International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I8P135

Volume 72 Issue 8, 378-388, August 2024
© 2024 Seventh Sense Research Group®

*Original Article*

# Optimizing Educational Interaction: An Advanced Real-Time Attention Recognition in Online Education Environment

Namita Shinde[1], Mayur Dilip Jakhete[2], Shekhar Shinde[3], Archana Vyas[4], Rajesh Ramdas Karhe[5], Neeru Malik[6], Pooja Deshmukh[7], Kedarnath Chaudhary[8]

[1,3,7]*Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India.*
[2]*Pimpri Chinchwad University, Pune, Maharashtra, India.*
[4]*Dr. Rajendra Gode Institute of Technology and Research Amravati, Maharashtra, India.*
[5]*Mauli College of Engineering, Shegaon, Maharashtra, India*
[6]*School of Engineering & Technology, Pimpri Chinchwad University, Pune, Maharashtra, India.*
[8]*COEP Technological University Pune, Maharashtra, India.*

[2]*Corresponding Author: jakhetemayur@gmail.com*

Received: 27 March 2024 Revised: 16 July 2024 Accepted: 31 July 2024 Published: 28 August 2024

***Abstract*** *- It is important to recognize a student's emotional condition in both traditional and virtual learning settings. In order to solve this difficulty, this research suggests a novel method that uses patterns of eye and head movements to infer emotions and measure learner engagement. We emphasize the need for enhanced system efficacy, value, and user interaction by using sophisticated emotion recognition algorithms in our approach. Our goal is to assess how deeply a student is involved in educational activities. In online learning environments, the suggested system not only detects and tracks students' attention in real time but also sets up a feedback mechanism for improved material delivery. We measure the degree of focus exhibited by a student by closely examining their head and eye movements. The system uses graphs to classify and display this data, providing insightful information about student interest and involvement. Subsequently, this information is relayed to educators as feedback, enabling them to optimize the learning environment for a more tailored and effective educational experience.*

***Keywords*** *- Image recognition, Face recognition, Education Tools, Convolutional neural networks, Computational intelligence.*

## 1. Introduction

As we know outbreak of COVID-19 has affected the education sector in many ways, compounding the education process from conventional face-to-face learning taking place in class to online education. It has been shown that there is a problem in assessing the student’s progress and their level of comprehension, especially when the course is delivered online. Current online learning analytics are largely focused on quantitative data; still, it include the amount of time which clients spend on educational websites, the number of interventions in the forum, etc., all of which cannot comprehend the emotional engagement of students. This research fills this important void by proposing an innovative system that, with the help of complex machine learning and computer vision methodologies, is capable of precisely ascertaining students’ eye and head activity, representing their emotions and engagement levels, respectively, in real-time. One of the major difficulties of applying virtual learning environments is the inability to quickly and qualitatively assess students’ interest in the learning process. It is impracticable to use traditional approaches to track small signs of emotions, such as lack of interest or yawning, which are fundamental indicators for improving the efficiency of students’ teaching and learning processes. These interactions are still lacking in modern technology-based learning solutions, although technology’s goal is to solely engage humans at the surface level of learning without considering facets of deep cognition and feeling. This research, therefore, seeks to close this gap by proposing a complex system that observes the students’ facial expressions and blinked eyes in a systematic manner and with consistency in order to establish a higher degree of comprehension of their engagement. The aforementioned pattern refers to the proposed system that will employ the newest computer vision approaches and machine learning methods to analyze and recognize the students’ head orientation and gaze direction. Intriguingly, by using deep learning architectures like CNNs and RNNs, the system is capable of learning slight changes in expressions and the movements of eyes expected at different levels of engagement, boredom or even drowsiness. This approach goes beyond the usual online learning analytics to monitor and give a detailed assessment of students' learning status based on their cognitive and emotional structures to give a comprehensive assessment instead of the general and ordinary assessment given by ordinary learning analytics. The first unique feature of this work is the likelihood of empowering educators since the app monitors students' activities in real time and continuously. The developed system produces graphic representations of attention values considering time; however, the generated data is sent by e-mail to educators. Such visualizations facilitate awareness by instructors, for instance, moments of either high or low level of student engagement, thus facilitating appropriate interventions and instructional modifications.

This real-time feedback mechanism does augment the learning effectiveness of web-based learning environments that educators foster as well as afford them the ability to create more effective web-based learning courses. Besides, in relation to the offered system, it is necessary to consider the pragmatic aspects of the proposed research. This paper analyses the practicality of the proposed model in terms of usability and user experience, where it considers the opinions of both the teachers and the students and the effect of teacher feedback on the amount of students' activity, as well as discusses the ethical issues and concerns related to constant monitoring. Implementing an intervention system known as the Student Engagement Detection System (SEDS), the authors aim to create individualized solutions based on real-time attention data to increase students' engagement during online classes.

In addition, this paper outlines guidelines on how the system will be operated in a responsible manner in the collection and use of the data to enhance students' learning without infringing on their privacy. This research proposed a new generation solution to the online learning engagement problem by IA-based emotion recognition algorithms and real-time monitoring uncertainties. It can be argued that the proposed system provides a solution that was missing in the educational technologies already in use and, at the same time, paves the way for building a better environment for learning. With detailed information and practical recommendations, this study tries to transform the online learning paradigm so that the processes would be more effectively aligned with the student's emotional and cognitive capabilities.

## 2. Motivation for the Research

The field of intelligent technologies, most especially computer vision, has enabled the examination of specificities of the student's learning behaviors in the online environment. It is possible to state that understanding the subtleties between facial expressions and head gestures can help benefit the community of online learners. Various systems that can be applied to the field of face feature detection, such as Gesture recognition systems – CNN (Convolutional Neural Network) and HOG (Histogram of Oriented Gradient), started competing. Although such improvements have been observed in the given technologies, some drawbacks have been pointed out by researchers in previous work. Most of the existing approaches, for instance, rely on static images and do not include necessary dynamic properties important for real-time monitoring of students' interactions in an online environment. However, the prior studies fail to recognize the context that affects gesture recognition, and there is a lack of a theoretical framework to link the identified gestures to learning outcomes. The purpose of this research stems from the gaps in knowledge relevant to effective online learning behaviors to be able to figure out to which extent gesture recognition algorithms can help, as outlined by the following questions. Thus, the study will seek to: Establish the steps that would be necessary for using such algorithms to identify online learning behaviors (RQ1), Identify the peculiarities of the students' gestures while learning in an online environment (RQ2), understand factors that might influence the appearance and/or growth of these gestures during connected online lessons (RQ3). In order to do this, we use a very successful deep-learning approach that combines head pose estimation with face landmarks recognition. This approach combines calculated and rational techniques, such as the algorithm with head movements and facial expressions, that are known to have resulted in performance issues in an online learning environment suited to the learners' requirements.

