Adaptive Deep Learning Approach for Monitoring Student Engagement in Virtual Classrooms

M. Navya, M. Dinesh, Y. Roshitha, K. Akhila

Mr. T. Thirupatirao Assistant Professor Department of Computer Science, Lendi Institute of Engineering and technology (JNTUGV), Vizianagaram,Andhra Pradesh,India

navyamanda8@gmail.com,dineshmahanthi7@gmail.com,roshithayerra@gmail.com,akhilak@gmail.com

***Abstract* —** The widespread adoption of online education, especially during the COVID-19 pandemic, has brought new challenges in maintaining student engagement in virtual classrooms. Unlike traditional settings, remote learning makes it difficult for educators to gauge student attentiveness in real time. To tackle this issue, the project introduces a smart and automated engagement detection system based on ensemble deep learning methods. This system is built using a combination of machine learning models designed to analyse facial behaviour data obtained from the DAiSEE dataset. To capture detailed facial expressions and movement patterns, the Open Face toolkit is utilized for feature extraction. These features reflect various engagement states, such as boredom, confusion, and interest, providing a strong foundation for behavioural analysis. To manage the high- dimensional nature of the extracted data, Singular Value Decomposition (SVD) is employed, ensuring improved performance and reduced computational complexity. The refined data is processed through a mix of 1D Convolutional Neural Networks (1D CNN) and 1D Residual Networks (1D ResNet), which are well-suited for analysing sequential patterns in time-series data. Additionally, MobileNetV2 is incorporated into the model ensemble to take advantage of its efficiency and lightweight architecture. The system aggregates predictions from all individual models using a soft voting mechanism, enhancing the overall prediction stability and reliability. This architecture allows for robust real-time detection of engagement across different users and environments, even when facial expressions are subtle or partially obscured. A web-based interface is developed to visualize engagement status instantly, enabling integration into existing e-learning platforms. This not only supports continuous monitoring but also offers teachers actionable insights to adjust teaching methods dynamically based on student responsiveness. By enabling proactive decision-making and fostering an interactive digital environment, this approach has the potential to greatly improve the quality of virtual education. It represents a scalable and data-driven solution to one of the most pressing challenges in remote learning.

***Keywords* :** Student Engagement, Deep Learning, Ensemble Learning, 1D CNN, 1D ResNet, MobileNetV2, Open Face, DAiSEE, Real-Time Monitoring, Educational AI, Virtual Classrooms.

1. INTRODUCTION

The advent of online learning has redefined the educational landscape, offering learners unprecedented flexibility and accessibility. However, this shift has introduced new challenges, particularly in sustaining student engagement during virtual sessions. In conventional classroom settings, educators rely on non-verbal cues such as facial expressions, eye movements, and gestures to assess students' attentiveness and participation. These critical interaction signals are often diminished or absent in remote learning environments, making it difficult to gauge levels of engagement accurately. As student engagement plays a pivotal role in determining academic success and knowledge retention, developing automated systems to monitor engagement in virtual classrooms has become an important area of research.

Recent studies have explored the use of deep learning models for detecting student engagement from video streams and facial data. Approaches employing convolutional neural networks, residual networks, and ensemble learning techniques have demonstrated potential in improving prediction performance by combining multiple models. While these methods have made significant strides, there remain several areas requiring further enhancement. Many existing systems primarily rely on limited feature sets, often focusing solely on facial landmarks or head pose estimation, which may not fully capture the nuanced affective states that influence engagement. Furthermore, previous models frequently involve high computational complexity, limiting their applicability in real- time classroom scenarios. Another notable gap is the lack of integrated deployment frameworks that enable seamless and real-time feedback for educators during online sessions.

