Non-Intrusive Classroom Attention Tracking System (NiCATS)

Andrew Sanders
Computer Science
Georgia Southern University
Statesboro, Georgia
as13770@georgiasouthern.edu

Bradley Boswell
Computer Science
Georgia Southern University
Statesboro, Georgia
bb05758@georgiasouthern.edu

Gursimran Singh Walia
Computer Science
Georgia Southern University
Statesboro, Georgia
gwalia@georgiasouthern.edu

Andrew Allen
Computer Science
Georgia Southern University
Statesboro, Georgia
andrewallen@georgiasouthern.edu

2021 IEEE Frontiers in Education Conference (FIE) | 978-1-6654-3851-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/FIE49875.2021.9637411

***Abstract*—This Innovative Practice Full-Paper presents a system for real-time accurate detection of classroom attentiveness using monitor-mounted webcams and eye trackers. Academic institutions and instructors cannot accurately assess the moment-to-moment attentiveness of students in classrooms where students' faces are obscured by computer monitors. This can cause the lectures of Computer Science, Information Technology, or other lab-based courses to be incorrectly paced, which leads to students having overall poorer grasps of the subject material. We present a system for accurate detection of classroom attentiveness using monitor-mounted webcams and eye trackers. To determine correlations for the attentiveness judging system, we compare an initial attentiveness score produced by trained labelers using an image of the student's face with a series of calculated eye metrics to determine a final attentiveness score. Because the student webcam images and eye coordinates are synchronously collected with the lecture, this final attentiveness score is used to provide post-hoc feedback to instructors on the status of their students via time-series graphs displayed on the instructor's computer monitor. The proposed system is invaluable for institutions seeking to improve student education, instructors striving to improve the flow of lectures, and students seeking a more accommodating learning environment. The primary source of innovation from this system comes from the correlation of extracted eye metrics with the face images labeled for attentiveness. Research exists about determining attentiveness using a convolutional neural network trained on face images and even determining attentiveness by correlating face-image-trained outputs, each of which we plan to incorporate to make our system real-time in the future. This novel research could prove helpful for the field of education.**

Keywords—Attention, Eyetracking, Eye Metrics, Education, Engagement

I. Introduction

Instructors know that effective instruction involves maintaining student attention. Research [8] has shown that attention levels in students generally decrease in learning environments, such as lectures, after about 10-30 minutes. While active learning techniques such as task-related activities can be used to break up a lecture to engage students [8] actively, the ability to accurately track the level of students' attention is an open research problem. This is especially true in Computer Science or other lab settings where computer monitors obscure student's faces).

The ability to accurately track the attention levels in real-time can guide the instruction design and lecture delivery, thus ensuring that objective feedback on students' engagement is being relied on. Attentiveness and engagement are often used synonymously and often in a competing manner. It is important to note that *attention* refers to the short-term application of the mind towards a topic. In contrast, *engagement* refers to the overall emotional commitment of a person's mind towards a subject. Though they are defined slightly differently, they are tracked and treated similarly in research designed to assess students' facial expressions visually. Our research focuses on measuring the perceived visual attentiveness of students, which can be used to track student engagement. We share the view [5] that professors use the perceived visual facial expressions to proxy for the underlying feelings and attitudes a student holds. Hence, a system designed to detect engagement automatically is desirable.

Automatically determining a student's attentiveness using machine learning and computer vision is a growing topic discussed and used in [2] and [5]. Additionally, the location and general metrics (for example, average fixation duration) of a person's eyes generally reflect the intent of the person's mind [1], which is a potentially valuable source of information regarding the tracking of attention. Using eye gaze data for assessing the engagement of students has been discussed in [4]. To our knowledge, there did not exist a dataset on facial images labeled by attentiveness. We generated our dataset by having participants sit through a video lecture while recording their facial images (which are labeled for attentiveness), eye gaze data (which are used for correlation analysis with the labeled facial images), and screenshots (which are used to generate heatmaps for use in comprehension analysis). Using the eye gaze data, we calculate fixations (the stabilization of the eye on the part of a stimulus for a period of time (200-300 ms) [1]), saccades (the quick and continuous eye movements within 40-50 ms from one fixation to another [1]), and specific metrics of both (such as average fixation duration; the rest are listed in III.B.b) Eye Data Extraction), which are used for correlation analysis with attentiveness. This dataset is being used to assess the usefulness of eye metrics in determining a student's attentiveness. It would be useful for educators to track students' attentiveness based on their facial image combined with their eye metrics. This work is a further iteration of "*Non-intrusive Identification of Student Attentiveness and Finding Their Correlation with Detectable Facial Emotions"* which uses facial images in conjunction with cloud-based emotion recognition to assess the role of emotions in determining attention.

