Chapter 1

Introduction

1.1 Motivation

MOOCs have gained a lot of popularity over the past years and are now being offered to millions of learners on various platforms such as Coursera, Udacity, and edX, among others. A major motivation behind MOOCs is the provision of ubiquitous learning to learners of all walks of life across the globe. In MOOCs and other web-based systems, students often register to download videos and materials but do not complete the entire course. As a result, the total number of activities a student engages in falls below the recommended threshold. MOOCs suffer from low levels of learner engagement and learner retention, as only a very small percentage of learners who start a course actually complete it successfully; on average $5-10\%$ of learners succeed, in extreme cases this metric can drop below 1%. Therefore, teachers must understand the engagement of their students.

In the traditional approach to education, teachers take various steps to appraise students' levels of performance, motivation, and engagement, such as conducting exams, checking student attendance, and monitoring studying via security cameras. However, in web-based platforms, there are no face-to-face meetings, and it is difficult to determine student engagement levels in online activities such as participating in discussion forums or watching videos. Therefore, in web-based systems, student data represent the only source through which instructors can assess student performance and engagement.

Due to the absence of face-to-face meetings, web-based systems face some challenges that need to be addressed. The first and most important is course drop out. In web-based systems, dropping out is the principal problem that research has attempted to solve. In web-based systems, 78% of students fail to complete their courses. The main reason students drop a MOOC course is the lack of student engagement, and the second most common reason is their inability to locate the requisite activities and materials for the next assessment.

Student engagement is the effort that a student spends on learning processes for the content of a specific course. The most recent definition of behavioral engagement involves students who take part in discussion forums and show interest in MOOC materials. Student engagement is an important research topic because a lack of student engagement affects the student's final grade, retention of material, and the course dropout rate. A student who engages more in discussion forums and other MOOC activities usually does not drop out.

In E-learning systems, a student’s degree of engagement in educational learning is lower than that in traditional education systems. Because the course involves web-based learning, often, no face-to-face interaction occurs between students and the instructor. In web-based systems, it is difficult to measure a student’s engagement using traditional methodologies (e.g., metrics such as class attendance, participation in discussions, and grades), because many of these predictors are not directly available in e-learning systems. Therefore, investigating students’ engagement in web-based learning is a challenging task.

To accomplish the goals, a predictive analytic model utilizing machine learning (ML) algorithms. ML is a field of artificial intelligence. ML algorithms can automatically find complex patterns from features extracted from existing data, enabling them to make smart decisions about current data. The main tasks of learning analytics in education are to collect data, analyze these data and provide appropriate suggestions and feedback to students to improve their learning. With the help of predictive analytics, an instructor can also discover what students are doing with the learning material and how a student’s assessment scores are related to that student’s engagement level. The cognitive ability of computers in some fields is still below that of humans, but due to ML algorithms, computer abilities are increasing quickly in domains such as e-learning, recommendation, pattern recognition, image processing, medical diagnosis, and many others. ML algorithms are trained using sample data as inputs and then tested with new data. Instructors can use ML algorithms to obtain student-related information in real time, which helps them intervene during early course stages. ML is often used to build predictive models from student data; ML techniques can address both numerical and categorical predictor variables.

In this report the state of the art of engagement detection methods in the context of online learning, and then it identifies the challenges of detecting engagement in online learning. The existing methods are classified into three main categories—automatic, semi-automatic and manual—considering the methods’ dependencies on learners’ participation. And, then the methods in each category are divided into subcategories based on the types of data used, e.g., audio, video, learner log data, etc. In particular, the computer vision-based methods in the automatic category that use facial expressions are examined because they are promising in an online learning environment, nonintrusive in nature, and cost-effective when considering the hardware and the software needed for capturing and analyzing video data.

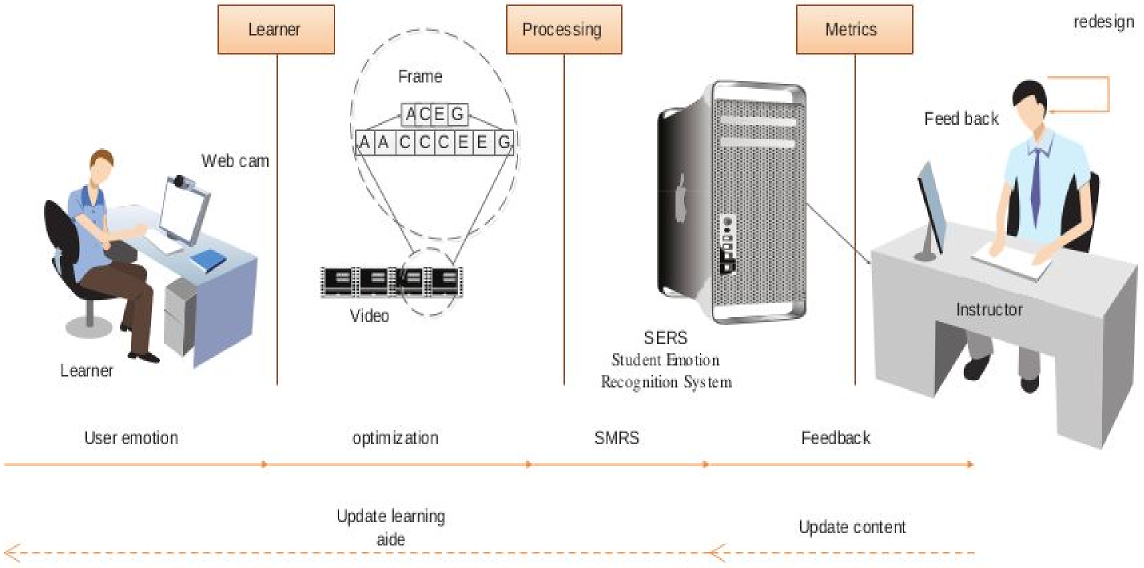


Figure 1.1: Proposed Method

Among the various possible behavioral states, engagement can be considered the most fundamental state in a learning environment as shown in Figure 1.1. Any other effective state, such as boredom, confusion, sleepiness, etc. gets reflected in the engagement levels of a student. Hence, this work contributes to the feedback system in e-learning environments by addressing the problem of engagement recognition of students while they watch online lectures by rating them on a scale of 2 engagement levels. One can conclude that the eye gaze can be associated as an indication of engagement, this approach to engagement determination has two major issues. First, the eye gaze determination is in itself an unsolved problem. Second, it can be noted that eye gaze is not a necessary condition for engagement, for example, a student may be analyzing and not making eye contact with the instructor. Engagement is very useful in e-learning, this work can also be applied in other application spheres, such as in advertising where it may help understand users' preferences better. Proposed methodology shows an improvement in performance of over 84% accuracy.

