

StayOnTrack: Student Attention Scoring Using Facial and Behavioral Cue Analysis

Prof. N. G. Bhojne¹, Mansi Ashok Gavhale², Richa Hemant Ashtekar³, Samiksha Ramesh Darekar⁴,
Srushti Bibhishan Vhare⁵

Dept. of Computer Engineering, Sinhgad College of Engineering (SPPU), Pune, India

Abstract — Online education offers flexibility, but maintaining student attentiveness during virtual sessions remains challenging. This paper presents StayOnTrack, an AI-based real-time student attention monitoring system using webcam input to analyze engagement. The system applies Computer Vision and Machine Learning techniques to evaluate facial expressions, eye movement, gaze direction, and head orientation. Features such as Eye Aspect Ratio (EAR), gaze tracking, and head pose estimation are used to detect drowsiness and distraction. A Random Forest classifier categorizes students as attentive or distracted in real time. The system also provides alerts for prolonged inattentiveness and dashboards for teachers and parents. Built using Python, OpenCV, MediaPipe, TensorFlow, Flask/Django, and React.js, the framework ensures efficient performance while maintaining privacy by avoiding raw video storage. Results demonstrate improved engagement monitoring in online learning environments.

Key Words: *Student Attention Detection, Artificial Intelligence, Machine Learning, Computer Vision, Eye Aspect Ratio, Gaze Detection, Head Pose Estimation, Real-Time Monitoring.*

1. INTRODUCTION

With the development of digital technologies, and of Internet access and use, the use of online learning platforms and virtual classrooms has increased. Although online learning is flexible, keeping students engaged is a critical issue because of the lack of supervision, long hours in front of a screen, the ability to multitask, and the lack of interaction that can have an adverse impact on learning.

The breakthroughs in Artificial Intelligence (AI), Machine Learning (ML), and Computer Vision (CV) for real-time behavioral analysis have made it possible to track attention through real-time facial and visual data such as facial landmarks, gaze estimation, head pose analysis, and drowsiness detection.

In this paper, the authors introduce an artificial intelligence-based real-time student attention monitoring system for virtual classrooms, called StayOnTrack. By analyzing information from the students' webcam camera, the system is able to identify various behavioral indicators such as eye movements, gaze direction, facial landmarks, and head orientation, which will be used to determine whether a student is focused or distracted. Off-screen behaviour is detected using eye aspect ratio (EAR) and the gaze tracking and head pose estimation. The extracted features are fed to a Random Forest classifier to get real time classification of attention level.

The framework also contains alerts to help students stay focused and teacher and parent analytics to monitor trends in engagement through the dashboard. The system is implemented in Python, OpenCV, MediaPipe, TensorFlow, Flask/Django, React.js and MongoDB/PostgreSQL, with a focus on scalable and efficient real-time performance, while being privacy-focused to enhance engagement in online learning environments.

2. LITERATURE REVIEW

Because of limited supervision in online classrooms, it is difficult to monitor students' attentiveness in online classrooms. Monitoring attentiveness of students in online classrooms is difficult with the development of online education, due to the limitations of supervision. Artificial Intelligence (AI), Machine Learning (ML) and Computer Vision have been used to analyse student engagement.

Praveen et al. employed facial expression, eye movement and gaze behavior as features to detect attention but did not provide real-time alerting and monitoring features. Khan et al. created a framework for facial landmark and head pose estimation with a webcam and OpenCV and TensorFlow that required unobstructed, constant facial detection. Setiawan et al. proposed an adaptive monitoring system for drowsiness and gaze changes that needed high computation

resources. Lin et al. introduced a CNN-based attention monitoring system that achieved higher accuracy, but needed a massive amount of data and greater computation. Patel and Sharma emphasized on behavioral monitoring instead of continuous learning engagement during online examinations.

Current gaze estimation systems are working very well with facial landmark detection, gaze analysis and behavior analysis, but they have some drawbacks that include illumination changes, the quality of the webcam, head motion, facial occlusion and high computational complexity.

To counter these problems, StayOnTrack has developed an attention monitoring system for real-time use that utilizes Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), gaze direction and head pose estimation for multiple modalities of behavioral analysis. The system leverages efficient Computer Vision and Machine Learning methods to deliver a real-time experience, sends alerts and dashboard analytics to teachers and parents, and maintains privacy by not storing video data.

3. METHODOLOGY

The proposed StayOnTrack system is designed to monitor and analyze student attentiveness during online learning sessions using Artificial Intelligence (AI), Computer Vision, and Machine Learning techniques. The methodology focuses on real-time behavioral analysis through webcam input to detect student distraction, drowsiness, and engagement levels. The complete framework consists of data acquisition, facial landmark detection, feature extraction, data preprocessing, attention classification, alert generation, and dashboard analytics modules.