In this paper, we propose and present a Non-intrusive Classroom Attention Tracking System (NiCATS). Its purpose is to provide post-hoc log-file analysis (with real-time analysis in the future) on the overall level of ***attention*** the students in a classroom are displaying. To validate the usefulness of the NiCATS system, a controlled experiment was conducted where we used non-invasive monitor-mounted eye trackers and webcams to collect information about students while they attended the lecture. We collected the facial images of subjects (every five seconds so they could be labeled for attentiveness) and compared them with a set of calculated eye metrics (generated from recorded eye gaze data) to check for underlying correlations. The results are strong. The advantage of using this approach is the lack of observational effect due to the passive nature of the system while having the similar attention-judging accuracy as domain expert humans. The passive nature of the system is aptly important as it can be applied to online classes that are ubiquitous as a result of the COVID-19 pandemic.

The remainder of the paper is organized as follows. In Section 2, we present an overview of the literature relevant to the automatic and semi-automatic measuring of attentiveness and the use of eye data in measuring attentiveness. Section 3 covers the design and methodology of NiCATS followed by the experiment design details in Section 4. Section 5 provides an analysis of the data collected during the experiment and is organized around the major research questions. Section 6 provides a discussion of study results followed by conclusion and future work in Section 7.

II. Literature Review

This section discusses the most relevant literature that motivated our work on measuring student attentiveness.

Veliyath et al. [4] used eye trackers to determine engagement of students in a computer lab environment. They concluded that eye gaze data can be used as a basis for a machine learning predictive model but their model lacked accuracy, perhaps due to self-reporting of student engagement. They also concluded that it may be worth utilizing the gaze data in conjunction with other features to better understand student engagement. While our work is focused on student attentiveness (at different points), it extends Veliyath's work by using gaze data in conjunction with labeled face images to produce better accuracy than gaze data alone.

Analysis of facial expression to recognize student engagement is not now. For example, Whitehill et al., [5] explored the development of real-time automated engagement assessment using students' facial expressions. Through the labeling of short 10-second clips, they trained a machine learning model to predict engagement with modest accuracy. They concluded that it is possible to develop real-time systems for judging engagement with similar accuracy to humans. They also found that while interobserver reliability is highest with 10-second clips, static expressions contain the majority of information needed to assess engagement. In our research, we are labeling facial images at every 5-second interval. Instead of a range of attentiveness (used in prior work), we are focussing on binary (attentive vs. inattentive) labels.

Our research does not currently use a machine learning model to predict engagement from facial expressions, but it remains a possible future step in our research. In terms of knowledge acquisition, researchers have often used interviews, observational assessments, and self-assessment when using eye trackers [11]. Based on the extensive literature review, while useful, self-reporting instruments are not capable of capturing the fluctuations in the cognitive information processing at different points as students acquire and process information. The review of prior findings motivated the need for a pre-post test that includes an objective assessment of the content being presented to students.

While there has been some prior work on using eye trackers to redesign the instructional strategies, Lai et al., [10], after a review of relevant literature focused on eye tracking technology, concluded that current research is inadequate and more work is needed to understand the knowledge acquisition can be garnered from eye tracking metrics to guide the instructional re-design. Our work is a step in that direction where we are utilizing eye tracking and face images as they correlate to the student's attentiveness to better understand the patterns of information processing exhibited in extreme cases (well understood vs hard to comprehend content areas).

In [7], saccadic eye movement was analyzed that provided useful insights into the relevance of saccades as it relates to student alertness. The results showed that velocity/duration ratio, normalized peak velocity, and normalized duration of saccadic eye movement were heavily correlated with brain wave amplitude, sleepiness, and subjective alertness. They conclude that the dynamics of saccadic eye movement can be used to assess alertness. While our work is focused on attentiveness and knowledge acquisition, these findings motivated the need to study eye metrics related to saccadic eye movement when assessing student attentiveness. In a similar vein, the researchers in [6] found that the peak velocity and duration of saccades had a high correlation with the time rate at which information is perceived within the brain through the retina along with the relative pupil diameter. In other words, when the peak velocity and duration of saccades are low, the perceptual performance of the subject is also low. This further reinforces the idea that metrics related to saccadic eye movement are worth studying (and are expected to be negatively correlated with attentiveness).

