System for Emotion and Engagement Recognition in Education (SEERE): An AI-Enabled System for Responsive Teaching

N P Subheesh
SRM University AP
Andhra Pradesh, India
npsubheesh@ieee.org

Sai Krishna Vishnumolakala
SRM University AP
Andhra Pradesh, India
saikrishna_vishnumolakala@srmap.edu.in

Sadwika Vallamkonda
SRM University AP
Andhra Pradesh, India
sadwika_vallamkonda@srmap.edu.in

Sobin C C
SRM University AP
Andhra Pradesh, India
sobin.c@srmap.edu.in

Prabhat Kumar
LUT University
Lappeenranta, Finland
prabhat.kumar@lut.fi

Randhir Kumar
SRM University AP
Andhra Pradesh, India
randhir.k@srmap.edu.in

***Abstract*—This paper presents the System for Emotion and Engagement Recognition in Education (SEERE), a cutting-edge advancement integrating computer vision and deep learning technologies to evaluate real-time student engagement through facial emotion recognition and eye tracking. SEERE, a transformative educational tool built on the robust YOLO V8 architecture, customizes the FER2013 dataset, making use of meticulously annotated emotion and eye position data. It goes further, establishing a unique 'concentration metric,' a quantitative index of student engagement, bridging a gap in modern responsive teaching approaches. Higher concentration metrics signal heightened student engagement, offering educators real-time data to adjust teaching techniques and feedback accordingly. The paper provides a thorough review of facial emotion recognition models, setting the stage for understanding the innovative strides made by SEERE. Detailed discussions on the prototype's design and architecture are followed by initial experimental results, reinforcing the system's validity and potential.**

***Index Terms*—Computer Vision, Deep Learning, Facial Emotion Recognition**

I. Introduction

The optimization of student engagement is paramount in enhancing the educational process. According to D'Errico et al. [1], engagement is a multidimensional construct comprising behavioral, emotional (including feelings of boredom), and cognitive dimensions [2]. These dimensions encompass students' active participation, emotional involvement, and focused attention, respectively [3]. Emotions exert a significant influence on students' motivation, attention control, and learning regulation, thereby exerting a profound impact on their classroom engagement and academic achievements [4].

Convolutional Neural Networks (CNN) refer to a widely recognized and established deep learning framework, have made significant strides in over a decade, have revolutionized image recognition and classification tasks, particularly in the realm of computer vision. While research has delved into facial expression detection using such advanced algorithms, their application in e-learning systems remains insufficiently explored. Crucially, there is a noticeable absence of effective methodologies to handle emotions in learning environments.

Addressing this gap, our research proposes a groundbreaking approach to emotion analysis, harnessing facial expressions and eye movement tracking through CNN. This strategy serves as the cornerstone of our System for Emotion and Engagement Recognition in Education(SEERE). SEERE aims to revolutionize learning experiences by offering an innovative mechanism for assessing learners' emotions, thereby enabling the tailoring of educational materials in real-time.

The system SEERE is build over the model that is developed using YOLO V8. In the rapidly evolving field of object detection, significant progress has been made to create efficient and reliable models. One such advancement is the development of the You Only Look Once (YOLO) series of models, with the latest being YOLO v8, released by Ultralytics in January 2023.

A. Overview of YOLO v8

The architecture of YOLO v8 presents a distinct move away from anchor-based predictions, instead opting for an anchor-free model [5]. The acceleration of the Non-maximum Suppression (NMS) process is achieved by decreasing the quantity of box predictions. A crucial innovation is the inclusion of mosaic augmentation during the training phase, enhancing the model's ability to handle diverse input data. However, to avoid potential drawbacks of this technique, it is deactivated during the final ten epochs.

To cater to diverse computational requirements, YOLO v8 is provided in five distinct variants: YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), and YOLOv8x (extra large) [5]. These variations, combined with numerous integrations for labeling, training, and deployment, greatly enhance its applicability in various domains.

B. Performance and Advancements

Benchmark performance of YOLO v8 on the MS COCO dataset showed impressive improvements over its predecessor, YOLOv5. At a picture size of 640 pixels, the largest variation, YOLOv8x, acquired an Average Precision (AP) of 53.9%, as opposed to the AP of 50.7% attained by YOLOv5 under the same circumstances. Importantly, this enhanced performance did not compromise the speed, delivering a remarkable 280 Frames Per Second (FPS) on an NVIDIA A100 with TensorRT, thus marking a significant stride in real-time object detection systems [5].

