International Journal of Multidisciplinary and Current Educational Research (IJMCER)
ISSN: 2581-7027 ||Volume|| 7 ||Issue|| 2 ||Pages 84-96||2025||

Student Interaction Recognizer Model (Sirm) :A Deep Learning Framework for Real-Time Student Engagement Monitoring In The United Arab Emirates

[1,]Samer Rihawi , [2,]Roznim Binti Mohamad Rasli, [3,]Shamsul Arrieya Bin Ariffin, [4,]Samar Mouti
[1,2,3,]Faculty of Computing and Meta Technology, Universiti Pendidikan Sultan Idris, Malaysia
[4,] Faculty of Engineering and Computing, Liwa College, United Arab Emirates

ABSTRACT: This study was meant to assess the relationship between Student Interaction Recognizer Model and a Deep Learning Framework for real time student engagement monitoring in the United Arab Emirates (UAE).The researchers employed the experimental research design to conduct the research study. Thirteen students were chosen using convenient sampling to participate in the study and videos were captured for three classes for the purpose of conducting the study. Subjects that were involved in the experiment were members of Management of Information Systems, Web design and programming. The timings were between 3.00pm and 8.00pm. Two models that were used in the study experiment included: a Convolutional Neural Networks (CNN) model used for face recognition and Engagement detection model for predicting the engagement type of the students during the classes. Digital camera was provided to capture the interactions of the students in their classes in a video and then, devided them into frames and predicted the interactions of the students within each frame. Data collected was analyzed using Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models, which predicted the student's level of attention. The educator received a report on the student's performance in the class for evaluation. The findings of the study included: Teaching a classroom poses a significant challenge for the educators when there are distractions such as fatigue, boredom, and smart phones. The researchers concluded that the Students Interaction Recognizer Model (SIRM) can be used to avert classroom challenges since it is based on Artificial Intelligence (AI) and Deep Learning. This model aims to provide educators with a tool to better track and predict the engagement of the students during real-time classes. It records and extracts the face and body poses landmarks and then predicts the engagement of the students. The researchers recommended that educators need to enhance classroom learning through the use of current digital technologies. This may go a long way in controlling students while in classroom in order to provide a better learning environment for better achievement. Educators need to mind much about the dress code of the students and adequate lighting in classroom so as to avoid disruption of the learning environment.

KEYWORDS: Artificial Intelligence, Monitoring, Deep Learning Framework, UAE.

I. INTRODUCTION

The field of Education has undergone significant changes, from the introduction of computers in classrooms to the adoption of online learning management tools like Blackboard and Moodle, and the development of educational tools that utilize Artificial Intelligence such as Chat GPT. However, physical classrooms still exist as they offer a crucial opportunity for human interaction between students and educators, as well as among peers. Nevertheless, teaching remains a challenging task for various reasons. For instance, the timing of classes, fatigue, boredom, and the use of smart phones can easily distract students, making it challenging for educators to notice and address the situation (Capone & Lepore, 2022). In-class student experience varies depending on several factors, including the teaching style of the educator, the learning environment, and the class dynamics (Pendy, 2023). While some students find the classroom to be an engaging and stimulating place where they can interact and learn with their peers, others may have trouble with the material, find it challenging to concentrate, or have little interest in the subject matter. In multicultural classrooms, educators must navigate cultural and linguistic differences while ensuring that all students can learn and participate effectively. According to Tahiru (2020), the increasing use of technology and smart phones can further distract students, making it challenging for educators to maintain their focus on the material. The use of CNN and LSTM models for real-time engagement monitoring is innovative, combining AI and deep learning effectively to provide actionable insights into classroom dynamics. AI has been applied to a variety of educational domains, including the automatic grading and assessment of assignments and tests, which save lecturers time by automating repetitive administrative tasks. This is limited to multiple choice exams, though. Regarding essay-style questions, researchers are currently investigating methods for grading written assessments. It has also been applied to the creation of clever instructional material. Popular apps that use AI to improve the coherence and usability of textbook content include Cram101. Personalized tutoring systems for students have been created using intelligent tutoring systems (ITS), such as the "Mike" software from Carneige Learning (Tahiru, 2020).

Education Data Mining (EDM) (Khan & Ghosh, 2021). ED focuses on evaluating the student performance in classroom setting as they addressed that the temporal aspect of prediction in class-based education has not been thoroughly explored. They reviewed over 140 education predictors, where they focused on identifying these predictors, methods, and the timing and goals of prediction. They found out that EDM achieved significant prediction efficiency during the tenure of the course. However, the prediction of the performance of the commencement of ta course is still a challenging area that requires special attention. A study by (Kim et al., 2022) addressed the integration of AI in K-12 education from the point of view of leading teachers in South Korea as AI has been used to design curriculum and enabled the structing of learning environments to facilitate Student-AI Collaboration (SAC), where a study was conducted by interviewing 10 teachers while focusing on identifying optimal learning goals and facilitating interdisciplinary learning. It has been found that subject-matter knowledge building and interdisciplinary and creative tasks are very important for supportive and creative learning environments.