To increase the accuracy of the features that needed to be identified, a pilot recording was made to capture lots of gestures and facial information. In the paper, the researcher has outlined five main motions, which include leaving, shaking, yawning, nodding, and blinking and also recommended detection methods for each of the motions. This approach is useful to eliminate the shortcomings of earlier research on the aspect of gesture recognition in an online learning environment by considering the dynamic and contextual perspective. Lastly, this project aims to address a gap in the literature by creating a recognition algorithm that provides valuable information regarding the students' online learning activity. Thus, due to the problematics identification and the consideration of the limitations of the prior research, as well as the introduction of novel ideas into studying the topic, the present study aspires to contribute to the shift of the future of online education, making it more efficient and adaptive.

### *2.1. Contributions of the Study*

Specifically, to analyze the objects and extract the online learning behaviors which were presented by college students in virtual learning environment learning activities, the present research introduced a new method. Using this technique, the developed programmer deftly combines head gesture detection with facial feature recognition. This approach presented a modified categorization system comprising five gestures: blinking, yawning, nodding, shaking the head, and leaving are the common behaviors that have been observed to express herself among the students. It also improved the adaptation of current algorithms to the online situation and the unique characteristics of the sampled population.

The following significant advances are made in the field of real-time online learning behavior research by this study (Contributions of the Study):

- Deeper Study of Online Learning Behavior: The study is the first to include the use of AI algorithms for the systematic collection of data on learning behavior and, thus, the understanding of students' interactions within online spaces.
- Building Blocks for Personalized Learning Design: The accumulated data on online learning behavior provides a starting point for creating teaching and learning plans.
- Enhancing Evaluation Techniques: The research Adds to the repertoire of techniques and metrics for evaluating online instruction. This, in turn has consequences for how frameworks and standards for online learning are constructed in the future.
- Extending Human-Machine Interaction in Education: The findings expand the use of human-machine interaction in the context of distance education. The efficient gathering, Pre-processing, and analyzing data on online learning behavior is made possible by this integration.

## 3. Related Work

The identification of the online learning status and behavior of students represents a critical challenge with profound implications for enhancing the effectiveness of online teaching. Despite the widespread adoption of real-time video lectures, there remains a notable dearth of comprehensive quantitative research in this domain. Addressing this gap is crucial in understanding and improving the quality of online teaching. Understanding the attitudes and intents of learners is beneficial for teaching aid systems to anticipate their requirements and deliver proactive services. Human bone gesture recognition data has been used to provide feedback on learning status in a variety of remote teaching and conference contexts.

Here, we add to the head gestures of students enrolled in online courses to investigate their behavior in a completely distant environment [7]. The integration of computer vision in education has long been a focal point of research. Previous studies have explored the utilization of various biological signals, such as facial expressions, pulse frequency, and breathing rate, to collectively estimate learners' emotional states and interest levels [8]. Additionally, web cameras have been employed to capture learners' facial expressions and movements, providing valuable insights into their interest levels [9]. Its alternatives consist of evaluating students' attention by computing and comparing physical and geometrical characteristics of the head gesture images through the means of regression iterative algorithms [10]. Despite these notable advancements, there exists a current gap in quantitative studies, particularly during the ongoing epidemic where real-time video is extensively employed in the education sector. Addressing this gap becomes imperative, presenting an opportunity for researchers to delve into the quantitative aspects of students' learning states and behaviors in the context of widespread real-time video usage in education.

In the study [11], Shen et al. used physiological data, including Heart Rate (HR), Skin Conductance Response (SCR), Blood Volume Pulse (BVP), and EEG sensors to identify emotions within the context of online learning. Their unique strategy was to use a Support Vector Machine (SVM) promising model specifically for the valence-arousal space. The authors were able to convincingly explain how emotional awareness technologies can improve students' ties and interactions.

However, real-time visualization was not included in the applied method, either. Tang [12] introduced a Convolutional Neural Network (CNN) model, experimenting with replacing the layer of softmax with a linear support vector machine. Although real-time visualization was absent, their approach demonstrated significant gains, achieving an accuracy of 69.3% on the FER-2013 dataset. In [13], the authors introduced an FPGA-based design utilizing a trained CNN for facial emotion identification on the FER-2013 dataset.

Thus, the focus was placed on furthering the development of hardware solutions for emotion recognition. The authors of [14] combine these fields to allow CDC to be customized to a student's level. Based on the facial expression states obtained from their CNN model and various parameters defining the responses of students, the fuzzy part of the proposed model chose the next level of learning. The effectiveness of this concept has been confirmed with the provision of the learners with an opportunity to engage in progressive learning that depends on the ability of the individual learners. Recently, Nezami et al. [15] introduced a new model designed to enhance the recognition of engagement by photos while addressing data scarcity issues. First, deep learning was utilized to train an FER model, which is converted to an engagement model.

This model was capable of paying proper attention and identifying students' engagement and disengagement based on the stipulations made above, which held a lot of positive contributions to the pedagogy. Understanding of student involvement. Bustos-López, Maritza, et al [16] concentrated on employing wearable sensors and machine learning techniques to monitor students' attention in real time. They aimed to utilise physiological indicators, such as skin conductance and heart rate, to gauge students' attention spans during lectures.

The study effectively shows that employing wearables for this function is feasible. Nonetheless, difficulties included the requirement that students wear their devices all the time, the possibility of discomfort, and restrictions on the kinds of physiological signs that could be accurately recorded in a classroom. Convolutional Neural Networks (CNNs) were utilised by S Kapil et al [19] to measure students' attention in classroom environments by analyzing their facial expressions.

