



Digital Detection of Attention and Distraction Behaviors

Omar Fahmy Hafez^a, Ann Nosseir^{a&b*}, Ralf Seepold^c, Natividad Martinez^d

Omar.kamal@bue.edu.eg; ann.nosseir@bue.edu.eg; or nosseir12@yahoo.co.uk; ralf.seepold@htwg-konstanz.de; natividad.martinez@reutlingen-university.de

^aThe British University in Egypt, El Sharouk, Cairo, Egypt

^bInstitute of National Planning, Salah Sallem, Cairo, Egypt

^cHTWG Konstanz – University of Applied Sciences, Konstanz, Germany

^dHochschule Reutlingen · School of Informatics, Reutlingen, Germany

Abstract

Paying attention helps us learn, advance in our careers, and build successful relationships, but when it's compromised, achievement of any kind becomes far more challenging. Causes of not paying attention can range from common factors like sleep deprivation, stress, or a mood disorder to health difficulties such as ADHD, OCD, or a thyroid problem that affects concentrating. This work extracts paying attention and not paying attention behavior patterns in the context of learning. In early work, our study identified attention and distraction behaviors using gathered video recordings of online classes. The work found ten paying attention behaviors and six distracted behavior patterns. In this paper, we use computer vision techniques to extract features related to these behaviors. These features are distance between hand and face, pitch yaw roll, eye-to-camera distance, hand-to-camera distance, iris direction, gaze tracking, mouth aspect ratio, eye aspect ratio, distance between face and frame side, and facial landmark configuration. This research also applied three types of machine learning—logistic regression, decision trees, and random forest—and the accuracy rates were 79%, 86%, and 89%, respectively. This result is better than relying only on two extracted features in our previous work.

© 2024 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 28th International Conference on Knowledge Based and Intelligent information and Engineering Systems

Keywords: Attention behaviour; computer vision; online sessions ;

1. Introduction

Attention, the skill of selecting and focusing on important information, is a crucial mental process guiding our responses to relevant stimuli. This ability is essential for everyday functioning. Measuring human attention is vital for understanding cognitive processes, assessing performance, monitoring attention-related disorders, and enhancing user experience. These measures encompass both behavioral and biological approaches. Behavioral observation entails

* Corresponding author.

E-mail address: nosseir12@yahoo.co.uk

real-time monitoring and recording of individuals' behaviors, using checklists, rating scales, or coding systems to document attention-related actions like eye movements and engagement with stimuli. In educational contexts, attention holds pivotal significance.

In the beginning of 2020, most countries and regions were affected by the outbreak of COVID-19. As a replacement for conventional teaching, many schools and colleges were promoting online learning [1][2]. Online learning platforms and virtual classrooms became the primary means of delivering educational content. This allowed students to access lectures, assignments, and resources remotely, enabling them to continue their studies from home. Despite the challenges posed by the sudden shift to online learning, it provided a feasible solution to mitigate the disruption caused by the pandemic [1]. As a replacement for conventional teaching, many schools and colleges are now using online learning [2][3]. Courses are delivered online across various channels, including the internal system of the universities, online classrooms, video conferencing, and free online education programmers. Many professors have been using online conference systems like ZOOM, Microsoft Teams, or instant messaging apps to Livestream their classes, play recorded videos, and coordinate online discussions. Online education offers crisis remedies, but the transition from physical to online classrooms has not been without problems [4].

Furthermore, as online learning becomes more mainstream, convenience and versatility are attracting a growing number of students to online learning courses [5]. In fact, however, most students experience serious difficulties and challenges that prevent them from successfully completing their courses [6][1]. In order to keep up with the workload, concentration was inadequate, easily distracted, and lacked self-discipline [7][8]. Some earlier studies claim that learning, including attention, understanding, perception, and organization [9], is an active process. Attention is the most important factor in the learning process, according to researchers, and it is connected to learning performance [10]. In addition, cognitive psychologists and educationalists have claimed that sustained attention to learning content is the basis of successful learning [11]. Researchers have found that the learning output of students is strongly linked to their attention [12]. In the field, there are several concerns with online teaching. Teachers do not understand whether they are slacking or paying attention to students. Consequently, they need to be solved for the issues of the students by proper initiatives or learning strategy.