The remainder of the report is organized as follows. In Chapter 2, a Literature Survey of engagement detection methods is discussed. Among the different methods, the computer vision-based methods in the automatic category are found to be beneficial and further detailed in Chapter 3. Benchmarking datasets, performance metrics, and evaluation strategies along with some results are discussed in Chapter 4. Chapter 5 provides conclusions and outlines future work.

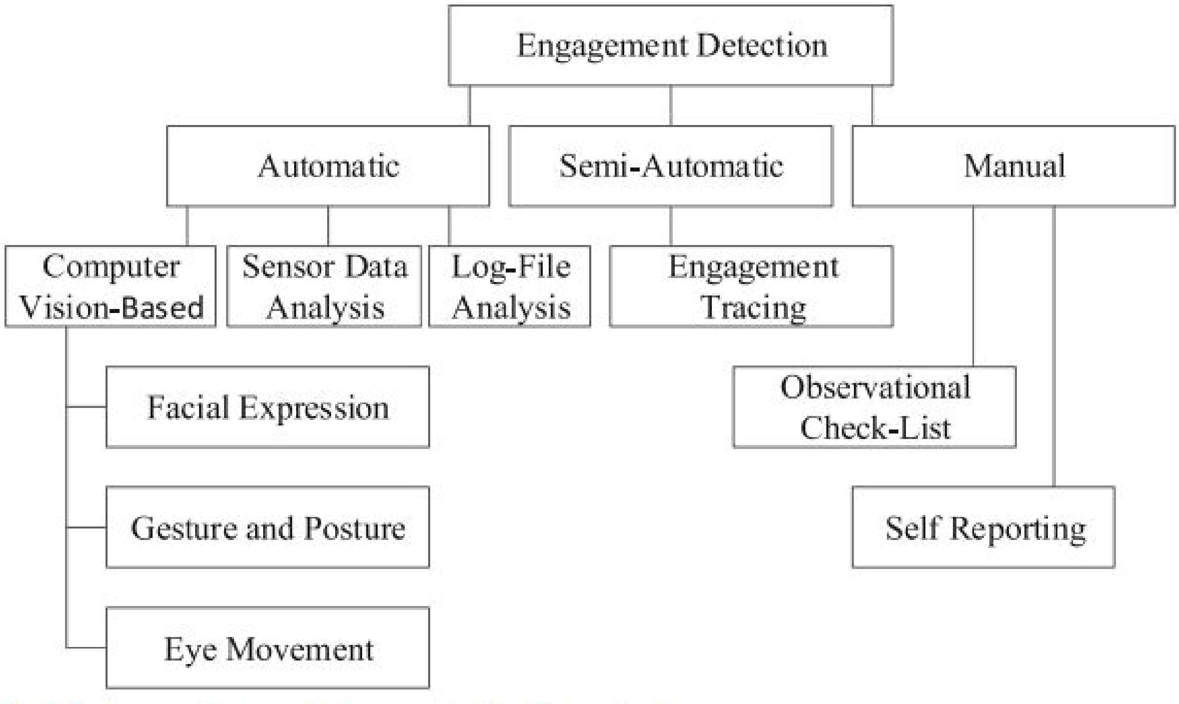


Figure 1.2: Different engagement detection methods

A few research contemplates on learners' engagement detection can be found in the literature as given in Figure 1.2. The existing methods for learners' engagement detection into three main categories — automatic, semi-automatic, and manual — based on the strategy and the type of users involvement in the engagement detection process. The manual methods are further divided into self-reporting and observational checklist categories. The methods related to engagement tracing are categorized as semi-automatic. The methods in the automatic category are divided into computer vision-based methods, sensor data analysis, and log-file analysis depending on the information that these methods process for engagement detection. The computer vision based methods are further divided into three sub-categories — facial expression, gestures and postures, and eye movement — based on the modalities they use for engagement detection.

Chapter 2

Literature Survey

2.1 Outcome of Literature Survey

References describe the details of the approaches of engagement detection. Following are the outcome of Literature survey given in Table 2.1.

TABLE 2.1: SUMMARY OF EXISTING WORKS

Author	Methodology	Remarks	Year
Amanjot Kaur, Aamir Mustafa, Love Mehta, Abhinav Dhall [1]	Facial expressions	Sequence network is powerful to detect high-level classification of engaged and not-engaged videos but is not so effective in fine grain classification of level 2 and level 3 videos.	2018
Tarmo Robal [2]	Eye movement and Facial expressions	The comparison between Tobii and TJS shows a relatively small performance gap between the Webcam-based eye tracker and the high-end device.	2018
Brandon M. Booth [3]	Facial expressions	Models trained and cross-validated on individual subjects, however, perform very well even when only a modest fraction of the annotated video frames are used.	2017
Pramodini A.	Eye movement	New eye Tracking Technology and its	2017

Punde [4]		applications	
Rajitha Navarathna [5]	Facial expressions and body motions	They proposed an automatic approach to predict movie ratings solely using audience behaviors. They showed that audience sentiment levels can be more predictive of overall movie rating than self-report measurements. They tested the utility of their approach using 30 movie sessions across more than 200 subjects.	2017
Abelardo Pardo [6]	Facial expressions	The paper highlighted the importance of analyzing learning experiences combining the insight gained with self-reported data based in well established theoretical frameworks such as self-regulation, with those obtained by methods such as recording the interactions between students and course events in an online platform.	2017
SungJin Nam [7]	Facial expressions	Their model performed significantly better than the majority-class baseline in predicting disengaged behaviors. The models were developed based on data-driven methods.	2017
Aditya Kamath [8]	Facial expressions	Results showed a 14% improvement on the dataset against traditional methods, and a 46% improvement when the most ambiguous class from the dataset is ignored, corroborating the promise of the method.	2016
Ciprian A Corneanu [9]	Facial expressions	They define a new taxonomy for the field, encompassing all steps from face detection to facial expression recognition, and describe and classify the state of the art methods accordingly.	2016
Hamed Monkaresi [10]	Facial expressions	Due to limits in analyzing head motions, and frequent face occlusions, their methods were not able to extract features from some video segments, thereby leading to data loss.	2013

2.2 Problem Statement

To analyze the impact of students' eye gaze detection for behavioral engagement analysis in the e-learning environment.

Objectives of the Project

1. To analyze the students' eye movement using Tobii eye tracker
2. To detect the facial expression for engagement analysis using existing literature such as haar, sift.
3. To statistically analyze and prove the impact of eye gaze detection over the other parameters such as facial expression.

Chapter 3

Methodology

A generic framework for a learner's perceived engagement detection using vision-based on based methods. The framework consists of five different functions that include detection, feature extraction, tracking, classification, and decision.