Tabassum et al. used computer vision, machine learning, and cloud-based facial emotion recognition to assess the attentiveness of a student and to find the relationship between attentiveness and emotions [2]. Their proposed methodology only used facial images in predicting student attentiveness. They also found facial emotions are strong predictors of attentiveness in students. This provides a strong foundation for enhancing our NiCats system by automating the face image labeling process and extending the prior work by analyzing eye tracking metrics and screenshots of information that is being presented to students.

III. NiCATS Design and Implementation

The goal of the Non-Intrusive Attentiveness Tracking System (NiCATS) is to provide instructors with real-time feedback on the attentiveness of students in their classroom. As a first step, this research explores different aspects of students' attentiveness and comprehension (e.g., facial images, eye movements) to understand how these inputs can be used to train a machine learning model (inbuilt in NiCATS) for predicting student attentiveness. To provide an overview, NiCATS utilizes webcams and eye trackers that can be mounted on student machines to collect webcam images, gaze points, and screenshots that can be analyzed for understanding student attentiveness and comprehension. To that end, Figure 1 shows a high level overview of NiCATS in terms of data collection, preprocessing and analysis (discussed in subsequent sections). While this paper is focussed on post-hoc analysis of data collected during the experiment, instructors and institutions can use this system to tailor their lectures and study material to best engage the students.

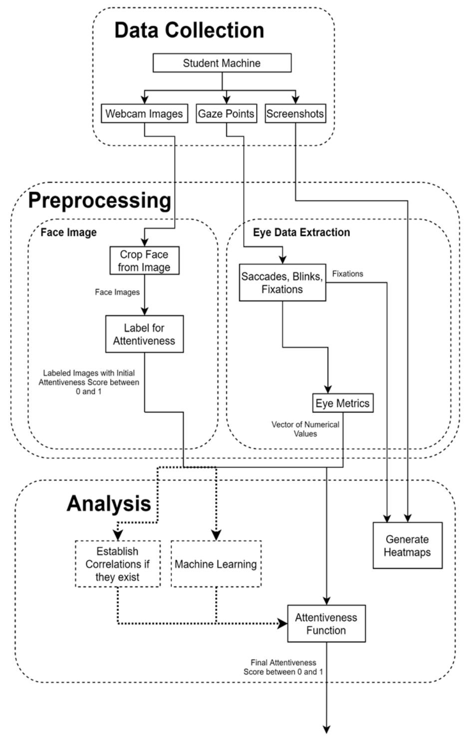

Fig. 1. High-Level Design of NiCATS

A. *Data Collection*

The NiCAT system is set up to capture three distinct data types about the student from the student's machine.

- Facial image - A computer monitor-mounted webcam periodically captures and sends images of the student's face to the server at a 5-second interval. A continuous video stream of the student's face could be collected, but this provides a marginal benefit compared to static images when determining attentiveness [5].
- Screen Capture - A screenshot capture will be triggered whenever the student interacts with their machine via keyboard/mouse input or, in the case where no input was received after the pre-defined time interval of 15 seconds.
- Eye movement - The gaze points of the student, with regard to their computer monitor, are captured continuously throughout the lecture. Anytime a screenshot is prepared to be sent to the server, we prepare the most recent chunk of gaze points since the last screenshot was transmitted to be sent along with the screenshot for pre-processing.

B. *Pre-Processing*

The preprocessing for each data item collected (facial images, eye movements, and screenshots) is explained in the following subsections. Each preprocessing step describes design decisions (e.g., how to best label images, how to create regions of interest) to be able to use the system in real-time and analyze the collected data in a post-hoc manner.

a) *Face image:* To capture isolated face images of the student for labeling, we crop a smaller image from the original, which contains only the student's face, by using Haar cascade classifiers [2]. To generate a set of attentive and inattentive images for comparison with the extracted eye metrics, we asked multiple labelers to label the face images based on the validated Behavioral Engagement Related to Instruction (BERI) protocol [3]. The human labeling was handled via the NiCATS mobile web application which allows human-labelers to swipe images right or left on their mobile devices to label images as being attentive or inattentive respectively. From the labeled image set, we arrive at the attentiveness score by summing all "attentive" labels on an image and dividing it by the total number of labels (attentive or inattentive) a face image receives.