This paper aims at identifying the image recognition features of paying and not paying attention behaviour. Section 2 is a literature survey. It gives an insight on the previous work done in identifying the features of paying and not paying attention. In Section 3, we describe methodology and shows its results. Final section concludes the work and notes the contributions.

2. Related Work

In the related work section, we explore various approaches and methodologies that have been employed to extract students' attention features and behaviors within the classroom setting. Our discussion primarily revolves around features based on camera devices. In recent years, researchers have delved into the potential of camera devices to detect attention levels during classroom settings. These devices have been utilized to capture indicators of attention, including face direction, eye opening, mouth opening (indicating yawning or dozing states), eye tracking, facial expressions, and emotions. One approach that researchers have taken is to analyze face direction as an indicator of attention. By tracking the orientation of the face, it is possible to determine whether an individual is actively engaged in the class or if their attention is diverted elsewhere [13][14]. In a paper authored by Daniel, Mauricio, et al. [15], the researchers proposed an approach to detect the level of attention using five face landmarks as shown in Fig 1 and Fig 2 [15].

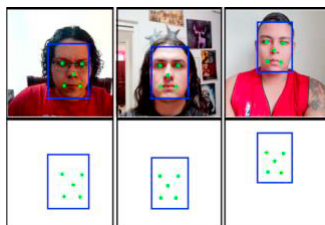


Fig 1. Three video frames are shown at the top while at the bottom the bounding boxes and five landmarks are shown [15].

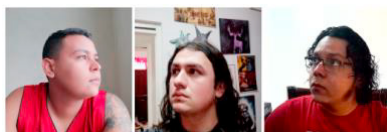


Fig 2. Examples of non-attention frames [15].

Another crucial aspect investigated is the measurement of eye opening, which can offer indications of students' alertness and engagement levels. This data is particularly useful for identifying instances of drowsiness or fatigue during classes. Mukul, Malathi, et al. [16] explored attention detection in students using mouth and eye opening as markers [16][17]. Moreover, camera devices can capture mouth openings, signaling yawning or drowsiness. Monitoring these states can help pinpoint reduced attention periods or signs of fatigue among students [16]. Eye tracking is also gaining traction in this field, accurately tracing eye movements to discern focus. This insight can evaluate teaching methods [18]. Additionally, cameras can capture facial expressions, revealing emotional states. Analysing these expressions can detect confusion, frustration, or engagement, informing tailored teaching strategies [19][17].

Shreya, et al. [17], examined various postures and facial expressions as attention indicators. They identified Person paying attention, Person not looking at the screen, Person writing notes, and Person leaning back as potential attention indicators. They also explored emotions and eye opening as attention markers. They suggested that eye opening could reliably signify students' focus, although their study's limitation was the separate analysis of these factors, potentially affecting the detection system's accuracy. Their work highlights the potential of postures, emotions, and eye opening in gauging student attention. However, future research should integrate these factors for more precise attention detection in educational settings.

Past studies on employing camera devices to monitor attention in classes have exposed the shortcomings of attention-related attributes. Yet, these investigations primarily emphasized utilizing these attributes in isolation rather than amalgamating them into a holistic system for precise student attention detection. The scrutinized attention-related attributes encompass face orientation, eye and mouth movements, eye tracking, facial expressions, and emotional cues. Nonetheless, prior research lacks in its utilization of a restricted set of attributes and the deficiency of a cohesive system that merges these attributes. Moreover, earlier efforts have not undertaken any examination or categorization of these attributes before their application.

3. Methodology

3.1. Attention and distraction behavior.

In earlier work Hafez et al [23], we designed a study to identify attention and distraction behaviours exhibited by participants, driven by the research question: "What are the behaviours indicating attention and distractions?" Nineteen participants, predominantly male, enrolled in the faculty of computer science at the British University in Egypt, were selected for this exploratory study, all actively engaged in a programming module. Approval for data collection was obtained from both module leaders and participants, who were briefed on the study's objectives and granted the freedom to withdraw without consequence. Utilizing ZOOM for online sessions and OBS Studio for recording participants' camera streams, the study meticulously captured participants' reactions and engagement levels. This setup, supplemented by dual screens, facilitated efficient teaching and comprehensive data collection, with quiz-based assessments employed to gauge attention levels at session intervals.