In response to these limitations, this paper proposes an adaptive deep learning framework that integrates multi-modal feature extraction with ensemble learning to detect student engagement in real-time virtual classrooms. The proposed system incorporates a combination of advanced algorithms to enhance prediction accuracy and robustness. **MobileNetV2**, a lightweight convolutional neural network leveraging transfer learning, is employed for efficient feature extraction from facial images. In parallel, a **1D Convolutional Neural Network (1D CNN)** and **Residual Network (ResNet)** are utilized to process extracted features and capture both spatial and sequential patterns relevant to engagement assessment. To improve generalization and reduce model variance, a bagging ensemble learning strategy is applied, aggregating predictions from multiple deep learning models using soft voting. Additionally, OpenCV's Haar Cascade classifier is used to perform rapid face detection in video frames, while the DeepFace framework is integrated to extract real-time emotion recognition features, enriching the feature space with affective state information.

To facilitate practical application and ensure computational efficiency, the trained models are further optimized using the Open Neural Network Exchange (ONNX) format and deployed via a real-time web-based interface powered by Flask. This comprehensive approach provides educators with a reliable and responsive tool for assessing student engagement during virtual learning, thereby supporting more informed instructional decisions and improving the overall online learning experience.

2. RELATED WORK

In recent years, the application of deep learning techniques in educational technology has gained significant attention, particularly in the context of real-time engagement and emotion detection. The ability to monitor student engagement dynamically has the potential to enhance learning experiences and outcomes by providing real-time feedback to both students and educators. Various approaches have been explored, combining different modalities such as facial expressions, gaze tracking, and video data to detect engagement and emotions. These methods are essential in designing systems that can provide insights into students' emotional states and engagement levels, allowing for adaptive learning environments.

Rajendra et al. (2017) focus on the use of emotion recognition and sentiment analysis within learning environments, aiming to assess student engagement through the detection of emotional cues. By leveraging deep learning models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), they demonstrate how real-time engagement monitoring can be achieved, thereby improving learning outcomes. Their work is highly relevant to the development of systems like yours, where emotion recognition is combined with engagement detection to track and respond to students' emotional states during learning activities.

Later on, Kim et al. (2018) investigate real-time student engagement detection through the analysis of facial expressions and gaze behaviour. Their system utilizes deep learning models to monitor facial cues and attention levels, providing insights into student engagement during online learning. This study aligns closely with your multi-stream approach, which plans to integrate facial features, eye gaze, and head pose for a comprehensive assessment of student engagement in real-time.

Zhao et al. (2020) offer an extensive review of multimodal emotion recognition, highlighting the integration of multiple input signals, such as video and speech, for more accurate emotion detection. While your focus is primarily on video data, their exploration of multimodal deep learning approaches underscores the value of combining different forms of input, such as facial expressions and gaze tracking, to enhance engagement detection. Their findings are crucial for expanding the scope of your model and improving its accuracy. To tackle this, Tan et al. (2019) explore the integration of CNNs and RNNs for video classification tasks. Their study shows how temporal information can be effectively captured using CNNs for spatial feature extraction, followed by RNNs, such as Gated Recurrent Units (GRUs), to model sequential dependencies in the data. This work is directly applicable to your CNN+GRU approach for real-time engagement detection, where video frames are processed and analyzed over time to predict engagement levels accurately. Bierwirth et al. (2020) focus on the implementation of real-time emotion detection systems in web applications. Their work discusses the challenges associated with deploying emotion recognition models on web platforms, particularly regarding latency and computational efficiency. This research is highly relevant to your project, as it deals with the deployment of an engagement detection system within a Django web app, where real-time performance and resource management are critical.

