A Review on Different Approaches for Assessing Student Attentiveness in Classroom using Behavioural Elements

Conference Paper · March 2022

DOI: 10.1109/ICAI55435.2022.9773418

5 authors, including:

Kainat ..
National University of Sciences and Technology | NUST , School of Mechanical and M…

Khawaja Fahad Iqbal
National University of Sciences and Technology

Muhammad Sajid
University College Dublin

A Review on Different Approaches for Assessing Student Attentiveness in Classroom using Behavioural Elements

Kainat
Department of Robotics and Artificial Intelligence, School of Mechanical and Manufacturing Engineering (SMME)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
kainatirfan1@gmail.com

Sara Ali*
Department of Robotics and Artificial Intelligence, School of Mechanical and Manufacturing Engineering (SMME)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
sarababer@smme.nust.edu.pk

Khawaja Fahad Iqbal
Department of Robotics and Artificial Intelligence, School of Mechanical and Manufacturing Engineering (SMME)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
fahad.iqbal@smme.nust.edu.pk

Yasar Ayaz
Department of Robotics and Artificial Intelligence, School of Mechanical and Manufacturing Engineering (SMME)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
yasar@smme.nust.edu.pk

Muhammad Sajid
Department of Robotics and Artificial Intelligence, School of Mechanical and Manufacturing Engineering (SMME)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
m.sajid@smme.nust.edu.pk

***Abstract*— Analyzing one's participation and attention may be useful in a variety of contexts, like work situations such as driving a car, defusing a bomb, and many learning environments. Increasing the student's involvement and participation in the classroom has been proven to improve learning results. Attention is core for effective learning, yet analyzing attention is a tricky task. People have been working on attention analysis for decades, and as a result, current learning systems contain methods for monitoring and reporting on students' attention states. Facial features and eye movements are some of the important behavioural features to access attentiveness. Approaches such as EEG signals, gaze detection, head and body posture detection are used in this context as they provide rich information about a person's behavior and thoughts. It also gives essential information for interpreting their nonverbal, cues. These are referred to be "honest signals" since they are unconscious patterns that reveal the focus of our attention. They give vital indications concerning teaching methods and students' responses to various conscious and unconscious teaching tactics inside the classroom. Examining verbal and nonverbal conduct in the classroom can give valuable input to the instructor. This paper will go through various approaches available for analyzing student attentiveness for effective learning in the classroom. Integrating different technical approaches with Machine learning and Deep learning models accuracy up to 90% can be observed in different research with minimum error.**

Keywords—Attention analysis, verbal skills, non-verbal skills, effective learning, engagement.

I. INTRODUCTION

Attention analysis is a popular area of research for decades. Attention is defined as the ability to focus one's mind on a related task eliminating distractions. In educational systems, student engagement refers to the degree of student's attention, interest, and participation in the class. Research shows students are attentive at the start of the lecture but end up losing attention after about 10 minutes [1]. Analyzing the student's attention also became a matter of interest due to serious issues like students' low academic performance, distancing from studies, and high dropout rates [2][3].

The purpose of this observation is to monitor the learning process and provide feedback to the teacher for effective learning. Feedback is useful if it is based on accurate data [4]. Educational systems are using modern techniques to monitor the student's and teacher's behaviour in the classroom and provide feedback [5]. Feedback is important for both students and the teachers as it helps students to understand where they lag and teachers to self-reflect on their teaching methodology. Most of the behavior is unconscious. Observing these unconscious cues can help to improve teaching and learning practices. This unconscious behaviour includes both verbal and non-verbal cues [5][6]. Analyzing the gaze pattern, auditory features, body and head pose are some most important behavioral elements considered as "Honest Signals" to students' point of attention [7]. Fredricks et al., [2] proposed that attention is divided into three types: behavioural, emotional, and cognitive. Engagement is used as an alternative to attention. Behavioural engagement is defined as one's participation in the class, while happiness, anxiety, interest, and

978-1-6654-6896-1/22/$31.00 ©2022 IEEE.

boredom fall under the category of emotional engagement. The term cognitive engagement is defined as being strategic or self-regulating which involves controlling behaviour, emotions, and thoughts. It means to stay focused while ignoring the distractions.

