EXECUTE: Exploring Eye Tracking to Support E-learning

Ahsan Raza Khan, Sara Khosravi, Sajjad Hussain, Rami Ghannam, Ahmed Zoha, and Muhammad Ali Imran
James Watt School of Engineering, University of Glasgow, United Kingdom
Email: {a.khan.9, s.khosravi.1}@research.gla.ac.uk,
{Sajjad.Hussain, Rami.Ghannam, Ahmed.Zoha, Muhammad.Imran}@glasgow.ac.uk

***Abstract*—The outbreak of the COVID-19 pandemic has caused unprecedented disruption to education and progressed remote teaching as a predominant model for delivering educational content. However, the online teaching and learning model has its own challenges, such as the lack of technological tools to quantify the student attention and engagement with the learning content. In this paper, we focus on developing an e-learning framework for capturing and analysing the students' attention during remote teaching sessions and subsequently profile their learning behaviour leveraging eye-tracking data. Our proposed eye-tracking solution deploys a webcam to capture and track raw gaze points that grant the user the freedom of natural head movement and scalability compared to conventional eye-tracking approaches. We derived various gaze metrics in conjunction with state-of-the-art machine learning (ML) models like logistic regression, support vector machine and polynomial regression to classify the student attention with an accuracy above 91%. Furthermore, our findings can help in the early detection and diagnosis of attention deficit hyperactivity disorder (ADHD) among students, and thus, support their learning journeys by creating an adaptive learning environment tailored to their needs.**

***Index Terms*—Gaze tracking, attention monitoring, attention-deficit hyperactivity disorder (ADHD), mind wandering.**

I. Introduction

Attention diversion is a natural phenomenon that occurs frequently in which the focus of an individual generally drifts away from the related task. This unintentional shift has a huge impact on the learning behaviour and academic performance of students [1]. In traditional classrooms, experienced teachers can easily judge the level of engagement through observation and intervene where necessary. Moreover, it also provides continuous visual feedback to teachers and assists them to understand if the student is following the content properly or not [2]. However, this flexibility has been lost in remote teaching and learning environments where no visual feedback is available, making it difficult to monitor the student attention. Furthermore, with digital media and smartphones, students are more prone to distraction due to multi-tasking, text messaging, which makes it more challenging to focus on lectures in remote learning [3].

The recent outbreak of the COVID-19 pandemic forced educational intuitions to adopt remote teaching and learning as a predominant model for delivering lectures. This method of teaching has its own challenges such as no visual feedback for teachers, lack of technological tools to monitor engagement level during lectures. Eye-tracking is a non-invasive technique to measure the visual attention of individuals. It is believed that gaze movement is connected with the attention level of participants which is used in various psychological studies in response to visual stimuli. This relation provides useful insights in various contexts including consumer preferences in online shopping [4], online reading, emotion processing and psychopathology i.e., autism and attention deficit hyperactivity disorder (ADHD) [5]. In recent times, eye tracking is also used in the academic environment to detect mind-wandering whilst reading the online content or by watching the online video lectures [3], [6]. In this regard, most of the existing research is relying on expensive commercial-grade eye trackers which are very expensive, obstructive and require additional equipment which prevents them from large-scale deployment. Furthermore, it is believed that gaze movement is also linked with the psychological and emotional state of individuals in response to visual engagement. Therefore, attention monitoring can also lead to early detection and diagnosis of ADHD among students by identifying the anomalous gaze pattern which can help them, and their universities to develop effective coping strategies that assist with concentration and memory.

