كلية علوم وهندسة الحاسب
College of Computer Science & Engineering

جامعة جدة
University of Jeddah

Monitoring Virtual Classroom In E-learning Environment Using Artificial Intelligence

Mehdhar Fadhol Saggaf	2140037
Salman Ahmed Alamoudi	2040198
Saud Abdullah Sahli	2140065
Waleed Ghazy AlRishi	2140033
Anas Elyas Mandourah	2142516

Supervised by

Dr. Hezam A. Albaqami

Contents

1. ABSTRACT:

More students today are taking online courses, but due to a lack of engagement, many feel that the quality of online courses is not as good as face-to-face classes. This project uses Artificial intelligence (AI) and deep learning to create a complete monitoring system for e-learning applications. The system identifies students through YOLOv8 for person detection and uses MediaPipe for face analysis to study facial expressions. A Convolutional Neural Network (CNN) model classifies students' behavior into two categories: 'focused' and 'distracted.' Manifestations of distracted behavior include multitasking, using other devices, and mind-wandering. Our proposed solution achieves 94% accuracy in real-time facial expression detection, enabling teachers to improve teaching methods while maintaining discipline in the classroom.

2. CHAPTER 1:

2.1. Introduction:

During the COVID-19 pandemic, a sudden transformation of the entire education system from traditional in-person learning to virtual classrooms, causing a dramatic shift in how teaching and learning were performed [1]. This rapid change posed significant challenges, including issues with student distraction, reduced attention spans, and poor concentration [2].

According to research, people experience mind wandering 30%–50% of the time in their daily lives [3]. As digital learning becomes more connected and educational technology more advanced, it is essential to address these challenges to ensure that students remain engaged, focused, and achieve high learning outcomes.

To improve the online learning experience, we propose the use of AI specifically computer vision and deep learning techniques to monitor student engagement. By tracking head position and movement, our system will detect instances of distraction, offering real-time feedback that can help instructors re-engage students when necessary. This approach not only enhances the quality of virtual learning but also promotes a more interactive environment between teachers and students. Ultimately, by combining technology with educational strategies, we aim to create a virtual learning experience that is more responsive to student needs and supportive of achieving academic success.

2.2. Problem Definition:

In virtual learning environments, student engagement is often lower than in traditional classrooms, leading to issues like distraction, reduced focus, and poor learning outcomes. Unlike physical classrooms where teachers can more easily observe and address disengagement, online instructors face challenges in identifying and responding to student distraction in real time. This lack of immediate feedback on student attention leads to a less interactive and effective learning experience.

2.3. Aim:

We aim to improve the online learning experience by using computer vision and deep learning to track student engagement. By monitoring head position and movement, the system will detect distractions and give real-time feedback to help instructors re-engage students. This approach enhances virtual learning quality and encourages more interaction between teachers and students. Ultimately, we want to create a learning environment that better meets student needs and supports academic success.

2.4. Objectives:

Our solution addresses the problem of distractions by focusing on specific types: multitasking, mind wandering, and using other devices. To help with this, Questions will appear Occasionally, requiring the student to provide a response.

These questions can be created by the teacher, or we will use a default question if the teacher does not add one. For mind wandering, we will track the student's head position to see if they are looking away from the virtual classroom.

3. CHAPTER2

3.1. literature review

In recent years, various systems have been developed to improve engagement monitoring, attention analysis, and behavior detection in e-learning environments and other settings. These systems leverage diverse methodologies, datasets, and models tailored to their specific purposes.

One system focused on multimodal facial cues for engagement detection in e-learning. The project used the FER-2013 and Wider Face datasets and employed the VGG19 model, achieving an accuracy of 92.58% [4].

Another system aimed to recognize multidimensional engagement in e-learners by integrating data from video/image sequences, eye movement data from an eye tracker, and clickstream data. This system utilized the USTC-NVIE and ImageNet datasets and tested models such as VGG16 with LSTM (76.08%), VGG16 without LSTM (71.54%), Inception-ResNetV2 with LSTM (75.47%), and Inception-ResNetV2 without LSTM (72.32%) [5].

In virtual classrooms, another project focused on student attentiveness, particularly analyzing distraction and emotion. Data was collected from seven participants, and the LSTM model achieved an accuracy of 90.2% in detecting student engagement [6].

