained attention. Learners who frequently engage in expressive reactions (e.g., smiling, raising eyebrows) tend to remain engaged in learning sessions.

**Negative Correlations (Indicators of Low Sustained Attention)**

- Looking away from the screen ($r = -0.67$, $p < 0.01$): Strongly associated with lower sustained attention, suggesting that frequent off-screen gazes indicate distraction.
- Emotional facial expression changes ($r = -0.49$, $p < 0.05$): Moderately associated with lower sustained attention. Learners who frequently shift emotions during learning sessions tend to lower sustained attention, likely due to external distractions or cognitive overload.
- Eye closure ($r = -0.39$, $p < 0.01$): Moderately associated with reduced attention levels, possibly due to fatigue or disengagement.
- Blink rate ($r = -0.31$, $p < 0.05$): Weakly associated with lower sustained attention. While a moderate blink rate is normal, an excessive blink rate suggests lower sustained attention, likely due to mental fatigue or loss of focus.

Further analysis indicates that the relationship between the selected behavioral features and sustained attention levels exhibit linearity. This suggests that changes in these behavioral features have a direct and proportional effect on sustained levels. Figure 4 illustrates these linear relationships between key features and sustained attention.

These findings highlight the importance of behavioral monitoring in online education, as tracking visual focus and facial engagement can provide real-time indicators of attention retention.



EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

Figure 4: Linear Relationship of Selected Features with Sustained Attention.

## 17.3 AI Model Performance Evaluation for Sustained Attention Monitoring

To assess the effectiveness of AI models in monitoring sustained attention, multiple machine learning algorithms were trained and evaluated. The evaluation metrics included: Accuracy, Precision, Recall, F1-score, ROC-AUC, and Latency. Table 5 summarizes the performance of the six AI models evaluated in this study.

Table 5 Performance Comparison of AI Models for Sustained Attention Prediction

| AI Model | Accuracy | Precision | Recall | F1-score | ROC-AUC | Latency |
|---|---|---|---|---|---|---|
| SVM | 71.48% | 0.71 | 0.69 | 0.70 | 0.83 | 230 mS |
| Logistic Regression | 62.96% | 0.61 | 0.63 | 0.62 | 0.77 | 10 mS |
| Random Forest | 84.07% | 0.83 | 0.84 | 0.83 | 0.86 | 760 mS |
| Gradient Boosting | 83.70% | 0.83 | 0.84 | 0.83 | 0.85 | 6.12 S |
| XGBoost | 82.22% | 0.82 | 0.82 | 0.82 | 0.86 | 310 mS |
| LightGBM | 83.70% | 0.83 | 0.84 | 0.83 | 0.85 | 660 mS |

The Random Forest model achieved the highest accuracy (84.07%) and the best overall performance with an F1-score of 0.83 and a ROC-AUC of 0.86, making it the most effective model for predicting sustained attention in online learning. However, its latency (760 mS) is slightly higher than SVM (230 mS), Logistic Regression (10 mS), XGBoost (310 mS), and LightGBM (660 mS), which is an important factor for real-time applications.

Although Gradient Boosting exhibited high accuracy and strong overall performance, it had the highest computational time (6120 mS), making it unsuitable for real-time tracking.

SVM and Logistic Regression performed moderately well and required significantly less computation time; however, their accuracy was lower than other models, limiting their effectiveness for sustained attention prediction.

XGBoost and LightGBM demonstrated competitive performance, achieving high accuracy and F1-scores while maintaining lower latency than Gradient Boosting. However, both models require higher computational costs than Random Forest, making them less efficient for large-scale online learning applications. Figure 5 illustrates the comparative performance of all models across different evaluation metrics.

EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

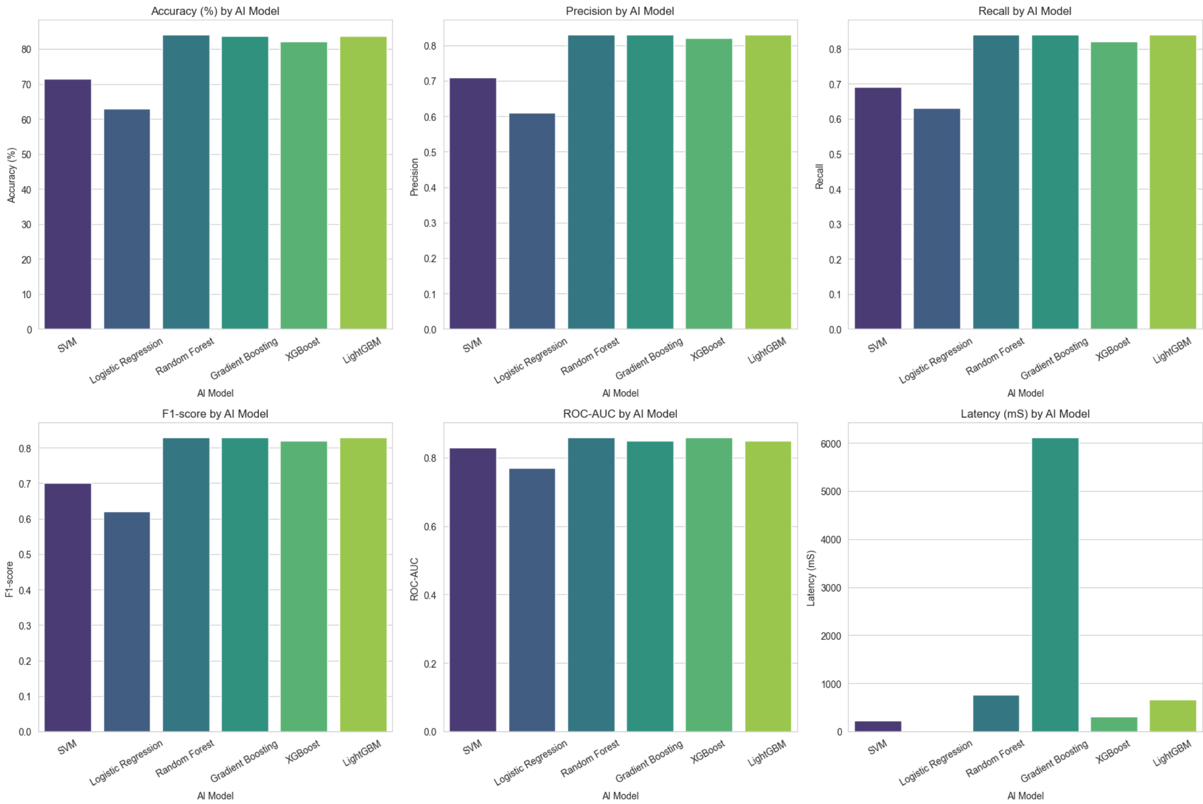



Figure 5 The Comparative Performance of AI Models.

For real-time monitoring in online learning, balancing model accuracy and computational efficiency is essential, if required high accuracy, Random Forest, Gradient Boost or LightGBM may be the best options. If low computational latency is required XGBoost provides a good balance of accuracy and faster processing. If simplicity and low resource usage are needed, Logistic Regression and SVM are the good choice but less accurate.

## 18 CONCLUSION AND DISCUSSION

This study demonstrates that Computer Vision-based AI can effectively monitor students' sustained attention in online learning environments. By identifying key behavioral patterns, it validates the effectiveness of AI-driven attention tracking, offering a scalable, automated approach to engagement monitoring.

Through correlation analysis, the study reveals that sustained attention is significantly influenced by visual and facial behavioral cues. The six key behavioral features identified are:

**High sustained attention** is associated with frequent screen focus, active facial expressions, and a balanced blink rate.

**Low sustained attention** is linked to frequent looking away from the screen, excessive emotional expression changes, frequent eye closure, and a high blink rate.

Similar to Miao et al. (2024), who integrated multimodal data for real-time attention detection, this study reinforces the importance of leveraging multiple behavioral cues for accurate attention assessment. The findings highlight that combining gaze tracking, facial expressions, and movement analysis enhances the effectiveness of Computer Vision-based AI models in monitoring concentration. Likewise, Zhou et al. (2024) emphasized that non-verbal cues play a crucial role in understanding student learning processes, further supporting the role of

EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

Computer Vision-based AI models in adaptive learning environments and personalized engagement monitoring.

