Development and Field-Testing of a Non-intrusive Classroom Attention Tracking System (NiCATS) for Tracking Student Attention in CS Classrooms

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

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

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

Gursimran Singh Walia
Computer Science
Augusta University
Augusta, USA
gwalia@augusta.edu

Md Shakil Hossain
Computer Science
Georgia Southern University
Statesboro, USA
mh34922@georgiasouthern.edu

Abstract: This Research to Practice Full Paper presents our Non-intrusive Classroom Attention Tracking System (NiCATS) and discusses the data collected through it. Academic instructors and institutions desire the ability to accurately and autonomously measure students' attentiveness in the classroom. Generally, college departments use unreliable direct communication from students, observational sit-ins, and end-of-semester surveys to collect feedback regarding their courses. Each of these methods of collecting feedback is useful but does not provide automatic feedback regarding the pace and direction of lectures. It has been widely reported that attention levels during passive classroom lectures generally drop after about ten to thirty minutes and can be restored to normal levels with regular breaks, novel activities, mini-lectures, case studies, or videos. Tracking these "drops" in attention can be crucial for the accurate timing of these change-ups in activities. This allows for maximal attention and a greater amount of deeply learned material. Autonomously collected data can also be used either in real-time or post-hoc to alter the design and presentation of lectures. Keeping track of student attention is vital to having confidence in delivering material. Even if lectures do not break up presentation slides with attention-raising activities, they can still show more important information during periods of high attention and less important information during periods of low attention. This area of research has applications both in in-person classrooms and online learning environments. The long-term goals of this research can prove invaluable for large in-person classrooms or classrooms where students' faces are obscured, such as behind computer monitors.

Keywords—Attention, Eyetracking, Eye Metrics, Education, Engagement, AI, Machine Learning

I. Introduction

It remains a challenge for instructors to accurately and automatically gauge the moment-to-moment level of attentiveness their students display as they perceive and process information in the classroom/labs, even more so for instructors with limited experience. This can lead to incorrectly-paced lectures and a lack of objective and timely means of identifying topics that may need to be reinforced or clarified. This is especially prominent in large lecture halls, online learning situations, and classrooms where computer monitors obscure students' faces (e.g., programming labs).

Feedback on instruction effectiveness is collected through end-of-course surveys, observer sit-ins, and analysis of course grades. While each of these is effective, they tend to be delayed or intrusive and, therefore, not applicable for tweaking the pace of lectures in real-time. More importantly, the summative nature of the feedback means instructors have to wait after the end of the semester to analyze issues and develop interventions to address these issues, assuming the next group of students would exhibit similar attention patterns.

Previous research in tool-assisted attention tracking has either been limited in scope, highly intrusive, or cost-prohibitive to scale up. Zhu et al. used wearable systems and sensors, which are highly intrusive and costly at scale.Whitehill et al. and Yun et al. used students' facial expressions to predict engagement levels. Veliyath et al. and Rosengrant et al. used students' eye gaze data to predict engagement levels. These were narrowly scoped in terms of the type of data collected and the purpose. While other researchers, such as Tabassum et al., used cloud-based facial emotion recognition services, which are costly at scale. Our system seeks to collect non-intrusive, multi-modal data that is scalable and not cost-prohibitive.

This paper presents the proposed system and discusses the data collected by our Non-intrusive Classroom Attention Tracking System (NiCATS). NiCATS was used to gather the subject's facial images and eye movements as students reviewed the information and provided feedback (real-time or replay) to the instructor. We evaluated the feasibility of the NiCATS to demonstrate that it 1) correctly captured subjects' facial images and eye metrics in different academic scenarios and 2) identified strong correlations between factors that can be used to predict student attentiveness in the future. We have

extended NiCATS's ability by extending the data collection and using a CNN model to train and test NiCATS' ability to predict student attentiveness using facial images. This paper will discuss our data collection and processing pipelines for NiCATS (version 2.0) and analyze the individual components that contribute to producing an attentiveness value that can be used to provide real-time feedback to professors.

