

AI-Based Student Emotion and Engagement Level Detection Framework

Chinar Deshpande ¹[0009-0000-9909-653X], a

¹Aditya English Medium School, Baner, Pune, Maharashtra 411045, India

^adeshchinar@gmail.com

Abstract. Class size and student-to-teacher ratio are the two most important factors determining the quality of teaching and learning in a classroom setting. In southern Asian countries, especially India, the class batch sizes are substantially large, leading to a very high student-to-teacher ratio of approximately 60:1. While the government plans to improve the availability of teachers via various policy measures, technological support to the existing teaching community is much needed in helping them raise the standards of education in India. This project proposes an emotion detection algorithm, which can be employed in a typical Indian classroom with a high student-to-teacher ratio. Currently, the Convolutional Neural Network designed as a part of the algorithm stands at 86% accuracy. The model successfully detects 7 primary emotions—happiness, sadness, disgust, surprise, anger, fear, and neutral. These are mapped to high, medium, and low engagement levels. The algorithm processes the real-time images of the students in a classroom using Facial Emotion Recognition (FER). It determines the emotions and then maps them to the appropriate engagement levels. The project has valuable implications for the teaching community. The teachers will be able to see students’ engagement reports class wise, or weekly/monthly, helping them identify student engagement trends, and employ appropriate interventions to improve student engagement and learning outcomes.

Keywords: learning, engagement, emotion, artificial intelligence, machine learning, education.

1 Introduction

Student engagement is the key to successful classroom learning. Measuring or analyzing the engagement of students is very important to improve learning and teaching [1]. In India, The Right of Children to Free and Compulsory Education (RTE) Act 2009 mandates a 30:1 pupil-to-teacher ratio (PTR). However, over 8% of Indian schools have one teacher per school, and often, students outnumber employed teachers greatly. When there are more students in a class, it is often difficult for teachers to assess the individual needs of the students, whether it’s academic or emotional. This project used machine learning and computer vision to help educators detect the emotional trajectories and engagement levels of students. This helped design appropriate and engaging curriculum materials as well as interventions.

Presently, multiple challenges are faced within the classroom. A prominent one is related to the identification of individual needs due to a high PTR. Contextualizing this in a post-COVID world where students returned to classrooms after a long time and felt a gap in their social growth, it becomes paramount for teachers to implement emotional regulation interventions. Historically, various measures have been used to identify if students are actively engaged in learning. These measures have predominantly focused on quantitative data such as attendance, standardized test scores, and truancy or graduation rates [2]. However, these fail to account for the subjectivity of a student. As students spend the majority of their time within the classroom, it is integral to understand their engagement in correlation to their emotional status. [3]

The project aims at detecting the engagement levels of students by identifying their emotions. The emotions are detected with Facial Emotion Recognition (FER) using Computer Vision. After the automatic identification of emotions, the algorithm presents a report to the teacher highlighting the engagement levels of a student. It also generates meaningful interventions and strategies based on different learning styles which can be immediately applied within the classroom. Working as an aid for teachers working in Indian classrooms in India, this project aims to make teachers' workload easier and more efficient by letting them focus on effective teaching.

2 Methodology

2.1 Innovation

Facial Emotion Recognition (FER) using Computer Vision is a vast area where much research has been done. However, rarely has this technology been used in a classroom setting to determine and improve the engagement level of the students. Similar to FER, there is research done to measure engagement levels manually. However, as an estimation, this project proposed a novel way to calculate the engagement levels automatically based on FER and plot a graph presenting the engagement level trends among students to the teachers in a classroom setting.

The scope of the study is limited to identifying the seven basic emotions. The FER program determined the following 7 emotions from an input image - happiness, sadness, disgust, surprise, anger, fear, and neutral emotion, as these are primary emotions that are easily identifiable. This project represented the combination of FER and engagement levels by melding them together to synthesize new Computer Vision applications and plot a graph to show the engagement level trend among the students to their respective teachers.