Another initiative sought to detect drowsiness in learners during online learning by monitoring heart rate, seat pressure, and facial recognition. The system used the SMOTE dataset and applied models like CatBoost, SVM, and RF, reaching an F1 score of 0.82 [7].

A system designed for real-time face detection and optimal face mapping in online classes used the ORL public dataset and a custom dataset collected with a laptop camera. This system employed CNN and LBPH models, achieving accuracies of 95% for CNN and 78% for LBPH [8].

Similarly, a project that utilized eye gaze data from an educational VR environment aimed to classify student distraction levels. The dataset, containing 3.4 million data points, was used to train a CNN-LSTM model, which achieved an accuracy of 89.8% [9].

Another project focused on monitoring students during online examinations, using face detection and recognition systems. The collected dataset was analyzed using the SVM model, which achieved an accuracy of 61% [10].

Additionally, a system designed to detect abnormal behaviors during online exams used descriptive data combined with various models such as CNN, VGG16, and MobileNetV3. The results showed accuracies ranging from 51% to 79%, depending on the model [11].

A system that monitored student attention in classrooms recognized behavior through facial expression analysis. This project used the AffectNet dataset and custom-collected data, achieving an accuracy of 87.7% using the YOLOv5 model [12]. In object detection, a project utilizing the YOLO series models (ranging from YOLOv1 to YOLOv10) achieved significant results, with YOLOv5 reaching an accuracy of 50% and YOLOv8 achieving 55% average precision (AP), while YOLOv10 surpassed 60% AP when tested on the COCO dataset [13].

In healthcare, the YOLOv5, MediaPipe, and CBAM models were used in a real-time motion assessment tool for spinal and shoulder conditions, achieving precision and recall rates as high as 92.3%-98.7%, with mAP values ranging from 90.4% to 98.6% [14].

For real-time American Sign Language (ASL) gesture recognition, the Lexset dataset from Kaggle and Roboflow were used with YOLOv8 and MediaPipe models, achieving a precision of 98%, a recall of 98%, an F1 score of 99%, and a mean average precision (mAP) of 98%, along with an mAP50-95 of 93% [15].

An optimized version of YOLOv8, called YOLOv8-CAB, was developed for real-time object detection, particularly for small and geometrically complex objects. Using the COCO dataset, the model achieved a mAP of 97% [16].

Another project that measured student engagement in virtual learning environments utilized facial landmarks and spatial-temporal graph convolutional networks, reaching accuracies of 71.24% on EngageNet and 83.15% on Online SE datasets [17].

Drowsiness detection for drivers was also a focus in several studies, with one system leveraging facial cues with MediaPipe and OpenCV. The CNN model applied to the MRL Eye and YawDD datasets achieved an accuracy of 84.53% on MRL and 96.42% on YawDD [18].

A distracted driving detection system that combined MHSA and CNN modules with the YOLOv8 model achieved an average precision of 81.4%, using the FER2013 dataset and custom-collected data [19].

A modified version of YOLOv8, ME-YOLOv8, was used to detect driver distraction and fatigue, achieving a mAP@0.5 of 97.9% and a mAP@0.5:0.95 of 75.3% on the DDFDD dataset [20].

The P-YOLOv8 system focused on real-time detection of distracted driving behaviors, using the State Farm dataset. The model achieved an average precision of 0.9942, recall of 0.9946, and an F1 score of 0.9944 [21].

Simple face detection was explored through the Haar Cascade algorithm, aimed at efficient real-time processing in low-resource environments using a custom dataset [22].

Another project compared face detection algorithms and their performance in real-time applications, using a custom dataset [23].

Similar research focused on adapting face detection models for mobile devices, ensuring low latency and compatibility with mobile hardware, using custom datasets [24].

Another system developed a low-power face detection model tailored for continuous mobile operation, optimizing battery efficiency, using a custom dataset [25].

An efficient face detection model for edge devices was developed, aimed at real-time detection with limited computational power. This system used the Wider Face and custom datasets, applying lightweight CNN models for improved efficiency [26].