In this work, we aim to develop a scalable, inexpensive eye-tracking solution EXECUTE (Exploring Eye Tracking Data to Support E-learning) which captures and analyses the gaze data to quantify the attention level of the individual learner during a remote learning session. For this purpose, instead of relaying the expensive commercial-grade eye tracker, our focus was to provide an inexpensive solution using ordinary webcams to capture the real-time on-screen eye-gaze locations during a remote lecture. The designed interface is privacy-aware since it runs entirely on the user computer, with no video recording. The data captured by the EXECUTE interface is the 2D gaze coordinates along with the time stamp in response to visual stimuli. Furthermore, we also proposed a novel attention score matrix to quantify the attention level of students leveraging the gaze metrics. To evaluate the performance of the proposed scheme, a small scale experiment with 25 participants is conducted. For raw gaze data collection, a short introductory video lecture of eight minutes is prepared as visual stimuli. Our proposed scheme employs the classical supervised machine learning (ML) algorithms like logistic regression, support vector machine (SVM) and polynomial regression to automatically classify the student attention using

the novel attention score matrix. The key contributions of this work:

- Developed a privacy-aware low cost eye-tracking interface that captures and analyse real-time eye-tracking data using an ordinary webcam to support remote learning.
- We proposed a novel attention score matrix to quantify the student engagement based on the gaze metrics.
- Developed a ML model leveraging the novel attention score matrix to automatically classify the student engagement.

The rest of the paper is organized as follows. Section II covers the related works. Section III discussed the proposed methodology for attention classification. In section IV, the results and discussion are given where as Section V concludes the paper.

II. Related Work

In recent times, the education of millions of students around the world was disrupted and they were forced to adopt the remote learning model due to the unexpected spread of the Covid-19 pandemic. In addition to many other associated issues in online learning, attention diversion or distraction caused by the surroundings is one of the major concerns for students as well as for the teachers. Although, the level of attention significantly affects the performance of a student, yet it is challenging to maintain a high degree of attention among students for a longer period of time in a remote learning environment [7].

The current online teaching and learning lack visual feedback which makes it very difficult to assess whether students maintain the concentration level during the learning activity [8]. Various methods to study the attention level of students have been used in traditional classrooms such as retention of course content [9], direct probes in class, and observation of distraction [10]. An experienced teacher usually use their skills to observe the student facial expression and determine whether the student is attentive or distracted. However, in an online scenario, students are more prone to distraction caused by their surroundings. Therefore, assessing the level of attention in remote learning is more difficult due to the lack of informal social interaction.

To overcome this challenge, gaze tracking is considered as one of the promising solutions to measure the level of engagement during online lectures. Apart from gaze aware systems, many wearable sensor-based attention monitoring systems are already available in literature. For instance, in [11], an efficient attention monitoring system is proposed using an electroencephalography (EEG) signal to support e-learning. In this work, SVM is used to automatically identify the state of a student i.e. either high or low in a video lecture with an accuracy of 89.52%. In [12], heart rate is used to detect the mind-wandering while students were engaged in watching a video lecture on their smartphones. In this work, the student response is recorded using a thought probe. Similarly, many researchers also used different sensor-based approaches to detect the student attention level during the academic environment. For instance, the authors in [13] used the Affectiva Q sensor to record the skin temperature and galvanic skin response while engaging with the reading task. The participant frequently reported their diversion using thought probes and provide valuable insight on attention level during a reading session. Similarly, other human behaviours like head pose tracking, face tracking, and heart rate variability are also used to track the attention level [3], [11]. However, these solutions involve specialized equipment for data collection which is not feasible in a remote learning environment on a large scale. Furthermore, face tracking and pose tracking involves the video capturing of students which poses a serious privacy concern and limits the applicability in a real remote learning scenario.

On the other hand, gaze tracking is perhaps one of the most promising modalities to detect the attention level due to the so-called eye-mind link which predicates a coupling between gaze movements and attention level. To exploit the gaze features, a supervised ML-based model is used to classify the student attention during a self-placed reading session [14]. In this study, a commercial eye tracker is used to collect the gaze sample and mind wandering is reported in response to thought probes triggered when the participant fixates on predefined areas on the screen and achieved an accuracy of 60%. The study in [15] provided clear evidence that gaze movement can be used to predict the attention level while watching a non-educational video. In this work, the participants were asked to watch a short video clip with two conditions i.e., without distraction and while performing the mental calculation during visual engagement with the content and achieved an accuracy of 80.6%. The eye-mind link has inspired my researcher to develop an automated attention classification system using gaze data. However, most of the recent studies used commercial eye trackers to capture the gaze data which are very expensive and difficult to deploy in a real remote learning session. Furthermore, it is also not feasible to ask the student to wear an eye-tracking device or other sensors during lectures. Therefore, a scalable non-invasive eye-tracking solution for attention classification is required to assist the teacher in gauging the student attention level and providing timely feedback to the potential at-risk students., Furthermore, it is believed that the gaze movement is directly linked with the emotional and mental state of an individual, hence it can also be used for early detection and diagnoses of psychological issues like ADHD by identifying the anomalous gaze pattern during the engagement with visual stimuli.