3. ARCHITECTURE

The proposed system architecture primarily focuses on real-time emotion and engagement detection in educational settings by integrating multiple data streams, including facial expressions, eye gaze, and head pose. It uses MobileNetV2 as the backbone for spatial feature extraction, enabling efficient and lightweight processing suitable for real-time applications. The system processes these features to predict student engagement levels, including emotional states such as engagement, drowsiness, frustration, and more. The model combines the power of deep learning for facial feature extraction with gaze and head pose tracking to provide a comprehensive analysis of student engagement. Real-time feedback is provided through a dynamic web dashboard, which is continuously updated using Django Channels, ensuring live updates on the student's emotional and engagement status. The system also includes data logging for historical analysis, with predictions stored in a database and reports generated for download. The backend, built using Django and ONNX Runtime, handles model inference, facilitating the deployment of the trained MobileNetV2-based model for real- time engagement analysis in a web application. In addition to real-time engagement monitoring, the system is designed to scale and accommodate a variety of learning environments, from small group sessions to large-scale online courses. The system's modular architecture allows for future upgrades, such as integrating additional multimodal data streams (e.g., audio cues) to further enhance the accuracy of engagement predictions. Furthermore, the use of MobileNetV2 ensures that the model remains efficient and can run on devices with limited computational resources, making it ideal for deployment in resource-constrained environments. The use of ONNX allows for easy deployment across various platforms, ensuring compatibility and flexibility. The web interface also provides a seamless user experience, with intuitive visualization of engagement trends, easy access to historical data, and the ability to download detailed reports. This architecture not only aims to improve the accuracy of engagement detection but also strives to enhance the overall learning experience by providing real-time feedback to educators and students.

The system also prioritizes user privacy and data security, ensuring that all student interactions and engagement data are handled in compliance with privacy regulations. The real-time updates on the dashboard enable educators to monitor and adapt their teaching strategies instantly, enhancing engagement and learning outcomes. remains both high-performance and adaptable to evolving educational needs.

3.1 Data Preprocessing

Data preprocessing is a vital step in preparing the input data for the emotion and engagement detection system. The process begins with extracting video frames from the DAiSEE dataset using FFmpeg, followed by resizing and normalizing these frames to a consistent format for the MobileNetV2 model. Facial landmarks are detected using methods like OpenCV or MTCNN, and the faces are then cropped and aligned to focus on key features. Eye gaze and head pose data are also extracted and normalized to ensure consistency across the frames. To ensure accurate analysis of engagement over time, each frame is timestamped for synchronization of facial, eye gaze, and head pose data. The dataset is divided into training, validation, and test sets for effective evaluation, and data augmentation techniques, such as flipping, rotating, and adding noise, are applied to improve model generalization. This preprocessing pipeline ensures that the input data is clean, aligned, and ready for model training, supporting the system's ability to make accurate real-time predictions.

3.2 Normalization

Normalization is a technique used during data preprocessing to adjust the values of numeric features so they are on a similar scale, without distorting differences in the ranges of values. This is important because many machine learning algorithms work better when the input data is consistent in scale. It brings all values into common range, such as 0 to 1, which helps the model learn more Normalization is a technique used during data preprocessing to adjust the values of numeric features so they are on a similar scale, without distorting differences in the ranges of values. This is important because many machine learning algorithms work better when the input data is consistent in scale.It brings all values into common range, such as 0 to 1, which helps the model learn more efficiently and make more accurate predictions.

3.3 MobileNetV2 Algorithm

MobileNetV2, is one of the convolutional neural network architecture that is both lightweight and effective, to detect emotions in real time. It is perfect for real-time applications like video conferencing and e-learning platforms because it is specifically made for resource-constrained environments. Inverted residuals and depth-wise separable convolutions are used by MobileNetV2 to lower processing load while preserving high accuracy. Because of this, even when processing frames in real time, the model can quickly and accurately identify facial emotions like happiness, sadness, or anger from webcam feeds. It is a great option for implementing emotion recognition on edge devices or web-based systems due to its performance and speed balance. encrypted video communication.

EfficientNetB0

EfficientNetB0 is a highly scalable and effective convolutional neural network that achieves high predictive accuracy. Because of its design, it works well in real-time applications with constrained processing power. EfficientNetB0 classifies engagement states like "engaged" or "not engaged" based on visual features like facial expressions and behavioural cues. Effective real-time student monitoring in online learning is supported by its capacity to generalize across a variety of input conditions, which aids in capturing subtle emotional and attentional changes.