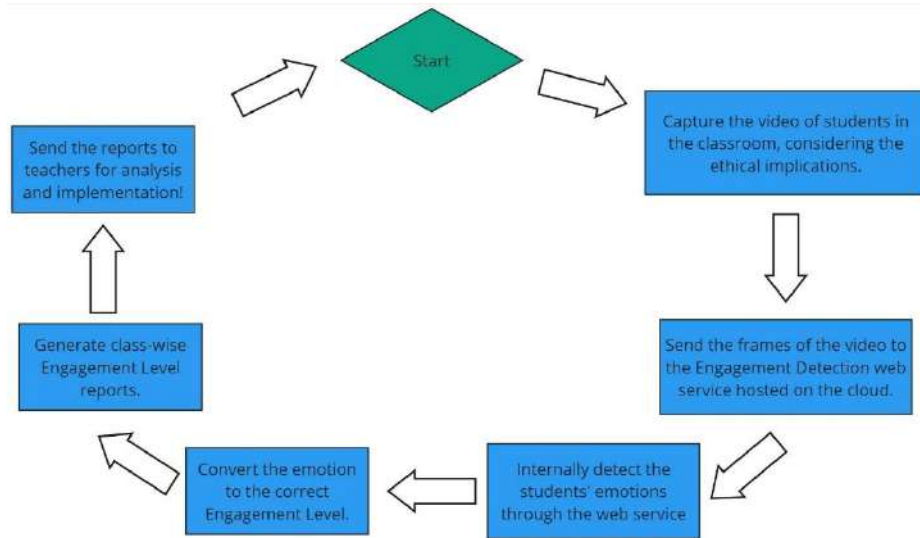


Fig. 1. Complete process of the project

The classroom student images were sent as input to the Engagement Detection web service, which was hosted on the cloud. As shown in Fig 1, the input to the web service is the classroom image of the student(s), and the output is the engagement level. Here is the way to access this web service using the curl command:

```
curl-XPOST-Ffile=@yourimagepath.jpg
http://emotion.excellonconnect.com/predict
```

The project used the engagement model proposed by Altuwairqi et al. [3] provided in Fig 2 below to determine engagement levels.

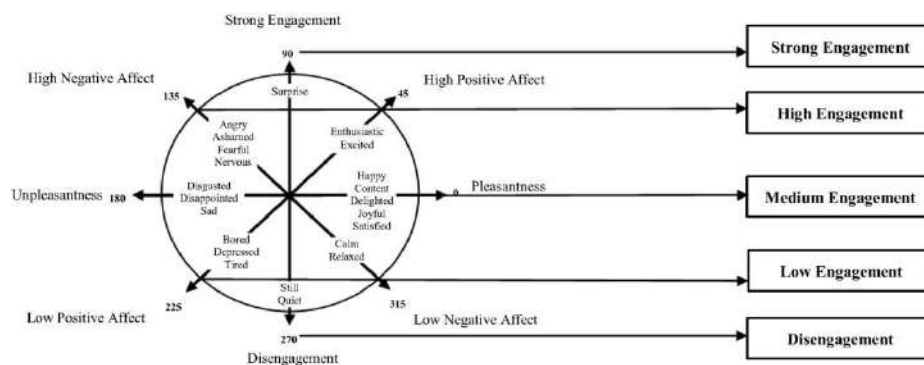


Fig. 2. Different Emotion Correlations to Engagement Levels

Fig. 2 shows the correlation between primary emotions and engagement levels. However, since we are detecting a subset of these emotions through our Engagement Detection model, the table below displays the engagement levels for the emotions that are initially detected:

Table 1. Emotion-Engagement Index

Emotion	Engagement Level	Engagement Level - numerical value
Surprise	Strong	5
Fearful, Excited, Anger	High	4
Happy, Sad, Disgust	Medium	3
Neutral	Low	2

Table 1 reflects the relationship between different emotions, their respective engagement levels and the numerical Engagement Level values. The numerical values of the Engagement Level were assigned based on a linear scale from 5 to 2, with 5 being the highest level of engagement and 2 being the lowest. The FER model currently does not measure the emotions associated with disengagement, hence the value of 1 is not in use.

Emotions with a high positive or negative effect reflect a high engagement or focus level on the class as the mind is stimulated. Based on this correlation, an emotion-to-engagement level mapping was created. Then, teachers accessed this data through class-wise reports. Based on the generated reports, the model also provided recommendations to increase engagement and/or facilitate the emotional regulation of students.

2.2 Method: Training and Optimizing the Model

The FER model used by the Engagement Level detection web service was trained using a variety of classification methods such as Random Forests, K Means Clustering, K Nearest Neighbours, and a Neural Network.