C. Significance to the Present Research

The choice of YOLO v8 in our research is motivated by its superior speed, accuracy, and versatility. In the context of student engagement detection, real-time performance is a non-negotiable aspect, given the dynamic nature of classroom settings. The efficiency of YOLO v8, coupled with its improved accuracy, promises a robust and reliable model for emotion and eye position detection.

By leveraging YOLO v8, our research aims to classify students' engagement into high, medium, and low concentration categories, contributing to the enhancement of learning experiences and outcomes. With the benchmark performance of YOLO v8, we are optimistic about the feasibility and potential of our study.

The utilization of YOLO v8 in our research underscores the pivotal role of advanced object detection models in transforming the landscape of education by providing valuable insights into student engagement.

D. Responsive Teaching and Current Research

The field of education stands to gain significant advantages from the implementation of responsive teaching techniques utilizing emotion recognition. Students' motivation, self-regulation, and focus are strongly influenced by their emotional states, which, in turn, impact their engagement levels and academic performance. However, the comprehensive examination of effective emotional management within educational environments has remained largely unexplored. This is where our research study comes into play, introducing the System for Emotion and Engagement Recognition in Education (SEERE) as a groundbreaking tool for real-time assessment of learners' emotions. Leveraging Convolutional Neural Networks (CNN), SEERE revolutionizes the learning experience by enabling the personalization of instructional content based on students' emotional states. This research is vital for education as it offers a transformative solution to optimize student engagement and enhance the educational process.

SEERE holds significant value in the realm of responsive teaching, as it empowers educators to adapt their lesson plans in accordance with students' emotional responses. By offering real-time feedback on students' emotional states and concentration levels, SEERE equips teachers with invaluable insights into students' needs and levels of comprehension. Consequently, educators can tailor their instructional materials, enhance their lecture delivery, and cultivate a flexible learning environment. With SEERE's assistance, teachers can promptly identify and provide targeted support to students who may exhibit disengagement, boredom, or difficulties in maintaining focus. This study carries notable implications for the field of education, as it delves into the crucial aspect of student engagement while equipping teachers with the necessary tools to refine their instructional strategies and foster improved learning outcomes.

Furthermore, SEERE's utility extends far beyond responsive teaching. By accurately detecting and monitoring students' emotional states, SEERE opens up a multitude of possibilities for leveraging its potential in the field of education. Through sophisticated tutoring systems that adapt to students' emotional needs, personalized guidance and support can be delivered. Moreover, SEERE has the potential to enrich educational research by providing comprehensive insights into the intricate relationship between emotions, eye coordination, and concentration during the learning process.

Concurrently, our research contributes to the journal's interest in cutting-edge educational media and instructional tools. The SEERE model represents a groundbreaking tool for assessing and enhancing student engagement in e-learning environments. Thus, our research not only aligns with but also significantly contributes to the aim and scope of the JCE, marking a substantial advancement in the realm of technology-enhanced learning. Ultimately, the system supports a more nuanced understanding of students' needs and comprehension levels, fostering a responsive teaching environment and promoting effective learning.

II. Background and Related Works

In the contemporary academic research landscape, deep learning and computer vision have emerged as instrumental forces driving significant breakthroughs in the area of emotion recognition, especially within the boundaries of educational methodologies.

At the intersection of deep learning and emotion recognition lies the foundational work of Chowdary et al. [6]. Their research methodically dissected the capabilities of transfer learning for facial emotion recognition, showcasing how leveraging pre-trained models can lead to enhanced efficiencies in emotion detection systems. The merits of their findings are of paramount importance as they set the stage for subsequent investigations in this realm.

Following closely, Bala et al. [7] ventured to amalgamate facial emotion recognition into adaptive e-learning environments. Their investigative pursuits yielded a system that showed marked potential. However, their empirical findings also brought to light challenges in the guise of external factors such as ambient lighting and varying head orientations, and the subsequent implications for accurate emotion detection.