In 2007 Vatahska et al., [8] leveraged the facial landmark, the face detector classifies the pose as frontal, left, or right. They then used an artificial neural network (ANN) to finally estimate the three continuous rotation angles used to model the head pose. In 2011 Fanelli et al., [9] did the pose estimation as a regression problem. They estimated head pose using depth data to train a random regression forest given their capability to handle large training datasets. That is publicly available. RRF is used to jointly estimate 3D coordinates of the nose tip and rotation angles ahead. In 2015 Raca et al., [10] discussed the approach based on principles of unobtrusive measurements and social signal processing. To improve the learning and teaching experience, researchers observed both students and teachers. They assume that students with different levels of attention will behave differently during the class. Cameras were used to record lectures and eye trackers to get gaze data. They also conclude the relationship between head orientation and gaze direction. The assumptions made by observing the video footage were supported by the questionnaire. The position of the teacher in class was linked with the head orientation of the students.

In 2017 Thomas & Jayagopi [11] proposed a predictive model to predict the state of the students in terms of engagement or distracted based on the classroom video recordings. They used OpenFace toolbox, SVM and Logistic regression models for analysis. Their controlled experiment dataset contains 2263 samples with 10 students. The 3 annotators labelled the dataset with engaged and distract labels. In 2019 Qiu et al., [12] proposed a facial expression recognition (FER) framework that relies solely on facial landmarks/action units. They successfully determined seven expressions, which include six basic expressions and one for the neutral face. These facial expressions were used to measure student engagement. They used distance vectors over the coordinates to design the algorithm for predicting the expressions accurately. The performance of this method was comparable to that of the known CNN architecture like the VGG-16 and ResNet (Residual Neural Network) which are the backbone for many computer vision tasks and are used in visual object recognition software research.

Researchers from the Intelligent Tutoring Systems (ITS) community have put significant effort into the development and evaluation of intelligent systems to enhance learners' learning experience. However, the focus was mainly on fostering engagement in an e-learning environment, not in classroom settings. The advantage is an early prediction of students' attention, dropouts, and withdrawals in an ongoing MOOC. They provide methods to timely assess the students to retain them, by suggesting suitable corrective strategies and policies, subsequently managing, and reducing attrition rates. [13]. In 2019 Hutt et al., [14] developed an ITS named GuruTutor which teaches biology topics to students in natural language. They also used the Tobii EyeX eye tracker which tracks students' eye gaze while they complete a 30–40 minutes learning session with GuruTutor.

In 2020 Dhingra et al., [6] focused on non-verbal communication by tracking eye gaze. Their system detected non-verbal communication for 28 participants. Using OpenFace, iView, and SVM, they created a system and compared its accuracies with SMI RED 250 eye-tracker. In 2021 Datta et al., [15] used image classification to introduce a better way of implementing the Convolutional Neural Network (CNN) models. They used two widely accepted datasets which include FashionMNIST which contains 60k images for training a model and MNIST for handwritten numbers. By utilizing parallel computation, the authors reduced operating time and increased efficiency.

In this paper, we have discussed several techniques used by researchers over several years for attention analysis in the classroom. As there is no standard method for estimating attention, many approaches have been developed to solve the problem. Researchers have divided these approaches into two main categories that can prove effective in different setups [16]. The first approach is **Intrusive** which includes physical contact with the subject's body like measuring the body temperature, heart rate, and brainwave readings. However, these measurements are extensive and can often be uncomfortable for the user, which in turn can lead to errors in the collected data. The second one is the **non-intrusive** approach, using acoustic features, facial expressions, gaze, head and body posture as a technique to analyze attention. Each of these approaches employed ranges widely in scope, cost, and complexity. Each technique has its own method to measure accuracy. Some of the examples are mentioned in table 1. In literature, several review papers cover work on these two techniques but separately. As for our knowledge, no review paper includes a discussion on both intrusive and nonintrusive techniques in the classroom,

II. APPROACHES FOR ANALYZING ATTENTION

In literature, we see various methods to analyze students' attention including both traditional as well as technological.