Finally, a comprehensive system for monitoring behavioral patterns of students during online examinations was developed to detect abnormal behaviors and ensure academic integrity. The project leverages facial landmark extraction using MediaPipe, a real-time computer vision framework, to track the positions of key facial features—such as the eyes, head, and lips—throughout the exam. The system uses a K-Nearest Neighbors (K-NN) classifier to classify head, eye, and lip orientations, detecting any potential signs of cheating, such as looking away from the screen or making suspicious movements. The system incorporates a penalty system that tracks abnormal changes in facial orientations. If orientations shift too drastically, a penalty is applied, and a cheating score is calculated for each student. A higher score indicates potential dishonesty, and if this score exceeds a predefined threshold, the system flags the student for further investigation.

The system also includes a visualization feature, providing proctors with real-time graphs of students' movements and cheating scores. This visual feedback allows for easy detection of suspicious patterns, such as group cheating or systematic distraction. Using a custom dataset, the system achieved 87.5% accuracy in detecting cheating and 100% accuracy in identifying non-cheating students. This advanced system provides a highly reliable solution for monitoring online exams, offering both real-time detection and preventive measures to combat academic misconduct [27].

3.2. Gap:

Research has addressed the use of a range of advanced techniques to analyze students' behaviors and improve the distance learning experience. For example, some studies have presented models such as VGG19 using FER- 2013 and Wider face to achieve high accuracy in recognizing facial expressions, with the aim of enhancing understanding of emotions and students' participation. However, this research focused on analyzing facial expressions only, without incorporating this into systems capable of monitoring dispersion in real time and continuously.

Other studies have used models such as LSTM to measure student participation or predict student attention levels, by collecting data on eye movements or behavior patterns. With some promising results, these systems still lack the ability to monitor dispersion in real time, which is an obstacle to their effective application in virtual learning environments. In addition, technologies such as MediaPipe have been used in different contexts, such as tracking facial expressions and eye movements, but have not been combined with advanced systems such as YOLO to achieve an immediate and rapid response.

Through our study of this research, it is clear that most current solutions either suffer from delayed response or lack complementarity between facial recognition and motion tracking, limiting their ability to provide accurate and sustainable monitoring of students' behavior. This shortage is a major challenge, especially in virtual learning environments that require flexible and effective solutions to ensure an ideal learning experience.

The opportunity here is to improve response and accurate monitoring of students' behavior, which previous research has not been able to achieve.

This Integration of technologies is a key factor in bridging the gap and providing advanced solutions in line with the requirements of virtual learning environments, providing a robust system capable of improving interaction and significantly reducing dispersion levels.

4. CHAPTER 3

4.1. Defining distraction:

Distraction is anything that takes your focus away from what you are supposed to be doing. It can happen when something else grabs your attention, whether it is something inside your head (like mind wandering) or from the environment (like noise or other people). When you get distracted, it becomes harder to pay attention, which can slow you down and reduce your performance, whether in studying, working, or other tasks Distractions can be internal, like thoughts that make you lose track of what you're doing, or external, like a phone ringing or someone talking nearby. When you switch your focus between tasks like checking social media while studying it can make you less effective because your brain must keep shifting attention, which reduces how well you can learn or complete work.

4.2. Types of distraction:

4.2.1. Multitasking:

This means students trying to do more than one thing at the same time, like listening to a lecture while checking emails, scrolling through social media, watching unrelated videos, or working on other assignments.

Multitasking splits attention makes it harder to think clearly and lowers the quality of learning because the brain has to keep switching between tasks and refocusing on the lesson. This can slow down work, cause mistakes, and lead to not fully understanding what is being taught [2,28,29].

4.2.2. Mind Wandering:

This happens when a student's thoughts drift away from the lesson or task, even if they seem to be paying attention by looking at the screen or taking notes. It often happens without them realizing it and can hurt focus, causing them to miss important information and lose track of the lesson. If it happens often, mind wandering can leave big gaps in understanding and the ability to remember what they have learned [2,3].

4.2.3. Using Other Devices:

This is when students use other devices, like a smartphone, tablet, or gaming console, while in a virtual class or studying. These devices often distract from the lesson, drawing attention away from what is important. Alerts, apps, or games can lower focus and make it harder for students to take part in class, which can hurt their school performance [2].