III. Proposed System Model

In this work, we aim to develop a scalable and privacy-aware interface to capture and analyse gaze tracking data using an ordinary webcam in a remote learning setting. The proposed model uses the classical ML approach to automatically classify the student attention levering the novel attention score matrix. For raw data, a small experiment is conducted with 25 participants who were asked to watch a short video lecture and

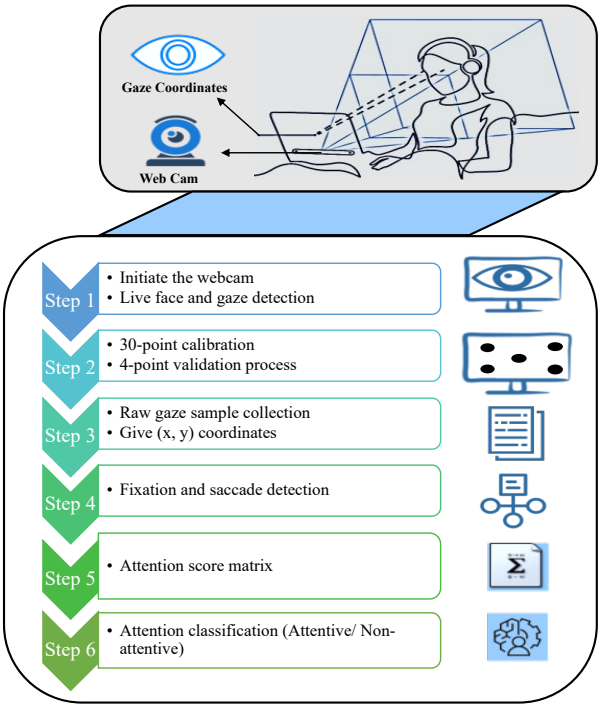


Fig. 1: System level architecture of gaze aware attention monitoring system

report distraction when occurs during the visual engagement. After the raw data filtering and feature extraction, a novel attention score matrix is proposed based on the gaze metric fixation and saccades to quantify the student attention. The system-level architecture of EXECUTE is shown in Fig. 1, and the details of each block are presented in the following subsections.

EXECUTE is a graphical user interface that captures the raw gaze data using an ordinary webcam. We build a web application using webgazer.js [16], an open-source eye-tracking library written in JavaScript which closely resembles the current massive open online course (MOOC) platform with additional capabilities. This platform is privacy-aware since it runs entirely on the client (student) browser without capturing any privacy-sensitive information. The data collected during the experiment session contains the raw 2D screen coordinates along with the timestamps. The entire setup is divided it six stages in which the first three stages are related to the interface setup for raw data collection, whereas the remaining deals with the gaze data processing and ML for automatic attention classification.

A. Setup initialization

The first step in attention monitoring setup is to run the EXECUTE interface using any web browser on the participant computer. Once the application is opened, it initializes the webcam after obtaining consent from the participant. After

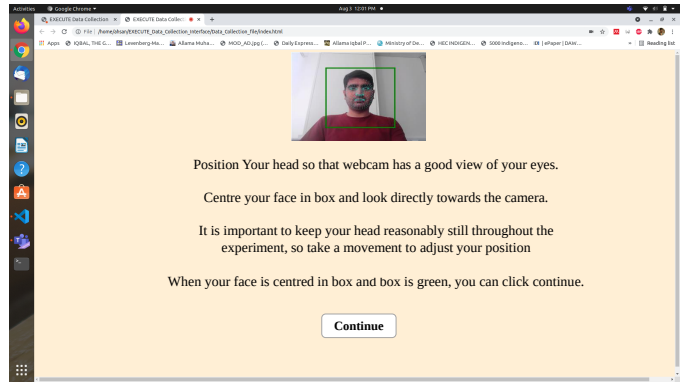


Fig. 2: EXECUTE GUI along with instruction to facilitate participants

webcam initialization, a small video container appears on the screen and live face and gaze detection are done. The video container is removed once the initial face and gaze detection is done to ensure user privacy. The GUI of EXECUTE is shown in Fig. 2.