(Swipe Left - Inattentive) (Swipe Right - Attentive)

Fig. 2. Attentiveness Labeling Mobile App

b) *Eye Data Extraction:* The gaze points are pre-processed to extract relevant eye metrics (e.g., fixation, saccades) that can be used to predict student attentiveness and comprehension. Using the gaze points, we calculate fixations (the stabilization of the eye on the part of a stimulus for a period of time (200-300 ms) [1]) and saccades (the quick and continuous eye movements within 40-50 ms from one fixation to another [1]). A subset of eye metrics (relevant to this work) that can be collected from fixation and saccade calculations are listed below:

- Fixation Count (total # of fixations). This can be collected for the entire lecture period or for specific lengths of time.
- Average Fixation Duration: This is measured by adding the durations of all fixations divided by the number of fixations.
- Number of fixations per second: Total number of fixations divided by the total duration of a recording session.
- Saccades Count (total # of saccades): This can be collected for the entire lecture period or for certain lengths of time.
- Average Saccade Duration: This is measured by adding the durations of all the saccades divided by the number of saccades.
- Saccades per second: Total number of saccades divided by the total duration of a recording session.
- Heat map: Visualization of students' gaze points superimposed on a particular area of interest.

The collection of eye metrics will then be compared with the results of the human-labeled face images to determine if any correlations exist between the eye metrics and a student's attentiveness level for that interval of time.

IV. Experiment Design

This section provides an overview of the experiment design, including research questions, variables, study participants and artifacts used, experiment procedure, and data collected during the experiment run.

A. *Research Questions*

Our experiment is focused on understanding how student attentiveness and knowledge acquisition can be measured using eye-tracking (e.g., gaze points), face images, and screenshots of the information provided to them. The following research questions were investigated during this experiment:

1. How can eye-tracking and classroom observational data be used to measure student attentiveness?
2. Can the information extracted from NICATS provide feedback to the instructor on students' ability to process the information presented to them?

B. *Independent and Dependent Variables*

a) *Independent Variables:* The following independent variables were manipulated

- Eye metrics: The eye metric data (e.g., fixations, saccades) collected during the experiment varied for different subjects.
- Number of screenshots: The number of screenshots varied depending on the user interaction during their recording session.

b) *Dependent Variables:* We measured the effect of independent variables on the following dependent variables:

- Attentiveness: The attentiveness scores were calculated that ranged between 0 (inattentive) and 1(attentive), representing the level of perceived attentiveness of the student. The face image data collected during the experiment was used to calculate the attentiveness score.

- Knowledge acquisition**:** Individual scores on the pre and post-test for each participant were compared to understand their knowledge acquisition.

C. *Subjects/Participants/Environment*

Computer science undergraduate and graduate students participated during the experiment run. Students volunteered to participate in the study. Before the actual experiment run, one of the researchers tested the NiCATS set-up to ensure that the data was being collected correctly. During the actual experiment, each experimental subject reviewed the presentation (on software errors) that was pre-recorded.

To simulate a computer lab-based classroom environment, the computers used for the experiments by the students were equipped with a 1080p webcam and a Tobii Eye Tracker 4c. The webcam was mounted in the top-center region of the student's computer monitor, while the eye tracker was mounted to the bottom-center region. The student computer had an i7-4790 CPU and 4GB of RAM, which is adequate for running the data collection program. Figure 3 presents the experimental setup.

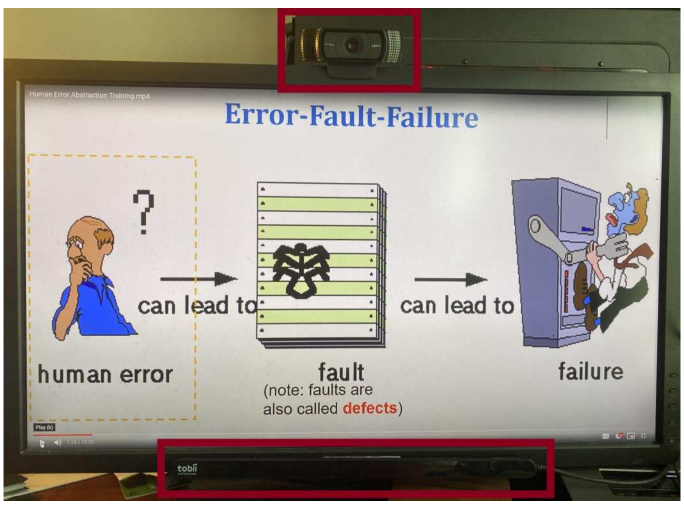

Fig. 3. Webcam and Eye Tracker mounted

D. *Artifacts and Instrument*

We used a 15-minute recorded lecture regarding Human Errors and their applications in everyday life. This lecture was selected because it was generic enough that the prior knowledge of CS subjects would not have a big impact on their engagement and knowledge acquisition. This video was prepared as part of an REU grant (by someone external to the research team - mitigating researcher bias) and was used to train the general public on the significance of human errors. The lecture included slides with varying font sizes, information, and visual aids to understand students' attentiveness patterns for different information types.