4.2.4. Other:

This includes other kinds of distractions not mentioned above. Examples are background noise from a TV, music, or people talking in the house; interruptions from family members, roommates, or pets; and personal challenges like feeling tired, hungry, or stressed. These distractions can make it hard for students to stay focused, join in the lesson, and fully understand the material [2].

4.3. Digital Distraction and Its Impact on Learning:

Digital distractions happen when students lose focus because of digital devices or online content, like social media, games, or notifications. It affects learning in both virtual and non-virtual classrooms by pulling attention away from lessons and interrupting the learning process [30,31].

In virtual classrooms, students often lose focus because they can easily switch to other apps or websites.

This can lead to missed information and poor understanding. Notifications and entertainment online lower interest in lessons, while multitasking—like chatting or browsing during class—makes it harder to learn and remember things. Time spent on distractions also leaves less time to complete tasks.

In non-virtual classrooms, students may use phones or tablets for unrelated activities, like texting or playing games. Notifications or calls can disrupt the class, pulling attention away from the teacher. Seeing classmates use devices can tempt others to do the same, creating more distractions. Splitting attention between lessons and devices makes it harder to understand and remember what is being taught.

Digital distractions make it harder to focus, stay engaged, and learn well, no matter what the classroom type. Managing device use is key to improving learning.

5. CHAPTER 4

5.1. Data collection:

The dataset for this study was compiled from four publicly available datasets [32–35], providing a total of 956 samples relevant to real-world scenarios of students in virtual classroom environments. It consists of two primary classes: Focus and Distracted. The Distracted class is further categorized into three labels: Looks Away, Using Phone, and Drowsy. Similarly, the Focus class includes two labels: Focused and a Special Case, which was specifically curated from Roboflow Universe. Figure 1 illustrates sample images from the dataset, while Table 1 provides a detailed breakdown of the sample distribution across the different classes.

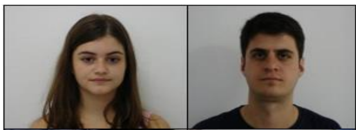
	*Class Focus - Focused*
	*Class focus - Special Case*
	*Class Distracted - Drowsy*
	*Class Distracted - Using Phone*
	*Class Distracted - Looks Away*

Table 1 sample distribution

Name classe	Distracted			Focused	
Name label	Drowsy	Look_away	Useing_phone	Focused	Special_Case
Total image label	158	248	99	308	161
Total image class	505			469	

Figure 1.2 sample of image

5.2. Data Preprocessing

The dataset was processed using the MediaPipe Face Detection module to extract face regions, which were subsequently resized to 224×224 pixels. Images from the "Focused" and "Distracted" directories were loaded, converted to RGB format, and labeled as 0 (Focused) or 1 (Distracted). These labeled images were combined into a unified dataset, normalized to a [0, 1] range, and split into training and testing subsets to facilitate model development.

5.3. Model:

In our project, we used the YOLOv8 model, which is used to detect if there is a person or not, and then used a combine between MediaPipe and CNN, which is used MediaPipe to segmentation and determine facial features then the CNN model will decide if the person is distracted or focus.

5.3.1. YOLOv8:

One of the released models in the YOLO (You Only Look Once) series of object detection models, developed by Ultralytics. It is an advanced deep learning framework designed for tasks like object detection, image segmentation, and image classification, It is train on COCO dataset (Common Objects in Context) contains 330K images, with 200K images having annotations for object detection, segmentation, and captioning tasks and the dataset comprises 80 object categories, including common objects like cars, bicycles, and animals, as well as more specific categories such as umbrellas, handbags, and sports equipment.

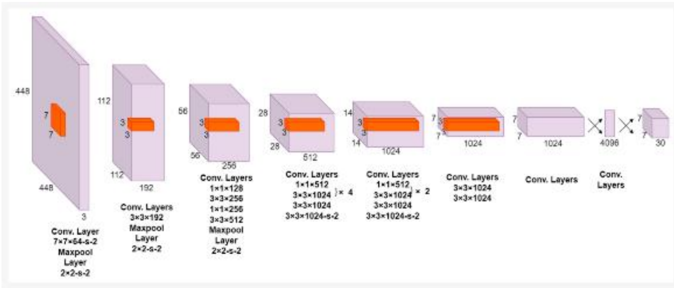


Figure1.2 : Architecture of the YOLO object detection model.