**Table 1. Comparative analysis of methods for recognizing and monitoring student engagement in online learning**

| Reference | Methodology Used | Algorithm Used | Features | Results | Limitations |
|---|---|---|---|---|---|
| [1] | Wearable device for PPG signal measurement | Decision Tree | Photo plethysmogram (PPG) | Attention level evaluation | Expensive as devices wearable devices involved |
| [3] | Intelligent cap with pen and gyroscope | Random Forest | Head motion | Engagement tracking | Practical discomfort, high cost |
| [5] | Body movements and eye gaze direction analysis | K-Nearest Neighbors (KNN) | Body movements, eye gaze | Real-time attention report | Unable to identify detailed behavioral patterns |
| [6] | Virtual reality systems | Convolutional Neural Network (CNN) | VR environment | Attention tracking in remote learning | Not suitable for classroom settings |
| [7] | Human bone gesture recognition | K-Nearest Neighbors (KNN) | Head gestures | Investigated student behavior in remote environments | Lack of real-time emotion analysis |
| [8] | Analysis of biological signals | Hidden Markov Model (HMM) | Facial expressions, pulse frequency, breathing rate | Estimated emotional states and interest levels | Lack of quantitative studies |
| [9] | Web camera analysis | Support Vector Machine (SVM) | Facial expressions and movements | Insights into interest levels | Limited by the absence of real-time visualization |
| [11] | Physiological data analysis | Support Vector Machine (SVM) | HR, SCR, BVP, EEG | Enhanced student connection and engagement | No real-time visualization |
| [13] | FPGA architecture | CNN | Facial expressions | Advanced hardware implementation of emotion recognition | Hardware-specific limitations |
| [15] | Engagement recognition from photos | CNN | Facial expressions | Improved engagement recognition | Data sparsity challenges |
| [16] | Wearable sensors | Random Forest | Skin conductance, heart rate | Real-time attention monitoring | Discomfort, limited physiological signals |
| [18] | Comparative analysis of ML techniques | Random Forest, SVM, k-NN | Various physiological signals | Evaluated attentiveness | Individual differences need for large-scale data |
| [20] | Real-time detection using facial expressions and EEG | Long Short-Term Memory (LSTM) | Facial expressions, EEG | Accurate attention detection | Challenges in interpreting facial expressions, EEG calibration |

The aim was to improve attention monitoring by using facial cue interpretation. The study proved how well CNNs could pick up on nuanced facial expressions that indicated attention levels. Nevertheless, obstacles encompassed possible discrepancies in distinct facial expressions and the requirement for sturdy training datasets to enhance the precision of the model. Gupta et al. [17] compared many machine-learning strategies for physiological signal-based student attentiveness monitoring. The aim of the study was to evaluate the accuracy with which various algorithms could estimate the levels of attentiveness. J. Santhosh and associates [18] used a multi-modal strategy, combining physiological sensors and eye tracking to provide thorough attention monitoring of students. The study effectively illustrated how eye tracking and physiological data might work together to improve attention assessment accuracy.

Challenges include the need for synchronized data collection and potential complexities in integrating diverse sensor modalities. Using facial expressions and EEG, Hassouneh, Aya, et al [20] research successfully created a real-time system for detecting students' attention. To overcome the existing limitations of identifying students’ attention, the authors used a system they created using facial expressions and EEG signals that analyze students’ attention correctly and objectively. In [1], a PPG signal obtained from the wearable device specifically has an average accuracy of 0. 69, thus offering a realistic front end for both the trainers and the learners as applicable in the context of a massive open online course. The study's findings [2] offer a useful tool for choosing and creating video courses for online education. Additionally, the system can be expanded to offer suitable modifications for ubiquitous learning environments as well as classroom settings.

Skeletal pose estimation and person detection algorithms. Their approach classified actions based on the number of pupils in a classroom by using a deep neural network and the Open Pose framework for skeleton identification. An intelligent cap that monitors head motion has also been utilized to gauge students' attention [3]. It was made out of a pen and gyroscope that tracked the degree of involvement. In real/long-term practices, wearing such devices during lectures is deemed uncomfortable, even though modern wearable-based technologies may yield correct findings [4]. These pricey techniques might not be feasible in the long run, especially when it comes to long-term equipment maintenance.

B. Ngoc Anh [5] calculated attention and produced a real-time report on classroom attention during lectures using body movements and eye gaze direction. However, the algorithm was unable to identify behavioral patterns like body Posture and facial expressions or to offer information like emotions. Virtual reality systems were utilized by D.M. Broussard [6] to track students' attention in a remote learning setting, but they weren't appropriate for classroom settings.

Table 1 highlights a comparative analysis of methods for recognizing and monitoring student engagement in online learning. One major drawback is that most methodologies lack real-time adaptability; they are static image analysis based, which does not evolve with g. time. Many studies do not take into consideration the surrounding context that influences gesture recognition— let alone come up with a framework that ties gestures to learning outcomes in a holistic manner. Even though some systems have been developed using advanced technology like wearable sensors, a lot of issues still remain unresolved due to their complexity — for example, discomfort in wearing sensors continuously because it leads towards large-scale data collection besides model training and many sensor modalities' integration. In the future, research should focus on overcoming these limitations by developing adaptive systems in real-time for monitoring and acknowledging contexts in which learning takes place: this includes demanding more extensive datasets from model developers about student engagement or using multimodal sources to provide a comprehensive picture of engagement scenarios.

## 4. Proposed Work

The conceptual framework of the proposed system, as depicted in Figure 1, unfolds in three key modules, each thoughtfully tailored for the dynamics of online class settings. The primary module orchestrates the seamless acquisition of real-time images through a camera, finely tuned to the nuances of the virtual learning environment. These captured images lay the groundwork for subsequent in-depth analysis. The second module takes center stage with the implementation of a sophisticated HAAR CASCADE-based system. This intelligent system is multifaceted, incorporating modules for action behavior detection, emotion analysis, and facial recognition, all calibrated to the distinctive context of virtual classrooms.

In the realm of online education, a dedicated camera is strategically positioned to capture the ebb and flow of virtual class interactions. The camera feed is intricately woven into a desktop computer setup, fostering continuous monitoring through the lens of the meticulously crafted model. The third and final module is dedicated to the meticulous evaluation of the system's efficacy within the online class milieu. Here, the trained model diligently processes the stream of captured images, giving rise to live reports that serve as a rich tapestry of insights. These reports, offering a granular view of students' actions, attention dynamics, and emotional states, are seamlessly transmitted to instructors. The system's outreach extends to both individual students and the collective class, ensuring a comprehensive understanding of the online learning landscape. In alignment with the virtual paradigm, the system's design places a premium on accessibility.

User-friendly interfaces take center stage, ensuring a seamless and intuitive experience for all participants engaged in the virtual educational journey. This holistic approach underscores the system's commitment to enhancing the online learning experience, fostering engagement, and providing educators with valuable tools for effective virtual instruction. Experiencing boredom, engaging in eating/drinking, laughter, reading, phone usage, getting distracted, and writing are classified as activities indicative of low attention levels. The automatic detection and real-time illustration of these behaviors, along with the instructor's actions, enable continuous awareness of the student's current attention status. This dynamic feedback mechanism empowers the instructor to make timely adjustments in their teaching approach, ensuring the student is brought back to a state of high attention, as shown in Figure 1.