E. *Experiment Step*

a) *Step 1:* The students were instructed to download the NiCATS client software to their machine, along with the Tobii Eye Tracking Core Software. Launching the NiCATS software prompts the student with a message describing the data that will be collected during the recording session and will allow them to "opt-in" or "opt-out" of the experiment. Agreeing to participate in the experiment by clicking "opt-in" initiates the data collection features of the NiCATS client until the recording session has ended.

b) *Step 2: Pre-test:* To understand the baseline of student knowledge on the topic that was presented to them, they were asked to take a small test. The pre-test included ten questions that covered questions related to content that was presented to them.

c) *Step 3:* Collecting Data during the experiment run: During this step, the researchers started the NiCATS session, as we instructed the students to begin watching the recording of the selected lecture. Throughout the session, raw data (related to their eye metrics, facial expressions, screenshots) was sent and stored on the server. The raw data collected during this step was later analyzed for calculating relevant eye measurements).

d) *Step 4*: At the conclusion of the lecture, the recording session was ended by the researcher, followed by the administration of the post-test. The post-test included the same questions that were asked during the pre-test. This step allowed researchers to be able to compare students' knowledge acquisition.

F. *Evaluation Criteria*

To gain insights into the relationship between the independent variables and measuring their impact on the dependent variable, we calculated the following using the data of the independent variables:

- *Calculation of relevant eye-metrics*: Based on the gaze points (the coordinates of eye movements), *fixations* and *saccades* can be calculated at different times. To calculate fixations (the stabilization of the eye on the part of a stimulus), we used a threshold of 200-300 ms as reported in the literature [1]). Similarly, the saccades were calculated based on the literature findings(the quick and continuous eye movements within 40-50 ms from one fixation to another) [1]).

- *Pre-processing of Screenshots*: While the same lecture recording was presented to the students, the screenshot capture (of the screen) varied depending on the user input (e.g., mouse click). Also, the NiCATS is programmed to periodically capture the screenshot after every 15 seconds in the absence of any user activity.

- *Facial Images*: Similar to the screenshots, face images at 5-second intervals were collected. Each facial image stored on the server is then labeled for attentiveness using the NiCATS mobile web application (refer

III.B.a)). Labellers use [3] as a guide for determining attentiveness.

TABLE I. PRESENTS THE AVERAGE OF THE MOST RELEVANT INDEPENDENT AND DEPENDENT VARIABLES COLLECTED DURING THE EXPERIMENT RUN. THE ANALYSIS OF THESE VARIABLES IS PRESENTED IN THE NEXT SECTION.

Data Item	**Calculation Method**	**Avg.**
Student attentiveness	For each student, all the “attentive” labeled face images were divided by the # of images (attentive/inattentive).	0.801
# Fixations	Average # of all fixations (summed up over the entire recording session)	1241
Fixations per second	Average # of all fixations (summed up over the entire recording session) divided by the total lecture time.	1.240
Average fixation duration (ms)	All fixation durations are summed up and divided by the number of fixations and then averaged per participant	510.5
# Saccades	All saccades are summed up over the entire recording session and then averaged per participant	2291
Saccades per second	The total number of saccades divided by the session length	2.342
Average Saccade Duration (ms)	All saccade durations are summed up and divided by the number of saccades	288.6
Pre-Test Score	Questions answered correctly out of 10	4
Post-Test Score	Questions answered correctly out of 10	8

V. RESULT AND ANALYSIS

The following section provides an analysis of data collected during the experiment run. The section is organized around two research questions.

a) *RQ 1*: How can eye-tracking and classroom observational data be used to measure student attentiveness?

To provide an overview of the results, Figure 4 provides an analysis of average fixation (left side of the Y-axis) and average saccade durations (right side of the Y-axis) for one subject throughout the entire lecture. The X-axis represents the screenshot for which the # fixations and saccades are being averaged upon. These averages were calculated for all 70 screenshots (as shown in Figure 4). Figure 5 provides the attentiveness score (based on the labeling of face images) for this subject for all 72 screenshots.

The following can be observed from Figure 4 and Figure 5:

- The subject demonstrated varying levels of fixation and saccades at different points of the lecture session. This pattern was consistent for each student and can be superimposed to understand the average fixations and saccades for the whole subject population. These peaks and troughs can be investigated further in terms of their correlation with student attentiveness (discussed later).
- There is no visible direct relationship between fixation and saccades. Average high fixation duration appears loosely correlated with lower average saccade duration. In general, saccades are associated with a large shift in gaze point coordinates, while fixations are associated with little to no change in gaze point coordinates over a duration.