Conducted over six sessions, each lasting 50 minutes, participants were informed of the research nature of the sessions without specific disclosure of the study's objectives. At the conclusion of each session, participants completed quizzes designed to cover session content, with correct answers signalling attentive behaviour and incorrect answers indicating non-attentiveness. Data analysis of quiz responses provided insights into participants' attention and comprehension levels throughout the sessions (See Table 1).

Table 1. attention and distraction behaviors

#	Attention Behaviours	#	Distraction Behaviours
1	Face approximately in the middle of the screen	1	Doing something else other than listening (using object e.g., mobile)
2	Iris in the middle	2	Speaking in a low voice
3	Reading text in the slide (iris is moving left to right and then down)	3	Looking beyond the laptop
4	Lean head on hands	4	Raise eyebrows up
5	Head movement	5	Talking to another person
6	Talk and asking	6	Head angle down by 60
7	Moving head up and down for agreeing		
8	Scratching		
9	crunch face		
10	Playing with hair		

3.2. Behavioral Mapping to Computer Vision Features

To map human behaviours into computer vision features, we utilized a combination of image processing and machine learning techniques. The process involved several steps, including data collection, pre-processing, feature extraction, and mapping.

- **Data Collection:** Recorded online session including student's camera feed, in which 19 students participated in 50 minutes' duration class. Each video attention and distraction behaviours, such as eye-opening, head movements, facial expressions, and body postures.
- **Pre-processing:** The video recordings were pre-processed to enhance the quality of the images and reduce noise using Salt and pepper noise removal. This technique replaces each pixel's value with the median value from the surrounding neighbourhood, effectively reducing the impact of outliers.

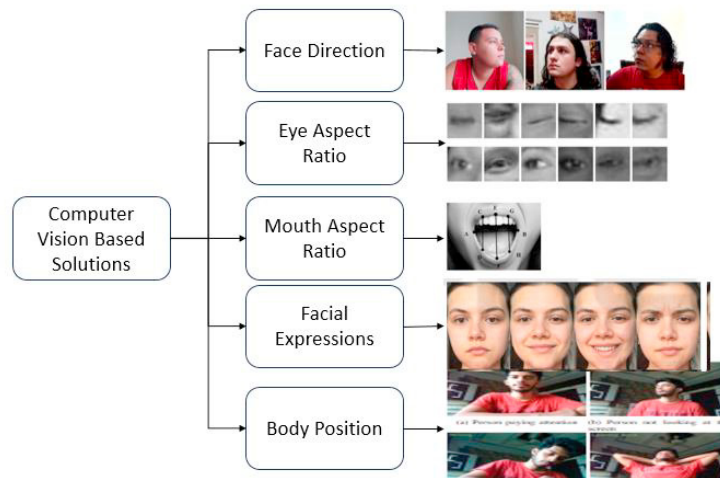


Fig. 3. Examples of computer vision features

- **Feature calculation:** Computer vision techniques were employed to extract relevant features from the pre-processed video frames. We used python and Media Pipe by google to identify the *hand landmarks* of the hands and the faces such as eye, head pose, iris position, and mouth. Fig 3 shows examples of these features.

1. Distance Between Hand and Face:

To extract the Distance Between Hand and Face feature, we employed a combination of hand and face detection algorithms. Initially, a hand detection algorithm identified the hand region in each frame. Subsequently, a face detection algorithm was applied to locate the facial region. By measuring the Euclidean distance between the centroids of the hand and face regions, the Distance Between Hand and Face feature was obtained (See Fig4).



Fig. 4. Hand to face distance.

2. Pitch Yaw Roll:

The Pitch Yaw Roll feature represents the orientation of the face in three-dimensional space. To extract this feature, a facial landmark detection algorithm was employed. By identifying specific facial landmarks, such as the nose, eyes, and mouth, the algorithm determined the pose of the face. The resulting pitch, yaw, and roll angles provided the necessary information for this feature.