Each algorithm was tested to determine which one could provide the highest accuracy. By passing a training and testing file through each model, data points were recorded in a detailed Excel sheet. The Neural Network demonstrated 60% accuracy, followed closely by the KNN Algorithm at 54%, and the Random Forests, K-Means Clustering algorithms at 50%. Thus, a Neural Network approach was finalized as visual data would accurately be processed through a Convolutional Neural Network (CNN), as shown in Fig 3.

Multiple datasets were then passed through the model to train it thoroughly. An intensive literature review was conducted which included surveys that analyzed different datasets in parallel. The primary dataset used was the Affectiva-MIT Facial Expression Dataset (AMFED) [4].

Then, the model was tuned to receive the dataset images from AMFED as input, and after training was completed, the accuracy was optimal at 220 epochs—74%. To improve accuracy, the number of convolutional layers was increased, and more images were fed to train the model. However, during model testing, the images were being identified wrong as the faces of the participants were partially obscured by the electrodes attached to measure signals from the brain. Hence, when given an image of a regular person, the model would fail since it wasn't trained with those types of images before. To resolve the bugs, a new dataset was found—DISFA [5]. While editing the model, only the input had to be changed to suit the structure of the new dataset, so the algorithm remained the same. To process high-resolution images (1280 x 720), the convolutional layers were set to 4 which provided more power to the model.

With DISFA and 4 convolutional layers, the model reached an accuracy of 75%. However, the model still failed to recognize emotions accurately as all images were identified as “Fear” when random images of actual people were passed through. The model was then trained extensively through VGG16, and more datasets such as DEAP [6] and SAVEE [7] were added. The training accuracy reached ~86%, however, emotions were still not accurate.

By adding one more dataset—FER-2013[8], the model was now trained on small, black-and-white images, which were accurately categorized into emotions. Additionally, there were a lot of participants who made up this dataset, which ensured diversity while training the model. By using Haar-Cascade Classifiers instead of regular CNN Classifiers, the accuracy had increased by 2% and the model was able to identify the correct emotions. Haar-Cascade Classifiers is a basic machine learning classifier to train a cascade function from a large number of images with faces (i.e., positive), and images without faces (i.e., negative), the model could detect emotions in other input images. Additionally, more random images were fed, and the model gave correct responses each time.

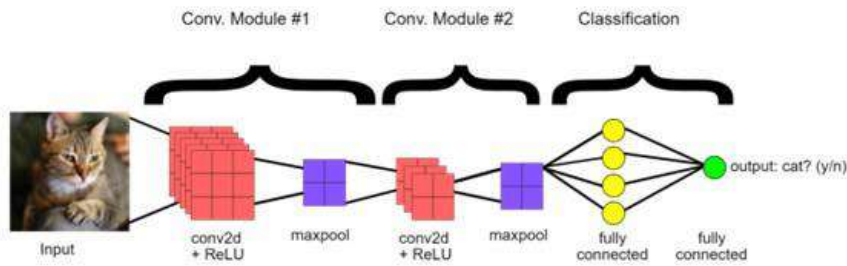


Fig. 3. Simplified Process of workings of a CNN

2.3 Creating Engagement Scale

After the successful detection of emotions, they were converted to a numerical Engagement Level value. Through a second round of literature review which consisted of research on detecting students' engagement in a classroom, the correlation of emotions to engagement was mapped. A New Emotion-based Affective model to detect

student engagement [3] proposed a system that ranked all the emotions, labelling them according to their engagement level. For example, surprise was associated with extremely high engagement, whereas stillness was associated with disengagement. This approach proved to be appropriate for the trained model.