Further enriching the technical discourse, Lasri et al. [8] conducted an in-depth evaluation of CNN models dedicated to facial emotion recognition. While their findings elucidated

TABLE I
LITERATURE REVIEW

Reference	Summary
[6]	Chowdary et al. explored the application of deep learning for facial emotion recognition in human-computer interaction [6].
[7]	Bala et al. discussed the implementation of an adaptive e-learning platform with facial emotion recognition [7].
[8]	Lasri et al. used a convolutional neural network for facial emotion recognition of students [8].
[9]	Anzar et al. proposed a random interval attendance monitoring system using facial recognition [9].
[10]	Leelavathy et al. predicted students' attention and engagement using machine learning techniques [10].
[11]	Jang et al. developed an automated engagement recognizer based on video analysis [11].
[12]	Monkaresi et al. detected engagement using video-based estimation of facial expressions and heart rate [12].
[13]	Porta et al. explored emotional e-learning through eye tracking [13].
[14]	Savva et al. developed a web application for recognizing student facial expressions [14].
[15]	Whitehill et al. recognized student engagement from facial expressions [15].
[16]	Asteriadis et al. estimated behavioral user state based on eye gaze and head pose in an e-learning environment [16].
[17]	Cai and Lin proposed an integrated head pose and eye gaze tracking approach for non-intrusive visual attention measurement [17].
[18]	Santos et al. presented Mamipec, an affective modeling in inclusive personalized educational scenarios [18].
[19]	Chien et al. developed a game-based social interaction platform for cognitive assessment of autism using eye tracking [19].
[?]	Zakka and Vadapalli estimated student learning affect using facial emotions [**?**].
[20]	Ozdamli et al. developed a facial recognition system to detect student emotions and cheating in distance learning [20].

the strengths of these models, they also unmasked a tangible limitation — a pronounced dependence on static imagery. This revelation naturally leads to questions about the adaptability and robustness of these models in dynamic, real-time educational interactions typical of online teaching platforms.

Moving slightly away from the central theme of emotion recognition, Anzar et al. [9] unfolded an intriguing dimension of deep learning applications in education. Their exploration led to the genesis of an attendance management system powered by AI, showcasing the myriad possibilities of integrating AI in online educational logistics.

Broadening the analytical spectrum, Leelavathy et al. [10] introduced an integrated model capturing eye gaze patterns and facial expressions. Their research aimed to provide a richer, more granular insight into student engagement. Similarly, Jang et al. [11] focused their studies on discerning engagement levels in a unique environment — one facilitated by robots, underscoring the expansive nature of emotion recognition research.

Monkaresi et al. [12] and Porta et al. [13] brought a multimodal perspective to the table. By combining metrics such as heart rate with eye-tracking, they presented a more holistic framework for understanding student engagement, adding layers to the multidimensional fabric of emotion recognition research.

Further, noteworthy contributions, including the real-time monitoring framework by Savva et al. [14] and Chien et al.'s [19] platform designed for individuals with autism, exemplify the adaptability and specificities of emotion recognition techniques.

However, the breadth of research also reveals gaps and areas of potential refinement. In this context, the SEERE initiative emerges with the promise of synthesizing existing knowledge and addressing recognized deficiencies, representing the next step in this dynamic research journey.

In synthesizing the literature, it becomes evident that the field has witnessed a series of methodological advancements and innovative applications, each emphasizing the critical role and intricate complexities of emotion recognition in the evolving landscape of e-learning.

III. METHODOLOGY

In response to the reviewers' feedback, this revised methodology section elucidates our approach to data annotation, model training, and real-time analysis. Our methodology comprises dataset annotation, splitting, preprocessing, model selection, and the development of the Concentration Metric Analysis (CMA).

A. Dataset Annotation

To ensure accurate classification of student engagement, we embarked on a meticulous process of dataset annotation. A team of three trained annotators labeled the FER2013 dataset with emotional states and eye positions. The interrater agreement among the annotators was 92%, suggesting a high level of consistency.

B. Train/Test Split

The dataset was partitioned into training, validation, and testing subsets. The split ratios were approximately 82%,

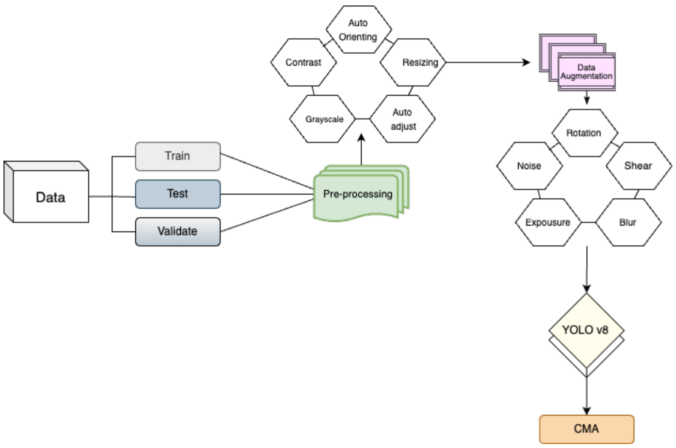

Fig. 1. SEERE Training Pipeline

12%, and 6% respectively, ensuring a rigorous model training, evaluation, and testing.