A. System Overview

The proposed **StayOnTrack** system continuously captures live webcam frames during virtual learning sessions to monitor student attentiveness and engagement. Each frame is processed using Computer Vision techniques to detect facial landmarks and extract behavioral indicators related to concentration, distraction, and drowsiness through eye movement, mouth activity, gaze behavior, and head orientation.

Extracted parameters, including eye movement, mouth activity, gaze direction, head orientation, and blink patterns, are analyzed using a Machine Learning classifier to determine whether a student is attentive or distracted. The system also incorporates alert mechanisms and role-based dashboards to help teachers

and parents monitor engagement trends and prolonged inattentive behavior.

The workflow begins with webcam-based frame acquisition, followed by facial landmark extraction using **MediaPipe Face Mesh**. Behavioral features such as **Eye Aspect Ratio (EAR)**, **Mouth Aspect Ratio (MAR)**, gaze direction, and head pose angles are extracted to create structured feature vectors. These features are then processed by a **Random Forest** classifier for real-time attention classification, after which alerts and dashboard analytics are updated for continuous monitoring.

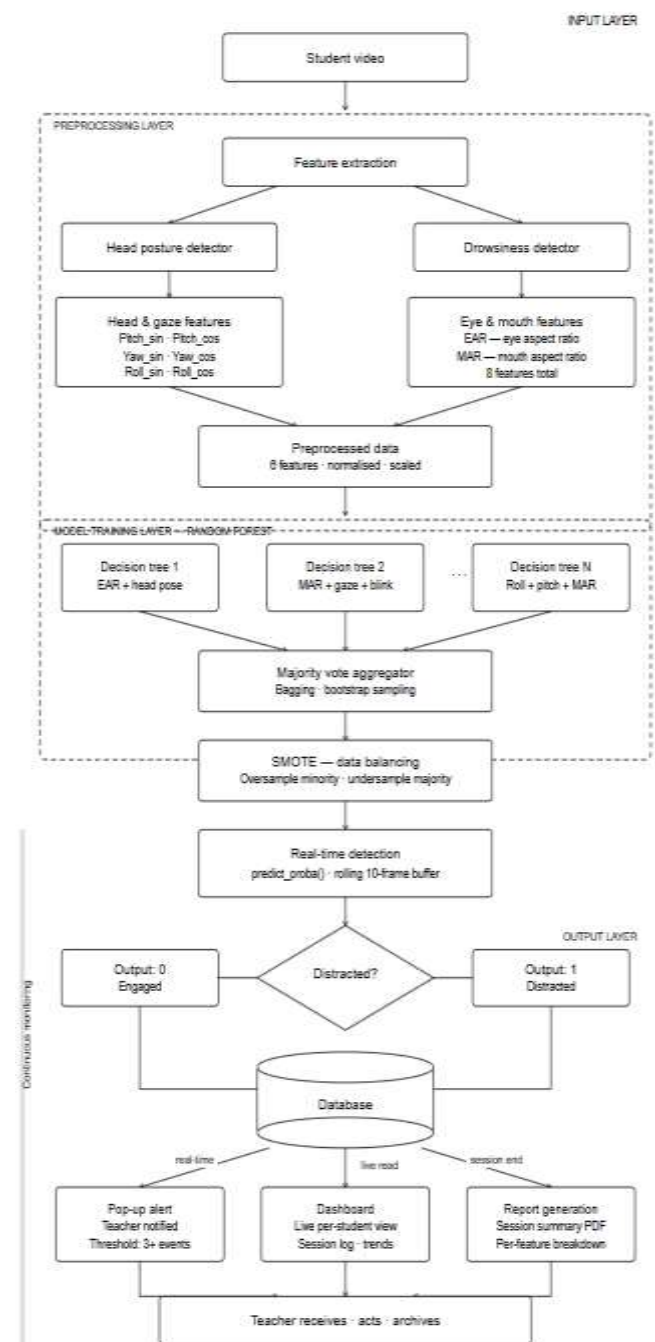


FIG 1: SYSTEM ARCHITECTURE

B. Data Acquisition and Video Processing

The first stage of the proposed methodology involves acquiring real-time webcam video streams from students during online classes. The video stream is processed frame by frame using the OpenCV library for image acquisition and preprocessing.

Initially, each webcam frame is resized to improve computational efficiency and converted from BGR to RGB format before being passed to the MediaPipe Face Mesh framework for facial landmark detection.