5.3.2. MediaPipe:

It is an open-source framework developed by Google that facilitates building multimodal applied machine learning (ML) pipelines, especially for real-time computer vision tasks. MediaPipe offers solutions for a range of applications, including face detection, hand tracking, pose estimation, and object detection, MediaPipe Pose Estimation is based on the Blazepose architecture capable of detecting a total of 33 keypoints, 16 more than the standard 17 keypoints defined by the Common Objects in Context (COCO) topology. These include keypoints for the face, fingers and feet, and these additional keypoints enable the capture of more intricate semantic information not possible with other models. The combination of these 2 capabilities makes BlazePose the preferred model for many real-time video applications such as the tracking of facial expressions and hand gestures, as well as for fitness applications like sports, dance and yoga trackers.

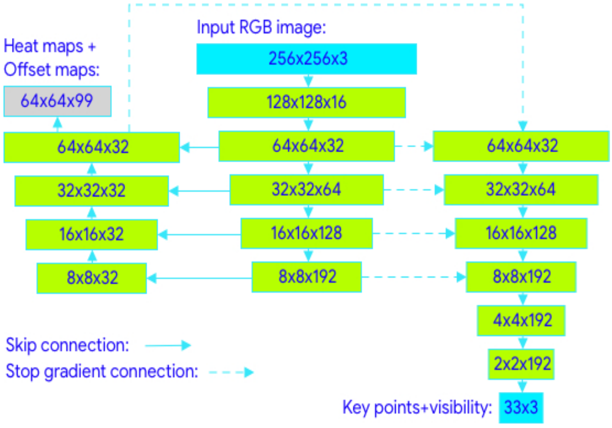


Figure1.3: Architecture of the Blazepose.

5.3.3. Convolutional Neural Network (CNN):

Convolutional Neural Networks (CNN) consist of 5 layers convolutional layers, Max-pooling layers, fully connected layers, and output binary sigmoid. automatically detect and learn features from images.

The CNN model was integrated with the MediaPipe model to enhance its performance. MediaPipe Segmentation identifies faces in the dataset images, provides the bounding box for the detected face, and the bounding box is used to crop the face region from the image, which is then resized to fit the CNN model input dimensions (224 × 224 pixels). The CNN model is trained on pre-processed face data and learns patterns in the face images to binary classification between two classes: “focus” and “distraction.” We tested the CNN model and achieved an accuracy of 94% - 96.4%.

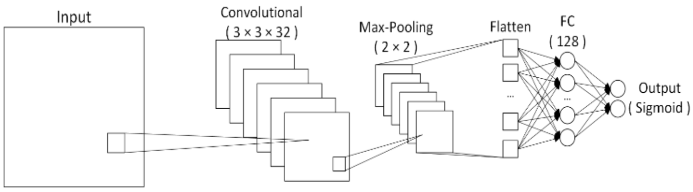
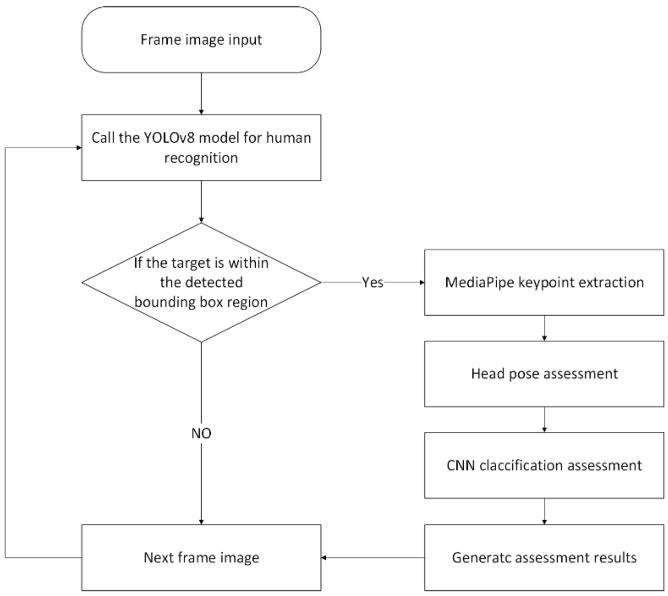


Figure1.4: Architecture of the Convolutional Neural Network (CNN).