In a computer vision based engagement detection system as given in Figure 3.1, video streams are captured using a webcam or a surveillance camera, where the camera provides a particular view of learners participating in a learning activity. The system seeks to detect the region of interests (ROIs) (e.g., face, gestures, postures or eye) of the learners in the live video stream. Typically, engagement detection in such system is performed with a track-and-classify approach. The system first performs segmentation to isolate the ROIs using a detection the module in each frame. For each ROI, features are then extracted in a feature extraction module and selected into patterns to initiate tracking and classification. A classification module is used to match input patterns against patterns extracted from training dataset and generates classification scores. A tracking module is designed for tracking the movement or changes in the ROIs in consecutive frames and

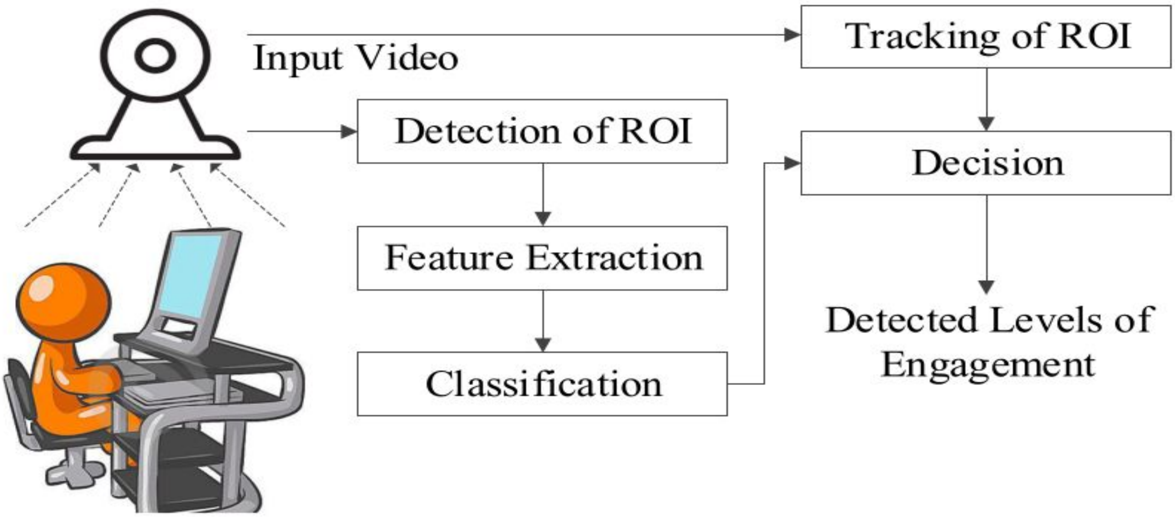


Figure 3.1: A generic framework for computer vision based engagement detection system

generates tracking trajectories. Finally, a decision module combines classification scores over trajectories to output a list of engagement levels of the learners in the input video stream as given in Figure 3.2.

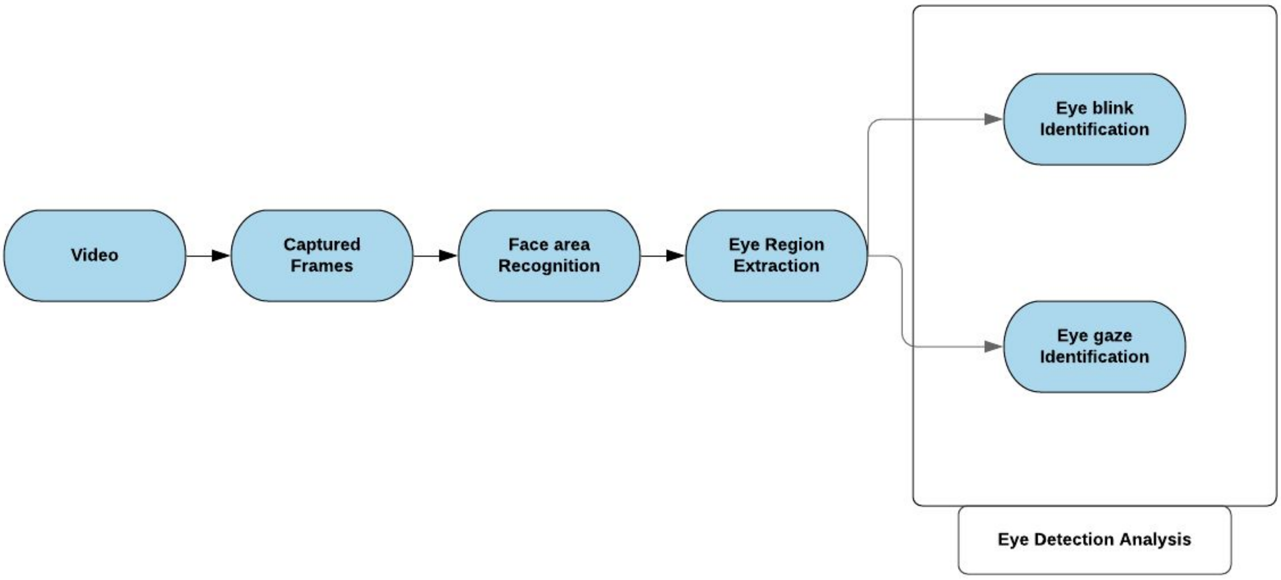


Figure 3.2: Eye movement and Facial expression Method

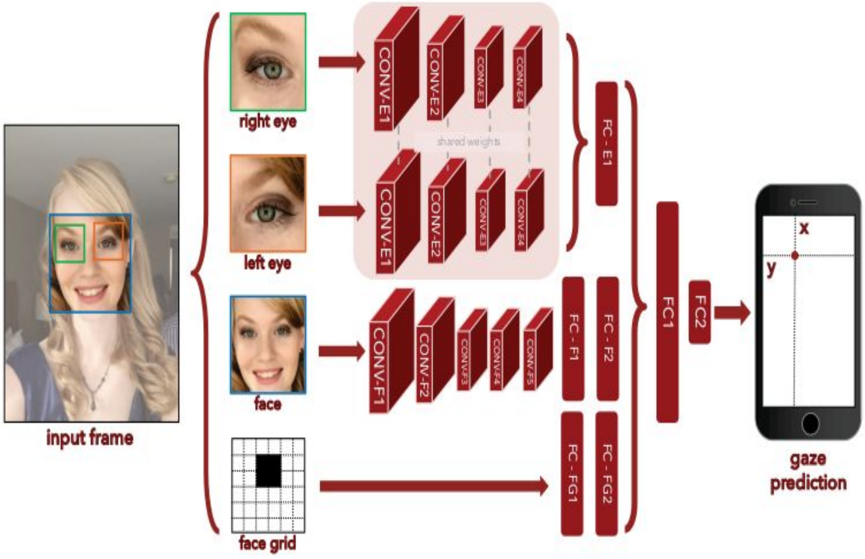


Figure 3.3: Eye Gaze Method

My goal is to design an approach that can use the information from a single image to robustly predict gaze as given in Fig 3.3. I choose to use deep convolutional neural networks (CNN) to make effective use of our large-scale dataset. Specifically, provide the following as input to the model: Firstly, the image of the face together with its location in the image (termed face grid), and secondly, the image of the eyes.