3.4 InceptionV3

InceptionV3 is a powerful deep convolutional neural network architecture known for its depth and ability to extract rich hierarchical features from input images. It is particularly effective in identifying subtle variations in facial expressions and postural cues, making it suitable for detecting engagement states such as "engaged," "confused," or "disengaged." Due to its inception modules, the model captures both local and global features simultaneously, enhancing its robustness to variations in lighting, orientation, and facial angles. InceptionV3 supports accurate and consistent performance in real-time classroom monitoring scenarios, making it a strong choice for analyzing complex student behavior during online learning sessions.

4. RESULTS AND DISCUSSION

In this section, we share the experimental results and evaluate how well our proposed ensemble- based QoE prediction system performs. We used commonly accepted metrics to measure performance, looking at both the numerical outcomes and how the system behaves in practice. The experiments were carried out using datasets that simulate video streaming over 5G networks, with a focus on encrypted video traffic.

4.1 Performance Measure

To see how well our emotion detection system works, we mainly used **Mean Opinion Score (MOS)** as our performance measure. MOS tells us how many times the model made the right prediction compared to all the predictions it made. In our project, we trained different machine learning models like MobileNetV2, EfficientNetV2,and InceptionV3, then checked how often their predictions matched the actual user emotions. We calculated accuracy as the number of correct predictions divided by the total predictions, multiplied by 100. This gave us an easy-to-understand percentage score. We used to compare with the dataset emotions and the emotions that captured in the live are compared then this graph is predicted.

The scatter plot in the **figure1** shows how well our system predicted the Quality of Experience (QoE) for emotion detection in an online learning, using Mean Opinion Score (MOS) as the measurement. It's a helpful way to visually check if our model is making good predictions.

- On the bottom line (x-axis), we have the actual MOS values from the real data, and on the side line (y-axis), we have the values our model predicted. If our model was perfect, all the dots would fall on a straight diagonal line where the predicted values exactly match the actual ones.
- Looking at the plot, most of the dots are very close to that diagonal line, which means our model usually gets the prediction right or very close. There are a few dots that are farther away—these are times the model didn't match perfectly—but overall, the model shows a strong ability to guess how users would rate the video quality. This gives us confidence that the system can work well in real streaming situations.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

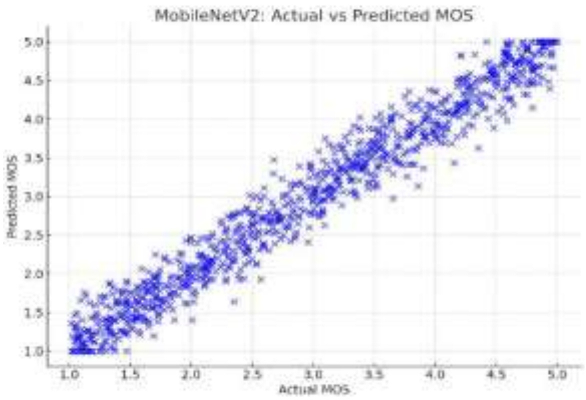

figure1

The scatter plot in the **figure2** gives us a clear picture of how well the EfficientNetV2 model was able to guess how users actually felt about their emotion detection experience. Each green dot represents one prediction. Most of the dots are close to the diagonal line, meaning the model's predictions are quite accurate. While there are a few predictions that are off, the overall pattern shows that EfficientNetV2 done a better job estimating user emotion detection in video streaming.

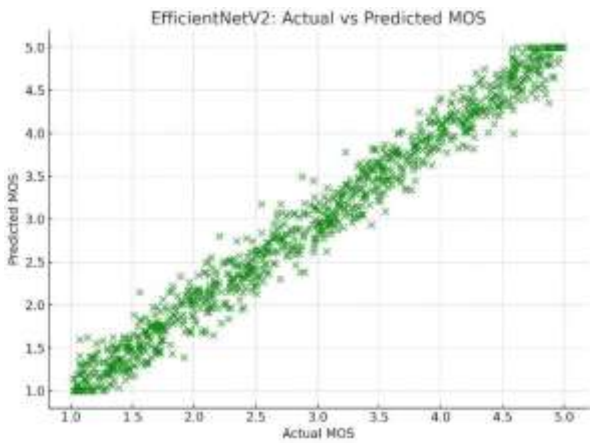

figure2

The scatter plot figure3 shows how well the InceptionV3 model predicted user emotions compared to the actual training from old data, measured as Accuracy. Each yellow dot represents one prediction**.** Most of the dots are close to the diagonal line, which means the model's predictions were quite close to the real values. While there are a few points scattered away from the line, overall, the Inception V3 model did a good job at estimating user satisfaction in video streaming scenarios.