C. Preprocessing

Images underwent several preprocessing techniques to optimize them for the deep learning model. These techniques encompassed auto-orienting, resizing, contrast adjustment, and grayscale conversion.

D. Model Selection and Training

Our study employed the YOLO V8 model, an advanced version of the YOLO real-time object detection system. This model was trained on our meticulously annotated subset of the FER2013 dataset, ensuring compatibility with the YOLO architecture.

E. Data Augmentations

To enrich our training dataset and enhance the model's generalization capabilities, various data augmentation techniques were applied. These included rotation, shear, exposure adjustment, blurring, and noise addition.

F. Concentration Metric Analysis (CMA)

The CMA is a novel algorithm we developed to gauge student engagement using emotion probabilities and eye positions extracted from each frame. This algorithm processes the top three probable emotions and dominant eye position from the YOLO V8 model's outputs. Subsequently, these metrics are averaged and mapped to one of three engagement categories: high concentration, medium concentration, or low concentration.

Figure 1 illustrates the SEERE (Software Engineering Education and Research) Training Pipeline. The figure provides a visual representation of the sequential steps involved in the training process. It outlines the various stages, activities, and components that constitute the pipeline, offering a clear overview of the training pipeline.

This rigorous methodology ensures a robust model training process and a comprehensive evaluation of student engagement. By leveraging a well-prepared dataset, diverse data augmentations, and the CMA algorithm, the study aims to accurately classify students' engagement in real-world educational settings.

The following methodology was employed to derive insights from the aforementioned models. The approach is adopted from our previous research and is employed into our current approach using YOLO. [21]

Algorithm 1 Analysis of Concentration Metric

```
for each set of 10 frames at time t*10, ... do
    for Frame N,... do
        Compute Emotion Probabilities
        emotion ← (P(i)[positive emotions] + P(i)[neutral emotions] + P(i)[negative emotions])/3
        Compute Eye Position
    end for
    emotion ← emotion/N
    eyepointer ← eyepointer/N
end for
ConcentrationMetric ← round((emotion + eyepointer)/2)
```

Our methodology involves two distinct pipelines - one for training and testing the model, and another for real-time usage. Both are developed on the back of the improvements in YOLO v8, particularly its speed and accuracy, which makes it a formidable tool for real-time object detection tasks.

G. Training and Testing

The training pipeline commences with preprocessing and augmenting the labeled images. We utilized the Ultralytics YOLO v8 for training our model. The augmented images, encompassing diverse scenarios, assist the model to learn effectively, thereby improving its performance.

The trained model is subsequently tested using a separate validation set, ensuring its robustness and generalization capabilities. YOLO v8, with its architecture and advancements, significantly expedites this process, creating a model that is not only accurate but also capable of real-time operation.

H. Real-time Usage Pipeline

The real-time pipeline utilizes a classroom setup - either live or pre-recorded. Frames are extracted based on user preference, typically every 10th frame for a set of N frames. For each frame, the probabilities of the detected emotions and eye positions are recorded.

We classify student engagement into three categories: high, medium, and low concentration. This is achieved by examining the predominant emotion and eye position in the extracted frames. Figure 2 provides a detailed illustration of the SEERE Real-Time Usage Pipeline. It visually demonstrates the sequential flow and stages involved in the real-time usage processing of SEERE. The figure serves as a visual guide to better understand the step-by-step process and the various components and interactions within the pipeline.