Based on this information, I design the overall architecture, as shown in Figure 3.4. The size of the various layers is similar to those of AlexNet. Note that the model includes the eyes as individual inputs into the network (even though the face already contains them) to provide the network with a higher resolution image of the eye to allow it to identify subtle changes, as shown in Figure 3.5.

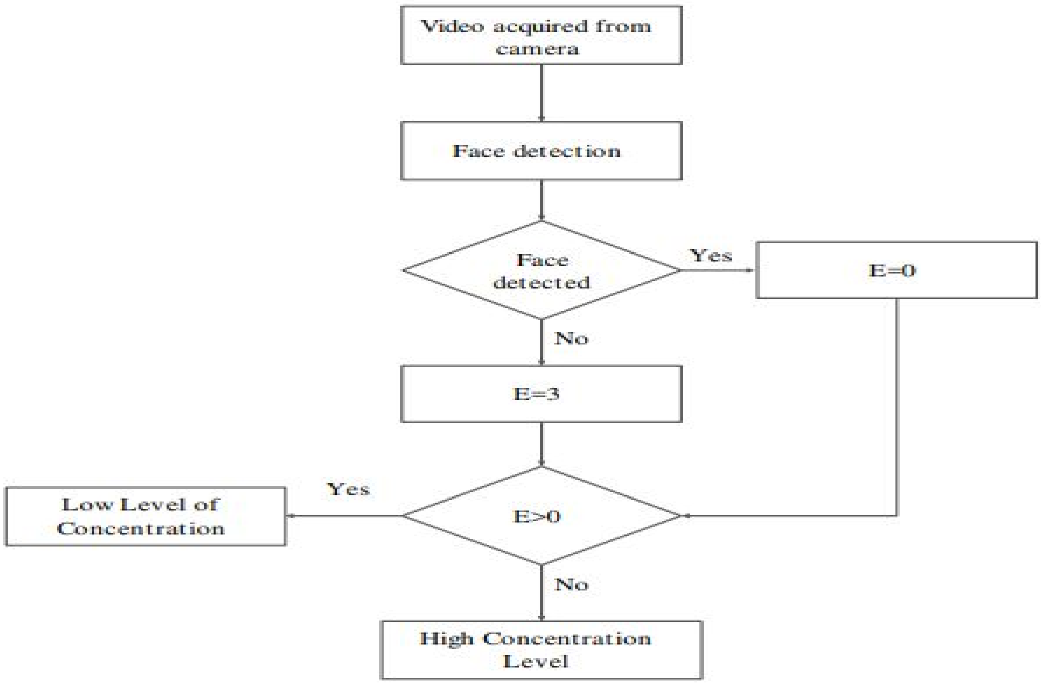


Figure 3.4: Low concentration level detection

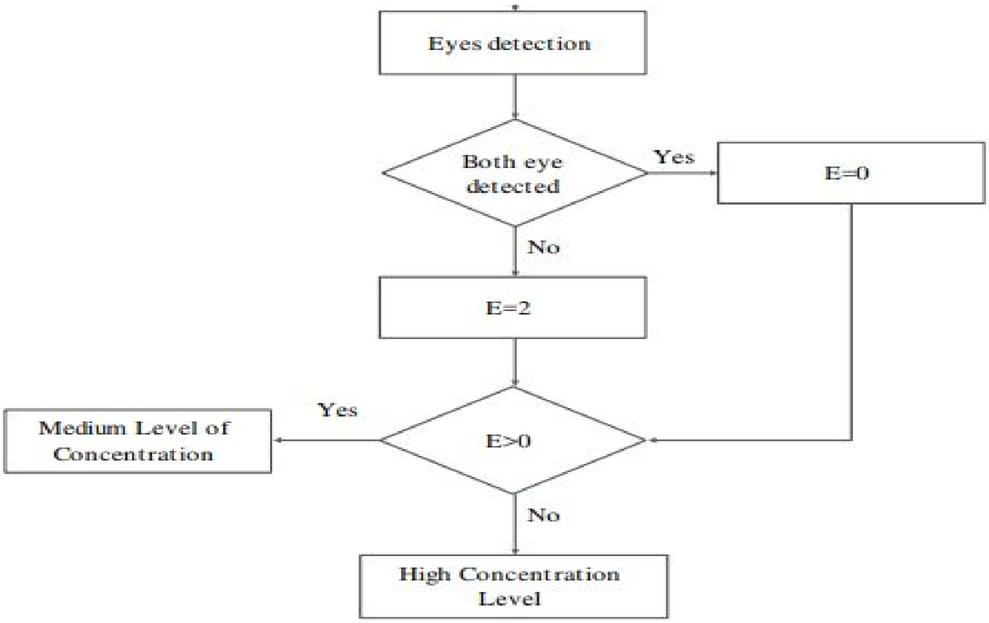


Figure 3.5: Medium concentration level detection

Chapter 4

Experimental Results and Analysis

The Confusion Matrix of different method is shown in Figure 4.1 and Figure 4.2.