2.1 The Traditional approach for measuring attention

In our educational system, there are several traditional ways to measure attention or students' engagement. Each of which has its own pros and cons.

2.1.1 Student Self Reporting

Self-reporting is the most common and easy method to assess student engagement in the classroom. In this approach, a questionnaire for the survey is provided to the student. The questionnaires contain general questions related to emotional and cognitive engagement. According to researchers' assessing emotional and cognitive engagement is important and self-report methods are particularly designed for it because they are not directly observable and need to be extrapolated from behaviour [3]. Some modern attention analysis systems also use this traditional approach digitally. The questionnaire is digital and once a standard is set so the results evaluate automatically which solved the problem of labor and time [17].

Another standard approach in the literature is thought auditory probes [18]. This technique is used if the student is taking a lecture online or using some ITS. During a lecture after a specific interval of time, an auditory-visual probe appears on the screen like a translucent overlay asking the student whether he/ she is attentive or not or his/her mind wanders or not. Inattentiveness can be intentional or unintentional [19]. However, the problem with the self-report approach is their low-quality results as students might not respond honestly [17].

2.1.2 Teacher Rating

Teacher rating of individual students' engagement is another approach. Teachers rated each student's engagement level in the class individually. An average of both teacher's rating and student's self-report based on the correspondence can give a better evaluation of student engagement in the class [20]. The teacher rating scale or checklist reflects questions on a multidimensional model of engagement which includes behavioural, emotional, and cognitive engagement. These studies show a strong correlation between the teacher and student reports on behavioural engagement as compared to emotional engagement. This is because students can mask their emotions [21][3].

2.1.3 Interviews

Few studies show the use of interview techniques in the educational system to assess engagement [22]. Interviews can provide a detailed description of students' experiences and how they are related to engagement. Students can give more detailed open-ended unstructured answers. However, the factors like knowledge, skills, and biases of the interviewer can affect the quality of the results. There is also the question of the reliability of the interview results [23]. Students' evaluation can also be done through feedback surveys where textual feedback is collected. Actionable information is extracted from the student's free-text response by using state-of-the-art supervised machine learning techniques and unsupervised clustering methods [24].

2.1.4 Observation

Observation methods are also applied at both the individual and whole classroom levels to measure attention. At an individual level, a student's behavior is observed to assess behavioral engagement. This technique is time-consuming because, for better understanding, a student needs to be observed in various academic settings like working individually, in a group etc. The observation method can also be used to cross-check the results we get from survey or interview methods. The major concern about the observation method is it highly depends on the observer and his/her ability to analyze [3].

All above mentioned traditional methods have their strengths and weaknesses. With the development in technology, there are other methods to analyze attention and engagement using wearable and non-wearable sensors by integrating both traditional and modern technology methods.

2.2 Technological approach to measure attentiveness

In literature measuring attention using technology is divided into two categories, Intrusive and Non-Intrusive [16] (also called invasive and non-invasive). These techniques are discussed in detail in the following sections.

2.2.1 Intrusive techniques for attention analysis

This is a technique in which a student's attention is measured by taking biological parameters using physiological sensors. These physiological sensors are integrated with wearable devices. They use brainwaves, heart rate, body temperature, gaze pattern, etc. as input.

In 2012 Yaomanee et al., [25] studied the use of Electroencephalography (EEG) signals to analyze attention. The proposed EEG Alpha and Beta waves could be coupled with the state of relaxation and attention supporting the previous studies. These two waves can be used to analyze whether the person is relaxed or attentive. The experiment was conducted on 10 people, divided into three sub-experiments. The total experiment time was 30 minutes divided into 3 parts, 10 minutes each. The EEG headset used for the study is the Emotive EPOK headset. The main purpose of the research was to use the low-cost commercial device and minimize the number of EEG channels. As a result, the suggested locations for detecting Alpha and Beta waves are (AF3) and channel 2 (F7) and channel 11 (FC6) and channel 13 (F8) respectively.