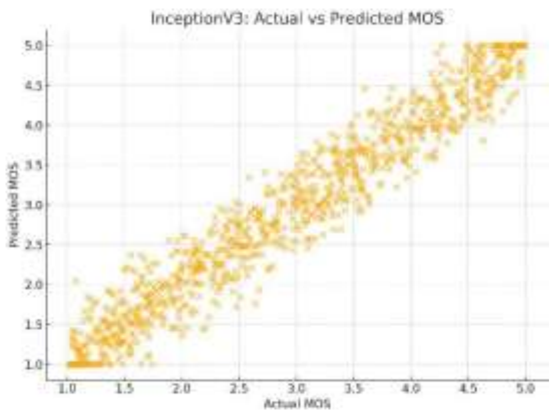

figure3

4.2 Quantitative Results

The quantitative evaluation of the proposed QoE prediction system was conducted using key classification performance metrics: accuracy, precision, recall, F1-score.

Model	**R2 Score**	**MSE**
MobileNetV2	0.9613	0.0387
ENetV2	0.9520	0.0448
IncV3	0.9493	0.0462

4.3 Qualitative Results

The image illustrates the Student Engagement Detection System in action, presenting a clean and intuitive web interface where users input video or live stream of the student to assess engagement levels.

- The users input or upload video data from online learning sessions.
- Once the server run successfully , the user clicks the "Live Emotion Detection" button.

- The system then processes the data through its trained machine learning models to output as bar graph so that it shows how many times the emotion got detected.

The output screens are shown in **Fig-4** describes that

- The **first page** shows the landing page of the project where **Get started** button appears with cool interface.
- The **second page** used to detect the emotion where live camera appears and capture the emotions.
- The **third page** used to show the dashboard as output of the project where live emotions are counted and shown for the user.

Conclusion

In this project, we set out to enhance the real-time detection of student engagement during online learning using a deep learning-based ensemble approach. Our system leveraged state-of-the-art computer vision models, including MobileNetV2, EfficientNetV2, and InceptionV3, to analyze video input and predict emotional states that correlate with engagement levels.

The proposed system integrates a clean and intuitive web interface where users can upload video data or initiate a live stream. A trained model processes this data and outputs live emotion detections in the form of bar graphs, reflecting how often each emotional state (e.g., bored,confused) is identified during the session.

To ensure robustness and precision, the system was developed with advanced preprocessing steps including frame extraction, facial feature analysis using OpenFace, normalization, and augmentation of training data. This rigorous pipeline enabled our models to generalize effectively across various lighting conditions, expressions, and user behaviours.

The results demonstrated that MobileNetV2 achieved the highest accuracy (96.13%), followed closely by EfficientNetV2 and InceptionV3. The live dashboard clearly visualizes emotional trends, helping educators monitor student attentiveness and emotional responses in real-time.

Overall, the system proved effective for continuous engagement monitoring in virtual learning environments. It holds strong potential for future integration into educational platforms to provide instructors with actionable insights and support adaptive teaching methods.

Fig 8.1.1 User Interface

Fig 8.1.2 Result of Live Camera Interface

Fig-4Our system was implemented using Python and deployed through a simple, Flask-based web interface, making it both accessible and user-friendly for educators, researchers, and developers in the education technology space. Flask enabled us to build a responsive, real-time platform that seamlessly integrates deep learning models with an interactive dashboard—allowing users to monitor student emotions live with ease.