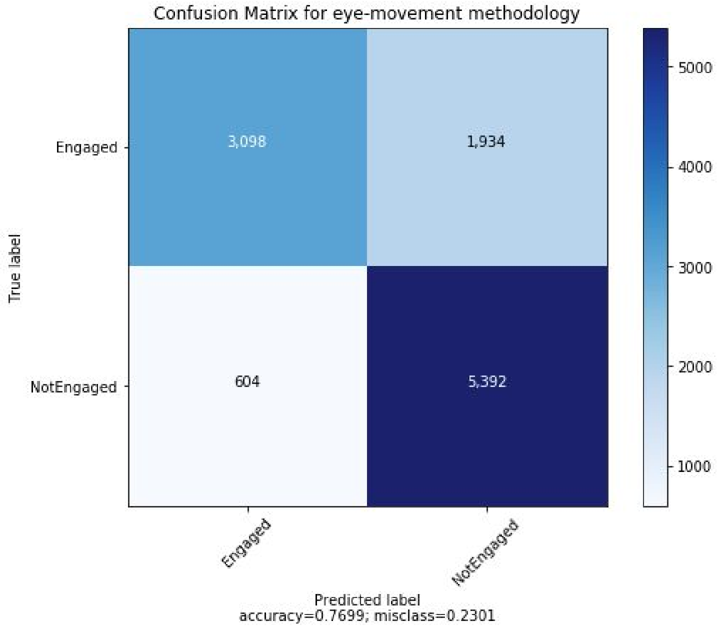


Figure 4.1: Confusion Matrix for eye-movement methodology

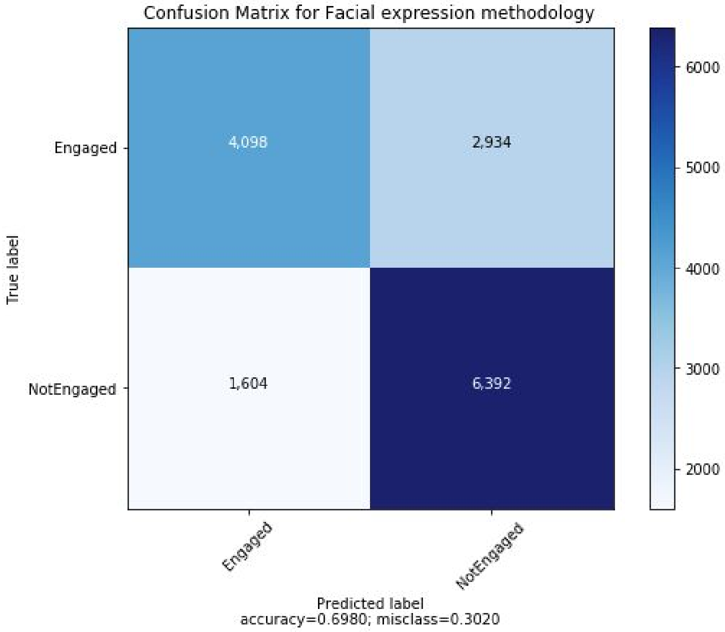


Figure 4.2: Confusion Matrix for Facial expression methodology

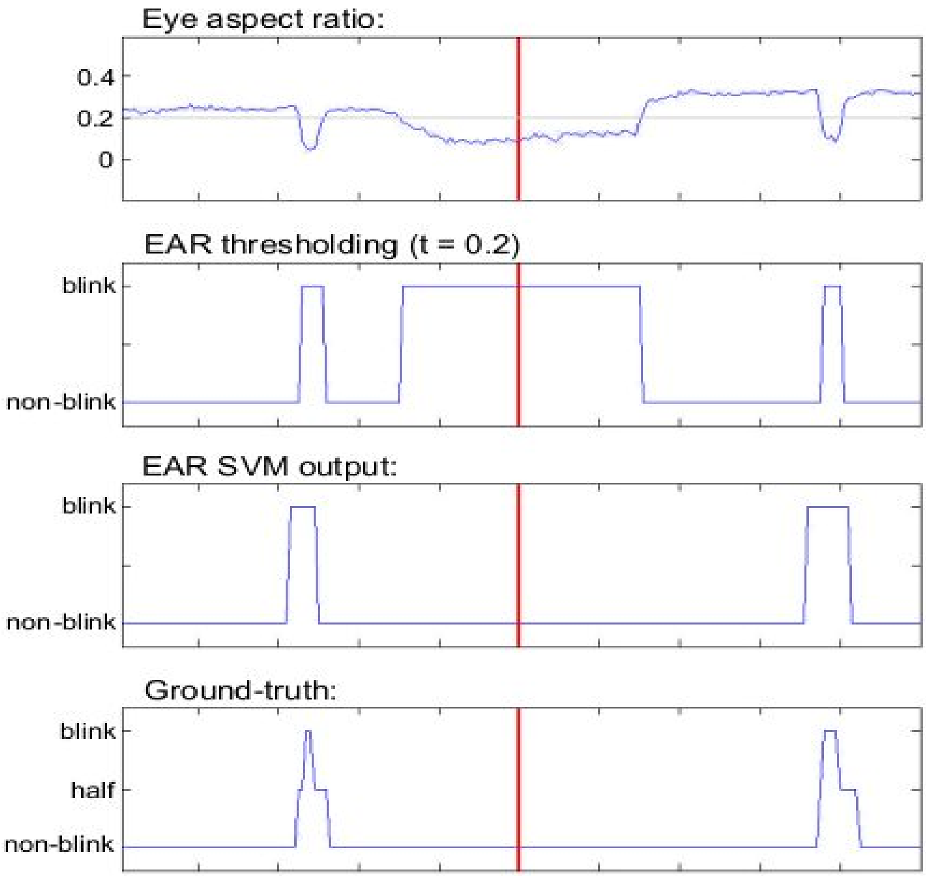


Figure 4.3: Example of detected blinks where the EAR thresholding fails while EAR SVM succeeds.

Table 4.1: Accuracy Table

Classification	Accuracy	Classifier	Window size
Engagement	74.40	CNN	12
Not Engagement	64.50	CNN	12

Concentration level can be detected by doing the analysis of the entire three components data together as shown in Figure 4.3 and the accuracy shown in Table 4.1. The concentration level can be categorized into three levels namely high, medium and low respectively. The data value 4 is considered as a medium, the data value 0 is considered as high because there is no variation and the data value 3 is considered the as high level of concentration. The count value 3 is for the head and thus if the head is not visible it means the student is not at all viewing the screen and thus the concentration level is considered as low. The table below represents the various level of concentration, as shown in Table 4.2.

Table 4.2: Concentration Table

Frame	Left Eye	Right Eye	Face	Concentration Level
1	1	0	0	Medium
2	0	0	0	Low
3	0	1	0	Medium

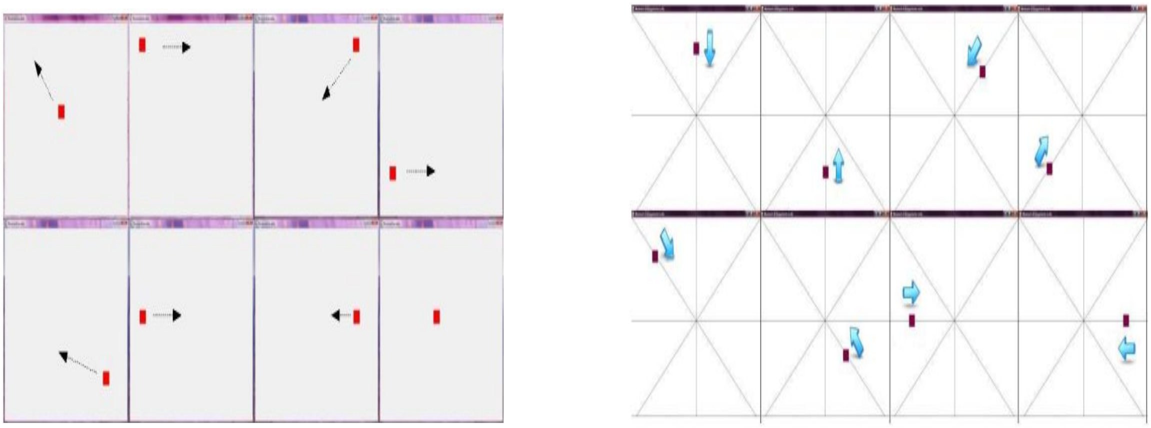


Figure 4.4: Movement of Tobii

The gaze movement of Tobii eye tracker is shown in Figure 4.4. Figure 4.5 shows the percentage of engagement level in a particular video of different frame size.