NiCATS 2.0 has been used in different student settings (e.g., during code review, lecture presentation, and timed exam settings) that demonstrate its ability to collect data in varying environments reliably. We believe that the NiCATS' ability to accurately capture student face images and eye data and use them to predict attentiveness is invaluable for institutions seeking to enhance student education, instructors striving to improve the flow of lectures, and students seeking a more accommodating learning environment.

II. Literature Review

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

Computer Vision, in combination with cameras, has been used to measure the facial expressions of students in classrooms and their relationship to attentiveness. Whitehill et al. used students' facial expressions to train a machine learning model to predict engagement levels (Whitehill et al., 2014). In a " Cognitive Skills Training " experiment, they collected data in video recordings (later synchronized with task performance) of the subject's faces from 34 undergraduate students in a "Cognitive Skills Training" experiment. They found that using binary classification; the automatic engagement detector had similar accuracy to humans. They found that both human and automatic engagement labels correlated decently with task performance.

Researchers have explored features like facial images and facial emotions for predicting attentiveness. Tabassum et al. proposed a methodology for predicting student attentiveness using these two features. They collected webcam recordings of the students in the classroom and extracted images from the videos. They assessed the attentiveness of a student and the relationship between attentiveness and emotions using cloud-based facial emotion recognition(Amazon Rekognition) software and a CNN model. Their CNN model classifies attentive and inattentive students in classroom environments(accuracy of 93%) and found correlations between the emotional states of the students and their attentiveness level.

Yun et al. used a pre-trained convolutional neural network (CNN) and transfer learning to create a useful model for recognizing engagement levels (Yun et al., 2020). They used VGG Face pre-trained networks and modified the models to recognize children's engagement. They proposed an automatic children engagement recognition method based on CNNs for the future.

Researchers have also employed eye-tracking to understand attentiveness. Veliyath et al. concluded that gaze data from a monitor-mounted eye tracker could work with other features to achieve higher accuracy. They used data collected from self-reporting and eye trackers as a non-intrusive means to predict student attention throughout a class. Using this data, they trained multiple machine learning models to be able to predict attention(peak accuracy of 77%).

Rosengrant et al. used Tobii Glasses to track student eye movements during a lecture and determine the causes of inattention (Rosengrant et al., 2012). They used eight volunteers and recorded what they were looking at during a physical science lecture. They identified some behavioral patterns related to student inattentiveness, such as instructor's movement, activities, and emotion, as well as distractions cause by other students' activities.

Researchers have used full-body motion sensors to detect attentiveness. Zaletelj et al. used the 2D, and 3D data obtained by a Kinect One sensor to build a feature set characterizing facial and body properties of students to train machine learning models that predict attentiveness (Zaletelj et al., 2017). They concluded that using full-body motion sensors for affordable tracking of attention is possible, which would help evaluate lectures. To predict "interest level" and "perception of difficulty", Zhu et al. used wearable systems and sensors for tracking hand motions and heart activity. They analyzed the data captured with smartwatches and concluded that using wrist-worn smartwatches provides excellent accuracy in attention detection(accuracy of 98.99% for interest and 95.79% for the perception of difficulty), and leveraging other physiological sensors could potentially improve the accuracy.

III. Proposed Approach: NiCATS

The long-term goal of the Non-Intrusive Attentiveness Tracking System (NiCATS) is to provide instructors with real-time feedback on students' attentiveness in their classroom. NiCATS utilizes webcams (for facial attentiveness) and eye trackers (for patterns related to cognitive activities) that can be mounted on student machines. NiCATS collects webcam images, gaze points of their eye movements, and screenshots (of their screen) that can be analyzed to understand student attentiveness. While other researchers have primarily focused on perceived facial attentiveness, this work utilizes student attentiveness informed by their facial expressions and statistics of their gaze patterns in real-time. Figure 1 shows the high-level design of NiCATS.