To better understand the relationship between these eye measurements (fixation and saccade durations) vs. attentiveness at each screenshot, correlation analysis was performed for each student as well as across all student populations.

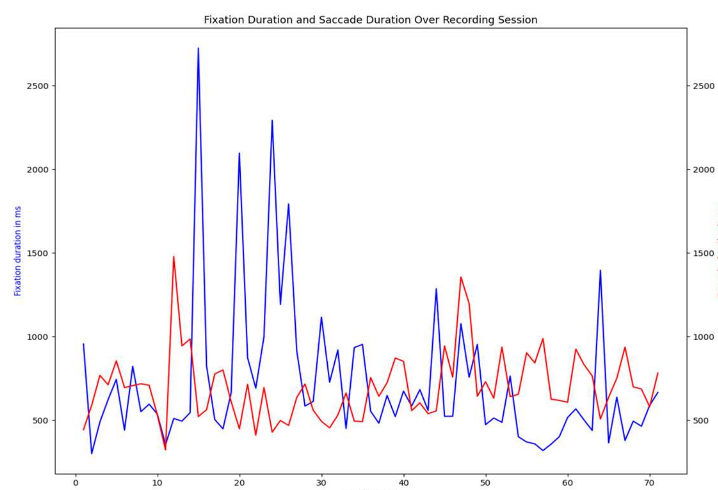

Fig. 4. Average Fixation and Saccade durations

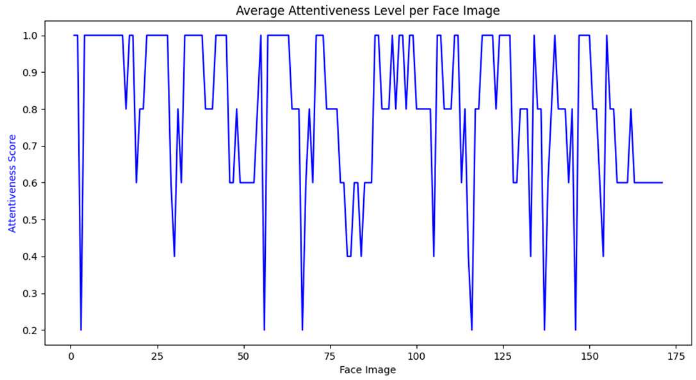

Fig. 5. Average attentiveness at different times

We analyzed the correlation between each independent variable (all eye metrics) and student attentiveness in terms of the correlation coefficient**,** the strength of correlation (R^2), and the p-value. Based on these analyses, the most strongly and statistically significant results are listed below:

- A*verage fixation duration* has a strong positive correlation with the student attentiveness (correlation coefficient = +0.0009; R2 = 0.84 and p-value of 0.028). This is expected as the fixation intervals have been shown to be when visual information is acquired [9].
- Average *saccade duration* is negatively correlated with student attentiveness (correlation coefficient = -0.000687; R2 = 0.87 and p-value of 0.019). Saccade duration represents the time spent searching for a visual target. We can infer from this correlation that longer durations would result in less visual information acquired.
- Number of saccades has a strong *positive* correlation with the student attentiveness (correlation coefficient = +0.00007665; R2 = 0.85 and p-value of 0.026). Since the length of the session was fixed, the larger number of saccades would translate to a lower average saccade duration (that has been shown to affect student attentiveness negatively).

Overall, based on these results, there is sufficient evidence to suggest that "time spent" fixating on or scanning for information can be valuable inputs when training a machine-learning algorithm to predict student attentiveness (the next logical step in this research).

b) RQ 2: Can the information extracted from NICATS provide feedback to the instructor on students' ability to process the information presented to them?

To answer this question, we performed qualitative analyses of gaze points for students (at individual and across the entire population) for specific slides based on the questions posed in the pre/post-test. To provide an example of the results, Figure 6 displays a heatmap (of students that gained knowledge) of the gaze points for an individual student when viewing a slide that contains a directly quoted question previously answered incorrectly by the student on the pre-test. By viewing the heatmap overlay on the multiple-choice question, "A physician misdiagnosing a patient when faced with an unfamiliar clinical situation," we can see the student had fixated on the answer, "*Mistake*," and the student proceeded to answer this question correctly on the Post-Test.