Student engagement was thus categorized into five levels: Strong engagement, high engagement, medium engagement, low engagement, and disengagement. All of these levels can be detected based on the emotion detected, as per Fig 4, given by Altuwairqi et. al [3]:

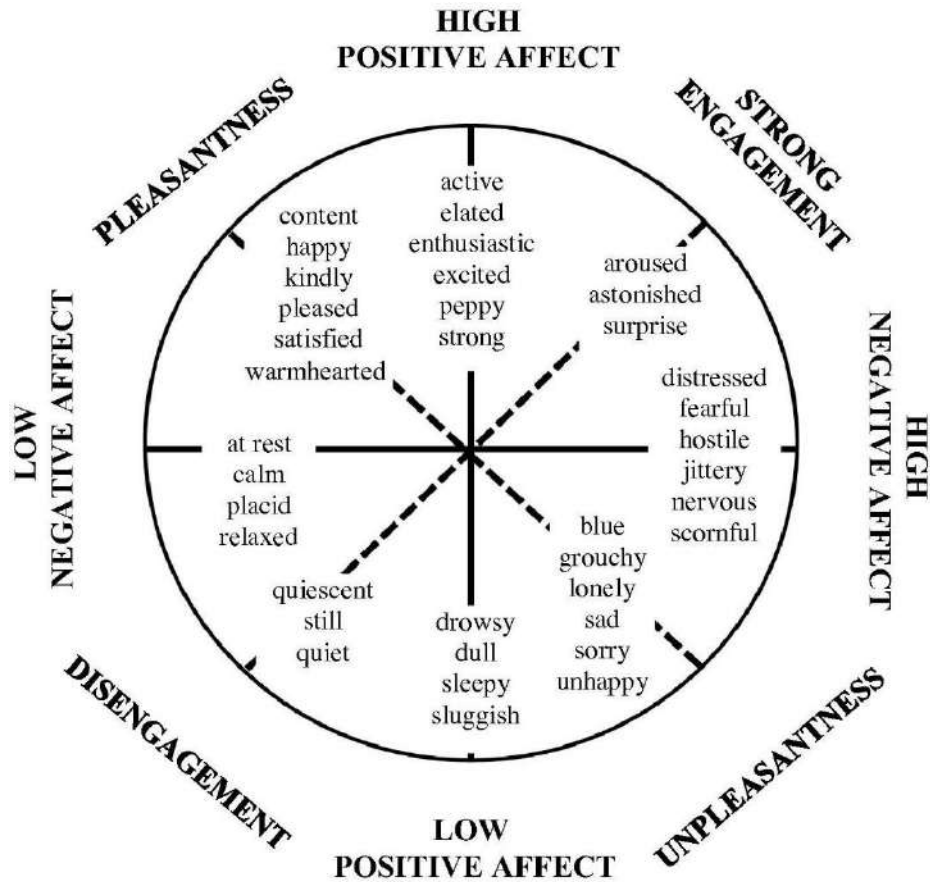


Fig. 4. Emotion to Engagement Chart showing all the possible maps.

This framework was then linked to the Engagement Level web service resulting in it successfully detecting the engagement level of a student from a photo. When deployed in a classroom, the images of the students will be captured at 2-3 minute intervals from the real-time video recording (delivering 15 - 20 data points on average) to mitigate any bias as a student might focus and refocus multiple times.

2.4 Algorithm

The algorithm for emotion detection uses the following concepts:

1. Convolutional Neural Network (CNN) - A recursive framework that divides an image into two parts repeatedly to find a certain feature in an image. Neural Networks are unique because they can be trained (assuming a comprehensive dataset is present) to identify trends in collections of images and can be used for a wide range of applications.
2. Action Units (AU) - Action Units are specific actions of facial muscles or groups of muscles. Since these are definite, the face can be divided into around 12 different AU's, which are simultaneously analyzed to identify a certain emotion.
3. Haar-Cascade Classifier - The Haar-Cascade Classifier is a faster version of a CNN classifier. It works based on Haar Cascade—an independent algorithm to detect different objects from images. These classifiers are used to detect the face of the people in the image fed to the FER model.

The algorithm for the emotion detection model works in the following way:

- The image is converted to a pixel array to allow important algorithms such as the Haar Classifier to modify and edit the image.
- The image is sent through the Haar-Cascade Classifier to identify the face of the people in the image.
- Each face is processed individually and run through the CNN to identify which emotion is being displayed and its intensity. (For eg: 75% happiness, 23% confusion, 2% anger).
- After the identification, the prominent emotion (in this case, happiness) is displayed.
- After the final detection of the primary emotion, the algorithm identifies the engagement level of a student.
- Then, the primary emotion of an image is correlated with the engagement level associated with the positive affect and negative affect model.
- Through this, an engagement score is calculated from 0-1 and displayed as a mean of 17 recordings of engagement throughout a 40-minute lecture.
- After a lot of learning and experimentation, the final model created had more than 1.2 million trainable parameters using 4+ Conv layers, 6+ Dense layers, and Max-Pooling, Flatten, and Dense layers as well. The model optimization took time and needed 150 epochs to finally reach the peak accuracy and give the best result as of now.