B. Calibration process

Once the initial face and gaze detection is done, the next step is to perform the calibration and validation process. The process of calibration is done to estimate the location of the gaze point on the screen when engaging with visual stimuli. In this step, 30 points appear from the top left to the bottom right corner of the screen on a predefined location. The participants calibrate the system by clicking and seeing the point at the same time and estimate the gaze coordinate location using ridge regression. Once the calibration process is done, 4 point validation is carried to check the accuracy of the real-time gaze estimation. For this purpose, 4 points appear randomly on the screen for a short period of time and the participant just stares at the dots and measure the accuracy based on the pre-defined area of interest (AOI). If the accuracy for each point is above then 70%, the experiment is proceeding forward to the next step and re-calibration is done if accuracy is below 70%.

C. Raw data collection

After the calibration and validation process, the next step is the collection of raw data samples which are the 2D screen coordinates of gaze point along with the time stamp. The sampling rate achieved using the webcam interface is 30 samples per second. To collect the data-set, 25 students from our research group were asked to watch a short introductory video lecture with a background voice that closely resembles the MOOC lecture format. The video lecture consists of various slides with three pre-defined AOI per slide to simplify the process of data filtering and gaze feature extraction. The attention level is measured in a self-caught method where a low tone acoustic bell is included with a feedback button to report any distraction. This bell rang when the instructor finishes the content on each AOI and the participants were asked to report distraction using the feedback button. This

study also has pre-and post-study questionnaires along with the engagement feedback of participants.

D. Data filtering and feature extraction

In the data filtering process, the 2D raw gaze points were converted into eye movements using the dispersion-based filter with an open-source tool pyGaze [17]. In the data filtering process, three most commonly used gaze features i.e., fixation, saccades, and eye blinks were extracted the using pyGaze library. Fixations are defined as the consecutive gaze point within the range of 25 pixels for a time duration of more than 100 ms. It is the shortest time where an eye can focus during the visual engagement. On the other hand, the saccade is the duration between two consecutive fixations. Eyeblink is the duration where the eye tracker loses the track of gaze for a short during of time which ranges between 83 ms to 400 ms. To motivate this approach, Fig. 3 visualize the gaze features through fixation plots and heatmaps with ground truth labels for attention on a particular slide. These are the plots of one of the participants from this experiment in which he reported distraction in AOI 1 and this method is used as ground truth for labelling the attention level for each AOI.

After performing the gaze data filtering, different features are extracted based on the relationship of fixation, saccades, AOIs and scan path. Each slide contains pre-defined AOI as shown in Fig. 3a and features in relation to fixation and saccade are extracted using the time window operation for each slide. The extracted features include the number of fixation in AOI, percentage fixation in AOI, percentage fixation duration in AOI, duration of largest fixation, fixation backtracks, fixation dispersion mean, first fixation duration and entry time of the first fixation in AOI.

E. Attention score matrix

The attention score matrix is the measure of how much a student is paying attention to the content covered in the lecture. It is temporal information of the student gaze stops on particular AOI while engaging with the visual stimuli. The attention scores are captured for each AOI based on time window operation. It has two levels of interaction with visual stimuli i.e., attentive and non-attentive. The information content on a particular slide is covered sequentially, as a result, each AOI has a pre-defined start and end time. Using the timestamps and location coordinates of AOI, the attention of the student is linked with the percentage of fixation in AOI, percentage fixation time in AOI, number of fixation in AOI, the maximum duration of highest fixation, entry time of the first fixation, and dispersion spread of two consecutive fixations. Furthermore, to measure if the student is following the content is done by plotting the scan path for 10 seconds when the information transition from AOI 1 to AOI 2 as shown in Fig. 4. This figure is showing the fixation number and the arrowhead is joining the two consecutive fixations. It is the plot of an attentive participant as the fixation moves from AOI 1 to AOI 2 during this time window. This information is also linked with the participant diversion report given by him using the feedback button during the lecture. This process is used to label the participant data into classes i.e. attentive and non-attentive to make it a supervised ML problem.