The final phase is the development of metrics that take the shape of graphs, which will then be plotted to determine the intensity and length of attention and actions, in particular, tracking the number of times an individual nods, blinks or yawns. This is done through a four-stage process that includes pre-processing, built-in processing, score computation and feedback. The first pre-processing stage plays a major role owing to the fact that real online classroom settings are uncontrolled, which results in issues like transformation of posture, variability of lighting conditions, and image quality degradation problems— an effective pre-processing can consequently address these issues. It will also be able to uplift the recognition accuracy as well as strengthen the resilience of the SEDS model.

### *4.1. Preprocessing*

The first action in the pre-processing phase is the extraction of frames from video data. OpenCV-python is used to accomplish this task by capturing particular keyframes from a video file or live stream via webcam. The process of frame extraction enables the recognition of significant points in the video which hold importance for additional scrutiny. Incorporating resizing and grayscale conversion: Resize the following generation of batteries to deliver energy densities greater than present technology (lithium-ion), as well as less chance of fire. Once the frames have been extracted, we resize them after the proper dimensions are established.

The resizing step resizes all frames to a consistent size that is appropriate for further processes in terms of quality and computational cost. After resizing, we transform the images from BGR to grayscale. This transformation converts the image from color to black and white only, making it less complex in nature— therefore, easier for processing without compromising information that can be used in image recognition applications.

Feature-Based Detection: In this particular sub-stage, the pre-processed frames are put through a feature-based detection. This can be done by introducing calibration of raw frames which helps to make data ready for more analysis. Grayscale transformation facilitates effective detection by the system to draw facial key features and gesture details; in addition, it rids noise plus disparities with the aim of producing clean data appropriate for deep learning and machine learning use cases. Figure 2 shows the basic design process of the system.

### *4.2. Database*

This research was carried out by collecting data from seven volunteers who were asked to record a 20-minute video in real-time with the help of their webcams. The participants were told to keep the camera in one spot and make sure that the upper part of their body was visible mainly the face and eyes. After the videos were collected, they were split into two-second clips for analysis.

This duration enabled human annotators to label the data without changing the behavior too much. To tackle class imbalance and thus boost model accuracy, we modified the dataset by redefining behaviors to include instances of participants talking while being focused or not being focused. Videos with unclear states were added to the dataset to keep the information clear. The dataset, which was 4.2 gigabytes in size, had 59,143 'Focused' and 45,962 'Not Focused' video samples, respectively. In Figure 3 we can see a taxonomy of actions and behaviors considered to categorise responses.