Generative AI tools such as Chat GPT have been widely used to create personalized learning paths as they offer real time feedback and foster an interactive learning environment. They can adapt to individual student needs and provide tailored education content assessments. Also they can engage students in dynamic discussions and simulate one-on-one tutoring scenarios, which can significantly enhance learning experiences and outcomes (Su & Yang, 2023). Those tools can also help to facilitate knowledge acquisition and support the students in writing tasks such as codes, essays, poems, and script. They can help the educators to identify gaps in students' learning and enable them to send timely feedback. However, they pose a challenge as it is hard to ascertain whether the projects presented by the students are truly novel or just copied from other websites or articles (Lim et al., 2023).

Another paper, titled Eye-tracking and AI to enhance motivation and learning aimed to understand and investigate the use of eye-tracking and AI to improve student motivation and learning in Massive Open Online Courses (MOOCs) (Sharma et al., 2020). This paper addresses how traditional analytics such as click-streams and keystrokes fail to capture the learning behaviors of the students, especially in passive activities like video watching. It shows that stimuli-based gaze variables, derived from eye-tracking data, can provide insight into students' content coverage, reading patterns, and levels of attention. Also, the mediation effects of reading pattens are associated with the students' motivation and their learning performance. This study combines eye-tracking data with AI techniques, where it highlights the use of gaze data and AI to personalize the feedback and improve the learning process. Students' engagement during the classes plays a vital role in achieving academic progress as it determines their overall performance during the courses and the exams. Checking on the understanding of the students is one of the main roles. Engagement in the classroom can be defined as a combination of behavioral, emotional, and cognitive aspects. It includes participation in the classroom activities, students' affective reactions in the classroom as well as the investment in learning and understanding (Disalvo et al., 2022).

II. METHODS AND MATERIALS

The implementation of the Student Interaction Recognizer Model SIRM discussed in the previous chapter was done by capturing videos of selected samples of students in different classes of different timings and different levels after obtaining their consent to participate in this research study. The difference in timing affected the experiments because of the lighting used in different classes and the timings were different between afternoon and evening. The researchers employed the experimental research design to conduct the research study. Thirteen students were chosen using convenient sampling to participate in the study and videos were captured for three classes for the purpose of conducting the study. Subjects that were involved in the experiment were Management of Information Systems, Web design and programming. The timings were between 3.00pm and 8.00pm. Two models used in the study experiment were: a CNN model used for face recognition and Engagement detection model for predicting the engagement type of the students during the classes. Digital camera was provided to capture the interactions of the students in their classes in a video and then, devide them into frames and predict the interactions of the students within each frame. Data collected was analyzed using Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models, which predict the student's level of attention. The educator receives a report on the student's performance in the class for evaluation.

III. RESULTS

The total number of students who participated in the experiments was 13 students and the videos were captured for over 3 classes.The experiments were composed of two parts: in the first part, the faces of individual students were captured to train the face recognition model using CNN. In the second part of the experiments, short videos were captured of the students while they were attending the classes after getting the approval of their lecturers. Those videos have been used to observe the engagement of the students during their classes by capturing the faces of the students and detecting their poses which have been used to predict the engagement level of the students to either active if the participant is paying full attention, bored if the eyes are partially closed and the face is down, distracted if the participant is looking somewhere else.

The model then has been trained and validated over different numbers of epochs: 5,10, and 15. The accuracy of the model has been measured by the steps below: The model makes predictions for the input data. It compares each prediction to the corresponding label. It counts the number of the correct predictions that match the corresponding label. It calculates the accuracy of the model by dividing the number of the correct predictions by the total number of the predictions.

The accuracy of SIRM can be measured using the following equation

$$A \quad cu \quad a \quad y = \underline{\qquad TP + TN}$$

$TP{+}TN{+}FP{+}FN$

Where:
TP: True Positives (Correctly predicted positive samples) TN: True Negatives (Correctly predicted negative samples)
FP: False Positives (Incorrectly predicted positive samples)
FN: False Negatives (Incorrectly predicted negative samples)

The results can be found in tables 1,2, and 3 for each experiment:

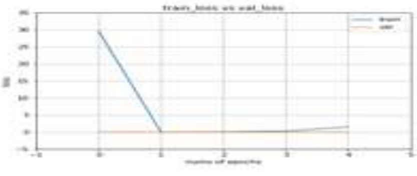
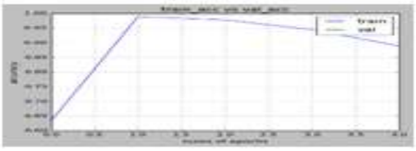
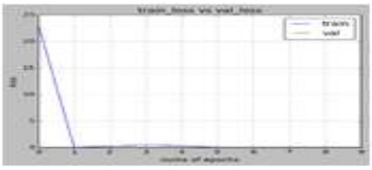
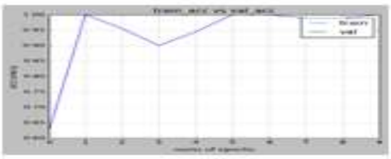
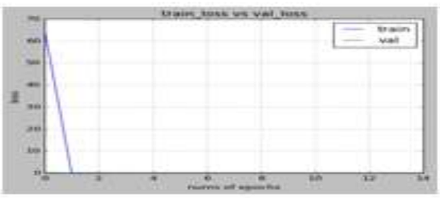
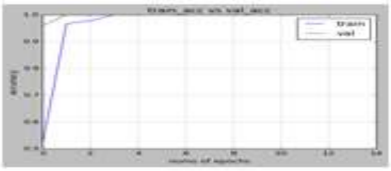
Epochs	Accuracy Rate	Loss	Accuracy
5	89%		
10	100%		
15	100%		

Table 1. Training the face recognition model for experiment 1

From table 1 above, we can find the following:

- At 5 epochs, the model achieved an accuracy of 89%, the loss value is 32%
- At 10 epochs, the model achieved a perfect accuracy of 100%, indicating that the model has likely converged and is now over fitting to the training data. The loss value is 6%.
- At 15 epochs, the model continues to achieve a perfect accuracy of 100%, the loss value is 6%.

As for the second experiment, the results of the face recognition training model can be found in table 2

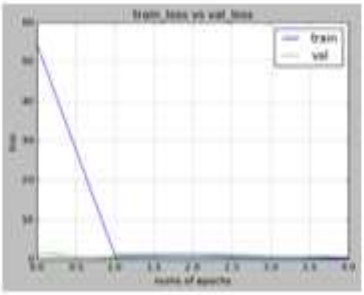
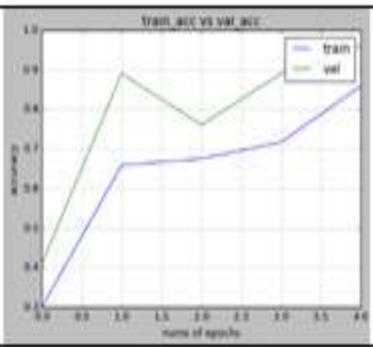
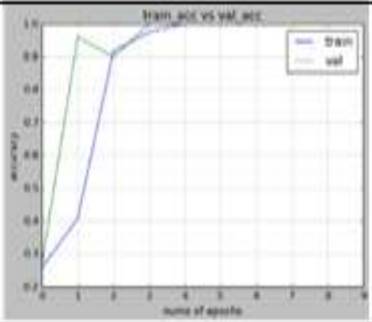
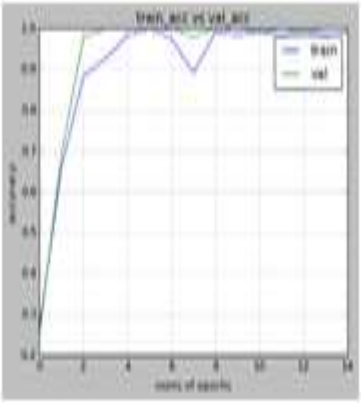
Epochs	Accuracy Rate	Loss	Accuracy
5	98%		
10	100%		
15	100%		

Table 2. Training the face recognition model for experiment 2

It seems that increasing the number of epochs from 5 to 10 or 15 led to an improvement in accuracy, with the model achieving 100% accuracy in both cases. Additionally, it may be helpful to monitor the loss over the epochs to see if the model is converging and to determine if early stopping is necessary to prevent over fitting.