3. Eye-to-camera Distance:

The Eye-to-camera Distance feature was obtained by utilizing a combination of facial landmark detection and camera calibration techniques. The algorithm detected the eyes' positions in the face and then estimated their distance from the camera using the known camera parameters.

4. Hand-to-camera Distance:

Like the eye-to-camera distance feature, the hand-to-camera distance feature was extracted using hand detection algorithms combined with camera calibration. The algorithm detected the hand region and estimated its distance from the camera using the camera parameters.

5. Iris Direction:

To extract the Iris Direction feature, iris tracking algorithms were employed. These algorithms identified the position and orientation of the iris within the eye region. By analysing the iris's displacement and orientation, the feature representing the iris direction was obtained.

6. Gaze Tracking:

Gaze tracking was achieved by combining the estimated eye positions and iris direction. By tracking the movement of the iris within the eye region over consecutive frames, the algorithm determined the direction in which the individual's gaze was focused.

7. Mouth Aspect Ratio:

The Mouth Aspect Ratio feature was obtained by analysing the mouth region's geometry. By measuring the mouth's width-to-height ratio, changes in mouth shape and openness were quantified, providing insights into attention-related behaviours.

8. Eye Aspect Ratio:

Like the Mouth Aspect Ratio, the Eye Aspect Ratio feature analyses the geometry of the eye region. By measuring the ratio between the eye's width and the distance between the upper and lower eyelids, the feature captured changes in eye shape, indicating attention-related behaviours.

9. Distance Between Face and Frame Side:

To extract the Distance Between Face and Frame Side feature, the face region was first detected. Subsequently, the distance between the face region and the frame's sides was measured, providing insights into the individual's proximity to the frame's boundaries.

10. Facial landmark configuration:

The MediaPipe's facial landmark detection model locates key facial landmarks such as the corners of the eyes, nose, mouth, etc. Once you have the coordinates of the facial landmarks, we define. For instance, the distance between the corners of the mouth and the midpoint of the eyes when the face is in a neutral position, and then compare it to the same distance when the face is crunched. The ratio of these distances could serve as a measure of facial crunch, with a higher ratio indicating a more pronounced crunch.

Mapping: These extracted features were then mapped to the corresponding human behaviours. The below table shows this mapping.

Table 2: The mapping behaviours to features using computer vision techniques.

Behaviour	Hand-to-Face distance	Head Mov. (pitch, yaw, roll)	Head-to-Cam distance	Hand-to-Cam distance	Gaze Tracking (includes iris direction)	Mouth-Aspect Ratio	Eye-Aspect Ratio	Face-to-Face-side distance	Finger Movement	Facial landmark configuration (includes Eyebrow points)
Attention Behaviours										
Face centred – implies head centred								✓		
Looking into the screen (Prev iris in the middle)		✓			✓					
Reading text in the slide		✓ (yaw)			✓					
Leaning head on hand(s)	✓	✓						✓		

Random head movement	✓	✓		✓
Talking			✓	
Agreeing behaviour (Move head up and down for agreeing)	✓	(pitch)		
Scratching	✓			✓
Crunch face				✓
Playing with hair	✓	✓		✓
Not Paying Attention Behaviours				
Doing something other than listening				
Speaking in a low voice			✓	
Looking beyond the laptop			✓	
Raises eyebrows up (wondering)				✓
Talking to another person	✓		✓	✓
Head angle down by 60		✓(roll)		

Table 2 shows the mapping from behaviours to features that can be measured using computer vision techniques. The table provides a structured representation of the identified behaviours and their corresponding measurable features, highlighting the intricate relationship between student actions and the extracted visual data.

3.3. Applying Machine learning

In the classification process, we discern between attention and distraction behavior and compare between three machine-learning algorithms results namely: Logistic Regression, Decision Trees, and Random Forests. These are the steps followed:

- **Data Organization:** Video sequences were saved in two folders: "attention" and "distraction" based on the labelled sequences.
- **Feature Calculation:** Videos loaded in python and 10 features were calculated for each frame. These values were saved into a CSV file, where columns represent the 10 features and rows represent calculated values for each frame. Total number of rows are 345600.