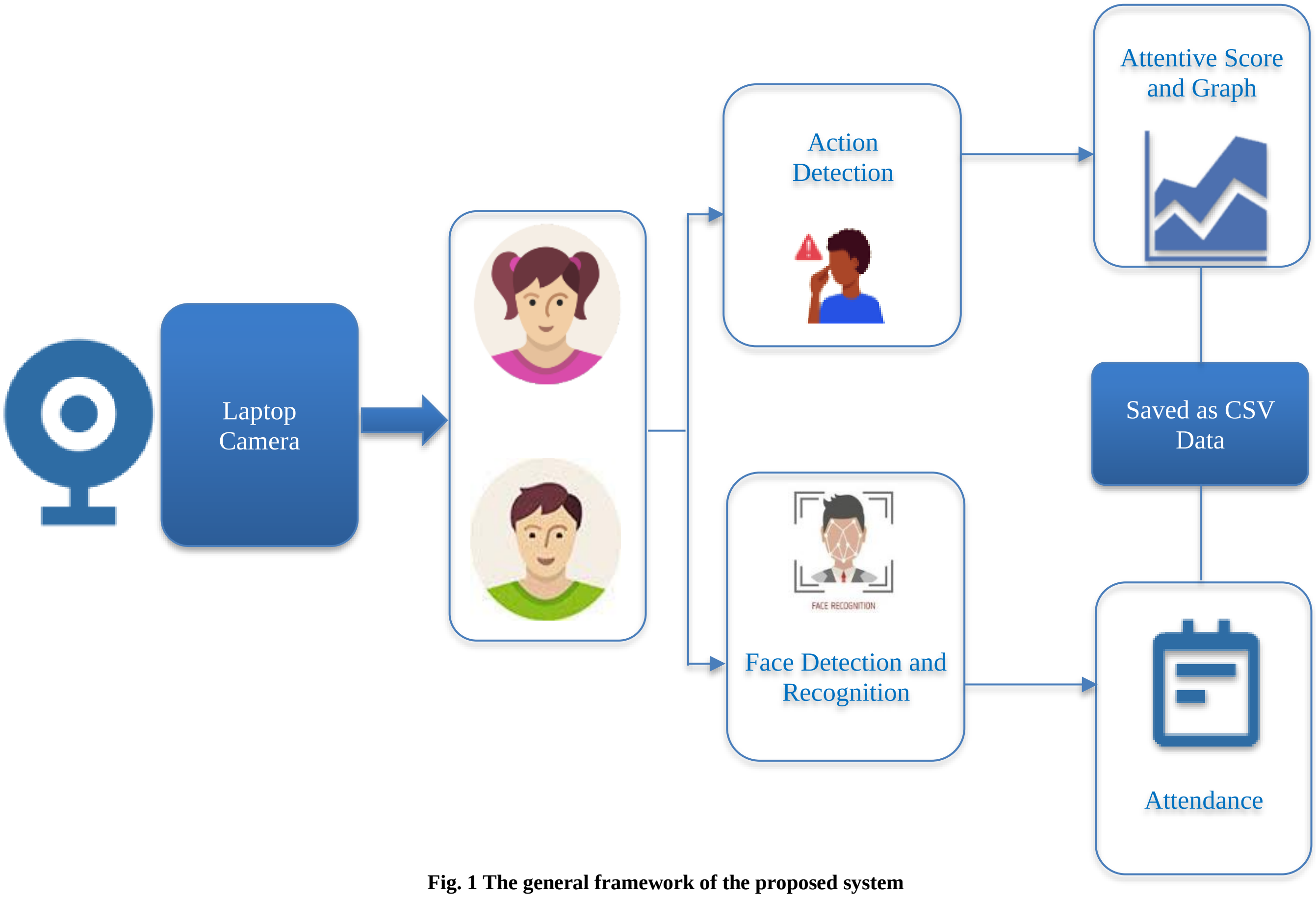


**Fig. 1 The general framework of the proposed system**

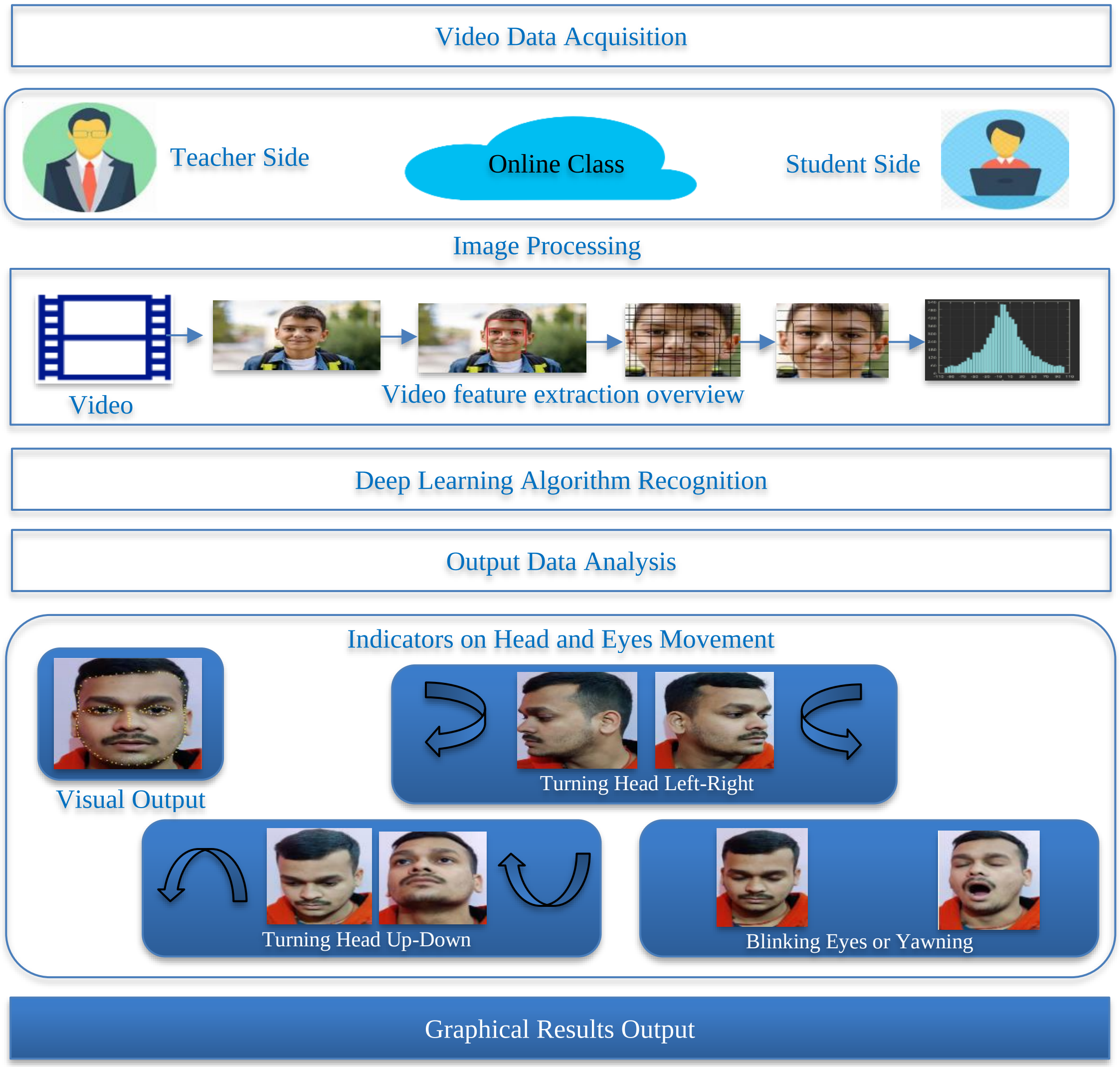


**Fig. 2 Design of the system**

### *4.3. Built-in SEDS System*

The combined data of the two states (Attentive state and non-attentive state) concentration can be determined. It is possible to divide the concentration level into two categories: high and low. We are assigning Attentive state a score of '1' and non- attentive state a score of '0' and finally adding it to get the score. The pre-trained SEDS model is trained and then delivered to the remote server utilizing the Haar cascade architecture. The proposed system has two modules. They are the Haar cascade eye and frontal face module to detect landmarks of the face. The suggested system's technique is based on OpenCV and Haar Cascades Classifiers for Face Detection. The detector engagement system leverages two algorithms which are the Haar Cascade Algorithm and CNN.

#### *4.3.1. Haar Cascade Algorithm*

The Haar Cascade algorithm is a method that extracts common features from photos with high efficiency and is widely used in object detection tasks. It is closely associated with Haar cascades, which are Haar-like features accomplished by applying the pattern feature to the training dataset within a sliding window in a given image. These algorithms must train a lot with both positive and negative images for very good performance. Once it learns anything it already has, it can now match similar features with new images based on the previous data. In this study, the Haar Cascade algorithm is employed to identify the students' frontal faces in the image and get the eye regions to the face.

#### *4.3.2. Convolutional Neural Network (CNN)*

CNNs distinctly differ from Conventional Artificial Neural Networks in their distinctive ability to code the important features of images directly from raw images, and thus offer more efficiency and fewer network parameters. In this project, a CNN was trained on eye images to classify whether a student is facing the webcam ("Focused") or not ("Distracted"), performing a binary classification. The CNN architecture comprises: Input Layer: 64x64 pixels that hold the raw image data.

Convolutional Layers: Using 3 times 3 filters to calculate the outputs of neurons that are connected to regions of local images. Pooling Layer: 2 times 2 size to get smaller size data. Fully Connected Layer: To find scores for each class.

*Eye Detection Module*

To establish the condition of the eyes, we must first identify the face in the gathered frame, followed by the eye region.

*Frontal Face Detection*

The presence or absence of both eyes can be used to determine whether the head is turned. When the eyes are hidden, the head is likely turned in a different direction.

### *4.4. Score Calculation and Feedback Stage*

In educational settings, it's crucial to monitor and assess students' attentiveness during class time to ensure effective learning outcomes. To achieve this, a mathematical approach is employed to quantify attentiveness levels based on continuous observations of students' states, categorized as either Attentive or Non-attentive**.** By analyzing the combined data of the two states (Attentive state and non-attentive state), concentration can be determined.