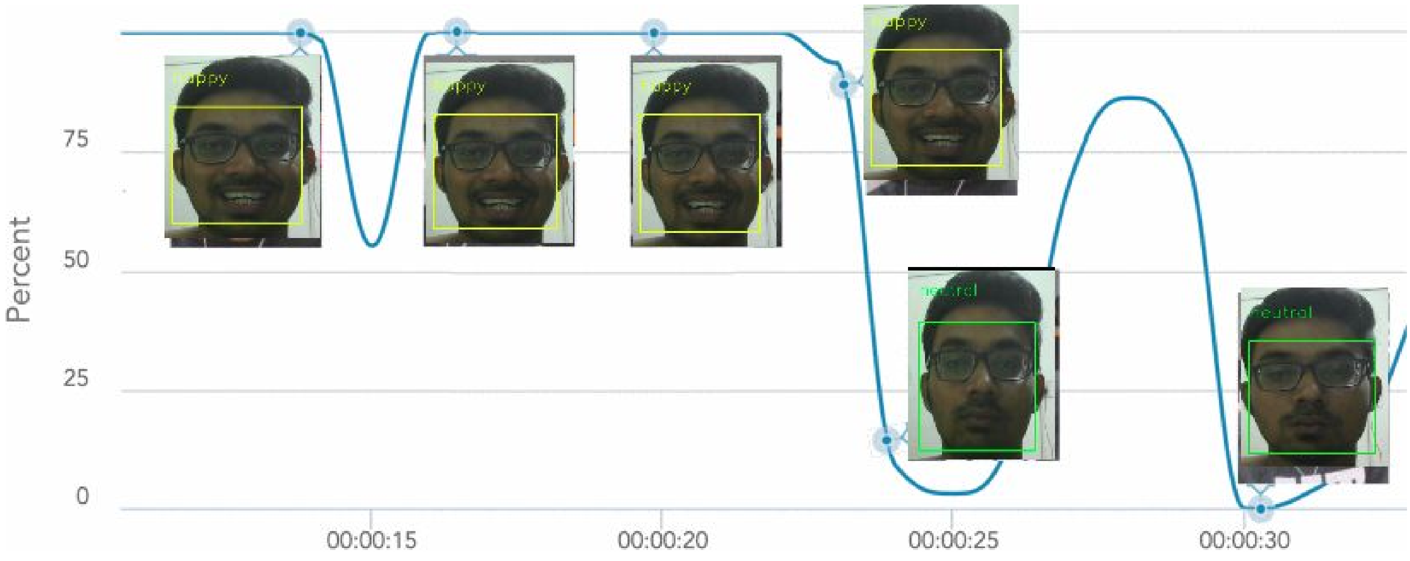


Figure 4.5: Engagement level (y-axis) as determined by Emotion

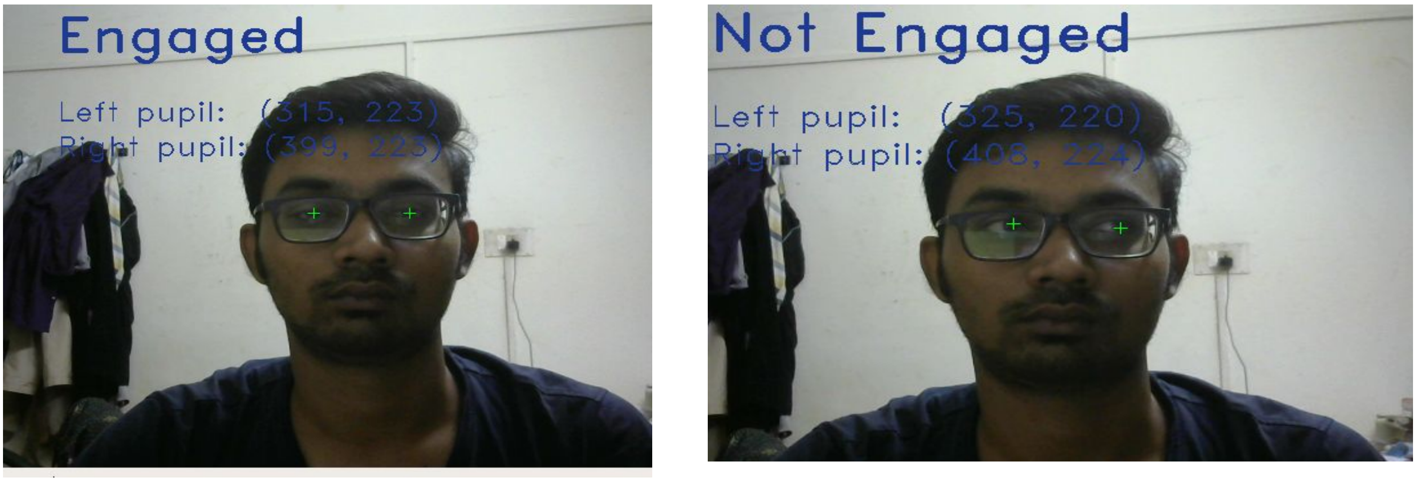


Figure 4.6: Engagement level using eye gaze

In Fig 4.6 Engaged or not engaged is detected using a webcam analyzing the eye gaze. If the student is looking at extreme left or right then it means that student is not engaged in the e-learning classes.

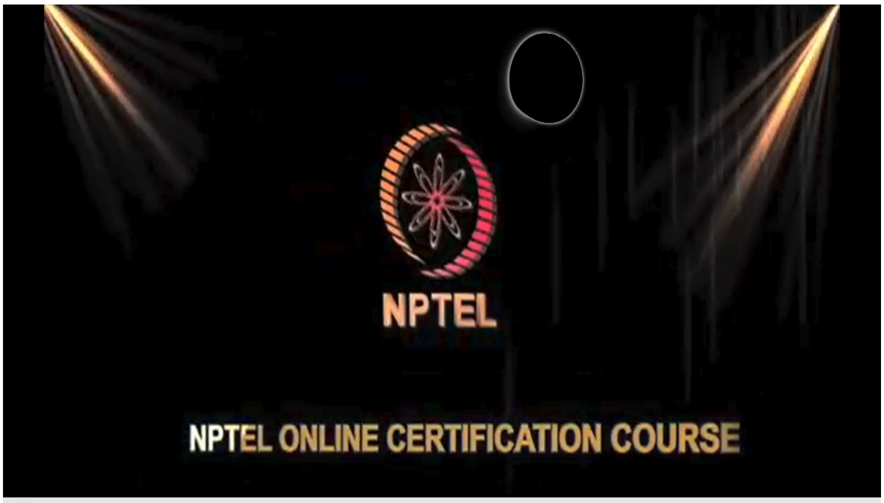
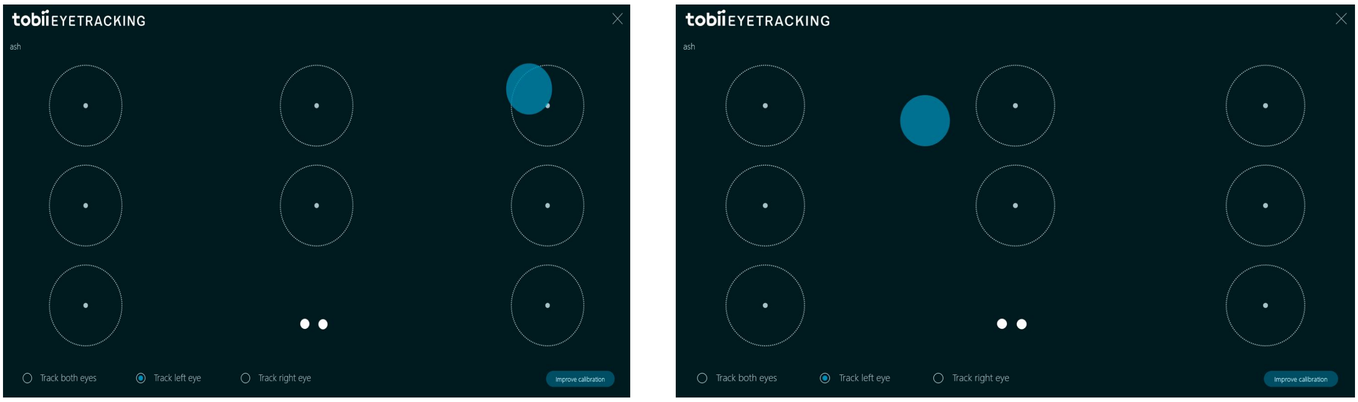