5.4. Methods Used:

The program starts at get the time (class time) to start setting the standing time and to calculate the total distraction time and to show random questions pop up at certain times to ask the person if he is present or not, if he does not answer within 15 seconds, he is considered distracted until he answers.

YOLOv8 model starts to detect if there is a person or not, if there is a person then it starts calling MediaPipe model inside the coordinates bounty box given by YOLOv8 to facilitate the face segmentation process, MediaPipe identifies the main facial points in a grid of points to determine the distraction by head movement, meanwhile CNN determines whether the person is distracted or focused by the data we trained on it, this process continues until the time runs out, and then a panel appears that determines the duration of the person’s distraction.



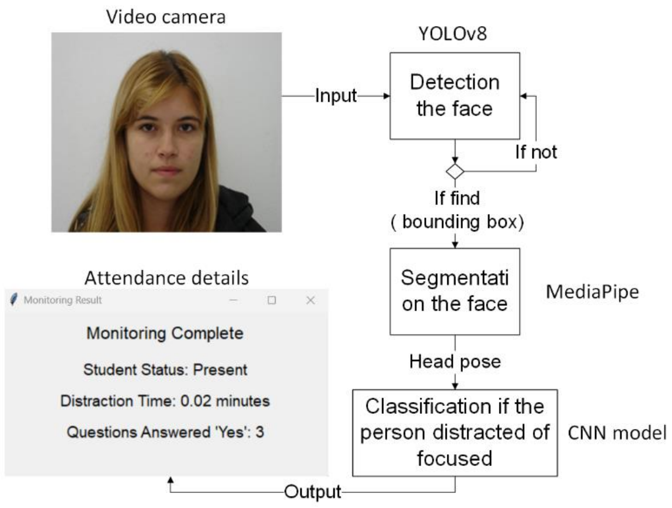


Figure1.4: The process of how the program works.

6. chapter 5

6.1. Result:

After we combine between mediapip and CNN model and train it in our dataset then we got this result for each class as the follow:

class/performance	precision	recall	f1-score
focused	0.90	0.99	0.94
distract	0.99	0.90	0.94

then we calculate the total performance of the combine between MediaPipe and CNN model is as follow:

model / performance	Accuracy	Recall	f1-score
mediapip & CNN	0.9427	0.9019	0.9435

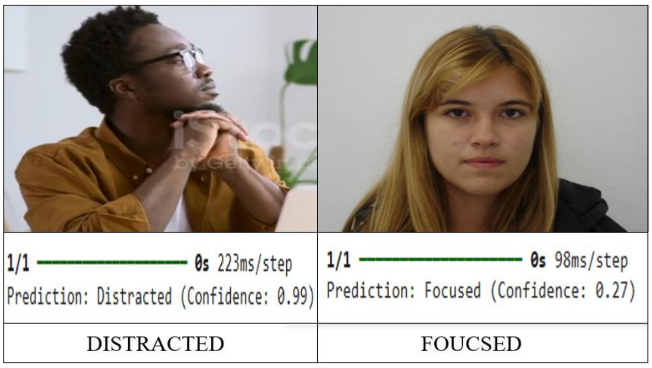


Figure: her there is some examples of uses

6.2. Discussion:

The sudden and widespread shift to e-learning platforms has created an urgent need to ensure that learners remain engaged and focused while studying online. Online education presents unique challenges, one of which is the difficulty in monitoring and assessing student activity to identify and address distractions. This project tackles this critical issue by introducing an innovative system designed to measure student activity and adjust for distractions effectively.

The proposed system provides a robust solution for real-time and frequent student monitoring by leveraging advanced deep learning methodologies and bets artificial intelligence tools. The system has demonstrated remarkable success, achieving a prediction accuracy of 94% in classifying a student's behavior as either “focused” or “distracted.” This impressive performance highlights the capability of the CNN model employed in the system to accurately analyze facial movement and expressions and behaviors.

The system is classification process is grounded in a reliable framework that integrates multiple advanced techniques. It begins with person detection facilitated by the powerful YOLOv8 model, ensuring accurate identification of students in the monitored environment. Following this, the MediaPipe framework performs detailed facial landmark analysis, providing precise data on facial features and hade movement . These inputs are then fed into the CNN model.