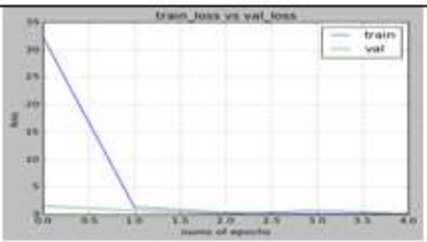
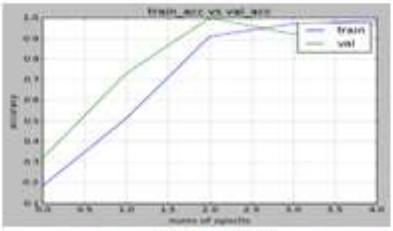
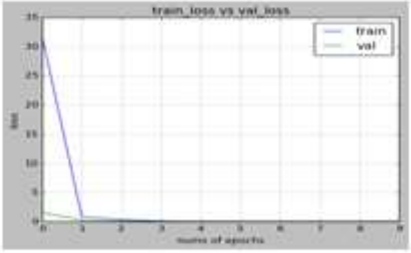
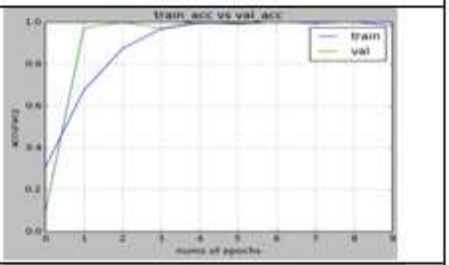
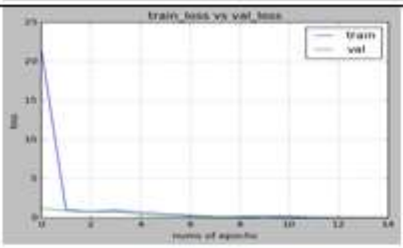
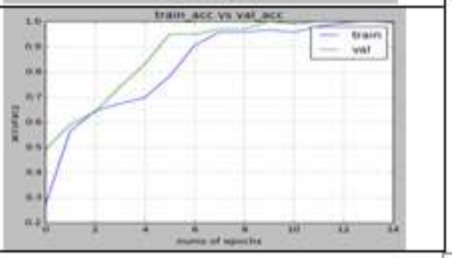
Epochs	Accuracy Rate	Loss	Accuracy
5	97%		
10	100%		
15	100%		

Table 3 shows the results of the training of the face recognition model for experiment 3.

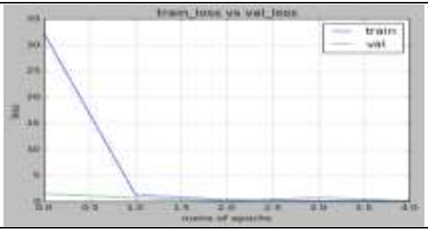
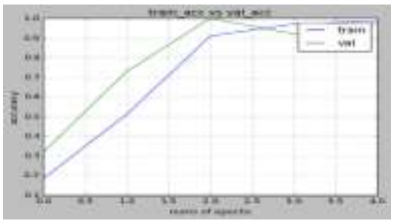
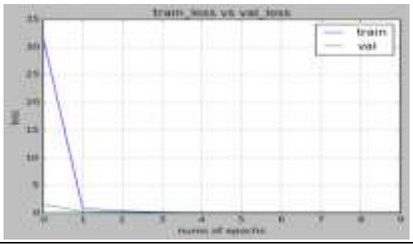
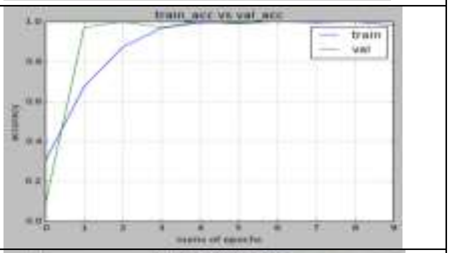
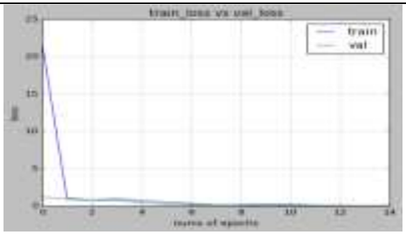
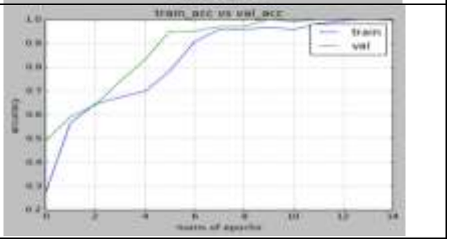
Epochs	Accuracy Rate	Loss	Accuracy
5	97%		
10	100%		
15	100%		

Table 3. Training the face recognition model for experiment 3

From table 5 above , it can be noticed that the model's accuracy improved as the number of epochs increased. At 5 epochs, the model achieved 97% accuracy, which increased to 100% at 10 and 15 epochs. The loss values were 4%, 6%, and 6% respectively. From the tables above, we can find that the increase in the number of epochs leads to an increase in the accuracy of the model. However, training the model for many times might cause the accuracy to reach 100% which may lead to the over fitting of the model as the same training data is trained many times. So the number of epochs will be 5 when training this model.