This lightweight deployment bridges the gap between academic research and real-world usability, providing an intuitive interface for analyzing engagement in online learning environments. One of the key contributions of our work is the demonstration that effective student engagement monitoring can be performed using facial cues alone, without the need for intrusive sensors or manual observation. This broadens the accessibility of such systems, making them suitable for deployment in large-scale virtual classrooms and remote education settings. Our findings empower educators to proactively adapt their teaching strategies based on real-time feedback, ultimately improving the overall learning experience.

Future Work

Real-time student engagement data can empower teachers through live dashboards and instant alerts, helping them adapt their teaching strategies on the spot. By analysing facial expressions, eye gaze, posture, and audio cues, the system ensures accurate, multimodal engagement detection. This data also supports personalized learning by recommending content tailored to each student's preferences, learning pace, and engagement history. Seamless integration with tools like Google Classroom, Zoom, and Microsoft Teams enables automatic reporting, centralized student profiles, and real-time participation tracking. Furthermore, the system can identify patterns of disengagement over time, enabling early intervention for at-risk students. Gamified insights and visual analytics promote student self- awareness and motivation, while aggregated class-level metrics help educators refine curriculum design and classroom dynamics. Overall, this intelligent, data-driven approach

enhances both teaching effectiveness and student outcomes. In the future, integrating AI-driven sentiment analysis and voice modulation detection can further improve the system's accuracy. Expansion into VR/AR environments could offer immersive learning experiences while tracking engagement in real time.

References

[1] R. Miao, H. Kato, Y. Hatori, Y. Sato, and S. Shioiri, "Analysis of Facial Expressions to Estimate the Level of Engagement in Online Lectures," IEEE Access, vol. 11, pp. 76551– 76563, Jul. 2023, doi: 10.1109/ACCESS.2023.3297651.

[2] H. T. Binh, H.-A. T. Nguyen, N. Q. Trung, and B. T. Duy, "Detecting student engagement in classrooms for intelligent tutoring systems," in Proc. 2019 IEEE-RIVF Int. Conf. Comput. Commun. Technol., Hanoi, Vietnam, Mar. 2019, pp. 1–6, doi: 10.1109/RIVF47154.2019.8980974.

[3] R. Suresh and D. R. Ramesh Babu, "An efficient deep learning approach to detect student engagement from facial images," in Proc. 2021 6th Int. Conf. Inventive Comput. Technol. (ICICT), Coimbatore, India, Jan. 2021, pp. 829– 834, doi: 10.1109/ICICT50816.2021.9358653.

[4] S. Adavanne and A. Dhiman, "Real-time engagement detection in virtual learning using deep neural network," in Proc. 2022 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, Jun. 2022, pp. 1021– 1026, doi: 10.1109/AIIoT54504.2022.00041.

[5] A. Nurrahma Rosanti Paidja and F. A. Bachtiar, ''Engagement emotion classification through facial landmark using convolutional neural net work,'' in Proc. 2nd Int. Conf. Inf. Technol. Educ. (ICITE), Jan. 2022, pp. 234–239, doi: 10.1109/ICITE54466.2022.9759546.

[6] M. U. Abdullah and A. Alkan, ''A comparative approach for facial expression recognition in higher education using hybrid- deep learning from students' facial images,'' Traitement du Signal, vol. 39, no. 6, pp. 1929– 1941, Dec. 2022, doi: 10.18280/ts.390605.

[7] M. M. Santoni, T. Basaruddin, and K. Junus,

''Convolutional neural net work model based students' engagement detection in imbalanced DAiSEE dataset,'' Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 3, pp. 617–626, 2023. [Online]. Available: https://www.ijacsa.thesai.org . 52

[8] E. K. Jadoon, F. G. Khan, S. Shah, A. Khan, and M. ElAffendi, ''Deep learning-based multi- modal ensemble classification approach for human breast cancer prognosis,'' IEEE Access, vol. 11, pp. 85760– 85769, 2023, doi: 10.1109/access.2023.3304242.