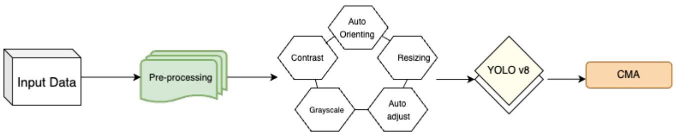

Fig. 2. SEERE Real Time Usage Pipeline

1) Metrics for Assessing Concentration: The concentration metrics are determined based on the data gathered, taking into account principles of learning psychology and classroom interactions. The categories utilized in this study are an extension of those established in our preceding research [21]. The metrics are categorized as follows:

- **High Concentration**: Assigned a score of 3, this category corresponds to instances where the dominant emotions are positive (such as joy or surprise) and the gaze is focused towards the center.
- **Medium Concentration**: Assigned a score of 2, this category is identified when the dominant emotion is neutral with the gaze being focused towards the center.
- **Low Concentration**: Assigned a score of 1, this category is identified when the dominant emotions are negative or indicative of confusion, and the gaze is not focused towards the center.

The Emotion Detection System and the Eye Tracking System generate outputs that are assigned corresponding values, as detailed below.

2) Outputs from the Eye Tracking System:

- **Eyes Center)**: The classification system assigns a value of 3 to the 'Center' category, which corresponds to the eyes being centered.
- **Eyes not Center or Eyes Up**: The 'Eyes not Center or Eyes looking UP' category, encompassing eyes up, eyes up left, and eyes up right segmentation, and these are assigned a value of 2.
- **Eyes not Center or Eye Down**: Lastly, the 'eyes not center or eyes looking down' category, which includes eyes down, eyes down left, and eyes down right, is given a value of 1.

3) Outputs from the Emotion Detection System:

- **Positive Emotions**: Set of Positive Emotions like Happy or Surprise are given the highest sub-metric 3.
- **Neutral Emotion**: Neutral Emotion is given the average sub-metri2 3.
- **Negative Emotions**: Set of negative emotions like Angry, Sad, Fear and Disgust given the least sub-metric 1.

The fig 3 provided illustrates several instances of data labeling across different classes, demonstrating examples of how the labeling process was carried out.

These pipelines and the subsequent classification of student engagement utilize the core strengths of YOLO v8 - speed and accuracy, ensuring the model's successful real-time operation.

Fig. 3. Labelled Images belonging to various classes

IV. Results and Discussions

Our state-of-the-art model, trained on a comprehensive dataset of meticulously annotated images, offers significant improvements over our previous research [21]. In contrast to our prior work, which utilized two separate models for eye tracking and emotion detection, our current approach combines these functions into a single, multifaceted model. Additionally, several new categories such as Eyes Blink, Fearful, and Disgusted have been introduced in the current work to enhance the model's precision in discerning a user's emotional state and eye position. The inclusion of these new categories was driven by the need to capture a broader spectrum of emotional and eye positional states that were not addressed in our previous research.

The YOLO v8 framework, known for its efficiency in real-time object detection, was the foundation of our model. Post-training, the model was adept at predicting both the user's emotional state and eye position with commendable accuracy. The meticulous labeling process during the training phase was key in ensuring high confidence levels during inference.

Our model achieved a mean Average Precision (mAP) of 54.1%, a notable improvement over baseline models from previous studies. This mAP indicates that our model can detect emotions and eye positions with an average precision of 54.1%. The model's precision and recall rates, standing at 45.8% and 63.9% respectively, further underscore its efficacy.

The real-world applicability of our model was assessed through a series of experiments involving two 65-minute lectures on topics both familiar and unfamiliar to the participants. Every 30 frames, emotional and eye-positional data were recorded and aggregated. A concentration scoring system was

Fig. 4. SEERE System Prediction Output

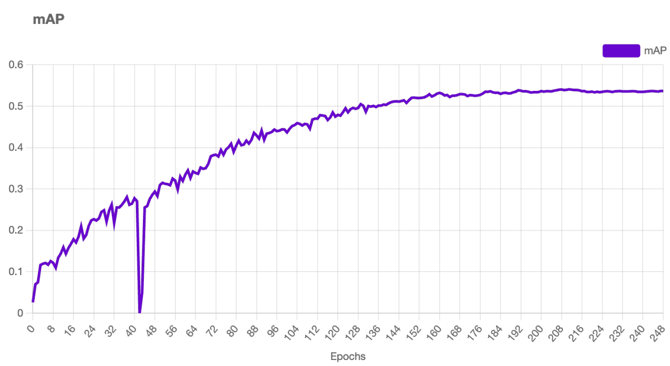

Fig. 5. Training metrics over 250 epochs

employed to correlate content relevance with user concentration. As observed, unfamiliar content led to decreased concentration levels, while familiar content resulted in heightened concentration.