The accuracy of the Face Recognition model can be calculated by finding the average of the accuracies for each experiment as follows: average accuracy= (89%+97%+98%)/3 = 94.6%, therefore the error rate can be calculated as follows: error rate = 100- average accuracy = 100-
94.6% = 5.4%

The LSTM model has been trained by using the engagement types I figure 5. Each type shows different facial expressions and body poses. When the student is active, he maintains a straight body pose and looks at the camera. When the student is bored, he lowers his body and keeps a bored facial expression such as half-opened eyes. When the student is distracted, he is looking away from the camera.

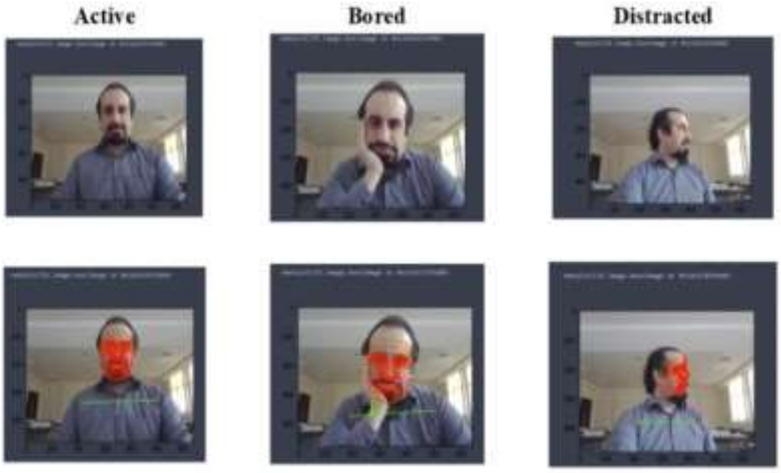


Figure 1: Engagement types of SIRM

The model has then been trained and validated over different numbers of epochs: 250, 300, 500, and 1000. The results are shown in table 4.

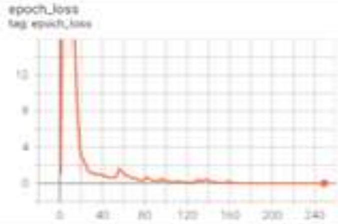
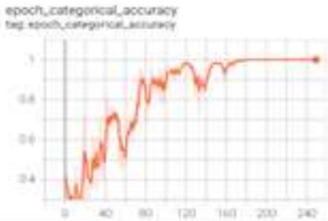
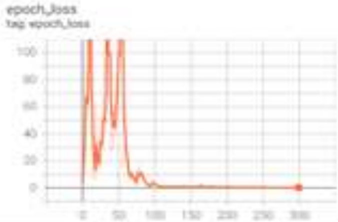
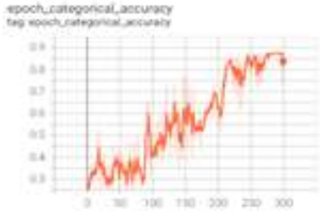
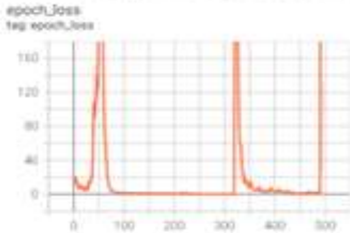
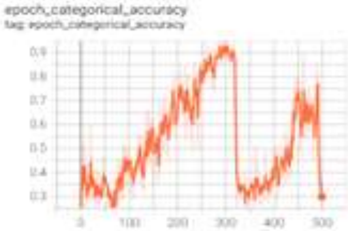
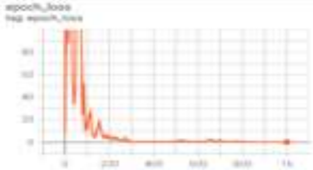
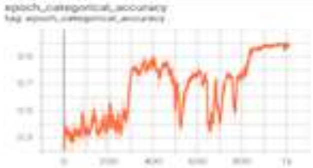
Epochs	Categorical Accuracy	Categorical Loss	Categorical Accuracy
250	100%		
300	87%		
500	30.5%		
1000	100%		

Table 6: Training the engagement prediction model

From this table 4 above , it can be noticed that the choice of epochs was done based on the number of times that the model needs training to achieve high accuracy. We can find out that the accuracy is high when the number of epochs is 100 and 1000, while it is best when the number of epochs is 300. After running those tests, we can conclude that the reason behind that is that when the number of epochs is low, it might lead to under fitting the model as it was not trained enough, while if the number of epochs is high, it might lead to over fitting the model as the model was trained too many times, therefore caused the over fitting of the model.The two models were tested on students in 3 classes as mentioned in the previous section. The experiments took place in the afternoon and evening to test the attention span of the students. First, the SIRM detects the faces of the students from the captured videos and then every 3 seconds.