Figure 4.7: Gaze Trace in windows using Tobii Eye Tracker

In Fig 4.7 the white bubble is the eye gaze on the e-learning video. It will move according to the eye gaze of a student in an e-learning environment and teacher can analyze where the student is actually looking at the particular frame.

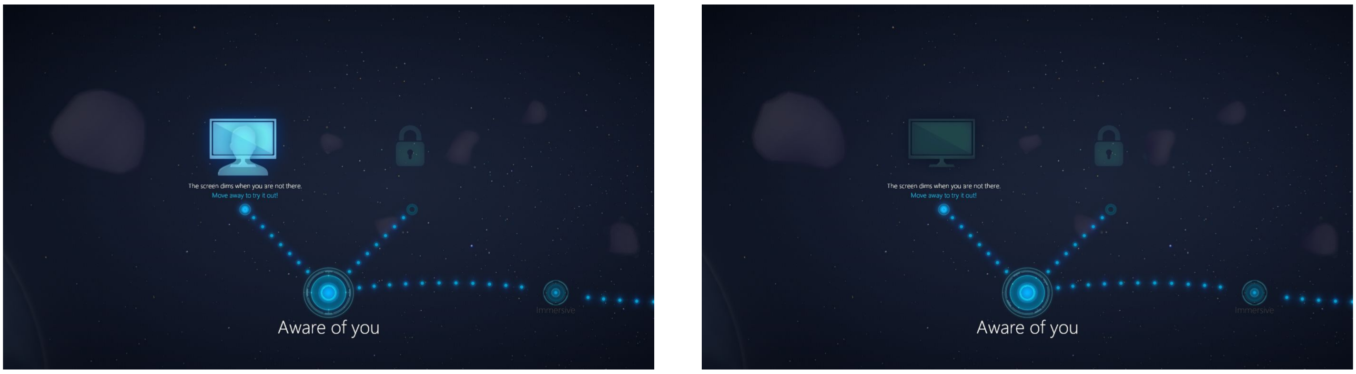


○ Track both eyes ● Track left eye ○ Track right eye

Figure 4.8: Gaze Trace using Tobii Eye Tracker

In Fig 4.8 the blue dot is the eye gaze on the screen. It will move according to the eye gaze of a student in an e-learning environment and teacher can analyze where the student is actually looking at the particular frame.

It also has the option of selecting a particular eye i.e., tracking only left eye, only right eye or tracking both the eye.



The screen dims when you are not there.
Move away to try it out!

Figure 4.9: Screen Brightness alarm based on presence in front of the screen.

In Fig 4.9 the two different screen when a student is present in front of the screen or when not present in front of the screen. The screen will produce different light on a different position.

In Table 4.3 the different task has been compared based on webcam eye gaze and Tobii eye gaze. The Tobii eye gaze is accurate in many cases as given below. But the cost of Tobii and Tobii software is more as compared to webcam-based eye gaze, as shown in Figure 4.10.

Table 4.3: Comparison of Accuracy Table

Tasks	Accuracy (%)	
	Webcam	Tobii
1. Cover the camera for 2 seconds	**45**	7
2. Cover your face with both hands for 5 seconds	67	**75**
3. Look what is under your table (3 sec)	64	**81**
4. Tilt your head to the right for 3 seconds	**59**	38
5. Check if there is an HDMI port on the laptop	56	**77**
6. Take a sip from the cup while turning away from the camera, return after the ding	40	**51**
7. Check your phone for 10 seconds	42	**89**
8. Open a browser and navigate to www.nitk.ac.in. Return after the ding. (15 sec)	**97**	80
9. Rest your eyes for 5 seconds (close them)	**84**	19
10. Stare at the camera for 3 seconds	**89**	88
11.Lean back and put your hands behind your neck for 5 seconds	**67**	63