In 2015 Moreno-Esteva and Hannula [26] studied student attention during mathematics lectures. In which the student wears a gaze tracking device, a glass frame equipped with miniature cameras placed to produce a video scene and keep track of the movement of the eyes. The experiment was performed for a single lecture on an eighth-grade student. The student followed the teacher's cues and gestures. As a result, it was observed that most of the time the student followed the teacher's cues.

In 2015 Wang and Cesar [27] proposed research using Galvanic Skin Response (GSR) sensor to measure the attention in the E-learning environment. GSR sensors measure the user's electrical conductance of the skin. The experiment was conducted on two groups of seventeen students. One group was physically present in the class and the other was remote. They used self-reports and GSR sensors to measure both groups' attention and comfort levels. Finally, an exam was conducted after the class and the results from the reports and sensor were crosschecked with the exam results.

In 2018 Sethi et al., [28] used Neurosky Mindwave, a single channel electrode headset to detect EEG signals and provide real-time attention-based neurofeedback on user's attention and performance. They consider a case study of an online SAT (Scholastic Aptitude Test) learning course. The first phase of the experiment was reading the paragraph and answering questions based on it. Participants wear the headset which was connected to the mobile device via Bluetooth. The headset recorded and stored the EEG and attention values at the rates of 512 Hz and 1 Hz respectively. After the break of 1 to 2 days the participants were engaged in the second phase of the experiment where they performed the same task but this time the headset also provides real-time attention-based neurofeedback to the users using a pop-up. The pop-up indicates that their attention level is low than the given threshold. The results of the experiment before and after the neurofeedback were calculated and it was concluded that for 36 out of 42 participants the attention level increased with the feedback.

In 2020 Gao et al., [29] presented a classroom sensing system named n-gage that uses the multidimensional model of engagement, heterogeneous data for engagement prediction, wearable sensor watches, and observing the data of the classroom environment. The experiment duration was 4 weeks, 144 classes over 11 courses. N-gage, use sensing data from two sources. One from the wearable devices capturing physiological and physical signals (e.g., EDA (electrodermal activity), HRV (heart rate variability), ACC (accelerometer)). The second source was indoor weather stations that capture environmental changes (e.g., temperature, sound). It was concluded that the n-gage worked best to predict emotional engagement.

2.2.2 Non-Intrusive techniques for attention analysis

In contrast, the researchers also adopt non-intrusive techniques to measure student attention. These devices measure facial features, verbal and non-verbal cues like gaze, head, and body movements. These unconscious cues are referred to as honest signals [7]. Nowadays researchers are using various non-invasive devices to extract these features such as cameras, eye-trackers, Kinect sensors, etc.

In 2008 Whitehill and Movellan [30] detected facial features such as the center of both eyes, the tip of the nose and center of the mouth. He then localized the face location frame-by-frame via pose tracker using an array of Viola-Jones-style classifier [31] by feeding the facial landmark into a classifier. This process is followed by linear regression for pose estimation. They measured the accuracy of the pose tracker on the GENKI-4K dataset. Bidwell et all., [32] used four Kinect sensors and five color cameras to capture students' gaze and model their attention.

In 2017 Zaletelj and Košir [33] used Kinect to detect facial and body features and applied machine learning models to predict the attentiveness of 18 students. A human observer is also part of the experiment to match the observed results. Data like face gaze, head pose, and skeleton joints taken from Kinect were combined with the computed features such as the point of gaze, lean back, eyes open or close. Finally integrating the results of the behavioral cues observed e.g., writing notes, supporting head with hand, yawing etc. The purpose of this research was to derive a set of body, gaze, and facial features used to analyze students' behavior.

In 2019 Veliyath et al., [34] used the Tobii Eye tracker to collect the eye-gaze data. Each student sits in front of the computer and take a lecture. Eye-tracker data was used to relate the location of gaze on the screen. Additional features like the local timestamp, name of foreground application on computer screen and coordinates, and ground truth, or self-reported scores were also saved. For the self-report survey, a question after every 5 minutes with a scale of 1-10 appears on the screen "How engaging were the previous 5 minutes of class?". On first use the software needs calibration. Machine learning models like Random Forest (RF), Support Vector Machines (SVM) etc. were used.