Next, it recognizes the faces of the students and saves those faces into images in folder named with the current date and current time.Then it draws the facial and pose landmarks for each of those faces, extracts the key points, and predicts the engagement type using the LSTM model.
Below are the case studies for each of the classes during which the experiments were conducted.

Experiment 1

The first experiment was conducted during the afternoon period. A student was distracted by his laptop as the other student is looking attentively at the lecturer as shown in figure 2.

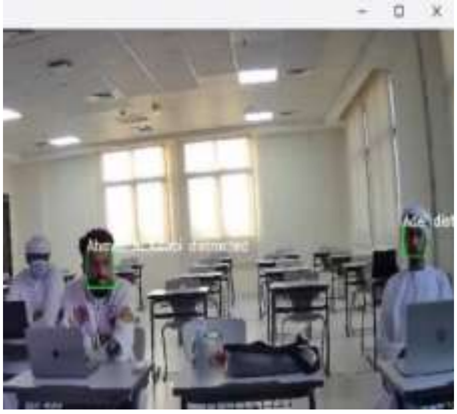


Figure 2: Results of SIRM in experiment 1

In this experiment, the identities of the students were recognized and then the interaction of each student was predicted and displayed. The model detected the faces of the students and predicted their engagement correctly. However, one of the students was wearing a face mask, which prevented the SIRM from detecting his face. Figure 3 shows the images of those students after being detected by SIRM and saved into separate files.

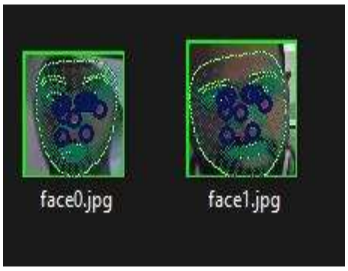


Figure 3: Facial Expression Detection (Experiment 1)

Figure 3 above shows the captured pictures of two students, as one of the participants was wearing a face mask, which prevented the SIRM from detecting his face.

Student Name	File Name	Engagement Type
Adel	face0.jpg	Distracted
Ahmed Al Naqbi	face1.jpg	Distracted

Table 5. The saved files by SIRM for experiment 1

The results in table 7 show that two faces have been detected with both the identities of the student recognized, but it shows that both are distracted as they were looking away from the camera.

Experiment 2 : The second experiment was conducted in the afternoon, around 5:15 pm as shown in the figure below. Consent was obtained for the students to participate in the experiment as shown in figure 3 and the camera was placed close to the students so that the model could recognize their identities. Figure 3 shows the results of the experiment

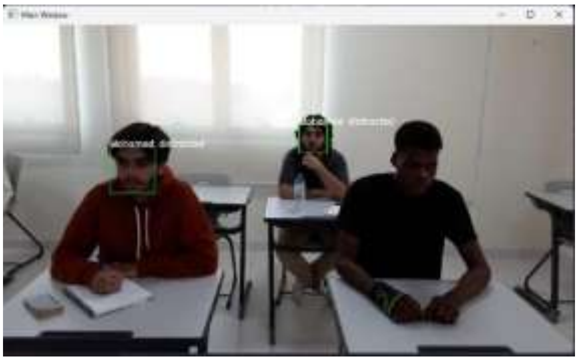


Figure 3: Facial Expression Detection (Experiment 2)

In this experiment, the faces of the students were detected and then their identities were recognized, and their engagement type was added. Figure 4 shows the faces of the students after being detected and saved by SIRM.

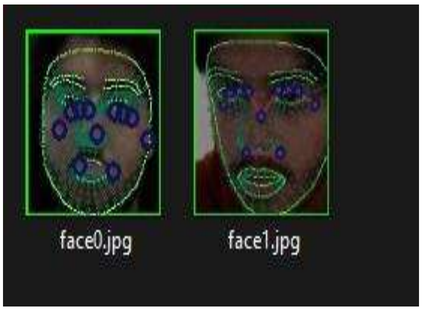


Figure 4: Facial Expression Detection (Experiment 2)

The pictures above show two brothers who were attending the same class, and the model was able to correctly recognize their identities. However, the third face wasn't detected because of the strong lighting

Table 6 shows the names of the students associated with the file names which were saved by SIRM.

Student Name	File Name	Engagement Type
Mohamed	face0.jpg	Distracted
Ahmed	face1.jpg	Distracted

Table 6: The files saved by SIRM for experiment 2

The results above show that two faces have been detected with both the identities of the student recognized but the third face was not detected due to the strong lighting of the room.

Experiment 3

The experiment was conducted in the evening time as the placement of the students was in one corner, so they were looking at the lecturer at the left of the camera as shown in figure 5.