- **Data Pre-processing:** it included encoding categorical data and normalizing and scaling numerical data. The data is split into 70% training data, 20% testing data, and 10% validation data.
- **Algorithms:** These algorithms were chosen based on their effectiveness in handling our dataset and their ability to provide accurate classification results.

Logistic Regression.

We employed logistic regression as a classification model to determine whether students were paying attention during online classes. The reason behind choosing logistic regression was its ability to handle binary classification problems effectively. As our objective was to classify students into two categories, paying attention or not, logistic regression was a suitable choice. This algorithm had an accuracy of 79% (see Fig 5).

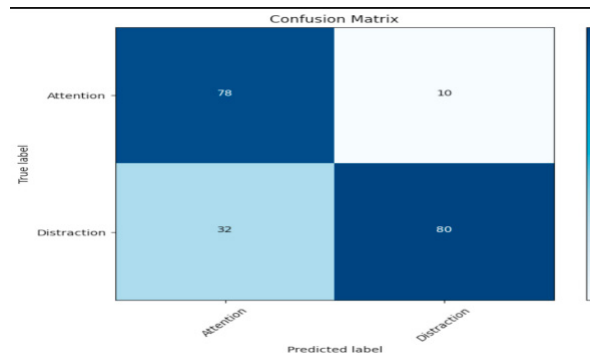


Fig. 5. logistic regression 79%

Decision Trees

To classify whether students are paying attention during online classes, we decided to employ a decision tree algorithm. This approach proved to be effective, as the decision tree model yielded an accuracy of 86%. We specifically chose the decision tree model due to its ability to handle non-linear relationships and interactions between variables, which are present in our dataset (see Fig 6).

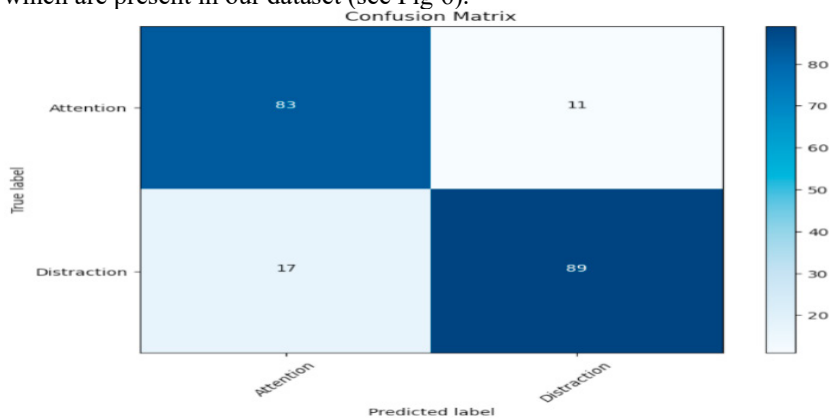


Fig. 6. Decision tree 86%

Random Forests

After careful consideration, our team made the decision to implement a random forest algorithm for our project. This approach proved to be highly effective, as it resulted in an impressive accuracy rate of 89% (see Fig 7).

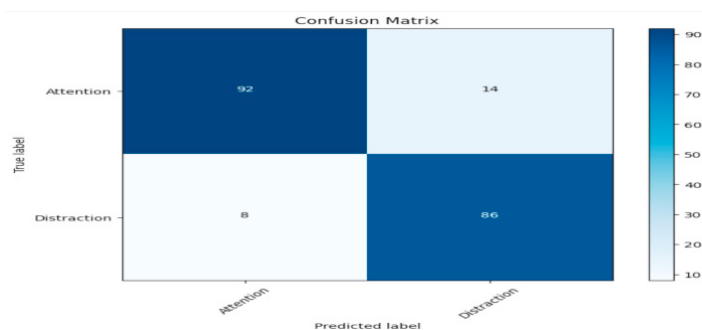


Fig. 7. Random forest 89%.

4. Discussion

This research extends prior work by enhancing the accuracy of attention detection through the inclusion of additional features related to attention and distraction. They are the distance between hand and face, pitch yaw roll, eye-to-camera distance, hand-to-camera distance, iris direction, gaze tracking, mouth aspect ratio, eye aspect ratio, distance between face and frame side, and facial landmark configuration. By incorporating these features, we aimed to refine the classification of students' attention levels during online classes.