A. Data Collection

The data collection block of the NiCATS system is primarily a lightweight application that resides on the students' computers. The lightweight application includes a consent prompt for students to opt in and is responsible for collecting, packaging, and submitting all student-related data. NiCATS can 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. LibFaceDetection, created by Feng et. al, was used to accurately detect if a person's face was in front of the webcam

(Feng, 2021). Less than 5% of images captured were blurry or unusable.

- *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* - Research commonly uses the eye-tracking terms that follow for which an eye tracker is generally used to measure. Eye gaze data is the immediate direction of a person's eyes and can be represented in XY coordinates when looking at a computer monitor. The gaze points of the student concerning their computer monitor are captured continuously throughout the lecture. Anytime a screenshot is prepared to be sent to the server, the most recent chunk (±5 seconds of the screenshot timestamp) of gaze points since the last screenshot was transmitted to be sent along with the screenshot for pre-processing.
    - Fixations stabilize the eye on the part of a stimulus for some time and are usually around 200-300ms (Sharafi et al., 2015).
    - Saccades are the quick and continuous eye movement between fixations and are usually around 50ms (Sharafi et al., 2015).

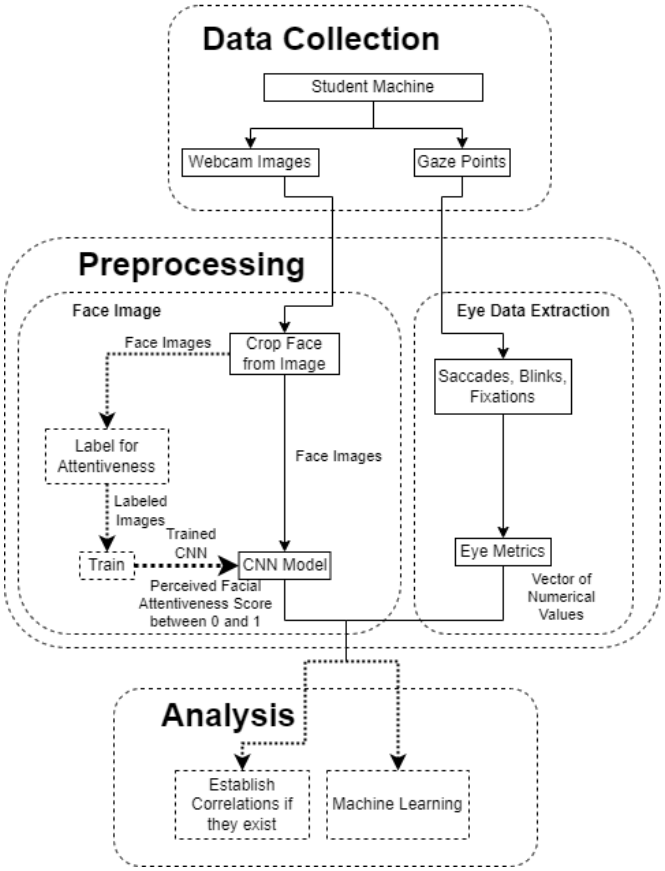

Figure 1. High-Level Design of NiCATS

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 and create regions of interest) to use the system to analyze the collected data post-hoc.

Face image: To capture isolated face images of the student for labeling, a smaller image is cropped from the original, which contains only the student's face. This was accomplished by initially using Haar cascade classifiers for the first experiment's data (James, 1910), then LibFaceDetection for the remaining experiments (Feng, 2021). To generate a set of attentive and inattentive images for comparison with the extracted eye metrics, multiple labelers were asked to label the face images based on the validated Behavioral Engagement Related to Instruction (BERI) protocol (Fredricks et al., 2004). The human labeling was handled via the NiCATS mobile web application, allowing human-labelers to swipe images right or left on their mobile devices to label images as attentive or inattentive, respectively (Figure 2). The attentiveness score was arrived at from the labeled image set by summing all "attentive" labels on an image and dividing it by the total number of labels (attentive or inattentive) a face image receives.