Although there have been approaches that use a pure engagement score, in addition to an emotion score, those approaches assume that the student will be learning in an online mode [9]. However, a pure engagement score is not ideal for classroom-based learning. This paper found that a many-to-one correspondence between emotions to engagement is highly appropriate for a live classroom.

3 Results and Conclusion

Currently, the model stands at 86% accuracy. The number of epochs used for the FER model was 150 and this was achieved using 6 Convolutional Layers, 6 Batch Normalization Layers, 6 Dropout layers, 3 Dense layers, 3 MaxPooling layers, and 2 Flatten Layers.

The model successfully detected 7 primary emotions—happiness, sadness, disgust, surprise, anger, fear, and neutral. These were mapped to high, medium, and low engagement levels (from a scale of one to five). However, the FER model will be trained further to be able to detect more emotions as the engagement level can be determined through various emotions other than the seven primary ones identified in this process. The Engagement levels were determined by mapping positive and negative emotions, so strong emotions like surprise or disgust, which have a high degree of positive or negative affect, were assigned an increased engagement score.



Fig. 5. Final Output of the Emotion Recognition Framework

Shown in Fig 5 is an example of the final output of the CNN and Facial Emotion Recognition Framework processed. As seen in the image, the face of the person is correctly identified, and the emotion is displayed next to them. The bounding box does not include unnecessary details of the image, but perfectly captures the facial features of the person. The algorithm has also been tested on images of multiple subjects, and correctly identifies each person's face, drawing precise bounding boxes. Through the final emotion detected, the engagement level is determined.

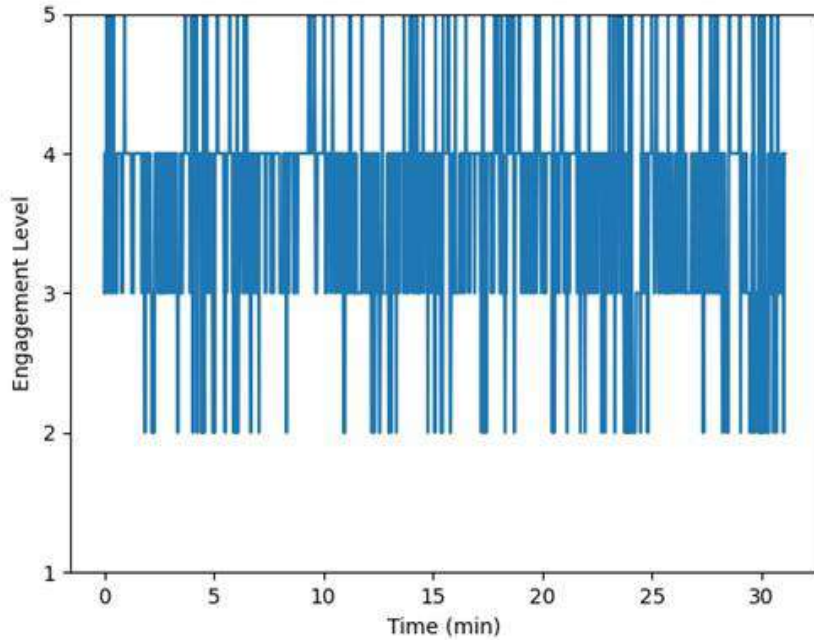


Fig. 6. Analysis of a Grade 7 Math Class at Aditya English Medium School, Pune

Fig 6 shows a graph plotted based on the data analyzed from the recording of a live academic class in Aditya English Medium School, Pune. It is from the Math class of Grade 7. To derive this analysis from the recorded video, frames at an interval of 1 second were passed to the Engagement Level detection web service. This was done to perform an in-depth analysis of the video. Thus 1800+ data points were gathered from this one recording, making it invaluable from the data collection perspective. As seen in the figure and supported by statistical calculations - the average engagement level of the students was 4 for 55% of the class time. It was at 3 for 29.5 % of the time, and below 3 for less than 5% of the time. Similarly, an engagement level of 5 was observed less than 2% of the time. The current recording of the live class was not of a specific student, but of an entire class. Hence, the engagement level computed is the mean of a few students' engagement levels at any time.