F. Attention classification using ML

To make the process of attention classification automatic, we converted it into a binary supervised classification problem. Before model training, multiple time-bounded features based on each AOI are extracted in relation to fixation and saccade. For data labelling, we derived an attention score matrix with the help of participant feedback and time-bounded feature and label the data of each participant in two classes i.e. attentive and non-attentive. To select the suitable classifier for attention classification we considered two requirements:

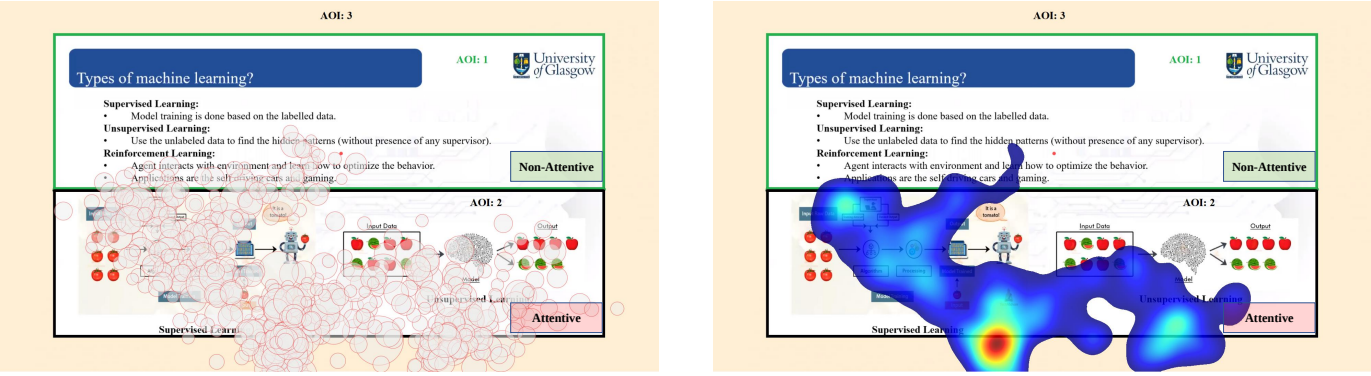
- The trained model has better parameter generalization which can be effectively used to infer attention classification in data of unseen students
- The trained classifier has low inference time so it can be easily integrated with the web application for real-time attention classification.

As this is the pilot study so we conducted this experiment with 25 students and have to consider a bias-variance trade-off of ML and the data size of our experiment. We used 7 classifiers and our requirements were met by logistic regression, SVM and polynomial regression. This result was expected because we have a very small dataset and the more advanced ML model tends to overfit due to the smaller training size. Furthermore, logistics regression, SVM and polynomial regression have very few inference steps, so they can easily be integrated with web applications.

IV. Experimental Setup

In this section, we will discuss the experimental setup to validate of results. Before the start of the experiment, each participant is briefed on how to use the interface and report distractions during the lecture. Furthermore, we also provide additional guidelines for each step in the interface. For gaze data collection, the participants were asked to sit in front of a laptop screen at the distance ranging from 50 to 60 cm. We used a standard laptop webcam with a screen size of 15.6 inches with a resolution of 1920 x 1080 to collect gaze location on the screen. Each participant was asked to watch a short video lecture of eight minutes and report distraction using the feedback button at the end of content from particular area of interest. A small acoustic bell was included in the experiment to notify the participant when information transitions between the AOIs.