The MediaPipe Face Mesh framework detects and tracks 478 facial landmarks across facial regions, including the eyes, mouth, nose, and facial boundaries. These landmarks provide geometric information required for behavioural analysis and attention monitoring. Continuous frame processing enables stable and real-time attentiveness detection.

C. Feature Extraction

Feature extraction represents one of the most important stages of the proposed system. Behavioral indicators associated with student attentiveness are extracted from facial landmarks to identify distraction and fatigue.

1) Eye Aspect Ratio (EAR)

The **Eye Aspect Ratio (EAR)** is used to detect eye closure and drowsiness. EAR measures the ratio between the vertical and horizontal distances of eye landmarks.

The EAR is calculated using the following equation:

$$EAR = \frac{\|P2 - P6\| + \|P3 - P5\|}{2 \times \|P1 - P4\|}$$

Where:

- (P1) and (P4) represent horizontal eye corner points
- (P2, P3, P5,) and (P6) represent vertical eye landmarks

A decrease in EAR indicates prolonged eye closure, which may represent fatigue or inattentiveness. The system continuously monitors EAR values and generates drowsiness alerts when the EAR falls below a predefined threshold for a specified duration.

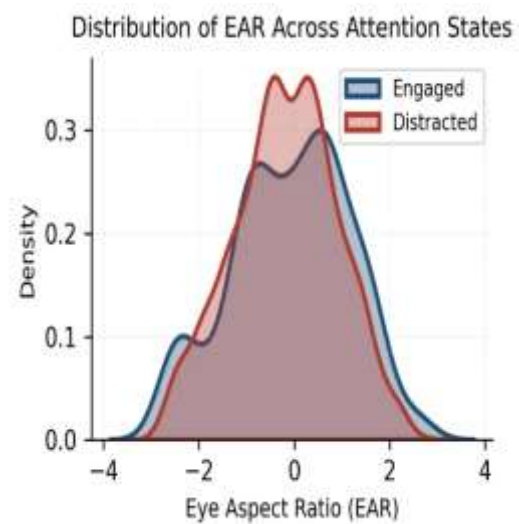


FIG 2: DISTRIBUTION OF EYE ASPECT RATIO (EAR) ACROSS ATTENTION STATES

2) Mouth Aspect Ratio (MAR)

The **Mouth Aspect Ratio (MAR)** is used to identify yawning behavior, which acts as an indicator of fatigue and reduced concentration.

The MAR is computed as:

$$MAR = \frac{\|P2 - P8\| + \|P3 - P7\| + \|P4 - P6\|}{2 \times \|P1 - P5\|}$$

Higher MAR values indicate increased mouth opening, which is associated with yawning activity. Repeated yawning behavior contributes to inattentive state detection.

3) Head Pose Estimation

Head Pose Estimation determines the orientation of a student's head in three-dimensional space. Facial landmarks such as the **nose tip, chin, eye corners, and mouth corners** are utilized with the **Perspective-n-Point (PnP)** algorithm to estimate head orientation.

The system computes three orientation angles:

- **Pitch** – Upward and downward head movement
- **Yaw** – Left and right head movement
- **Roll** – Head tilt angle

These angles help determine whether the student is focused on the screen or looking away, thereby assisting in distraction detection