These results were further discussed with the Grade 7 Math teacher. Her initial feedback was that she believed the engagement levels were higher. However, after examining the recorded video, she started noticing the class from a different perspective. We then mapped the video that produced Fig. 6 results by pausing some frames at the specific time when the engagement level was determined. After discussing the specific frame and another 25-30 frames, we were able to verify the engagement levels. In few cases, our model was able to prove to the teacher that few students who were perceived to be disengaged by the teacher were actually engaged. This was accurate as the student responded to queries asked by the teacher and even participated

actively in class discussions at different intervals. While talking about improving the model, the teacher did point out that the engagement levels of students in certain frames were not correctly captured as they were writing in their notebook and/or not looking at the camera/teacher.

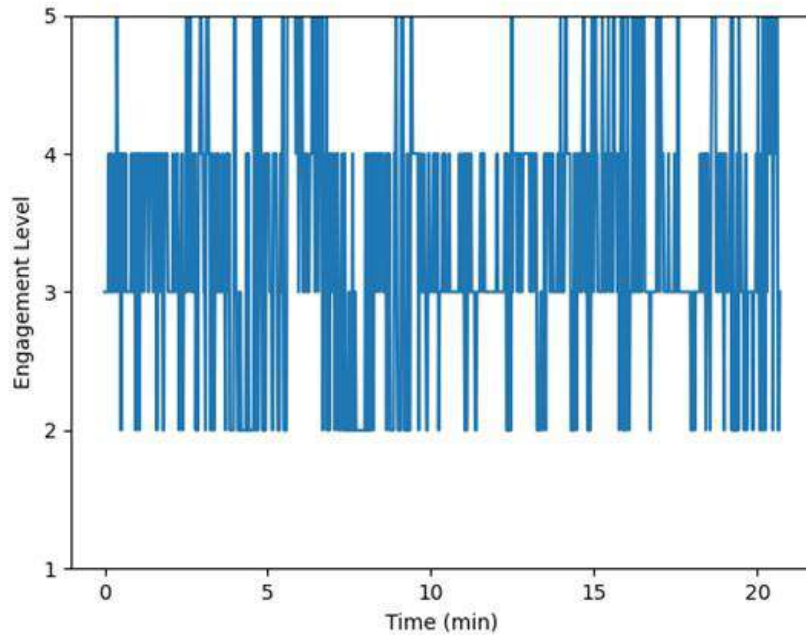


Fig. 7. Analysis of a Grade 9 Math Class at Aditya English Medium School, Pune

Displayed in Fig 7 is another analysis of the footage from the same school. This time, it was a recording of Grade 9, since it is interesting to see how the engagement level changes across grades. During the class, the average engagement level of the students was primarily 3, for 40% of the time. 30% of the time, it rose to 4, and the engagement level was at 19% and 10% for 1 and 5, respectively.

In both lectures, the teaching style of the professor was different. In the first lecture, the professor discussed problems with the students, asking questions frequently and keeping the students engaged. The professor also cracked some jokes, talked kindly, and had a friendly nature to her teaching.

However, in the Grade 9 lecture, the textbook was focused more on. Concepts were read out from the book and the professor moved in between the classroom rows, checking to see that all the students were keeping up with the class. Different students were called to stand up and read from their books, keeping the engagement through the content, instead of through social means.

Hence, there were a lot of factors that determine the engagement level of students in a classroom, such as the age of the students, the topic being discussed, the teaching style being utilized, and many more.