In the whole experiment, no privacy-sensitive data like video or images of the participant is captured. This interface only collects the 2D coordinate of eye movement along with the time stamps estimated by webgazer.js and the open-source library to estimate gaze position. One of the most challenging tasks of gaze movement capturing using a webcam was the low sampling rate. In this experiment, we achieved the average sampling rate of 29 samples per second which is almost half of the commercial eye tracker like tobii pro. To overcome



(a) Fixation plots on slide with attention label

(b) heatmaps of gaze data on slide with attention label

Fig. 3: Fixation plot and heatmap of gaze data along with distraction report in AOI 1. The AOI is represented by the bounding box around the lecture content. In this figure their are 3 AOI i.e. green box represents AOI 1, black box AOI 2 and outside the slide is the AOI 3

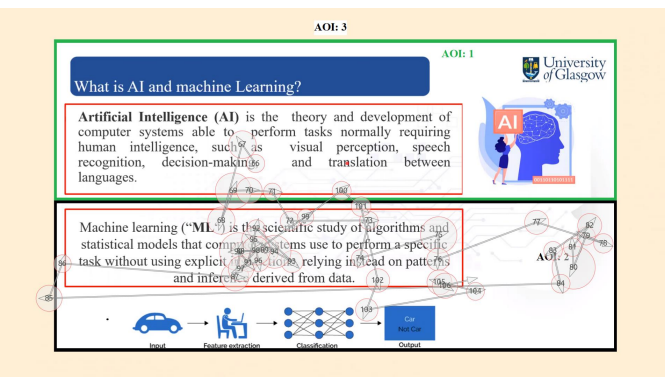


Fig. 4: Scan path gaze during the transition of information

this challenge we extracted the time-bounded features and link them with the distraction report to label the data of each participant. For automated attention classification, we trained 7 different classifiers and used grid search for hyperparameters and evaluate performance in terms of accuracy, precision, recall and F1 score.

A. Performance metrics

The baseline performance metric for attention classification is accuracy, however, we have added precision and recall for each class to avoid the accuracy paradox which means the classifier with low predictive power has a tendency to report higher accuracy. For instance, when the number of true positives is less than false positives, the accuracy of the model increases by changing the classification rules. Similarly, when the number of true negatives is less than false negatives, the same will happen by changing the classification rules. Therefore, precision-recall and F1 scores are the additional metrics to evaluate the performance of the classifier. These performance scores are given by

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

and

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

where TP means true positive, TN true negative, FP false positive, and FN false negative.

V. Results and Discussion

In this study, we converted the attention classification into a supervised binary classification where the labelled data is used to train the ML models. The accuracy is the first metric to evaluate the performance of the trained model. As we have a very smaller data-set, we used 10 fold cross-validation method to provide a better generalization capability to the classifier. The accuracy of the 7 classifiers used in this study is given in Tab. I. From the results, it is found that the logistic regression, SVM and polynomial regression have the highest accuracy of 91.05%, 92.6% and 89.5% respectively. These results were expected as the simple classifier has low variance compared to others and provide a better fit for a smaller data-set as ours. Further, the confusion matrix of the top 2 classifiers is shown in Fig. 5 in which the diagonal gives the accuracy of each class. It is used to compute another performance metric like precision, recall, and F1-score to avoid accuracy paradox as discussed in Section IV-A. It is clearly evident for results that classifiers with low variance perform much better in our case. However, we believe that a more complex ML model can provide better generalization but it requires a large amount of data for model training. Furthermore, the adoption more complex ML model might limit the scalability of the system as it requires a short inference time for real-time attention monitoring.

From this experiment, we found that a robust ML can be developed to classify the student attention using a webcam for a remote learning setting. Using the setup, the teacher can monitor the attention level of students and intervene when needed. The system can be deployed in the real classroom