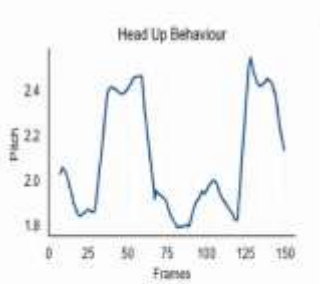


FIG 3: HEAD UP

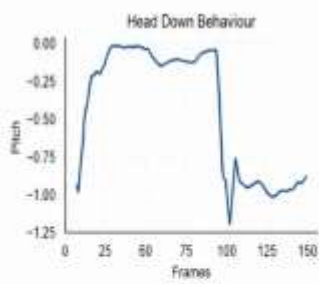


FIG 4: HEAD DOWN

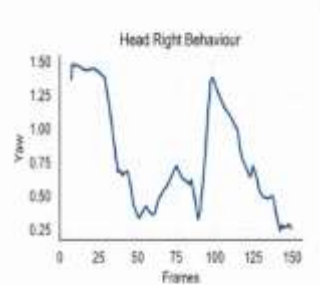


FIG 5: HEAD RIGHT



FIG 6: HEAD LEFT

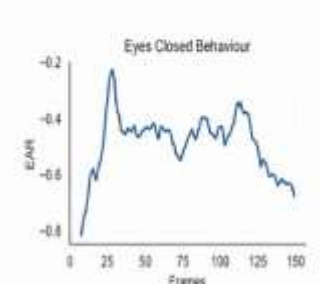


FIG 7: EYES CLOSED

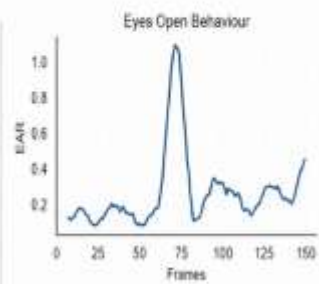


FIG 8: EYES OPEN

4) Gaze Direction Analysis

Gaze direction analysis is performed to determine whether students are visually focused on the screen. Eye landmark coordinates obtained from MediaPipe Face Mesh are analyzed to estimate eye orientation and viewing direction.

If a student continuously looks away from the screen for an extended duration, the system marks the behavior as inattentive. Gaze deviation serves as an important indicator of distraction during online learning sessions.

D. Data Preprocessing

After feature extraction, the behavioral feature vectors undergo preprocessing to ensure consistency and improve classification performance.

The extracted features include:

- Eye Aspect Ratio (EAR)
- Mouth Aspect Ratio (MAR)

- Gaze direction
- Head pose angles (pitch, yaw, roll)

Initially, the dataset is processed through **normalization and label mapping** to standardize feature representation.

Since the dataset contains class imbalance, the **Synthetic Minority Over-sampling Technique (SMOTE)** is applied to improve minority class representation and prevent model bias.

The processed dataset is then divided into **training and testing datasets using an 80:20 stratified split ratio**, ensuring balanced class representation.

E. Attention Classification

After preprocessing, the extracted behavioral feature vectors are supplied to the **Random Forest classifier** for student attention classification.

Random Forest is selected due to its **high classification accuracy, robustness to noise, ability to handle multiple behavioral features, and reduced overfitting characteristics**.

Based on extracted behavioral features, the system classifies students into two attention states:

- Attentive
- Distracted

The trained model performs **real-time attention detection** using webcam-derived behavioral data.

F. Alert and Notification Mechanism

To improve student engagement, the system incorporates an intelligent alert mechanism. Whenever continuous distraction or drowsiness is detected, immediate on-screen notifications are generated to encourage students to regain focus.

Teachers and parents can also receive alerts through the dashboard interface when inattentive behavior persists for extended periods.

G. Dashboard and Analytics Module

The proposed framework includes role-based dashboards for **students, teachers, and parents** to monitor attention-related activities.

The dashboard module provides:

- Real-time attention status
- Weekly engagement reports
- Alert history
- Student performance analytics

- Attention trends and behavioral statistics

These analytics assist teachers and parents in identifying students requiring additional academic attention and behavioral support.

H. Data Security and Privacy

Since the developed system processes webcam-based behavioral information, privacy and security are essential considerations.

The framework avoids storing raw webcam video streams and instead stores only extracted behavioral metrics and attention-related statistics. Secure authentication and encrypted communication mechanisms are implemented to ensure privacy-preserving operation and protect user information.

The proposed methodology combines lightweight Computer Vision techniques, Machine Learning classification, and real-time behavioral analytics into a scalable framework capable of improving attentiveness monitoring and learning effectiveness in online education environments.

4.DATASET AND ENVIRONMENTAL SETUP

A. Dataset Description

The proposed StayOnTrack framework utilizes the DAiSEE Dataset, a publicly available dataset designed for analyzing student engagement in online learning environments. It was selected due to its relevance to virtual classrooms, as it contains webcam-based student recordings captured during e-learning activities, making it suitable for real-time attention monitoring.

The dataset contains over 9,000 video snippets from approximately 112 users under unconstrained learning conditions. It captures natural student behavior with variations in illumination, head movement, facial expressions, camera angle, and posture, improving model robustness under realistic classroom scenarios.

The original dataset includes annotations for engagement, boredom, confusion, and frustration. For the proposed framework, these labels were reorganized into binary attention categories: Attentive (higher engagement) and Inattentive (distraction, low engagement, fatigue, or reduced concentration) to simplify real-time classification.

For reliable evaluation, the dataset was divided using a stratified 80:20 train-test split to maintain class distribution. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied only to the training set, improving classifier

generalization and reducing bias toward majority classes.

Additionally, behavioral diversity in gaze deviation, facial orientation, lighting, and webcam positioning helps evaluate the robustness of the proposed framework in realistic online learning environments.