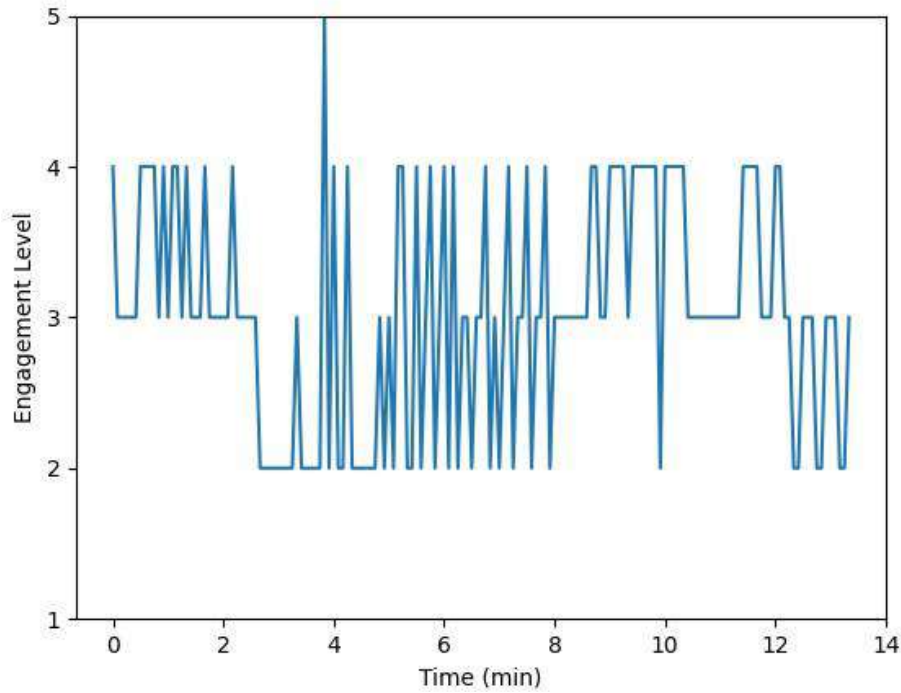


Fig. 8. Analysis of a Class based on data provided by Eagle Robot Lab, Bangalore

Eagle Robot Lab (a company based in Bangalore, India) aims to develop AI technology and humanoid robots to co-exist with humans through an indigenous principle of the collaborative learning model. This company has shown a keen interest in utilizing the Engagement Level detection service for academic betterment and is currently working with different schools across India to incorporate their humanoid robot in different classrooms. As a demo test for 15 minutes, the robot used the Engagement Level web service to determine the engagement level in a classroom.

Fig 8 represents the data that was created from a 15-minute recording provided by Eagle Robot Lab. As seen in the graph, the overall engagement level index was mostly between 2 to 4. Based on the data, the conclusion was that the robot used for teaching was able to maintain a good engagement level in the class. More can be done to further improve the engagement level to achieve results close to 4 and 5. This is currently being discussed with Eagle Robot Lab.

In conclusion, the project aims to assist teachers in Indian classrooms with the use of accurate and regular engagement reports. Aligned with the UN's Sustainable De-

velopment Goal 4 of promoting quality education for all, as well as building and upgrading safe, inclusive schools, the project would be a great tool for the teaching and learning community.

3.1 Disadvantages and Limitations

There can be some potential ethical challenges with FER in an Indian classroom setting. Due to the high number of students in one classroom, it is essential to have the consent of both, students and their parents/guardians to record the session. This is a crucial step in ensuring that facial recognition is conducted ethically with data being sourced with complete transparency. Additionally, accurately determining the engagement levels is only possible if the images are clear and their quality is consistent.

One known limitation in determining the engagement level using FER is related to the backlight from windows or improper lighting in the classroom while capturing the video/image. Often, the data can be skewed if the video is not recorded with the classroom lighting in mind. As the Engagement Level Detection web service is designed to help teachers, teacher training becomes an important step in the process. It was innovated as a helping tool for our educators, and the technology needs to be explained clearly for them to use it optimally.

Another possible disadvantage can be in the form of student placement within the classroom. To record the session and for the model to accurately determine emotions, the student must face the teacher, or the camera placed in front of the class. However, sometimes students tend to look down in case of taking notes or reading their course book. During such times, the facial expression cannot be captured and hence it becomes difficult to predict the engagement level for these specific cases. To accommodate this challenge, the algorithm processes images frequently.

4 Future Scope

Another organization that has shown interest in the Engagement Level Detection service is Pune Knowledge Cluster (PKC), a not for profit organization established by the Office of the Principal Scientific Adviser to the Government of India. PKC is currently working with more than 25 schools in the state of Maharashtra in India to plan the digital curriculum content for teachers. The number of schools will increase in the future. The impact of PKC's Digital Literacy Program will be analyzed using the AI-Based Student Emotion and Engagement Level Detection Framework. The analysis will help in further improving the curriculum content. Initially, 5 schools will run a pilot program to determine the class engagement levels.