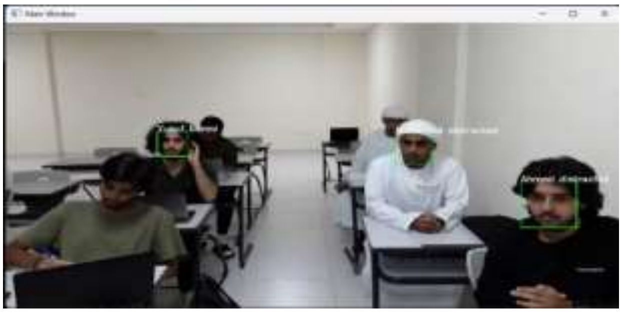


Figure 5: Facial Expression Detection (Experiment 3)

In the third experiment, 3 faces were detected and captured by the SIRM model. SIRM was able to capture the faces of the students, recognize their identities correctly, and then display the engagement type to each of the students.

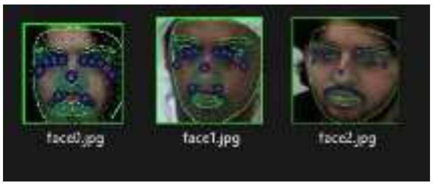


Figure 6: Facial Expression Detection (Experiment 3)

SIRM was able to capture the faces of three students as they were looking at the educator while other students were looking down, so their faces were not detected.

Table 7 shows the names of the students associated with the file names which were saved by SIRM.

Student Name	**File Name**	**Engagement Type**
Yusuf	face0.jpg	Bored
Mohammed	face1.jpg	Distracted
Ahmed	face2.jpg	Distracted

Table 7. The saved files by SIRM for experiment 3

The results above show that three faces were detected, and the identities of the students were recognized. However, some students were sitting far away from the camera, so their faces were not detected. As shown in the previous section, 3 experiments were conducted on the students during their classes where their faces have been detected using the SIRM, which recognized each one of those faces. Then the engagement detection model drew facial and pose landmarks for each of those faces and predicted the engagement type for each student with high accuracy.

However, a few challenges were faced during the experiments, as one of the students was wearing a face mask so his face wasn't detected by SIRM. In the second experiment, the class was conducted during the afternoon so there was strong lighting which caused the SIRM not to be able to detect the face of one of the students. In the third experiment, one of the students was looking down at his notebook, so his face wasn’t detected by SIRM.The videos were used to study and test the SIRM model and every 3 seconds an individual folder was created with the current date and time to enable a continuous application of the SIRM model. However, some issues occurred as the SIRM to detect the faces was unable to detect the faces of some students because of lighting and face masks which hid the faces of some participants. Table 8 gives a detailed comparison between SIRM and the other models which were discussed in this paper in terms of accuracy and performance.

Study/Model	**Methodology**	**Accuracy**	**Key Features**
SIRM (Proposed Model)	CNN and LSTM for facial and pose landmark detection	94.6% (Face Recog.) 87% (Engagement Detection)	Real-time student engagement monitoring and Analysis
ResNet-50 (Facial Emotion Recognition)(Gupta et al., 2023)	Deep learning model for engagement detection based on facial expressions	92.32%	Focused on facial emotion recognition for engagement
Inception-V3	Deep learning model for facial emotion recognition	89.11%	Lower performance compared to ResNet-50
VGG19	Deep learning model for facial emotion recognition	90.14%	Comparable but slightly lower accuracy than ResNet-50
CatBoost (Machine Learning for Engagement)	Ensemble learning for classroom engagement prediction	92.23%	Based on student activity data like clicks and forum interactions
ASSISTments	AI-based Intelligent Tutoring System (ITS) using predictive modeling	Not specified	Focused on knowledge tracing and skill prediction
Children’s Education Auxiliary System	CNN for recognizing speech/picture input and providing interactive feedback	Not specified	Designed for children's education with interactive features
Haar Classifier and Fisherface Algorithm	Facial expression and gesture recognition	Not specified	Used for driver drowsiness detection and other gesture tasks
Adaptive Learning Tools (Moodle and Blackboard) (Clark & Kaw, 2020)	Adaptive learning platforms for assessing engagement	Not specified	Widely used for adaptive learning and engagement tracking

Table 8. The files saved by SIRM for experiment 2

IV. DISCUSSION

SIRM performs better than the other models in terms of face recognition (94.6%), but the accuracy is slightly lower in engagement detection (87%) in comparison with other models like ResNet-50 (92.32%) and CatBoost (92.23%), while other models like Inception-V3 and VGG19 slightly fall behind in accuracy for engagement detection. SIRM provides real-time monitoring and analysis, which makes it better than the other models that focus only on post hoc analysis of data. It uniquely combines CNN and LSTM to analyze both facial and pose landmarks, offering a holistic view of student engagement. Other models like