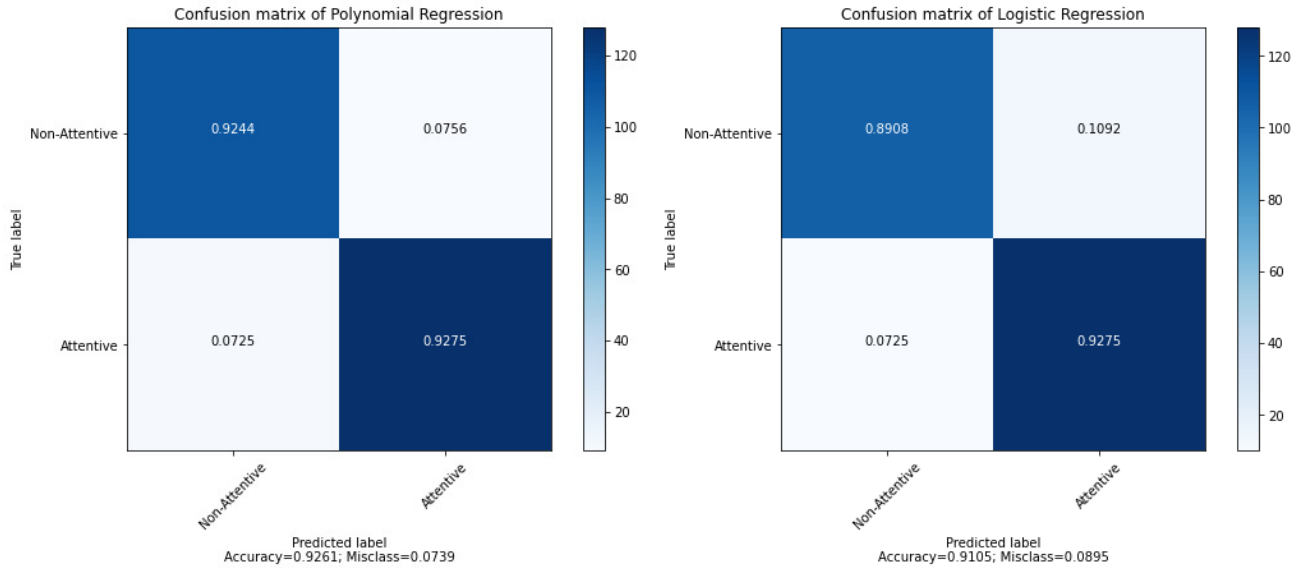


Fig. 5: Confusion matrix for top 2 classifiers

TABLE I: Results of ML classifiers

Classifier	Accuracy	Precision	Recall	F1-Score
Polynomial Regression	92.6	0.91	0.89	0.92
Logistics Regression	91.05	0.89	0.91	0.88
SVM	89.94	0.89	0.90	0.88
Decision Tree	80.02	0.81	0.79	0.80
Random Forest	85.22	0.84	0.81	0.85
Extra Trees	82.22	0.83	0.81	0.85
Multi-layer Perceptron	85.64	0.82	0.85	0.83

without any requirement of additional equipment which is an expensive and unsuitable remote learning setup. Another advantage of this system is the privacy by design where the webcam only captures the 2D gaze coordinates along with time stamps. Furthermore, the gaze movements are linked with the psychological and emotional state of individuals. Therefore, this study can be enhanced to identify the anomalous pattern of gaze movement which can leads to the early detection of ADHD and their associated symptoms like the anxiety and depression.

VI. Conclusions

Attention monitoring is very crucial for teaching and learning. An attention monitoring system that can detect the engagement level of students and redirect their attention could be very beneficial in remote learning. Attention monitoring, however, required visual feedback which is absent in online learning. In recent times, eye tracking is used to measure attention in the academic environment using expensive commercial-grade eye-tracking devices which limits the study in labs only. In this paper, we presented a scalable and privacy-aware attention classification system using a standard webcam. In this setup, gaze data collection is done in a non-invasive manner without any specialized equipment. Furthermore, we presented a novel attention score matrix based on the fixation, saccades, scan path in conjunction with the pre-defined AOI to quantify the level of attention. This experiment offers a unique opportunity to develop a large scale remote learning study to analyze and evaluate the visual perception for new perspectives to improve the instructional design of multimedia learning and teaching. As future work, this study will be enhanced for early detection and diagnostics of ADHD by identifying and analysing the anomalous gaze pattern and providing assistance by recognizing their individual needs. The Universities can also provide made-to-measure educational methods to assist students with ADHD with exceptional help to ease their education, provide permission to record the lecture, extra time on the exam, and help with any symptoms that usually accompany ADHD, such as anxiety.