Information from overlain heatmaps like this can provide feedback to the instructor on the effectiveness of their presentation slides. In combination with a pre/post-test, eye metric data, and facial images, educators can determine where students are looking on a given slide and decide if the information is presented to be advantageous for learning. A possible application of this is deciding to reword or rearrange a slide if students' eye gaze data focuses on a point of information that is commonly misunderstood.

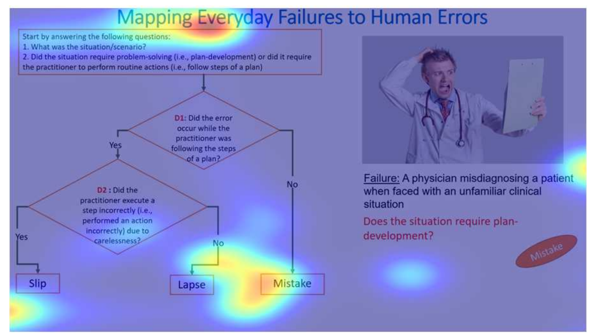

Fig. 6. Heatmap for an individual subject that acquired knowledge (from pre to post)

Similarly, Figure 7 shows a student that showed no improvement (from pre vs. post-test) and answered the question, "What type of requirement fault is the following: Some information in the software artifact contradicts information in the requirements document or the general domain knowledge" incorrectly on pre and post-test. Viewing the heatmap, we can see that the student briefly glanced at the content and subsequently answered the same question on the post-test incorrectly again. Furthermore, the bottom three rows of data displayed on the lecture slide were not viewed by the student, indicating the lack of engagement. While the concentration of gaze points at the periphery may suggest that students are fixating, no new knowledge is acquired.

It is important to note that gaze point concentrations are related to the number of fixations/saccades (non-significant metric based on our analysis in Section 5, *RQ 1*). This also provides more evidence to support our findings that the average duration of fixations/saccades is more critical when measuring student attentiveness and, especially, their knowledge acquisition. This is a novel finding *that average durations of* *fixations/saccades in conjunction with facial images are a bigger determinant than the face image expressions alone.*

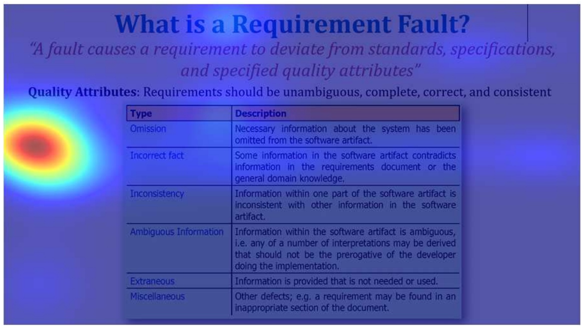

Fig. 7. Heatmap for an individual subject that did not gain knowledge (from pre to post)

While figures 6 and 7 generated heatmaps for individual students, this analysis can be extended for the entire student population as shown in Figure 8 (as shown below). To provide some context to the slide selected for heat map analysis, the slide in Figure 8 was where most students gained knowledge from the presented material. In the pre-test question relevant to this slide, students repeatedly answered incorrectly at a correct score rate of 20%. In the post-test question, students answered correctly at a much higher rate of 60%. Analyzing slides like these would be very beneficial to instructors in determining the best way to present classroom material.

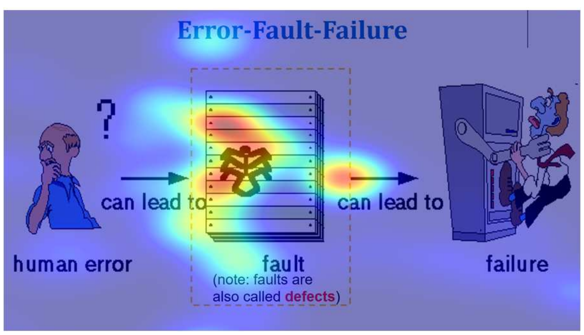

Fig. 8. Heatmap for all subjects

VI. Discussion of Results

Based on the results, Average Saccade duration, Average Fixation duration, and Number of saccades all independently show a strong relationship with attentiveness. Average saccade duration represents the time it takes for an eye to make a rapid movement from one place to another. If the duration is higher, that means the eye either moves slower from one place to another, or the distance traveled is further. For both possibilities, a higher value can be representative of an inattentive student.

Average fixation duration represents the time an eye focuses on a single point/small area. If the duration is higher, the eye spends more time focusing on single points, possibly indicating more focused attention. The number of saccades is the raw amount of saccades calculated during a student's recording session. While the coefficient will not extrapolate to longer sessions (ours were 15 minutes), the exact value can indicate, in conjunction with a lower average saccade duration, more focused attention. In the future, we plan to reevaluate the correlation analysis with longer recording sessions.