In 2021 Ahuja et al., [35] developed a 3D classroom digital twin to capture the 6-DOF head rotation and gaze of the student along with the instructor using ArUco markers and two cameras. The purpose of this digital twin was to allow teachers to practice before they stand in front of students in the classroom via a VR headset.

In 2022 Xu et al., [36] propose head pose estimation using a single depth image, deep neural network, and 3D point cloud. The abstraction layer of PointNet++ [37] was implemented to extract features from the 3D point cloud and regression was performed with that classification loss to get the head pose estimation.

2.2.3 Combination of Intrusive & Non-Intrusive Techniques

Research including both intrusive and non-intrusive techniques was performed [26]. The student was wearing gaze tracking glasses and observed with the camera installed in the class. The results of both camera and the glasses were used for the conclusion.

In 2016 Monkaresi et al., [16] proposed a system that analyzes students' engagement while completing a writing activity. They followed the study of Whitehill et al., [38] who utilized a video-based method to detect students' engagement while playing cognitive training games. The 23 students were observed using a Microsoft Kinect along with the sound recording. A computer vision video-based heart rate sensing technique was used to detect the engagement. Its results were compared with that of the heart rate detector attached to students. Their research showed the possibility of detecting engagement from remote sensing of heart rate.

III. Feature Extraction

3.1 Intrusive Technique Features

Intrusive techniques use biological parameters using physiological sensors. These sensors are embedded into devices. The feature extracted to measure attention are EEG signals, HRV, ACC, GSR and EDA. EEG works with brain signals and the rest of them with skin conductance. By applying a low constant voltage, the change in skin conductance can be measured [39].

EDA has a long history of being used in psychological research. Researchers use EDA to observe activities like emotion detection [40], depression and engagement [41]. EDA refers to the variation of the electrical properties of the skin in response to sweat excretion. HRV is controlled by the autonomic nervous system and it is a parameter used to measure stress levels. HR is a good indicator to observe change between different affective states such as HR is higher in situations like fear, and anger than in happiness, and surprise [42]. HRV data is obtained with a variety of sensors placed on the chest or wrist.

GSR was used in several studies to detect stress levels by measuring skin conductance [43]. GSR has been connected to a participant's emotional state and arousal level. EEG is widely used to differentiate between alertness vs. drowsiness [44]. EEG is measured by recording the voltage of the electrodes on the scalp. The electrodes are placed at designated positions allocated on the head [45]. Now EEG is for educational research by many researchers and has achieved good results [46].

Moreover, it has been found that students' engagement is affected by their thermal comfort level in the classrooms [47]. The thermal comfort level is influenced by many factors such as indoor temperature, humidity, skin temperature, sound, CO_2 level, etc. [29][48]. For example, students feel sleepy in the warm and more relaxing environment of the classroom during winter.

3.2 Non-Intrusive Technique Features

Non-intrusive techniques extract features like head orientation, body pose, eye-gaze, facial expressions, and emotions. Emotions have a significant impact on human cognitive functions such as observation, attention [49], learning, and problem-solving skills. Emotion detection starts from face detection which is the first step for many applications, such as face detection, recognition, facial expression, and landmarks detection. Cameras are used as a continuous and non-intrusive way to capture images. The facial information can be used to understand certain aspects of the student's current state of mind, and in literature, several techniques are used to automate this process [50]. Ekman and Friesen 1978 proposed a coding system for facial actions (FACS) which highlights the expressive aspects of emotions. According to him, there are six basic emotions happy, sad, surprise, disgust, anger and fear [51].

One of the very natural unconscious cues to measure attention is a person's gaze. According to Khorrami et al., the location of one's gaze focus and its duration are useful indicators of attention [52]. Sharma and Abrol [53] proposed a research survey of eye gaze estimation techniques. If the eye gaze is not directly observable so the head pose is taken as an indication of gaze. Stiefelhagen [54] concluded 87% of the time person's eye gaze agreed with the orientation of his/her head. Additional information can be gained by integrating the teacher's cues and observing all the students at the same time. Systems like EduSense which use multiple cameras to observe students and the teacher is effective in this scenario [5]. Gaze can also be detected preciously using eye trackers but these devices are expensive, and not suitable for sustained use in the classroom [55].