The results of the project underscore the effectiveness of combining detection and classification technologies to deliver a comprehensive solution for student monitoring. By utilizing the strengths of YOLOv8, MediaPipe, and CNNs in a synergistic manner, this system sets a new benchmark for addressing the challenges of maintaining student engagement and focus in online learning environments. This integration demonstrates how modern AI tools can be applied creatively and effectively to meet pressing needs in education, ensuring that the technologies used are harnessed in the most impactful way possible.

7. CONOLICTION

We have implemented a monitoring system that uses AI to detect student deviations during virtual classes. Using YOLOv8 for person detection, MediaPipe for face analysis, and a CNN model for behaviour classification can classify students as focused or distracted with an accuracy of 94% or more. Through this solution teachers can get useful information on student participation in classes and the efficacy of virtual classes. Also, it aids in evaluating the extent of student distraction and ascertaining their presence or otherwise that leads to better discipline and interactivity.

Reference

1- Wieland, N., & Kollias, L. (2020). Online learning before, during and after COVID-19: Observations over 20 years. International Journal of Advanced Corporate Learning, 13(2), 84.

2- Wang, C. (2022). Comprehensively summarizing what distracts students from online learning: A literature review. Human Behavior and Emerging Technologies, 2022(1), 1483531

3- Hollis, R. B., & Was, C. A. (2016). Mind wandering, control failures, and social media distractions in online learning. Learning and Instruction, 42, 104-112.

4- Gupta, S., Kumar, P., & Tekchandani, R. (2023). A multimodal facial cues based engagement detection system in e-learning context using deep learning approach. *Multimedia Tools and Applications*, *82*(18), 28589-28615.

5- Yue, J., Tian, F., Chao, K. M., Shah, N., Li, L., Chen, Y., & Zheng, Q. (2019). Recognizing multidimensional engagement of e-learners based on multi-channel data in e-learning environment. *IEEE Access*, *7*, 149554 149567.

6- Kaewkaisorn, K., Pintong, K., Bunyang, S., Tansawat, T., & Siriborvornratanakul, T. (2024). Student attentiveness analysis in virtual classroom using distraction, drowsiness and emotion detection. *Discover Education*, *3*(1), 1-14.

7- Caucasian, K., Pintong, K., Bunyang, S., Tansawat, T., & Siriborvornratanakul, T. (2024). Student attentiveness analysis in virtual classroom using distraction, drowsiness and emotion detection. *Discover Education*, *3*(1), 1-14.

8- Archana, M. C. P., Nitish, C. K., & Harikumar, S. (2022). Real time face detection and optimal face mapping for online classes. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012063). IOP Publishing.

9- Asish, S. M., Hossain, E., AK, K., & Borst, C. W. (2021, January). Deep learning on eye gaze data to classify student distraction level in an educational vr environment. In *International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments (ICAT-EGVE)*.

10- Geetha, M., Latha, R. S., Nivetha, S. K., Hariprasath, S., Gowtham, S., & Deepak, C. S. (2021, January). Design of face detection and recognition system to monitor students during online examinations using Machine Learning algorithms. In *2021 international conference on computer communication and informatics (ICCCI)* (pp. 1-4). IEEE.

11- Ngo, D. A., Nguyen, T. D., Dang, T. L. C., Le, H. H., Ho, T. B., Nguyen, V. T. K., & Nguyen, T. T. H. (2024). Examining Monitoring System: Detecting Abnormal Behavior In Online Examinations. *arXiv preprint arXiv:2402.12179*.

12- Trabelsi, Z., Alnajjar, F., Parambil, M. M. A., Gochoo, M., & Ali, L. (2023). Real-time attention monitoring system for classroom: A deep learning approach for student's behavior recognition. *Big Data and Cognitive Computing*, *7*(1), 48.

13- Wang, C. Y., & Liao, H. Y. M. (2024). YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems. *APSIPA Transactions on Signal and Information Processing*, *13*(1)

14- Zhang, W., Li, Y., Cai, S., Wang, Z., Cheng, X., Somjit, N., ... & Chen, F. (2024). Combined MediaPipe and YOLOv5 range of motion assessment system for spinal diseases and frozen shoulder. *Scientific Reports*, *14*(1), 15879.