Additionally, this study validates Computer Vision-based AI models as a non-intrusive and scalable solution for automated attention monitoring. The proposed Learner Sustained Attention (LeSA) AI model, powered by Random Forest, achieved the highest accuracy (84.07%), demonstrating its effectiveness in predicting sustained attention levels. Unlike traditional manual observation, self-reported surveys, or wearable sensors, the Computer Vision-based approach enables real-time, automated tracking using only a webcam, making it a practical tool for online learning.

This study aligns with Ma et al. (2022), who explored the use of Computer Vision in analyzing student engagement, confusion, gaze following, emotional expressions, and head movement, demonstrating its effectiveness in supporting online teaching. Similarly, Alruwais & Zakariah (2024) highlighted that AI-driven computer vision enables real-time student behavior recognition, allowing instructors to tailor learning experiences and ultimately improve engagement, learning outcomes, and instructional support. By analyzing behavioral patterns such as gaze direction, facial expressions, and attention shifts, Computer Vision-based AI provides real-time insights that enhance personalized learning experiences and engagement monitoring in both individual and collaborative learning settings.

As educational technology continues to evolve, AI-driven behavioral tracking offers an opportunity to redefine online learning experiences. This study demonstrates that Computer Vision-based AI models can provide real-time insights into student attention, improving engagement monitoring and retention strategies. The findings contribute to an expanding field of research on AI in education, offering a scalable, objective, and automated solution for tracking sustained attention in online learning environments.

Although this study provides strong evidence for Computer Vision-based attention tracking, several limitations must be acknowledged. Environmental factors, such as lighting conditions and webcam quality, can impact the accuracy of facial feature extraction, potentially affecting real-time tracking. Generalization challenges also exist, as real-world applications may require adaptive AI models that continuously learn from individual variations in student behavior, despite the dataset being balanced using SMOTE. Additionally, the AI model has yet to be tested in fully operational online learning environments, requiring future validation in real-world classrooms to assess its practical effectiveness and integration with existing learning platforms.

Future research should focus on deploying the AI model in live online learning environments to evaluate its practical effectiveness in monitoring sustained attention. Conducting live classroom trials would help validate the model's ability to provide real-time feedback and support adaptive teaching strategies. Additionally, further exploration of multi-modal behavioral analysis, integrating keystroke dynamics, speech tone, and body posture, could improve prediction accuracy by capturing a broader range of engagement indicators.

Incorporating adaptive AI systems that dynamically adjust attention thresholds based on individual learning styles could further enhance engagement tracking, making AI-powered learning analytics more personalized and responsive. By addressing these areas, future advancements in Computer Vision-based AI can continue to revolutionize online education, making learning environments more adaptive, data-driven, and student-centered.

## REFERENCES

[80] Achuthan, K., Kolil, V. K., Muthupalani, S., & Raman, R. (2024). Transactional distance theory in distance learning: Past, current, and future research trends. *Contemporary Educational Technology*, *16*(1), ep493. https://doi.org/10.30935/cedtech/14131

[81] Afzal, D. M. T., Ashfaq, M. S., & Mushtaq, M. S. (2024). Revolutionizing Distance Learning through Technology: A Case Study of Allama Iqbal Open University. *Pakistan Journal of Society, Education and Language (PJSEL)*, *10*(2), Article 2.

[82] Ahuja, K., Shah, D., Pareddy, S., Xhakaj, F., Ogan, A., Agarwal, Y., & Harrison, C. (2021). Classroom Digital Twins with Instrumentation-Free Gaze Tracking. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3411764.3445711

[83] Al Machot, F., Ali, M., Ranasinghe, S., Mosa, A. H., & Kyandoghere, K. (2018). Improving Subject-independent Human Emotion Recognition Using Electrodermal Activity Sensors for Active and Assisted Living. *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, 222–228. https://doi.org/10.1145/3197768.3201523

[84] Alruwais, N. M., & Zakariah, M. (2024). Student Recognition and Activity Monitoring in E-Classes Using Deep Learning in Higher Education. *IEEE Access*, *12*, 66110–66128. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3354981