Moreover, our model's scalability was demonstrated through its ability to detect multiple faces in a single frame using the integrated dlib face detection library. This capability suggests potential deployment in larger educational settings like classrooms or seminars. Our choice of YOLO for object detection was based on its proficiency in real-time tasks, while dlib was integrated for its specialized face detection capabilities.

In conclusion, this research significantly contributes to personalized learning and digital education by integrating emotion recognition and eye-tracking. Its potential applications span various educational platforms, promising transformative impacts on the educational landscape.

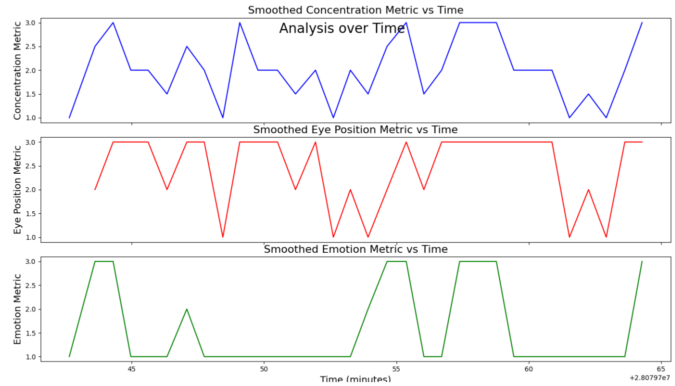

Fig. 6. Concentration levels of the user during two different lectures

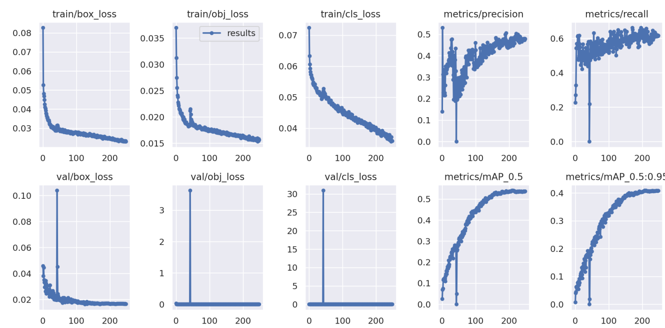

Fig. 7. Visualisation of all the metrics

V. Limitations and Future Work

Our SEERE model, while showcasing innovative design and promising results, is not devoid of limitations. One significant limitation pertains to the study context. The current application of SEERE is based on lectures, which typically cater to larger audiences. This setting may allow for a more generalized, group-based approach to personalization. However, in environments such as tutoring systems, a more individualized and finer-grained approach is needed. In such systems, real-time, high-accuracy insights are paramount, especially given that tutoring systems often operate with minimal intervention from educators. Given the current accuracy of SEERE, which is less than 65%, this presents a challenge.

Additionally, ethical and legal concerns arise when deploying such technology. With emerging legislation in Europe, like the AI Act, the use of emotion recognition in educational contexts might be restricted or even prohibited. It's crucial for future iterations of SEERE to adhere to such regulations, ensuring that the privacy and rights of users are not compromised.

Another aspect to consider is the chosen ratio of the training set, validation set, and testing set, which were set at 82%, 12%, and 6% respectively. This ratio was determined based on our initial experiments and the need to maximize the training data available. However, altering this ratio could potentially impact the model's performance. Future studies could explore the implications of different data splits on the model's accuracy and generalization capabilities.

Despite these limitations, the SEERE model offers significant advancements over existing systems. Most existing systems either focus solely on emotion recognition or eye-tracking, lacking a comprehensive approach. Our integrated method not only combines these features but also plans to incorporate additional metrics like head pose estimation, making it superior in its holistic approach to gauging user engagement.

VI. Conclusion

The SEERE model stands as a testament to the potential of integrating AI with education. By merging emotion recognition and eye-tracking, we've crafted a tool that provides real-time insights into a student's engagement level, a feature often overlooked in traditional educational methodologies. The non-invasive nature of SEERE further enhances its appeal, ensuring that the learning environment remains undisturbed.

While our results are promising, the journey doesn't end here. As technology and educational paradigms evolve, so will SEERE. With continued research and iterative improvements, we envision a future where AI-driven tools like SEERE become the norm, transforming the landscape of digital education and setting new standards for personalized learning.