References

[1] A. Vehlen, I. Spenthof, D. Tönsing, M. Heinrichs, and G. Domes, "Evaluation of an eye tracking setup for studying visual attention in face-to-face conversations," *Scientific reports*, vol. 11, no. 1, pp. 1–16, 2021.

[2] A. F. Abate, L. Cascone, M. Nappi, F. Narducci, and I. Passero, "Attention monitoring for synchronous distance learning," *Future Generation Computer Systems*, vol. 125, pp. 774–784, 2021.

[3] K. K. Loh, B. Z. H. Tan, and S. W. H. Lim, "Media multitasking predicts video-recorded lecture learning performance through mind wandering tendencies," *Computers in Human Behavior*, vol. 63, pp. 943–947, 2016.

[4] W. Wong, M. Bartels, and N. Chrobot, "Practical eye tracking of the ecommerce website user experience," in *International Conference on Universal Access in Human-Computer Interaction*. Springer, 2014, pp. 109–118.

[5] D. Rojas-Líbano, G. Wainstein, X. Carrasco, F. Aboitiz, N. Crossley, and T. Ossandón, "A pupil size, eye-tracking and neuropsychological dataset from adhd children during a cognitive task," *Scientific data*, vol. 6, no. 1, pp. 1–6, 2019.

[6] J. Madsen, S. U. Julio, P. J. Gucik, R. Steinberg, and L. C. Parra, "Synchronized eye movements predict test scores in online video education," *Proceedings of the National Academy of Sciences*, vol. 118, no. 5, 2021.

[7] N.-H. Liu, C.-Y. Chiang, and H.-C. Chu, "Recognizing the degree of human attention using eeg signals from mobile sensors," *sensors*, vol. 13, no. 8, pp. 10 273–10 286, 2013.

[8] N. Dietrich, K. Kentheswaran, A. Ahmadi, J. Teychené, Y. Bessière, S. Alfenore, S. Laborie, D. Bastoul, K. Loubière, C. Guigui *et al.*, "Attempts, successes, and failures of distance learning in the time of covid-19," *Journal of Chemical Education*, vol. 97, no. 9, pp. 2448–2457, 2020.

[9] K. K. Loh, B. Z. H. Tan, and S. W. H. Lim, "Media multitasking predicts video-recorded lecture learning performance through mind wandering tendencies," *Computers in Human Behavior*, vol. 63, pp. 943–947, 2016.

[10] C. Mills, R. Bixler, X. Wang, and S. K. D'Mello, "Automatic gaze-based detection of mind wandering during narrative film comprehension." *International Educational Data Mining Society*, 2016.

[11] C.-M. Chen, J.-Y. Wang, and C.-M. Yu, "Assessing the attention levels of students by using a novel attention aware system based on brainwave signals," *British Journal of Educational Technology*, vol. 48, no. 2, pp. 348–369, 2017.

[12] P. Pham and J. Wang, "Attentivelearner: improving mobile mooc learning via implicit heart rate tracking," in *International conference on artificial intelligence in education*. Springer, 2015, pp. 367–376.

[13] C. Mills and S. D'Mello, "Toward a real-time (day) dreamcatcher: Sensor-free detection of mind wandering during online reading." *International educational data mining society*, 2015.

[14] K. Rayner, "Eye movements in reading and information processing: 20 years of research." *Psychological bulletin*, vol. 124, no. 3, p. 372, 1998.

[15] S. D'Mello, J. Cobian, and M. Hunter, "Automatic gaze-based detection of mind wandering during reading," in *Educational Data Mining 2013*, 2013.

[16] A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays, "Webgazer: Scalable webcam eye tracking using user interactions," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*. AAAI, 2016, pp. 3839–3845.

[17] E. S. Dalmaijer, S. Mathôt, and S. Van der Stigchel, "Pygaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments," *Behavior research methods*, vol. 46, no. 4, pp. 913–921, 2014.