We initially employed logistic regression as our classification model due to its effectiveness in handling binary classification tasks. Given our objective of categorizing students into attentive and distracted groups, logistic regression seemed a suitable choice. However, it is imperative to acknowledge the inherent limitations of logistic regression. Despite achieving a respectable accuracy rate of 79%, logistic regression assumes a linear relationship between independent variables and the log odds of the dependent variable. This assumption may not hold true when the relationship is nonlinear, potentially leading to reduced accuracy. Additionally, logistic regression assumes independence among observations, which may not be valid in certain contexts, thereby impacting performance.

Recognizing the need to address these limitations, we explored alternative models capable of capturing nonlinear relationships and dependencies. Traditional linear models may not adequately capture the complexity of attention dynamics in online classes. Thus, we investigated decision trees as a viable alternative. Decision trees recursively partition data based on various features, allowing for the capture of nonlinear decision boundaries. Although decision trees offered an improvement in accuracy, achieving 86%, they were not without their drawbacks, notably overfitting and instability.

To mitigate these issues, we turned to ensemble learning techniques, particularly random forests. By aggregating predictions from multiple decision trees and introducing randomness in feature selection, random forests offer enhanced robustness and generalization to unseen data. This approach yielded an impressive accuracy of 89%, surpassing both logistic regression and decision tree accuracies. Random forests proved particularly adept at overcoming the limitations of linear models and decision trees by introducing randomness in model creation, reducing overfitting, and providing more reliable predictions.

5. Conclusion

In conclusion, our research demonstrates the effectiveness of incorporating additional features and employing advanced classification techniques in enhancing the accuracy of attention detection during online classes. The inclusion of features such as hand-to-face distance and iris direction, along with the adoption of ensemble learning methods like random forests, represents a significant step forward in accurately discerning students' attention levels in educational settings. Further research may explore additional features and techniques to continue improving the precision and robustness of attention detection systems in online learning environments.