The Haar Cascade algorithm is set up first, providing a binary choice of "Distracted" or "Focused." The student is then analyzed for facial expressions only when he is classified as "Focused." The CNN Face Emotion Detector then classifies the facial emotions data and outputs the DEP (Dominant Emotion Probability) score. The emotions involved in this evaluation are Neutral, Happy, Surprised, Sad, Disgust, Anger, and Scare. To get the final Concentration Index, we multiply the DEP value with the respective Emotion Weight as shown below Equation 1and the calculated emotion weights are shown in Table 2.

$$CI = DEP \times EW \qquad (1)$$

Let: A (t) be the number of instances where the student/user was in the Attentive state at time t.

NA (t) is the number of instances where the student/user was in the non-attentive state at time t.

t be the total duration of observation (e.g., class time).

The score calculation at time t can be mathematically represented as in Equation 2:

$$\text{Score } (t) = A\ (t) + NA\ (t)\ A\ (t) \qquad (2)$$

To determine the overall score over the observation period, the score function is integrated over the entire duration T as given in Eq.3:

$$\text{Overall Score} = \int 0TA\ (t) + NA\ (t)\ A\ (t)dt \qquad (3)$$

The threshold for determining attentiveness is calculated as in Equation 4.

$$\text{Threshold} = \int 0TNA\ (t)\ dt \int 0TA\ (t)\ dt \times 100 \qquad (4)$$

If the calculated threshold exceeds 125%, then the result is classified as non-attentive. Otherwise, it is classified as attentive. These mathematical expressions provide a comprehensive and objective framework for quantifying attentiveness levels based on continuous observations, enabling accurate assessment and feedback in the education field.

**Table. 2 The emotion weights**

| Dominant Emotion | Emotion Weight |
|---|---|
| Neutral | 0.9 |
| Happy | 0.6 |
| Surprised | 0.6 |
| Sad | 0.3 |
| Disgust | 0.2 |
| Anger | 0.25 |
| Scared | 0.3 |

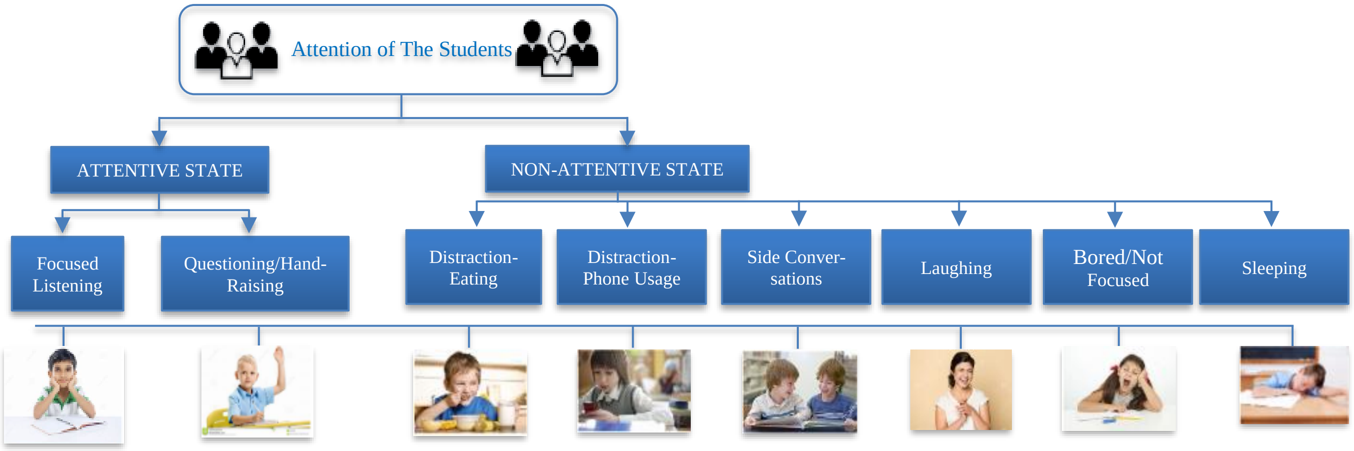
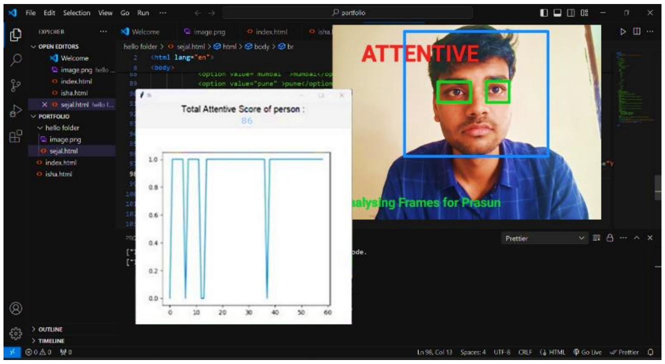
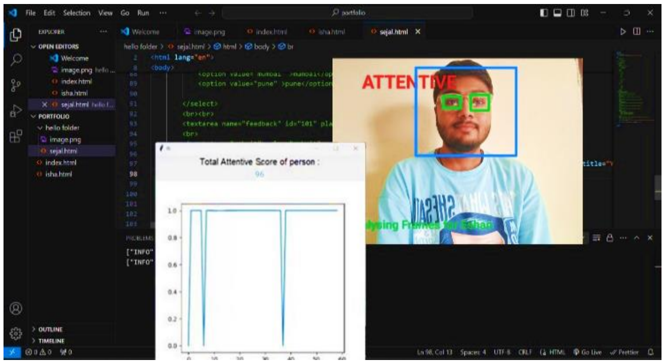


**Fig. 3 Taxonomy of actions and behaviours: Caterozing responses in accoradance with attention level**



Total Attentive Score of person :
86

Fig. 4 Attentive Score obtained after monitoring the student for 60 Seconds



Total Attentive Score of person :
96

Fig. 5 Attentive Score obtained after monitoring the student for 200 seconds

The real-time concentration outputs are continuously stored, and a final report is generated, which determines whether the student/user was attentive during class time. The predetermined threshold for determining the final result is If the non-attentive state exceeds the ratio to the attentive state by 25%, the result is said to be non-attentive; otherwise, the report states attentive. Finally the report is sent to the teacher's email id as the feedback.