B. Data Preprocessing

Several preprocessing operations are applied to prepare the dataset for training and testing:

1. Video frame extraction from webcam recordings.
2. Facial landmark detection using MediaPipe Face Mesh.
3. Feature extraction for EAR, MAR, gaze direction, and head pose estimation.
4. Removal of non-numeric and irrelevant attributes.
5. Feature normalization and scaling.
6. Binary label mapping for attention classification.
7. Dataset balancing using Synthetic Minority Over-sampling Technique (SMOTE).

SMOTE is applied only to the training dataset to handle class imbalance between attentive and distracted samples. This improves the classifier's ability to generalize across different behavioral conditions.

C. Experimental Environment

The proposed system is implemented using Python and integrated with several Computer Vision and Machine Learning libraries. The experimental setup includes both software and hardware components to support real-time behavioral analysis and attention monitoring.

Software Configuration

The following software technologies and frameworks are used:

- Python
- OpenCV
- MediaPipe
- TensorFlow/Keras
- Scikit-learn
- Flask/Django
- React.js
- MongoDB/PostgreSQL

Hardware Configuration

The experiments are conducted on a system with the following specifications:

- Intel Core i5 Processor or higher
- 8 GB RAM
- HD Webcam (720p or higher)
- Windows/Linux Operating System

The system is designed to process video streams in real time with minimal latency while maintaining stable performance during continuous monitoring sessions.

D. Machine Learning Model Configuration

Random Forest Classifier for Attention Detection

The proposed system uses behavioral and facial features extracted from the live feed of the web camera as the input to a Random Forest classifier for detecting student's attention. Random forest is an supervised ensemble machine learning algorithm which is the combination of multiple decision trees by the majority voting method to enhance the classification accuracy, robustness and generalization as compared to single decision tree.

Classifier: Bootstrapping (bagging) and feature randomness - random sample of samples and features to train each tree. This will help to decrease variance and prevent overfitting, and this is effective for real-world behavioral data that contain variations in facial expressions, lighting, posture, camera angle, and head movement throughout online learning.

The system extracts multiple behavioural features from the frames captured from the webcam, such as Eye Aspect Ratio (EAR) for blink and eye-closure analysis, Mouth Aspect Ratio (MAR) for yawn detection, head pose angles (pitch, yaw, roll) analysis, gaze direction, blink frequency and facial movement pattern. These feature vectors are fed into the Random Forest Classifier which distinguishes between the states of attentive or distracted.

The Random Forest classifier was chosen because it has a high classification accuracy, supports multiple behavioral features, is robust to environmental variability, has reduced overfitting, and it can also effectively handle class imbalance.

The model is set up as follows:

Stable ensemble classification: 100 trees.

2. Maximum Tree Depth: 12 levels (complexity control and reduction of overfitting).

3. Balanced Class Weights (applied to better detect minorities).

Stratified 4. Train Test Split (80:20) to maintain class distribution.

5. Parallel Processing: Multi CPU Cores for efficient Training.

Following feature extraction, the feature vectors are used for training classifier and real-time student attention detection from live camera input. Random Forest is a suitable model for monitoring students' engagement and distraction because of its accuracy, robustness, and efficiency.

E. Evaluation Metrics

The performance of the proposed system is evaluated using standard Machine Learning evaluation metrics:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

These metrics help measure the effectiveness of the model in correctly classifying attentive and distracted behavioral states. Real-time processing delay and system responsiveness are also analyzed to ensure smooth operation during live online learning sessions.

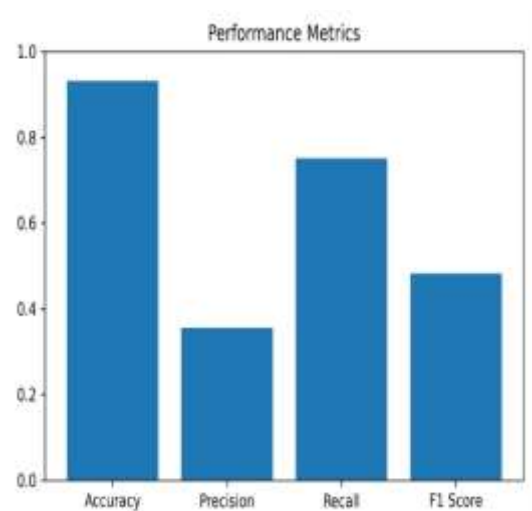


FIG 9: PERFORMANCE METRICS

F. Experimental Procedure

The experimental workflow follows these steps:

1. Capture webcam video streams.

2. Extract facial landmarks using MediaPipe.
3. Compute behavioral features such as EAR, MAR, and head pose angles.
4. Generate feature vectors for each frame.
5. Train the Random Forest classifier using processed dataset features.
6. Perform real-time attention detection.
7. Generate alerts and dashboard analytics based on detection results.