[85] Balalle, H. (2024). Exploring student engagement in technology-based education in relation to gamification, online/distance learning, and other factors: A systematic literature review. *Social Sciences & Humanities Open*, *9*, 100870. https://doi.org/10.1016/j.ssaho.2024.100870

[86] Bedenlier, S., Wunder, I., Gläser-Zikuda, M., Kammerl, R., Kopp, B., Ziegler, A., & Händel, M. (2021). "Generation invisible?. Higher Education Students' (Non)Use of Webcams in Synchronous Online Learning. *International Journal of Educational Research Open*, *2*, 100068. https://doi.org/10.1016/j.ijedro.2021.100068

[87] Cha, S., & Kim, W. (2015). Concentration analysis by detecting face features of learners. *2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, 46–51. https://doi.org/10.1109/PACRIM.2015.7334807

[88] Dewan, M. A. A., Lin, F., Wen, D., Murshed, M., & Uddin, Z. (2018). A Deep Learning Approach to Detecting Engagement of Online Learners. *2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 1895–1902. https://doi.org/10.1109/SmartWorld.2018.00318

[89] Di Lascio, E., Gashi, S., & Santini, S. (2018). Unobtrusive Assessment of Students' Emotional Engagement during Lectures Using Electrodermal Activity Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(3), 103:1-103:21. https://doi.org/10.1145/3264913

[90] D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, *70*(5), 377–398. https://doi.org/10.1016/j.ijhcs.2012.01.004

[91] Guo, Y. R., Goh, D. H.-L., Luyt, B., Sin, S.-C. J., & Ang, R. P. (2015). The effectiveness and acceptance of an affective information literacy tutorial. *Computers & Education*, *87*, 368–384. https://doi.org/10.1016/j.compedu.2015.07.015

[92] Hakimi, M., Katebzadah, S., & Fazil, A. W. (2024). Comprehensive Insights into E-Learning in Contemporary Education: Analyzing Trends, Challenges, and Best Practices. *Journal of Education and Teaching Learning (JETL)*, *6*(1), 86–105.

[93] Hassib, M., Schneegass, S., Eiglsperger, P., Henze, N., Schmidt, A., & Alt, F. (2017). EngageMeter: A System for Implicit Audience Engagement Sensing Using Electroencephalography. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 5114–5119. https://doi.org/10.1145/3025453.3025669

[94] Hutt, S., Mills, C., Bosch, N., Krasich, K., Brockmole, J., & D'Mello, S. (2017). "Out of the Fr-Eye-ing Pan": Towards Gaze-Based Models of Attention during Learning with Technology in the Classroom. *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 94–103. https://doi.org/10.1145/3079628.3079669

[95] Jiang, H., & Cheong, K. W. (2024). Developing teaching strategies for rural school pupils' concentration in the distance music classroom. *Education and Information Technologies*, *29*(5), 5903–5920. https://doi.org/10.1007/s10639-023-12056-1

[96] Khasawneh, Dr. N. A. S., Khasawneh, A. J., Khasawneh, Dr. M. A. S., & Khasawneh, Dr. Y. J. A. (2023). Improving Arabic Content Delivery on Cloud Computing Platforms for Jordanian E-learning Environments. *Migration Letters*, *21*(S1), 575–585. https://doi.org/10.59670/ml.v21iS1.6181

[97] Kosmyna, N., & Maes, P. (2019). AttentivU: An EEG-Based Closed-Loop Biofeedback System for Real-Time Monitoring and Improvement of Engagement for Personalized Learning. *Sensors (Basel, Switzerland)*, *19*(23), 5200. https://doi.org/10.3390/s19235200

[98] Ma, S., Zhou, T., Nie, F., & Ma, X. (2022). Glancee: An Adaptable System for Instructors to Grasp Student Learning Status in Synchronous Online Classes. *CHI Conference on Human Factors in Computing Systems*, 1–25. https://doi.org/10.1145/3491102.3517482

[99] Marland, P. W., & Store, R. E. (1993). Some instructional strategies for improved learning from distance teaching Materials. In *Distance Education: New Perspectives*. Routledge.