Body posture detection is also important to estimate the attention state of the students. Lean back and dull body posture point towards low attention. Whereas an attentive posture, hand raise shows student's participation [56] and points to active state and high attention level. Liu et al., [57] surveyed different body posture estimation techniques.

IV. Conclusion

Despite the extensive research in attention analysis, there are still a lot of challenges. This paper presents a review based on the latest literature for measuring attention. Several invasive, as well as non-invasive techniques, have been used. Both methods have their pros and cons. The decision for adopting a technique depends upon the situation. Invasive or intrusive techniques can be used in the lab or controlled environments where the priority is to take accurate readings, however, these techniques use costly hardware. Moreover, the intrusive technique also results in the discomfort of the subject. In the case of non-intrusive techniques, as the hardware has no physical contact with the participant's body therefore they feel more comfortable, and results are declared more natural.

Table 1

Author	Technology	Participants	Accuracy
Sethi et al. [28]	Intrusive, EEG headset	participants: 42. 14 female & 28 males in an online course	Attention accuracy: 53.792
Gao et al.[29]	Both, E4 wristbands, indoor weather station	Participants: 23 students of grade 10, & 6 teachers.	MAE: 0.788 & RMSE: 0.975
Veliyath et al. [31]	Non-intrusive, Tobii eye tracker	Participants:10 in a computer lab-like environment.	Gaze tracking accuracy: 77%
Ahuja et al.[35]	Non-intrusive, RGB cameras	Participates: 8 in classroom	gaze estimation accuracy: 90.03%
S.Hutt [14]	Non-intrusive, ITS and Eye-tracker	Individual testing: 9 students Small group testing: 7 students Classroom pilot: 35 students	Gaze tracking accuracy: 75% (both eyes) and 95% (one eye)

References

[1] N. A. Bradbury, "Attention span during lectures: 8 seconds, 10 minutes, or more?," *Adv. Physiol. Educ.*, vol. 40, no. 4, pp. 509–513, 2016.

[2] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," *Rev. Educ. Res.*, vol. 74, no. 1, pp. 59–109, 2004.

[3] J. A. Fredricks and W. McColskey, "The measurement of student engagement: A comparative analysis of various methods and student self-report instruments," in *Handbook of research on student engagement*, Springer, 2012, pp. 763–782.

[4] K. T. Brinko, "The practice of giving feedback to improve teaching: What is effective?," *J. Higher Educ.*, vol. 64, no. 5, pp. 574–593, 1993.

[5] K. Ahuja *et al.*, "EduSense: Practical classroom sensing at Scale," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–26, 2019.

[6] N. Dhingra, C. Hirt, M. Angst, and A. Kunz, "Eye Gaze Tracking for Detecting Non-verbal Communication in Meeting Environments," in *15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020)*, 2020, pp. HUCAPP20--RP.

[7] A. Pentland, *Honest signals: how they shape our world*. MIT press, 2010.

[8] T. Vatahska, M. Bennewitz, and S. Behnke, "Feature-based head pose estimation from images," in *2007 7th IEEE-RAS International Conference on Humanoid Robots*, 2007, pp. 330–335.

[9] G. Fanelli, J. Gall, and L. Van Gool, "Real time head pose estimation with random regression forests," in *CVPR 2011*, 2011, pp. 617–624.

[10] M. Raca, "Camera-based estimation of student's attention in class," 2015.