## 5. Results and Discussion

The rigorous testing of the proposed system on a diverse set of students has yielded valuable insights into its performance, shedding light on its effectiveness in real-world monitoring scenarios and its potential impact on educational settings.

### *5.1. Attention Monitoring Scores*

Figures 4 and 5 present the Attentive Scores obtained after monitoring students for varying durations, providing a nuanced view of how attention levels fluctuate over time. The scores serve as a quantitative representation of students' engagement during the observation period, offering a detailed and dynamic assessment of their attentiveness. In Figure 6, a collective analysis showcases the Attentive Scores of 20 students over a 200-second monitoring interval, contributing to a broader understanding of group dynamics. Individual attention reports can be checked, as shown in Figure 7.

### *5.2. Detailed Reports*

To extend the analysis of the system's operation, Figure 7 shows complete Attentive Score reports and the related emails sent to teachers. These reports give the educators the full picture of the student's needs and the amount of attention given to all students and every student in particular. The above-differentiated breakdowns provide teachers with specific data regarding the cases where students pay much or little attention, which can be useful for providing interventions that would help apply a more focused and, thus, productive teaching approach.

Moreover, the fact that attention can be measured over time in school settings makes it possible to determine trends in the students' behavior patterns, which can then be used to refine teaching techniques. Ideally, when teachers are in a position to constantly receive feedback from the students, they are better placed to offer faster responses to the students, hence improving the kind of learning environment provided.

### *5.3. Performance Metrics*

A meticulous comparison between the Haar Cascade Algorithm and the HOG algorithm is showcased in Figures 8 and 9. Figure 8 provides a visual representation of time consumption, offering a comparative analysis of the processing efficiency of both algorithms across diverse video scenarios.

Meanwhile, Figure 9 delves into the accuracy evaluation, highlighting the strengths and limitations of the Haar Ca cade Algorithm in contrast to the LBP algorithm. These performance metrics contribute to a comprehensive understanding of the system's algorithmic choices and their implications for real-world implementation.

### *5.4 Graphical Representation*

The graphical representations in Figures 8 and 9 not only serve as visual Aids but also provide an in-depth analysis of the algorithmic nuances. These visualizations enable a more nuanced interpretation of the comparative performance, offering insights into the trade-offs between time consumption and accuracy, crucial considerations for system optimization and deployment.

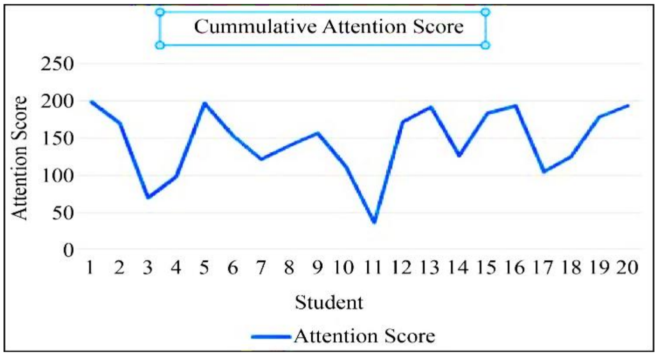
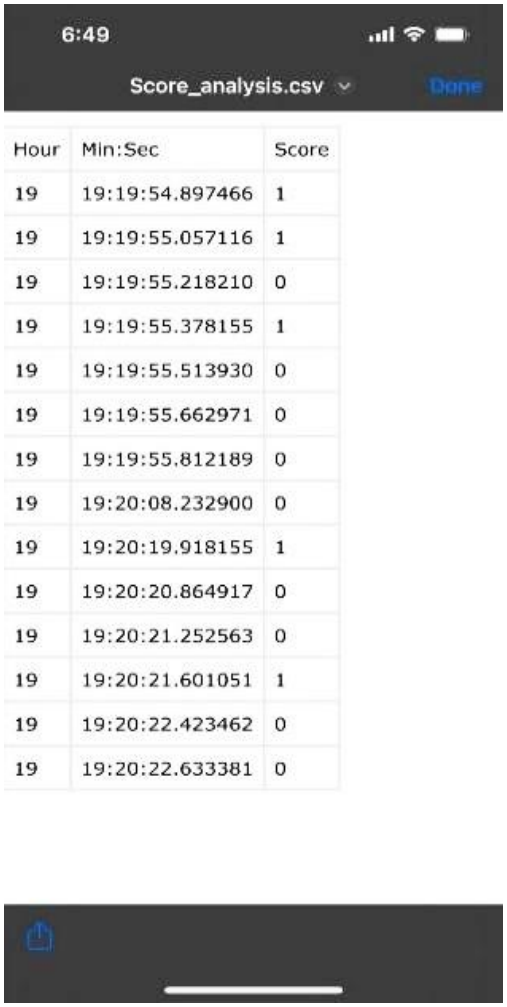


Fig. 6 Attentive score obtained after monitoring 20 students for 200 Seconds



| Hour | Min:Sec | Score |
|---|---|---|
| 19 | 19:19:54.897466 | 1 |
| 19 | 19:19:55.057116 | 1 |
| 19 | 19:19:55.218210 | 0 |
| 19 | 19:19:55.378155 | 1 |
| 19 | 19:19:55.513930 | 0 |
| 19 | 19:19:55.662971 | 0 |
| 19 | 19:19:55.812189 | 0 |
| 19 | 19:20:08.232900 | 0 |
| 19 | 19:20:19.918155 | 1 |
| 19 | 19:20:20.864917 | 0 |
| 19 | 19:20:21.252563 | 0 |
| 19 | 19:20:21.601051 | 1 |
| 19 | 19:20:22.423462 | 0 |
| 19 | 19:20:22.633381 | 0 |

**Fig. 7 Attentive score report**

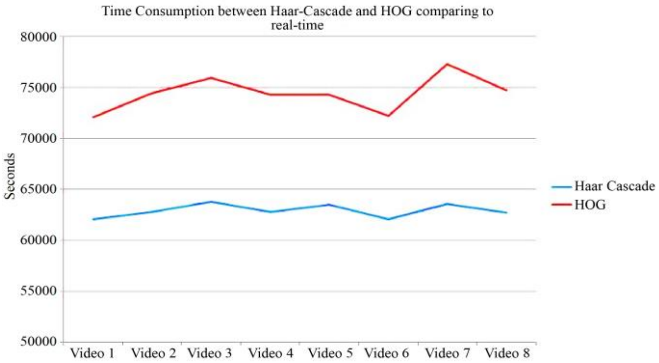


**Fig. 8 Graph comparing time consumption between the Haar cascade algorithm and HOG algorithm to process different videos**