References

- [1] Y. Lee and J. Choi, "A review of online course dropout research: Implications for practice and future research," *Educ. Technol. Res. Dev.*, vol. 59, no. 5, pp. 593–618, 2011, doi: 10.1007/s11423-010-9177-y.
- [2] C. Pneumonia, "Some Thoughts on Analytical Chemistry Teaching under New Coronavirus Pneumonia," *Univ. Chem.*, vol. 35, no. 5, pp. 7–9, 2020, doi: 10.3866/PKU.DXHX202002039.
- [3] A. C. W. Fung and J. Ledesma, "Classes suspended but learning continues " Initiative during SARS 2003 in Hong Kong / Extending the Classroom : The Virtual Integrated Teaching a Extending the Classroom The Virtual Integrated Teaching and Learning Environment," Springer, no. April 2020, 2005.
- [4] C. Rapanta, L. Botturi, P. Goodyear, and L. Guàrdia, "Online University Teaching During and After the Covid-19 Crisis : Refocusing Teacher Presence and Learning Activity," *Postdigital Sci. Educ.*, pp. 923–945, 2020, doi: <https://doi.org/10.1007/s42438-020-00155-y>.
- [5] K. Jordan, "Initial Trends in Enrolment and Completion of Massive Open Online Courses," *IRRODL*, vol. 15, no. 1, 2014.
- [6] K. Rita, "The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course," *Int. Rev. Res. Open Distance Learn.*, vol. 12, pp. 19–38, 2011.
- [7] N. Michinov, S. Brunot, O. Le Bohec, J. Jehel, and M. Delaval, "Procrastination, participation, and performance in online learning environments," *Comput. Educ.*, vol. 56, no. 1, pp. 243–252, 2011, doi: 10.1016/j.compedu.2010.07.025.
- [8] C. H. Wang, D. M. Shannon, and M. E. Ross, "Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning," *Distance Educ.*, vol. 34, no. 3, pp. 302–323, 2013, doi: 10.1080/01587919.2013.835779.
- [9] S. Kalyuga, P. Chandler, and J. Sweller, "Managing split-attention and redundancy in multimedia instruction," *Appl. Cogn. Psychol.*, vol. 25, no. SUPPL. 1, pp. 351–371, 2011, doi: 10.1002/acp.1773.
- [10] C. M. Chen and C. H. Wu, "Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance," *Comput. Educ.*, vol. 80, pp. 108–121, 2015, doi: 10.1016/j.compedu.2014.08.015.
- [11] A. Manna et al., "Neural correlates of focused attention and cognitive monitoring in meditation," *Brain Res. Bull.*, vol. 82, no. 1–2, pp. 46–56, 2010, doi: 10.1016/j.brainresbull.2010.03.001.
- [12] R. Shadiey, T. T. Wu, and Y. M. Huang, "Enhancing learning performance, attention, and meditation using a speech-to-text recognition application: evidence from multiple data sources," *Interact. Learn. Environ.*, vol. 25, no. 2, pp. 249–261, 2017, doi: 10.1080/10494820.2016.1276079.
- [13] B. N. Anh, N. T. Son, P. T. Lam, and L. P. Chi, "A Computer-Vision Based Application for Student Behavior Monitoring in Classroom," 2019.
- [14] D. Umarale, S. Sodhani, A. Akhelikar, and R. Koshy, "Attention Detection of Participants during Digital Learning Sessions using Edge Computing," *SSRN Electron. J.*, no. Iicinis, pp. 575–584, 2021, doi: 10.2139/ssrn.3769810.
- [15] D. F. Terraza Arciniegas, M. Amaya, A. Piedrahita Carvajal, P. A. Rodriguez-Marin, L. Duque-Munoz, and J. D. Martinez-Vargas, "Students' Attention Monitoring System in Learning Environments based on Artificial Intelligence," *IEEE Lat. Am. Trans.*, vol. 20, no. 1, pp. 126–132, 2022, doi: 10.1109/TLA.2022.9662181.
- [16] M. L. Roy, D. Malathi, and J. D. D. Jayaseeli, "Students Attention Monitoring and Alert System for Online Classes using Face Landmarks," 2021 IEEE 4th Int. Conf. Comput. Power Commun. Technol. GUCON 2021, pp. 1–6, 2021, doi: 10.1109/GUCON50781.2021.9573793.
- [17] A. Revadekar, S. Oak, A. Gadekar, and P. Bide, "Gauging attention of students in an e-learning environment," 4th IEEE Conf. Inf. Commun. Technol. CICT 2020, 2020, doi: 10.1109/CICT51604.2020.9312048.
- [18] T. Robal, Y. Zhao, C. Lofi, and C. Hauff, "Webcam-based attention tracking in online learning: A feasibility study," *Int. Conf. Intell. User Interfaces, Proc. IUI*, no. April, pp. 189–197, 2018, doi: 10.1145/3172944.3172987.
- [19] N. Gerard et al., *Detection of Subject Attention in an Active Environment Through Facial Expressions Using Deep Learning Techniques and Computer Vision*, vol. 1201 AISC, no. December. Springer International Publishing, 2021.
- [20] B. N. Manu, "Facial features monitoring for real time drowsiness detection," *Proc. 2016 12th Int. Conf. Innov. Inf. Technol. IIT 2016*, pp. 78–81, 2017, doi: 10.1109/INNOVATIONS.2016.7880030.
- [21] R. Garg, V. Gupta, and V. Agrawal, "A drowsy driver detection and security system," 2009 Int. Conf. Ultra Mod. Telecommun. Work., 2009, doi: 10.1109/ICUMT.2009.5345430.
- [22] H. Khalaf Salih Juboori and L. Kulkarni, "Fatigue Detection System for the Drivers Using Video Analysis of Facial Expressions," 2017 Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2017, pp. 1–9, 2018, doi: 10.1109/ICCUBEA.2017.8463437.
- [23] O. Hafez, A. Nosseir, and G. McKee (2023) "The Features of Students Paying and Not Paying Attention in Online Classes" : 2023 International Conference on Computer and Applications (ICCA), IEEE Nov. 2023, doi: 10.1109/ICCA59364.2023.10401506.