In addition to convincing the school authorities to use this technology, the main challenge to overcome is the consent from parents to collect this data for determining engagement levels. While consent from Eagle Robot Lab, Aditya English Medium School and PKC schools was obtained, scaling up to get data from another 25 schools will need to be planned.

The collection of data and constantly training the algorithm with more and more data is one of the most important actions planned for the future. This will be achieved by collaborating with companies like Eagle Robot Lab and not-for-profit organizations like PKC, who are trying various methods to help in the educational field, such as providing the schools with digital teaching resources.

However, once this project is extended to more schools in India, the teachers can use suggested techniques to raise the engagement levels of high-frequency classes. Currently, a few more schools have been shortlisted as candidates to deploy the Engagement Detection project, so this project is one of few that can keep on expanding its work and impact, as long as there are schools to reach!

5 References

1. Thomas, C., Jayagopi, D.: Predicting Student Engagement in Classrooms using Facial Behavioral Cues. MIE 2017-Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education, 33–40 (2017). Association for Computing Machinery (2017).
2. Taylor, L., Parsons, J.: Improving Student Engagement. *Current Issues in Education*, vol. 14(1), 2011
3. Altuwairqi, K., Kammoun, J., Allinjawwi, A., Hammami, M.: New Emotion-based Affective model to detect student's engagement. *Journal of King Saud University – Computer d Information Sciences*, vol. 33, 99-109 (2021)
4. Facial Expression Dataset (AM-FED) – Affectiva, www.affectiva.com/facial-expression-dataset.
5. DISFA+, mohammadmahoor.com/disfa.
6. DEAP: A Dataset for Emotion Analysis Using Physiological and Audio-visual Signals, www.eecs.qmul.ac.uk/mmv/datasets/deap.
7. Surrey Audio-Visual Expressed Emotion (SAVEE) Database, personal.ee.surrey.ac.uk/Personal/P.Jackson/SAVEE.
8. FER-2013, www.kaggle.com/datasets/msambare/fer2013.
9. Shen, J., Yang, H., Li, J., Cheng, Z.: Assessing Learning Engagement Based on Facial Expression Recognition in MOOC's Scenario. *Multimedia Systems*, vol. 28, 469–78 (2021).
10. Vinola, C., Vimaladevi, K.: A Survey on Human Emotion Recognition Approaches, databases and applications. *ELCVIA: electronic letters on computer vision and image analysis*, 24-44 (2015).
11. Stock-Homburg, R.: Survey of Emotions in Human–Robot Interactions: Perspectives from Robotic Psychology on 20 Years of Research. *Int J of Soc Robotics* (14), 389–411 (2022).
12. El Hammoumi, O., Benmarrakchi, F., Ouherrou, N., El Kafi, J., El Hore, A.: Emotion recognition in e-learning systems. 6th International Conference on Multimedia Computing and Systems (ICMCS). IEEE (2018).
13. Spezialetti, M., Placidi, G., Rossi, S.: Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives. *Frontiers in Robotics and AI* 2020, vol. 7
14. Breazeal, C.: Function meets style: insights from emotion theory applied to HRI. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*. vol 34.2, pp. 187-194, IEEE, (2020).

15. Sun, J., Pei, X., Zhou, S.: Facial emotion recognition in modern distant education system using SVM. 2008 International Conference on Machine Learning and Cybernetics, vol. 6, IEEE, 2008.
16. Liu, Z., Wu, M., Cao, W., Chen, L.: A facial expression emotion recognition-based human-robot interaction system. IEEE CAA J. Autom, vol 4.4, pp. 668-676, IEEE (2017).
17. Devillers, L., Tahon, M., Sehili, M., Delaborde, A.: Inference of Human Beings' Emotional States from Speech in Human-Robot Interactions. *International Journal of Social Robotics* 7(4), 451–463 (2015).
18. Lackey, S., Barber, D., Reinerman., Badler, N.: Defining next-generation multi-modal communication in human robot interaction. *Proceedings of the human factors and ergonomics society annual meeting 2011*, vol. 55, 461-464. CA: SAGE Publications (2011).