[100] Miao, Q., Li, L., & Wu, D. (2024). An English video teaching classroom attention evaluation model incorporating multimodal information. *Journal of Ambient Intelligence and Humanized Computing*, *15*(7), 3067–3079. https://doi.org/10.1007/s12652-024-04800-3

[101] Muhammad, R. N. (2024). Insights from Students on the Use of Asynchronous and Synchronous Approaches during Digital Learning in Thailand's Primary Schools. *DIDAKTIKA : Jurnal Pemikiran Pendidikan*, *30*(1), Article 1. https://doi.org/10.30587/didaktika.v30i1.7387

[102] Murali, P., Hernandez, J., McDuff, D., Rowan, K., Suh, J., & Czerwinski, M. (2021). AffectiveSpotlight: Facilitating the Communication of Affective Responses from Audience Members during Online Presentations. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3411764.3445235

[103] Raca, M., & Dillenbourg, P. (2015). Translating Head Motion into Attention—Towards Processing of Student's Body-Language. *Proceedings of the 8th International Conference on Educational Data Mining*, 320–326. https://infoscience.epfl.ch/record/207803

EDC Proceedings 2025 ISBN-978-86-87043-96-1 DOI: doi: 10.52987/edc.2025

[104] Rothkrantz, L. (2017). An Affective Distant Learning Model Using Avatars As User Stand-in. *Proceedings of the 18th International Conference on Computer Systems and Technologies*, 288–295. https://doi.org/10.1145/3134302.3134314

[105] Sadanala, G., Xu, X., He, H., Bueno-Vesga, J., & Li, S. (n.d.). Enhancing distance learning with virtual reality: Investigating learners' engagement and outcomes. *Distance Education*, *0*(0), 1–28. https://doi.org/10.1080/01587919.2024.2351929

[106] Samandar, Z. (2024). UNDERSTANDING THE COGNITIVE PROCESSES OF DISTANCE LEARNING STUDENTS: IMPLICATIONS FOR EDUCATION AND INSTRUCTIONAL DESIGN. *World Bulletin of Social Sciences*, *32*, 71–73.

[107] Sun, J., Guan, N., Wang, X., Jin, C., & Chi, Y. (2019). Real-Time Scheduling and Analysis of Synchronous OpenMP Task Systems with Tied Tasks. *Proceedings of the 56th Annual Design Automation Conference 2019*, 94:1-94:6. https://doi.org/10.1145/3316781.3317891

[108] Torres, R. M. (2011). Lifelong learning: Movingbeyond Education for All (EFA). In *Conceptual evolution and policy developments in lifelong learning—UNESCO Digital Library*. https://unesdoc.unesco.org/ark:/48223/pf0000192081

[109] Vieira, D. (2019). Lifelong Learning and its Importance in Achieving the Sustainable Development Goals. In W. Leal Filho, A. M. Azul, L. Brandli, P. G. Özuyar, & T. Wall (Eds.), *Quality Education* (pp. 1–9). Springer International Publishing. https://doi.org/10.1007/978-3-319-69902-8_7-1

[110] Yang, H., Wang, C., Che, X., Luo, S., & Meinel, C. (2015). An Improved System For Real-Time Scene Text Recognition. *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, 657–660. https://doi.org/10.1145/2671188.2749352\

[111] Zeng, H., Shu, X., Wang, Y., Wang, Y., Zhang, L., Pong, T., & Qu, H. (2019). EmotionCues: Emotion-Oriented Visual Summarization of Classroom Videos. *IEEE Transactions on Visualization and Computer Graphics*, 1–1. https://doi.org/10.1109/TVCG.2019.2963659

[112] Zhou, Q., Suraworachet, W., & Cukurova, M. (2024). Detecting non-verbal speech and gaze behaviours with multimodal data and computer vision to interpret effective collaborative learning interactions. *Education and Information Technologies*, *29*(1), 1071–1098. https://doi.org/10.1007/s10639-023-12315-1

[113] Wichuda, R. (1999). Web-Based Learning: A New Alternative for Educational Technology in Thailand. *Journal of Education*, *27*(3), 29–35.