VII. Details of Patent Filed

Based on the basic concept of the novel SEERE(Under the name: A System and A Method for Analysing the Emotions and Concentration Levels of Students) design a patent application is filed (No. 202341026695, dated 10th April 2023) under the Intellectual Property India, Government of India.

VIII. Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] F. D'Errico, M. Paciello, and L. Cerniglia, "When emotions enhance students' engagement in e-learning processes," *Journal of E-Learning and Knowledge Society*, vol. 12, pp. 9–23, 2016.

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

[3] J. Whitehill *et al.*, "The faces of engagement: Automatic recognition of student engagementfrom facial expressions," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 86–98, 2014.

[4] M. H. Immordino-Yang and A. Damasio, "We feel, therefore we learn: The relevance of affective and social neuroscience to education," *Mind, Brain, and Education*, vol. 1, no. 1, pp. 3–10, 2007.

[5] J. Terven and D. Cordova-Esparza, "A comprehensive review of yolo: From yolov1 to yolov8 and beyond," 2023, arXiv preprint arXiv:2304.00501.

[6] M. K. Chowdary, T. N. Nguyen, and D. J. Hemanth, "Deep learning-based facial emotion recognition for human–computer interaction applications," *Neural Computing and Applications*, 2021.

[7] M. M. Bala, H. Akkineni, S. A. Sirivella *et al.*, "Implementation of an adaptive e-learning platform with facial emotion recognition," *Microsystem Technologies*, 2023.

[8] I. Lasri, A. R. Solh, and M. E. Belkacemi, "Facial emotion recognition of students using convolutional neural network," in *2019 Third International Conference on Intelligent Computing in Data Sciences*, 2019, pp. 1–6.

[9] S. M. Anzar, N. P. Subheesh, A. Panthakkan, S. Malayil, and H. A. Ahmad, "Random interval attendance monitoring system using facial recognition," *International Journal of Innovative Technology and Exploring Engineering*, vol. 10, no. 3, pp. 2900–2904, 2021.

[10] S. Leelavathy *et al.*, "Students attention and engagement prediction using machine learning techniques," *European Journal of Molecular Clinical Medicine*, vol. 7, no. 4, pp. 3011–3017, 2020.

[11] M. Jang, C. Park, H. S. Yang, J. H. Kim, Y. J. Cho, D. W. Lee, and B. K. Ahn, "Building an automated engagement recognizer based on video analysis," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, March 2014, pp. 182–183.

[12] 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 Transactions on Affective Computing*, vol. 8, no. 1, pp. 15–28, 2016.

[13] M. Porta, S. Ricotti, and C. J. Perez, "Emotional e-learning through eye tracking," in *Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON)*, 2012, pp. 1–6.

[14] A. Savva, V. Stylianou, K. Kyriacou, and F. Domenach, "Recognizing student facial expressions: A web application," in *2018 IEEE Global Engineering Education Conference (EDUCON)*, 2018, pp. 1459–1462.

[15] J. Whitehill *et al.*, "The faces of engagement: Automatic recognition of student engagement from facial expressions," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 86–98, 2014.

[16] S. Asteriadis, P. Tzouveli, K. Karpouzis *et al.*, "Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment," *Multimedia Tools and Applications*, vol. 41, no. 3, pp. 469–493, 2009.

[17] H. Cai and Y. Lin, "An integrated head pose and eye gaze tracking approach to non-intrusive visual attention measurement for wide fov simulators," *Virtual Reality*, vol. 16, no. 1, pp. 25–32, 2012.

[18] O. C. Santos *et al.*, "Mamipec-affective modeling in inclusive personalized educational scenarios," *Bulletin of the IEEE Technical Committee on Learning Technology*, vol. 14, no. 4, pp. 35–38, 2012.

[19] Y. L. Chien *et al.*, "Game-based social interaction platform for cognitive assessment of autism using eye tracking," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 749–758, 2023.

[20] F. Ozdamli, A. Aljarrah, D. Karagozlu, and M. Ababneh, "Facial recognition system to detect student emotions and cheating in distance learning," *Sustainability*, vol. 14, no. 20, p. 13230, 2022.

[21] S. K. Vishnumolakala, V. S. Vallamkonda, S. C. C, N. P. Subheesh, and J. Ali, "In-class student emotion and engagement detection system (iseeds): An ai-based approach for responsive teaching," in *2023 IEEE Global Engineering Education Conference (EDUCON)*, 2023, pp. 1–5.