[11] C. Thomas and D. B. Jayagopi, "Predicting student engagement in classrooms using facial behavioral cues," in *Proceedings of the 1st ACM SIGCHI international workshop on multimodal interaction for*

education, 2017, pp. 33–40.
[12] Y. Qiu and Y. Wan, "Facial Expression Recognition based on Landmarks," in *2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2019, vol. 1, pp. 1356–1360.
[13] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Comput. Human Behav.*, vol. 104, p. 106189, 2020.
[14] S. Hutt *et al.*, "Automated gaze-based mind wandering detection during computerized learning in classrooms," *User Model. User-adapt. Interact.*, vol. 29, no. 4, pp. 821–867, 2019.
[15] D. Datta and S. B. Jamalmohammed, "Image classification using CNN with multi-core and many-core architecture," in *Applications of Artificial Intelligence for Smart Technology*, IGI Global, 2021, pp. 233–266.
[16] H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D'Mello, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," *IEEE Trans. Affect. Comput.*, vol. 8, no. 1, pp. 15–28, 2016.
[17] J. J. Appleton, S. L. Christenson, D. Kim, and A. L. Reschly, "Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument," *J. Sch. Psychol.*, vol. 44, no. 5, pp. 427–445, 2006.
[18] J. Smallwood, M. McSpadden, and J. W. Schooler, "When attention matters: The curious incident of the wandering mind," *Mem. \& Cogn.*, vol. 36, no. 6, pp. 1144–1150, 2008.
[19] P. Seli, E. F. Risko, and D. Smilek, "On the necessity of distinguishing between unintentional and intentional mind wandering," *Psychol. Sci.*, vol. 27, no. 5, pp. 685–691, 2016.
[20] E. Skinner, C. Furrer, G. Marchand, and T. Kindermann, "Engagement and disaffection in the classroom: Part of a larger motivational dynamic?," *J. Educ. Psychol.*, vol. 100, no. 4, p. 765, 2008.
[21] A. Wigfield *et al.*, "Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes," *Psychol. Sch.*, vol. 45, no. 5, pp. 432–445, 2008.
[22] P. Blumenfeld *et al.*, "School engagement of inner-city students during middle childhood," in *Developmental pathways through middle childhood*, Psychology Press, 2006, pp. 157–182.
[23] M. M. McCaslin and T. L. Good, *Listening in classrooms*. HarperCollins, 1996.
[24] R. Nawaz *et al.*, "Leveraging AI and Machine Learning for National Student Survey: Actionable Insights from Textual Feedback to Enhance Quality of Teaching and Learning in UK's Higher Education," *Appl. Sci.*, vol. 12, no. 1, p. 514, 2022.
[25] K. Yaomanee, S. Pan-ngum, and P. I. N. Ayuthaya, "Brain signal detection methodology for attention training using minimal EEG channels," in *2012 Tenth International Conference on ICT and Knowledge Engineering*, 2012, pp. 84–89.
[26] E. G. Moreno-Esteva and M. Hannula, "Using gaze tracking technology to study student visual attention during teacher's presentation on board," in *CERME 9-Ninth Congress of the European Society for Research in Mathematics Education*, 2015, pp. 1393–1399.
[27] C. Wang and P. Cesar, "Physiological Measurement on Students' Engagement in a Distributed Learning Environment.," *PhyCS*, vol. 10, p. 0005229101490156, 2015.
[28] C. Sethi, H. Dabas, C. Dua, M. Dalawat, and D. Sethia, "EEG-based attention feedback to improve focus in E-learning," in *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*, 2018, pp. 321–326.
[29] N. Gao, W. Shao, M. S. Rahaman, and F. D. Salim, "n-Gage: Predicting in-class Emotional, Behavioural and Cognitive Engagement in the Wild," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 3, pp. 1–26, 2020.
[30] J. Whitehill and J. R. Movellan, "A discriminative approach to frame-by-frame head pose tracking," in *2008 8th IEEE International Conference on Automatic Face \& Gesture Recognition*, 2008, pp. 1–7.
[31] M. J. Jones, P. Viola, and others, "Robust real-time object detection," in *Workshop on statistical and computational theories of vision*, 2001, vol. 266, p. 56.