Based on the multiple regression, Average Saccade Duration and Average Fixation Duration are the best predictors of students' shown level of attentiveness. Similarly, the combination of Average Saccade Duration and Average Fixation Duration using multiple linear regression seems to show a strong relationship between certain eye metrics and the level of attentiveness the student is showing. Generally, as the fixation duration increases and saccade duration decreases, the more attentive the student appears to be.

VII. Conclusion and Future Work

This paper presents the design, analysis, and results of exploratory work investigating significant indicators of students' attentiveness. The experimental design collects students' eye data, facial images, and screen captures as input. The results of the experiments indicate a significant positive correlation with fixation duration, indicating a correlation with visual knowledge acquisition and attentiveness, and significant negative correlations with saccade duration, indicating a negative correlation between visual scanning and attentiveness. Additionally, the results demonstrated a significant correlation between fixation, attentiveness, and knowledge acquisition.

Motivated by the results, the authors plan to replicate this work for a more prominent subject population that can, in turn, be used to train a machine learning model. We also plan to automate the labeling process as part of the NiCATS system (in real-time) and validate the real-time tracking of student attentiveness. We are encouraged by the positive results in terms of statistically significant findings on eye metrics and attentiveness and are motivated to expand on the current work to develop a fully functioning NiCATS tool that can be used by educators (post-hoc or real-time) to assess and improve the instructional design of the information that is being presented to students.

References

[1] Z. Sharafi, T. Shaffer, B. Sharif and Y. Guéhéneuc, "Eye-Tracking Metrics in Software Engineering," 2015 Asia-Pacific Software Engineering Conference (APSEC), 2015, pp. 96-103, doi: 10.1109/APSEC.2015.53.

[2] Tasnia Tabassum, Andrew A. Allen, and Pradipta De. 2020. Non-intrusive Identification of Student Attentiveness and Finding Their Correlation with Detectable Facial Emotions. In Proceedings of the 2020 ACM Southeast Conference (ACM SE '20), pp. 127–134. DOI:10.1145/3374135.3385263

[3] Lane, Erin S., and Sara E. Harris. "A New Tool for Measuring Student Behavioral Engagement in Large University Classes." Journal of College Science Teaching 44, no. 6 (2015): 83-91. Accessed May 21, 2021. http://www.jstor.org/stable/43632000.

[4] Narayanan Veliyath, Pradipta De, Andrew A. Allen, Charles B. Hodges, and Aniruddha Mitra. 2019. Modeling Students' Attention in the Classroom using Eyetrackers. In Proceedings of the 2019 ACM Southeast Conference (ACM SE '19), pp 2–9. DOI:10.1145/3299815.3314424

[5] J. Whitehill, Z. Serpell, Y. Lin, A. Foster and J. R. Movellan, "The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions," in IEEE Transactions on Affective Computing, vol. 5, no. 1, pp. 86-98, 1 Jan.-March 2014, doi: 10.1109/TAFFC.2014.2316163.

[6] A. Ueno and Y. Uchikawa, "Relation between human alertness, velocity wave profile of saccade, and performance of visual activities," The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2004, pp. 933-935, doi: 10.1109/IEMBS.2004.1403313.

[7] Ueno, A., Tateyama, T., Takase, M. and Minamitani, H. (2002), Dynamics of saccadic eye movement depending on diurnal variation in human alertness. Syst. Comp. Jpn., 33: 95-103. https://doi.org/10.1002/scj.1145

[8] Young, M. S., Robinson, S., & Alberts, P. (2009). Students pay attention!: Combating the vigilance decrement to improve learning during lectures. Active Learning in Higher Education, 10(1), 41–55. https://doi.org/10.1177/1469787408100194

[9] Rucci, M., & Poletti, M. (2015). Control and Functions of Fixational Eye Movements. Annual review of vision science, 1, 499–518. https://doi.org/10.1146/annurev-vision-082114-035742

[10] M.L. Lai, M.J. Tsai, F.Y. Yang, C.Y. Hsu, T.C. Liu, S.W.Y. Lee, et al. A review of using eye-tracking technology in exploring learning from 2000 to 2012 Educational Research Review, 10 (2013), pp. 90-115

[11] P. Rodrigues, P. Rosa, “Eye-tracking as a research methodology in educational context: C. Was, F. Sansost, B. Morris (Eds.), Eye tracking technology applications in educational research,” Information Science Reference, Hershey PA, USA (2017), pp. 1-26