The proposed experimental setup enables efficient real-time student attention monitoring while ensuring scalability,

privacy, and reliable behavioral analysis for virtual learning environments.

5. RESULT AND DISCUSSION

To test the effectiveness of the proposed StayOnTrack framework, it was tested in a real time student attention monitoring during virtual learning sessions. The evaluation was directed to the accuracy of classification, the strength of the classification in different classroom environments, computational time and the effect of single behavioral indicators on the accuracy of the classification. Experimental results show that the combination of multimodal behavioural analysis can enhance the reliability of the attentiveness detection results, compared to single feature based approaches. The proposed framework involves two stages: the first is to extract the Eye Aspect Ratio (EAR), gaze direction analysis, mouth movement monitoring and head pose estimation features from the video stream, and the second is to classify student attentiveness in real-time based on these features. Standard Machine Learning evaluation metrics such as accuracy, precision, recall, F1 score, confusion matrix analysis, standard robustness testing and real-time execution performance of the system were used to evaluate the performance.

This graph, Fig. 10 is a Feature Correlation Heatmap. It is a simple measure of the relationship between the features that you extracted.

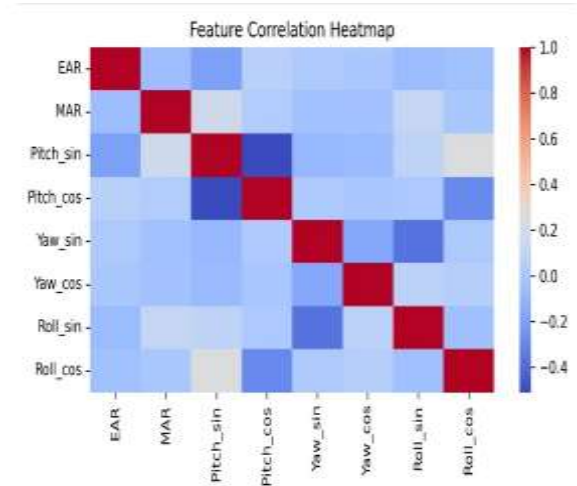


FIG 10: FEATURE CORRELATION HEATMAP

A. Quantitative Performance Evaluation

The proposed StayOnTrack framework achieved an overall classification accuracy of approximately 85% on the testing dataset, demonstrating strong capability in distinguishing attentive and inattentive student behavior during virtual learning sessions. The model exhibited particularly strong performance in identifying attentive states due to consistent behavioral characteristics such as stable gaze direction, proper head orientation, and regular eye activity. To evaluate classification performance in greater detail, standard Machine Learning metrics including precision, recall, and F1-score were analyzed. The obtained classification report is presented.

TABLE 1: PERFORMANCE EVALUATION

Class	Precision	Recall	F1-Score
Attentive	0.99	0.94	0.96
Distracted	0.36	0.76	0.48
Macro Average	0.67	0.85	0.72
Weighted Average	0.96	0.93	0.94

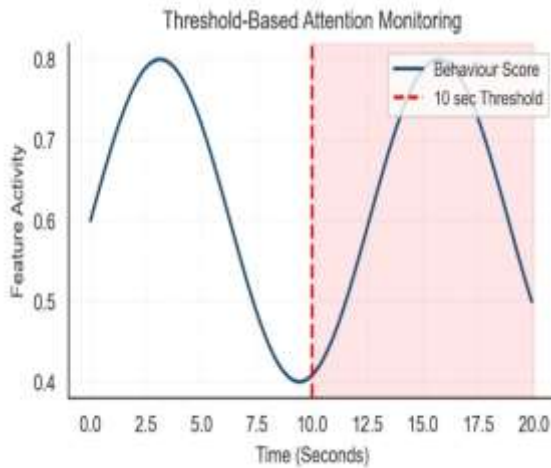


FIG 11: THRESHOLD BASED ATTENTION MONITORING GRAPH

The obtained results indicate that the proposed framework effectively recognizes attentive behavior with high precision and recall values. The recall observed for inattentive behavior remained comparatively strong, demonstrating that the system successfully detects distracted student activity in most scenarios. However, the comparatively lower precision for inattentive detections indicates the occurrence of occasional false-positive classifications.

B. Confusion Matrix Analysis

To evaluate detection reliability, confusion matrix analysis was performed to assess the framework's ability to classify attentive and inattentive states.

TABLE 2: CONFUSION MATRIX

Actual / Predicted	Attentive	Distracted
Attentive	2789	1789
Distracted	321	990

The results show that the proposed framework correctly classified a large proportion of attentive and distracted student samples, demonstrating reliable attention detection. Some misclassification occurred due to short-term distractions, temporary off-screen gaze, and natural head movements that resembled inattentive behavior.