[32] J. Bidwell and H. Fuchs, "Classroom analytics: Measuring student engagement with automated gaze tracking," *Behav Res Methods*, vol. 49, no. 113, 2011.
[33] J. Zaletelj and A. Košir, "Predicting students' attention in the classroom from Kinect facial and body features," *EURASIP J. image video Process.*, vol. 2017, no. 1, pp. 1–12, 2017.
[34] N. Veliyath, P. De, A. A. Allen, C. B. Hodges, and A. Mitra, "Modeling students' attention in the classroom using eyetrackers," in *Proceedings of the 2019 ACM Southeast Conference*, 2019, pp. 2–9.
[35] K. Ahuja *et al.*, "Classroom Digital Twins with Instrumentation-Free Gaze Tracking," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–9.
[36] Y. Xu, C. Jung, and Y. Chang, "Head pose estimation using deep neural networks and 3D point clouds," *Pattern Recognit.*, vol. 121, p. 108210, 2022.
[37] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *arXiv Prepr. arXiv1706.02413*, 2017.
[38] J. Whitehill *et al.*, "Towards an optimal affect-sensitive instructional system of cognitive skills," in *CVPR 2011 WORKSHOPS*, 2011, pp. 20–25.
[39] D. C. Fowles, M. J. Christie, R. Edelberg, W. W. Grings, D. T. Lykken, and P. H. Venables, "Publication recommendations for electrodermal measurements," *Psychophysiology*, vol. 18, no. 3, pp. 232–239, 1981.
[40] F. Canento, A. Fred, H. Silva, H. Gamboa, and A. Lourenço, "Multimodal biosignal sensor data handling for emotion recognition," in *SENSORS, 2011 IEEE*, 2011, pp. 647–650.
[41] E. Di Lascio, S. Gashi, and S. Santini, "Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–21, 2018.
[42] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010.
[43] J. Zhai, A. B. Barreto, C. Chin, and C. Li, "Realization of stress detection using psychophysiological signals for improvement of human-computer interactions," in *Proceedings. IEEE SoutheastCon, 2005.*, 2005, pp. 415–420.
[44] L. Chaabene, S., Bouaziz, B., Boudaya, A., Hökelmann, A., Ammar, A., & Chaari, "Convolutional Neural Network for Drowsiness Detection Using EEG Signals.," *Sensors*, vol. 5, p. 1734, 2021.
[45] L. E. Nacke, S. Stellmach, and C. A. Lindley, "Electroencephalographic assessment of player experience: A pilot study in affective ludology," *Simul. \& Gaming*, vol. 42, no. 5, pp. 632–655, 2011.
[46] R. Shadiev, 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.
[47] H. Jiang, M. Iandoli, S. Van Dessel, S. Liu, and J. Whitehill, "Measuring Students' Thermal Comfort and Its Impact on Learning.," *Int. Educ. Data Min. Soc.*, 2019.
[48] M. S. Rahaman *et al.*, "An Ambient--Physical System to Infer Concentration in Open-Plan Workplace," *IEEE Internet Things J.*, vol. 7, no. 12, pp. 11576–11586, 2020.
[49] P. Vuilleumier, "How brains beware: neural mechanisms of emotional attention," *Trends Cogn. Sci.*, vol. 9, no. 12, pp. 585–594, 2005.
[50] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, 2008.
[51] P. Ekman and W. V Friesen, "Facial action coding system," *Environ. Psychol. \& Nonverbal Behav.*, 1978.
[52] P. Khorrami, V. Le, J. C. Hart, and T. S. Huang, "A system for monitoring the engagement of remote online students using eye gaze estimation," in *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2014, pp. 1–6.
[53] A. Sharma and P. Abrol, "Eye gaze techniques for human computer

interaction: A research survey," *Int. J. Comput. Appl.*, vol. 71, no. 9, 2013.

[54] R. Stiefelhagen, "Tracking focus of attention in meetings," in *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, 2002, pp. 273–280.

[55] H. Jarodzka, I. Skuballa, and H. Gruber, "Eye-tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom," *Educ. Psychol. Rev.*, vol. 33, no. 1, pp. 1–10, 2021.

[56] K. A. Rocca, "Student participation in the college classroom: An extended multidisciplinary literature review," *Commun. Educ.*, vol. 59, no. 2, pp. 185–213, 2010.

[57] Z. Liu, J. Zhu, J. Bu, and C. Chen, "A survey of human pose estimation: the body parts parsing based methods," *J. Vis. Commun. Image Represent.*, vol. 32, pp. 10–19, 2015.