15- Alsharif, B., Alalwany, E., & Ilyas, M. (2024). Transfer learning with YOLOV8 for real-time recognition system of American Sign Language Alphabet. *Franklin Open*, *8*, 100165.

16- Talib, M., Al-Noori, A. H., & Suad, J. (2024). YOLOv8-CAB: Improved YOLOv8 for Real-time object detection. *Karbala International Journal of Modern Science*, *10*(1), 5.

17- Abedi, A., & Khan, S. S. (2024). Engagement Measurement Based on Facial Landmarks and Spatial-Temporal Graph Convolutional Networks. *arXiv preprint arXiv:2403.17175*.

18- Reddy, A. K., Khan, I. A., Neethu, C. T., Thampi, L. L., Aswathy, M. A., Kumar, A., & Kumar, S. (2024). Smart Driver Assistance: Real-Time Drowsiness Detection Leveraging Facial Cues with MediaPipe and OpenCV.

19- Ma, B., Fu, Z., Rakheja, S., Zhao, D., He, W., Ming, W., & Zhang, Z. (2024). Distracted Driving Behavior and Driver's Emotion Detection Based on Improved YOLOv8 with Attention Mechanism. *IEEE Access*.

20- Debsi, A., Ling, G., Al-Mahbashi, M., Al-Soswa, M., & Abdullah, A. (2024). Driver distraction and fatigue detection in images using ME-YOLOv8 algorithm. *IET Intelligent Transport Systems*.

21- Elshamy, M. R., Emara, H. M., Shoaib, M. R., & Badawy, A. H. A. (2024). P-YOLOv8: Efficient and Accurate Real-Time Detection of Distracted Driving. *arXiv preprint arXiv:2410.15602*.

22- Face Detection Using Haar Cascade Algorithm: Applying the Haar Cascade algorithm for efficient face detection, focusing on real-time processing in low-resource settings.

23- Younus, S. M., Bhardwaj, V., Jain, A., Sharma, V., & Reddy, T. J. (2023, August). A Comparative Analysis of Face Detection Algorithms and Real-Time Facial Recognition. In *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 173-177). IEEE.

24- Khabarlak, K. (2022). Face detection on mobile: Five implementations and analysis. *arXiv preprint arXiv:2205.05572*.

25- Lee, H., Jeong, W., Ryu, D., Je, H., No, A., Kim, K., & Chun, S. Y. (2023). Fully Quantized Always-on Face Detector Considering Mobile Image Sensors. *arXiv preprint arXiv:2311.01001*.

26- Efficient Face Detection on Edge Devices Using Lightweight CNN Models: Developing an efficient face detection model for edge devices, achieving real-time detection with limited computational power.

27- actions.Ferdosi, B. J., Rahman, M., Sakib, A. M., & Helaly, T. (2023). Modeling and Classification of the Behavioral Patterns of Students Participating in Online Examination. *Human Behavior and Emerging Technologies*, *2023*(1), 261380

28- Lepp, A., Barkley, J. E., Karpinski, A. C., & Singh, S. (2019). College students' multitasking behavior in online versus face-to-face courses. Sage Open, 9(1), 2158244018824505.

29- Parlangeli, O., Mengoni, G., & Guidi, S. (2011, September). The effect of system usability and multitasking activities in distance learning. In Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction: Facing Complexity (pp. 59-64).

30- Rosen, L. D., Carrier, L. M., & Cheever, N. A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, *29*(3), 948-958.

31- Junco, R., & Cotten, S. R. (2011). Perceived academic effects of instant messaging use. *Computers & Education*, *56*(2), 370-378.

32- (name: Test Computer Vision Project, https://universe.roboflow.com/facehenry/test-5wztq)

33- (name: Attention_Detection_1 Computer Vision Project, https://universe.roboflow.com/summer-project-2/attention_detection_1).

34- (name: RIMAS Computer Vision Project, https://universe.roboflow.com/project-fetpf/rimas-oonjc).

35- (name: drowsy detection Computer Vision Project, https://universe.roboflow.com/madness/drowsy-detection-83unc/browse?queryText=&pageSize=50&startingIndex=0&browseQuery=true