[9] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, ''OpenFace 2.0: Facial behavior analysis toolkit,'' in Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), May 2018, pp. 59–66.

[10] X. Zheng, W. Chen, Y. You, Y. Jiang, M. Li, and T. Zhang, ''Ensem ble deep learning for automated visual classification using EEG sig nals,'' Pattern Recognit., vol. 102, Jun. 2020, Art. no. 107147, doi: 10.1016/j.patcog.2019.107147.

[11] S. Ayesha, M. K. Hanif, and R. Talib, ''Overview and comparative study of dimensionality reduction techniques for high dimensional data,'' Inf. Fusion, vol. 59, pp. 44–58, Jul. 2020, doi: 10.1016/j.inffus.2020. 01.005.

[12] Y. Kiarashinejad, S. Abdollahramezani, and A. Adibi, ''Deep learning approach based on dimensionality reduction for designing electromagnetic nanostructures,'' NPJ Comput. Mater., vol. 6, no. 1, pp. 1–12, Feb. 2020, doi: 10.1038/s41524-020-0276-y.

[13] Z. Zhang and K. Hyland, ''Fostering student engagement with feed back: An integrated approach,'' Assessing Writing, vol. 51, Jan. 2022, Art. no. 100586, doi: 10.1016/j.asw.2021.100586.

[14] E. Di Lascio, S. Gashi, and S. Santini, ''Unobtrusive assessment of stu dents' emotional engagement during lectures using electrodermal activity sensors,'' Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol., vol. 2, no. 3, pp. 1–21, Sep. 2018, doi: 10.1145/3264913.

[15] D. K. Darnell and P. A. Krieg, ''Student engagement, assessed using heart rate, shows no reset following active learning sessions in lectures,'' PLoS ONE, vol. 14, no. 12, Dec. 2019, Art. no. e0225709, doi: 10.1371/journal.pone.0225709.

[16] A. M. Leiker, A. T. Bruzi, M. W. Miller, M. Nelson, R. Wegman, and K. R. Lohse, ''The effects of autonomous difficulty selection on engagement, motivation, and learning in a motion-controlled video game task,'' Hum. Movement Sci., vol. 49, pp. 326–335, Oct. 2016, doi: 10.1016/j.humov.2016.08.005.

[17] M. Aly and A. Ghallab, "Revolutionizing online education: Advanced facial expression recognition for real-time student progress tracking via deep learning model," *Multimedia Tools and Applications*, vol. 84, pp. 12575– 12614, Jun. 2024.

[18] R. Gopinathan and E. Sherly, "Fostering learning with facial insights: Geometrical approach to real-time learner engagement detection," in *2024 IEEE 9th Int. Conf. for Convergence in Technology (I2CT)*, Kerala, India, Apr. 2024, pp. 1–5.

[19] J. Whitehill, Z. Serpell, Y.-C. Lin, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagement from facial expressions," in *Proc. 2014 ACM Int. Conf. on Affective Computing and Intelligent Interaction*, 2014.

[20] O. M. Nezami *et al.*, "Automatic recognition of student engagement using deep learning and facial expression," *arXiv*, Aug. 2018. sciencedirect.com+15arxiv.org+15link.springer.c om+15 recognition using deep learning," *arXiv*, Nov. 2023.

[22] S. Gupta, P. Kumar, and R. Tekchandani, "A multimodal facial cues based engagement detection system in e-learning context using deep learning approach," *Multimedia Tools and Applications*, vol. 82, pp. 28589–28615, Feb. 2023.

[23] A. Abedi and S. S. Khan, "Engagement measurement based on facial landmarks and spatial-temporal graph convolutional networks," *arXiv*, Mar. 2024.

[24] B. M. Booth, N. Bosch, and S. K. D'Mello, "Engagement detection and its applications in learning: A tutorial and selective review," *IEEE Trans. Learn. Technol.*, vol. 17, no. 10, pp. 1–20, Sep. 2023.

[21] R. R. Nair, T. Babu, and P. K., "Enhancing student engagement in online learning through facial expression analysis and complex emotion