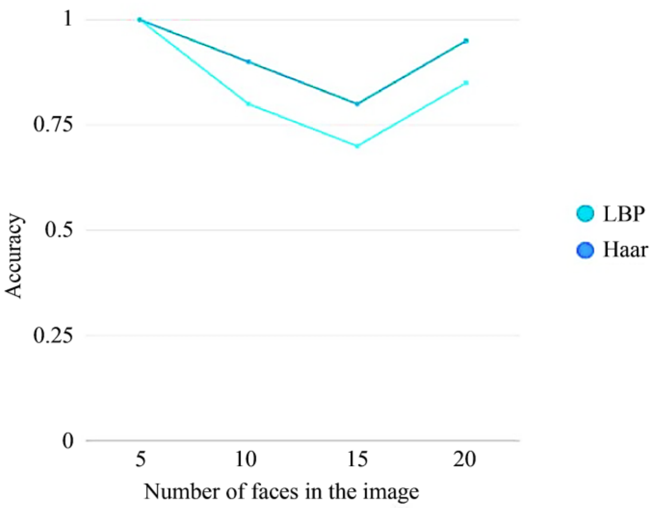


**Fig. 9 Graph comparing accuracy between Haar cascade algorithm and LBP algorithm to process videos**

## 6. Conclusion

The detailed examination of the system's performance underscores its robustness in gauging and reporting students' attention levels. The integration of detailed reports, graphical representations, and algorithmic comparisons contributes to a holistic evaluation of the system's capabilities. The results not only validate the system's effectiveness in real-time monitoring but also provide valuable information for potential optimizations and future enhancements. The findings have significant educational implications. The system's ability to provide granular attention assessments empowers educators to tailor their teaching strategies to individual and group needs, fostering a more engaging and responsive learning environment. The detailed reports can inform instructional design, allowing educators to identify patterns, trends, and areas of improvement.

## References

[1] Qing Li et al., "A Learning Attention Monitoring System via Photoplethysmogram Using Wearable Wrist Devices," *Artificial Intelligence Supported Educational Technologies*, *Advances in Analytics for Learning and Teaching*, pp. 133-150, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Feng-Cheng Lin et al., "Student Behavior Recognition System for the Classroom Environment Based on Skeleton Pose Estimation and Person Detection," *Sensors*, vol. 21, no. 16, pp. 1-20, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Xin Zhang et al., "Analyzing Students' Attention in Class Using Wearable Devices," *2017 IEEE 18th International Symposium on a World of Wireless, Mobile and Multimedia Networks*, Macau, China, pp. 1-9, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[4] Marcela Hernandez-de-Menendez, Carlos Escobar Díaz, and Ruben Morales-Menendez, "Technologies for the Future of Learning: State of the Art," *International Journal on Interactive Design and Manufacturing*, vol. 14, pp. 683-695, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Bui Ngoc Anh et al., "A Computer-Vision Based Application for Student Behavior Monitoring in Classroom," *Applied Science*, vol. 9, no. 22, pp. 1-17, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] David M. Broussard et al., "An Interface for Enhanced Teacher Awareness of Student Actions and Attention in a VR Classroom," *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*, Lisbon, Portugal, pp. 284-290, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Zhang Hong-yu et al., "Depth Image-Based Gesture Recognition for Multiple Learners," *Computer Science*, vol. 42, no. 9, pp. 299-302, 2015. [CrossRef] [Publisher Link]

[8] K. Nosu, and T. Kurokawa, "A Multi-Modal Emotion-Diagnosis System to Support E-Learning," *First International Conference on Innovative Computing, Information and Control - Volume I*, Beijing, China, pp. 274-278, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[9] Kumiko Fujisawa, and Kenro Aihara, "Estimation of User Interest from Face Approaches Captured by Webcam," *Virtual Mixed Reality, Lecture Notes in Computer Science*, vol. 5622, pp. 51-59, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[10] Liu Yuanyuan, "*Research and Application of Head Pose Estimation Method in Natural Environment*," Ph.D Thesis, Central China Normal University, Wuhan, Hubei, China, pp. 1-98, 2015. [Google Scholar] [Publisher Link]

[11] Liping Shen, Minjuan Wang, and Ruimin Shen, "Affective E-Learning: Using 'Emotional' Data to Improve Learning in Pervasive Learning Environment," *Educational Technology & Society*, vol. 12, no. 2, pp. 176-189, 2009. [Google Scholar] [Publisher Link]

[12] Yichuan Tang, "Deep Learning Using Linear Support Vector Machines," *arXiv*, pp. 1-6, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[13] Hanh Phan-Xuan, Thuong Le-Tien, and Sy Nguyen-Tan, "FPGA Platform Applied for Facial Expression Recognition System Using Convolutional Neural Networks," *Procedia Computer Science*, vol. 151, pp. 651-658, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Mohammed Megahed, and Ammar Mohammed, "Modeling Adaptive E-learning Environment Using Facial Expressions and Fuzzy Logic," *Expert Systems with Applications*, vol. 157, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Omid Mohamad Nezami et al., "Automatic Recognition of Student Engagement using Deep Learning and Facial Expression," *Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science*, vol. 11908, pp. 273-289, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Maritza Bustos-López et al., "Wearables for Engagement Detection in Learning Environments: A Review," *Biosensors*, vol. 12, no. 7, pp. 1-30, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Swadha Gupta, Parteek Kumar, and Rajkumar Tekchandani, "A Machine Learning-Based Decision Support System for Temporal Human Cognitive State Estimation during Online Education Using Wearable Physiological Monitoring Devices," *Decision Analytics Journal*, vol. 8, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Jayasankar Santhosh, David Dzsotjan, and Shoya Ishimaru, "Multimodal Assessment of Interest Levels in Reading: Integrating Eye-Tracking and Physiological Sensing," *IEEE Access*, vol. 11, pp. 93994-94008, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Kapil Sethi, and Varun Jaiswal, "PSU-CNN: Prediction of Student Understanding in the Classroom through Student Facial Images Using Convolutional Neural Network," *Materials Today Proceedings*, vol. 62, no. 7, pp. 4957-4964, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Aya Hassouneh, A.M. Mutawa, and M. Murugappan, "Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG Based on Machine learning and Deep Neural Network Methods," *Informatics in Medicine Unlocked*, vol. 20, pp. 1-9, 2020. [CrossRef] [Google Scholar] [Publisher Link]