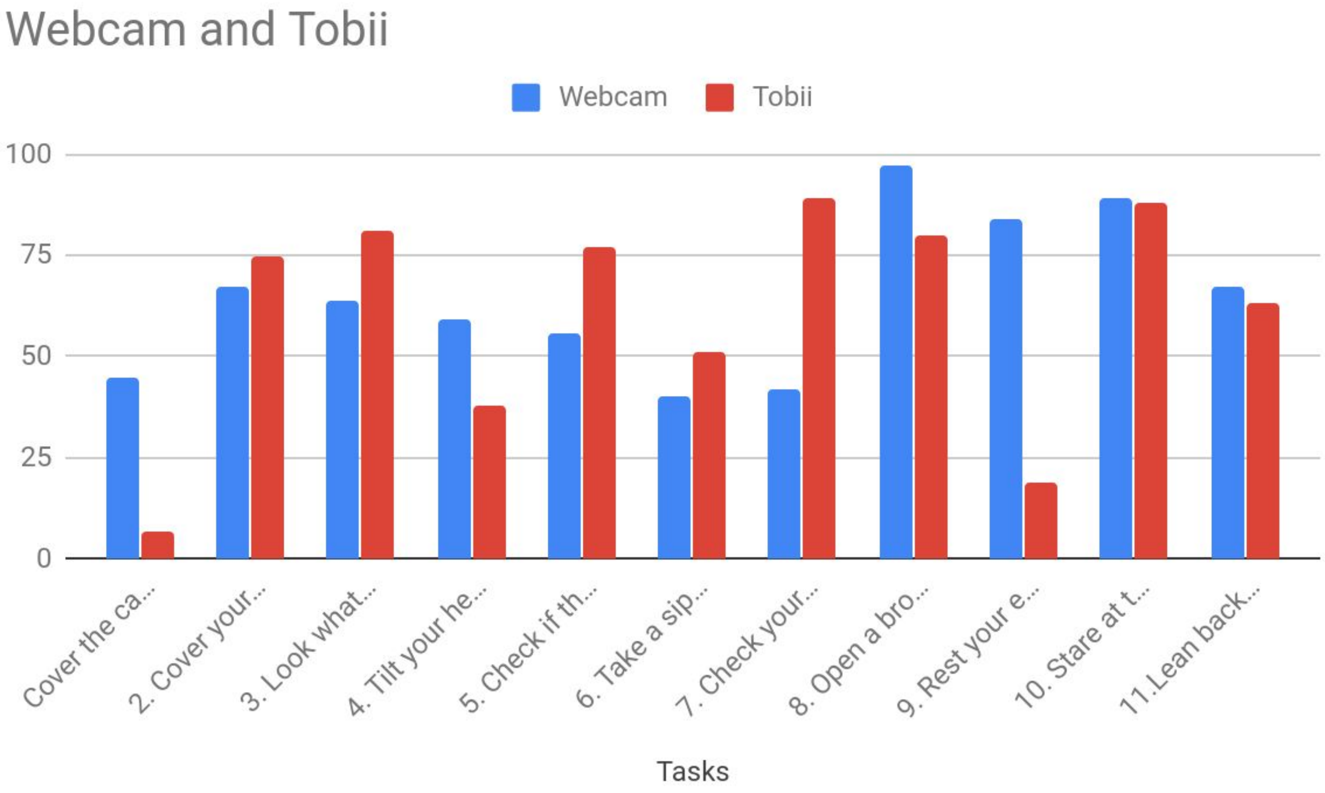


Figure 4.10: Histogram for tobii vs webcam

Chapter 5

Conclusion and Future work

This seminar has presented a review of engagement detection methods in the learning context. Although the computer vision-based methods are found to be promising in engagement detection, they do have some limitations. Automatic gathering and analyzing the behavioral data in naturalistic scenarios is still challenging for computer vision-based methods. For example, the existing algorithms face challenges to analyze head motion and facial occlusions. In such a situation, these algorithms are not able to extract features from some video segments, thereby leading to data loss. Another challenge is to extract robust features from the region of interests due to segmentation error. Although a lot of attention has been given towards deploying facial expression analysis, the challenges encountered in these endeavors are not only in terms of technical issues.

This is not clear enough how frequently the decision on engagement detection should be made – frame by frame, a short fragment of a video or an entire video clip? In case of a short fragment, what the length of a video clip is suitable to assign a single level? During labeling training data, it is unclear what exactly should be the standard for deciding what emotions a learner is truly having. Should it be the learner or the trained judges? Although the highest interrater reliability was obtained between the trained judges, it might nothing more than an artifact brought on by the training. This is also not clear what environmental constraints are needed to be considered while capturing videos for engagement detection in the context of online learning. As also many other researchers suggested in their research studies that combining different modalities can help to improve the accuracy of engagement detection. For example, facial expression, eye tracking, body parts motion, ocular parameters, gestures, postures, voice, and gaze are needed to experiment with biometric information (e.g., galvanic skin response, heart rate, electromyography of the jaw, respiration rate, respiration amplitude) collected from learners' smartwatch and brainwave-sensing eyeglasses. Features extracted from the engagement tracing, self-reporting, and observational checklist can also be experimented with the above automatically extracted features to improve in engagement detection results.

Addressing the above challenges can contribute to advance the research of automatic engagement detection in a computerized educational environment and lead to more effective learning and more engaging experience for learners. Along with these, make the following recommendations for further improvement in this research field.

Recent advances in machine learning tools, such as CNN, require more data volumes than currently available. Collecting and analyzing behavioral data in naturalistic scenarios is itself a challenging issue. Learner engagement detection systems cannot be useful unless that can address the issues related to environmental constraints. Some of the challenges include illumination variation, occlusions, head poses, objects appearing too far or close, and so on. So far, very limited attempts have been taken to resolve this problem.

Future study should also investigate what, how, when and why learners' get disengaged and how to re-engage them effectively. Future research should go more detail into the temporal domain and investigate at what frequency an engagement expression appears and how quickly it goes away. Further effort should also be given to examining how engaged/disengaged behaviors are associated with learning outcomes.

REFERENCES

[1] Amanjot Kaur, Aamir Mustafa, Love Mehta, and Abhinav Dhall. (2018). “Prediction and Localization of Student Engagement in the Wild”. In Proceedings of ACM Woodstock conference, Jennifer B. Sartor, Theo D’Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, Article 4, 9 pages.

[2] Yue Zhao, Christoph Lofi and Claudia Hauff and Tarmo Robal. (2018). “Webcam-based Attention Tracking in Online Learning: A Feasibility Study”. In 23rd International Conference on Intelligent User Interfaces, Pages 189-197.

[3] Brandon M. Booth, Asem M. Ali, Shrikanth S. Narayanan, Ian Bennett, and Aly A. Farag. (2017). “Toward Active and Unobtrusive Engagement Assessment of Distance Learners”. In Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), Pages 2156-8111.

[4] Pramodini A. Punde, Mukti E. Jadhav and Ramesh R. Manza. (2017). “A study of Eye Tracking Technology and its applications”. In 1st International Conference on Intelligent Systems and Information Management (ICISIM).

[5] Rajitha Navarathna, Peter Carr, Patrick Lucey, and Iain Matthews. (2017). “Estimating Audience Engagement to Predict Movie Ratings”. IEEE Transactions on Affective Computing (2017).

[6] Abelardo Pardo, Feifei Han, and Robert A. Ellis. (2017). “Combining University Student Self-Regulated Learning Indicators and Engagement with Online Learning Events to Predict Academic Performance”. In IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, Volume: 10, Pages: 82-92.

[7] SungJin Nam, Gwen Frishkoff, and Kevyn Collins-Thompson. (2017). “Predicting Students’ Disengaged Behaviors in an Online Meaning-Generation Task”. In IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, Volume: 11, Pages: 362 - 375.

[8] Aditya Kamath, Aradhya Biswas, and Vineeth Balasubramanian. (2016). “A Crowdsourced Approach to Student Engagement Recognition in e-Learning Environments”. In IEEE Winter Conference on Applications of Computer Vision (WACV).

[9] Ciprian A Corneanu, Marc Oliu, Jeffrey F Cohn, and Sergio Escalera. (2016). “Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications”. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[10] Hamed Monkaresi, Nigel Bosch, Rafael A. Calvo, and Sidney K. D'Mello. (2013). “Automated Detection of Engagement Using Video-Based Estimation of Facial Expressions and Heart Rate”. In IEEE Transactions on Affective Computing, Volume: 8, 2017 Pages, 15 - 28.