Despite these challenges, the findings indicate that multimodal behavioral analysis improves classification stability and robustness, reducing major detection errors compared to single-feature approaches during continuous online monitoring.

C. The ROC Curve Analysis

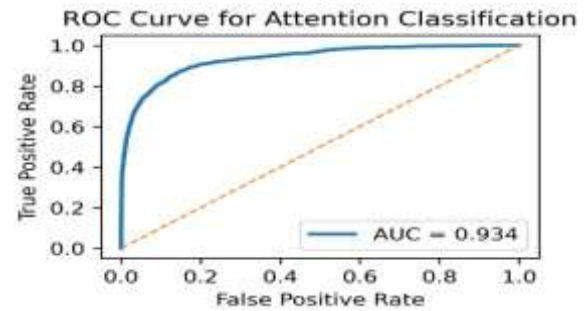


FIG 12: THE ROC CURVE

The **Receiver Operating Characteristic (ROC) Curve** evaluates the model's ability to distinguish between **attentive** and **distracted** student states by showing the relationship between the **True Positive Rate (TPR)** and **False Positive Rate (FPR)** at different thresholds.

The proposed **Random Forest** classifier achieved an **Area Under Curve (AUC)** score of **0.934**, indicating strong classification performance and effective discrimination between attention states. Since the ROC curve lies close to the upper-left corner, the model maintains a high true positive rate with a low false positive rate, confirming reliable attention detection.

D. Robustness Evaluation Under Practical Conditions

The proposed framework was tested in real-world online classroom environments to gain insights into the robustness of the proposed framework with respect to variations in the environment and the behavior of the students.

TABLE 3: SYSTEM PERFORMANCE UNDER DIFFERENT TESTING CONDITIONS

Testing Condition	Observed Performance
Normal Lighting	Stable attention detection
Low Lighting	Slight reduction in gaze estimation accuracy
Students Wearing Glasses	Minor facial landmark variation observed
Moderate Head Movement	Stable classification performance
Temporary Gaze Diversion	Occasional false inattentive detection

Testing Condition Observed Performance

Different Webcam Quality Moderate impact on landmark precision

The performance of the system is shown in Table 3 for various testing conditions. To verify the reliability of the framework the tests were conducted in different lighting environments, webcams of different quality, head movement, gaze deviation and spectacle usage. Results showed that it was consistently stable at moderate variations, but occasionally low illumination and sudden movements had some impact on the consistency of detection. The multimodal framework, however, diminished reliance on a single behavioural component, making for better overall robustness.

E. Comparative Analysis with Existing Systems

A comparative evaluation was conducted between the proposed StayOnTrack framework and existing student attention monitoring approaches reported in recent literature. The comparison considers methodology, classification accuracy, and real-time deployment capability.

TABLE 4: COMPARATIVE ANALYSIS

System	Accuracy	Real-Time	Multimodal	Lightweight	Dashboard Support
Praveen <i>et al.</i>	88%	Partial	Yes	No	No
Khan <i>et al.</i>	90%	Yes	Partial	No	No
Lin <i>et al.</i>	91%	Limited	Yes	No	No
Patel and Sharma	89%	No	Partial	No	No
StayOnTrack	85%	Yes	Yes	Yes	Yes

Although several existing approaches achieved comparatively higher classification accuracy, many rely on computationally intensive architectures and limited deployment flexibility. In contrast, the proposed StayOnTrack framework emphasizes real-time execution, lightweight computation, and multimodal behavioral monitoring, making it more suitable for practical virtual classroom environments.

F. Real-Time Performance Evaluation

The proposed framework was evaluated under real-time classroom conditions using an HD webcam and mid-range computing hardware. The system maintained

smooth monitoring performance during continuous execution.

TABLE 5: REAL TIME PERFORMANCE EVALUATION

Parameter	Observation
Frame Processing Rate	24–28 FPS
Alert Generation Latency	Less than 1.5 seconds
CPU Utilization	Moderate
Dashboard Update	Stable and Responsive

The obtained results confirm that the framework supports real-time deployment without noticeable latency, enabling continuous attentiveness monitoring during virtual learning sessions.

G. Ablation Analysis

To analyze the contribution of individual behavioral indicators, separate experiments were conducted using isolated features.

TABLE 6: ABLATION ANALYSIS

Configuration	Accuracy
Only EAR	81%
Only Head Pose	84%
Only Gaze Detection	83%
Multimodal Model	85%

The results demonstrate that the multimodal framework consistently outperforms single feature approaches, improving detection reliability and reducing misclassification through combined behavioral analysis.