CatBoost and Adaptive Learning Tools follow a broader engagement tracking using various types of input like clicks and fora interactions, but they lack precision in analyzing students' behavior at a granular level. ASSISTments and Children's Education Auxiliary System are customized for specific contexts like tutoring and children's education, making them less versatile than SIRM. While SIRM demonstrates promising accuracy, certain challenges need to be addressed. Lighting variations and face coverings, such as masks, can reduce detection accuracy, particularly in diverse classroom environments. Additionally, ethical considerations regarding the use of facial recognition technology, including privacy concerns and data security, must be prioritized. Scalability in larger classrooms and potential biases in facial recognition algorithms are also critical areas for further research.

V. CONCLUSION

In this paper, by leveraging CNN and LSTM, the SIRM model introduces an innovative framework for real-time engagement monitoring, combining the strengths of AI and deep learning to enhance classroom teaching. The SIRM model uses AI and Deep Learning, including CNN and LSTM to recognize the faces of the students during the classes and at the same time, predict their engagement type seamlessly without causing any disruption to the flow of the class. During the experiments, individual photos of each student were taken. The selected sample of the students includes students who covered their heads as part of the UAE traditional dress, students who had facial hair, and even two brothers who had similar facial features. The experiments were conducted on different levels of the students during different times of the day to study and analyze the change in their attention levels during the classes. SIRM was able to capture the faces of the students and recognize their identities with a high accuracy of 94.6% and error rate of 5.4% and their engagement with an accuracy of 87%. While effective, improvements in robustness to varying conditions (e.g., lighting, face masks) and scalability for larger classrooms are needed. Future work will focus on integrating SIRM with Learning Management Systems for real-time feedback and personalized interventions, enhancing its impact in diverse educational settings. The results of this research can contribute to identifying the strong and weak aspects of using predictive modeling in the learning process as it helps the teachers to study and understand the facial gesture expressions of the students, which will allow them to correctly evaluate how their students are learning, thus aiding them to achieve their learning objectives. By using deep learning in predicting the engagement of the students in the classroom. The teacher can gain some insight in their teaching methodologies during the classroom hours, so they don't have to be burdened with following up with the engagement of the students, thus they can focus on covering the course syllabus.

It can also help the schools/institutions to know more about the feedback of the students without having to conduct surveys, which generally return inaccurate and subjective responses from the students.

VI. RECOMMENDATIONS

The researchers recommended that educators need to enhance classroom learning through the use of current digital technologies. This may go a long way in controlling students while in classroom in order to provide a better learning environment for better achievement. Educators need to mind much about the dress code of the students and adequate lighting in classroom so as to avoid disruption of the learning environment

REFERENCES

1. Capone, R., & Lepore, M. (2022). From Distance Learning to Integrated Digital Learning: A Fuzzy Cognitive Analysis Focused on Engagement, Motivation, and Participation During
2. COVID-19 Pandemic. *Technology, Knowledge and Learning*, *27*(4). https://doi.org/10.1007/s10758-021-09571-w
3. Disalvo, B., Bandaru, D., Wang, Q., Li, H., & Plötz, T. (2022). Reading the Room - Automated, Momentary Assessment of Student Engagement in the Classroom: AreWe There Yet?
4. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*,
5. *6*(3). https://doi.org/10.1145/3550328
6. Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and Information*
7. *Technologies*, *26*(1). https://doi.org/10.1007/s10639-020-10230-3
8. Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: perspectives of leading teachers for AI in education. *Education and Information Technologies*, *27*(5). https://doi.org/10.1007/s10639-021-10831-6
9. Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative

10. AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *International Journal of Management Education*, *21*(2).
11. https://doi.org/10.1016/j.ijme.2023.100790
12. Pendy, B. (2023). Artificial Intelligence: The Future of Education. *Jurnal Indonesia Sosial*
13. *Sains*, *2*(11). https://doi.org/10.59141/jiss.v2i11.801
14. Sharma, K., Giannakos, M., & Dillenbourg, P. (2020). Eye-tracking and artificial intelligence to enhance motivation and learning. *Smart Learning Environments*, *7*(1). https://doi.org/10.1186/s40561-020-00122-x
15. Su, J., & Yang, W. (2023). Unlocking the Power of ChatGPT: A Framework for Applying
16. Generative AI in Education. *ECNU Review of Education*, *6*(3), 355–366. https://doi.org/10.1177/20965311231168423
17. Tahiru, F. (2020). AI in Education. *Journal of Cases on Information Technology*, *23*(1), 1–https://doi.org/10.4018/jcit.2021010101