H. Discussion and Limitations

The experimental results demonstrate that the proposed StayOnTrack framework effectively identifies student attentiveness in virtual learning environments while maintaining stable real time performance. The integration of EAR, gaze detection, and head pose estimation improves behavioral understanding and attentiveness detection. However, factors such as lighting variation, webcam quality, facial occlusion, and

temporary gaze deviation occasionally affected detection consistency.

Despite these limitations, the proposed framework provides a practical, lightweight, and scalable solution for student attention monitoring with real-time alerts and dashboard-based engagement tracking.

6. CONCLUSIONS AND FUTURE WORK

A. Conclusion:

This paper suggested the attention monitoring system named as StayOnTrack that monitors student attention in real-time for improving student engagement in online learning environments using artificial intelligence. The framework uses Computer Vision and Machine Learning methods to make inferences from the behavioral clues of the eye movements, gaze direction, facial landmarks, and head pose based on the input from a webcam. The use of Eye Aspect Ratio (EAR), gaze detection and head pose estimation provided a good measure of attentive, distracted and drowsy states. Experimental results demonstrated good real-time performance for attention classification under low computational demands and also efficient and practical performance for virtual classrooms using the Random Forest classifier. There are also alert systems and dashboard analytics for teachers and parents to track engagement trends in the system. Also, no raw video data are saved, using a privacy-focused approach, only behavioral metrics. In conclusion, StayOnTrack shows that AI-driven behaviour analysis can be a useful tool for enhancing attentiveness monitoring in virtual classrooms.

B. Future Work :

Several improvements can be incorporated in future versions of the developed system. Advanced Deep Learning models such as YOLOv8, Vision Transformers (ViT), and CNN-based architectures can be integrated to improve detection accuracy under complex conditions.

Future work may also include multi-student classroom monitoring, speech and emotion recognition, and integration with Learning Management Systems (LMS) such as Google Classroom and Moodle. Cloud-based deployment and mobile application support can further improve accessibility and scalability.

Additionally, personalized AI-based learning analytics can help generate individual attention profiles and provide adaptive learning recommendations for students. These enhancements can transform

StayOnTrack into a more intelligent, scalable, and comprehensive educational monitoring platform for next-generation virtual classrooms.

7. REFERENCES

- [1] M. A. Khan, A. Raza, and M. Ahmad, "Student Attentiveness Monitoring: A Surveillance System for Virtual Classes," *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 7, no. 5, pp. 122–129, 2023.
- [2] K. Praveen, R. Anitha, and S. Ghosh, "Machine Learning Applied to Student Attentiveness Detection Using Emotional and Non-Emotional Measures," in *Lecture Notes in Networks and Systems*, Springer, 2023, pp. 645–654.
- [3] A. Setiawan and R. Nurul, "An Adaptive System for Predicting Student Attentiveness in Online Classrooms," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 1, pp. 91–100, 2023.
- [4] Z. Lin, H. Zhao, and J. Chen, "Real-Time Attention Monitoring System for Classroom: A Deep Learning Approach for Student Behavior Recognition," *Big Data and Cognitive Computing*, vol. 7, no. 1, pp. 31–39, 2023.
- [5] H. Patel and R. Sharma, "An Automated Online Proctoring System Using Attentive-Net to Assess Student Mischievous Behavior," in *Advances in Artificial Intelligence and Data Engineering*, Springer, 2023, pp. 819–829.
- [6] A. Alqahtani and J. Smith, "Deep Learning for Cognitive Attention Recognition in E-Learning Environments," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 2, pp. 78–85, 2023.
- [7] A. Kumar and P. Das, "Attention Detection Using Facial Features and Eye Tracking in Virtual Classrooms," *Journal of Artificial Intelligence Research*, vol. 82, no. 4, pp. 555–566, 2023.
- [8] V. Reddy and K. Raj, "Enhancing Virtual Learning with Real-Time Attention Detection Using Deep Learning and Computer Vision," *IEEE Access*, vol. 12, pp. 4332–4345, 2024.
- [9] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, vol. 25, no. 11, pp. 120–126, 2000.
- [10] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition

and Clustering,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 815–823.

[11] T. Baltrušaitis, P. Robinson, and L. Morency, “OpenFace: An Open Source Facial Behavior Analysis Toolkit,” in Proc. IEEE Winter Conf. Applications of Computer Vision (WACV), Lake Placid, NY, USA, 2016, pp. 1–10.

[12] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems (NIPS), vol. 25, 2012, pp. 1097–1105