Figure 2. Attentiveness Labeling Mobile App

Eye Data Extraction: The gaze points are pre-processed to extract relevant eye metrics that can be used to predict student attentiveness. Using the gaze points, fixations and saccades are calculated. A subset of eye metrics (relevant to this work) that can be collected from fixation and saccade calculations is 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 saccades' durations divided by the number of saccades.
- *Saccades per second*: Total number of saccades divided by the total duration of a recording session.

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.

C. Convolutional Neural Network

The convolutional neural network (CNN) is being utilized in this research to predict student attentiveness based on their facial images alone. The CNN model is being trained on facial images that were manually labeled according to the Behavioral Engagement Related to Instruction protocol (Lane et al., 2015), and then results produced by the CNN model (on student attentiveness) are validated against the ground truth (manual labeling of student attentiveness). Additionally, the output of the CNN model is also being used to understand correlations with eye metrics to better understand the independent variables that can improve the performance of the CNN model in the future.

D. Convolutional Neural Network Model Architecture

The CNN model was trained on a set of face images labeled on perceived facial attentiveness. The dataset consisted of 630 equally split images (315 images labeled inattentive, 315 images labeled attentive). The original dataset contained 2,122 images, but it was not evenly split. It was shrunk down using random sampling so the dataset could be balanced. Each image was 350 pixels wide, 350 pixels tall, and was RGB. The CNN model consisted of several sequential layers. The layers consisted of 2D Convolution, Activation, 2D Max Pooling, Dropout, Flatten, and Dense layers. Figure 3 shows the architecture of the CNN model. Our loss function was binary cross entropy and our optimizer was AdaDelta at a learning rate of 0.01.

The model reached a max of 77% accuracy when predicting attentiveness from facial images, but more data will need to be collected for accuracy. From our best-collected data, we reached a precision of 85% for the attentive label and 16% for the inattentive label. The recall was 88% for attentive and 13% for inattentive. Based on the classification report analysis, the machine learning model is better at identifying attentive students than identifying inattentive students correctly.

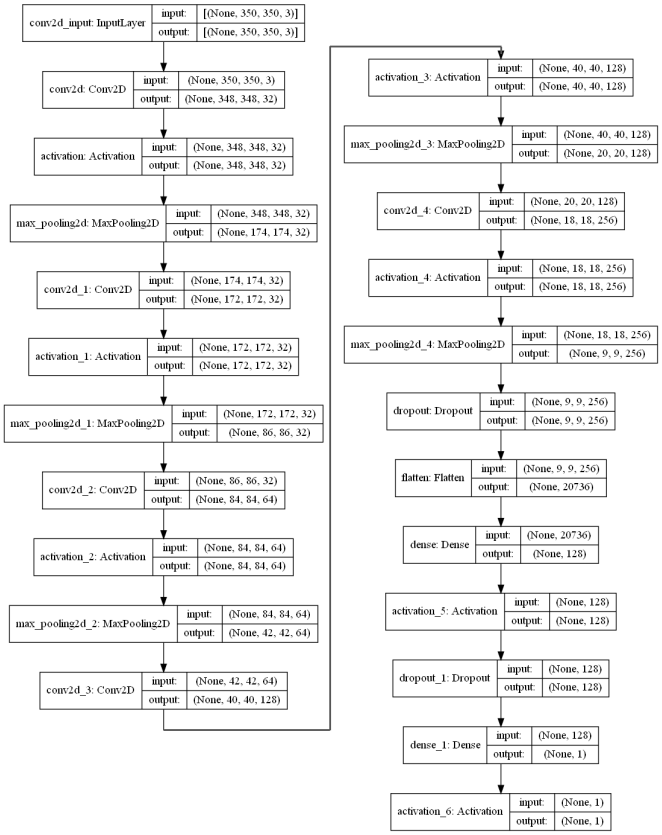

Figure 3. *CNN Model Architecture*

E. Design Decisions

Several design decisions were made during the creation of NiCATS.

The data for the assessment of facial attentiveness can be collected in one of two ways, a continuous video stream of the student's face or selective capture of static images of the student's face. The video stream method incurs a higher cost in terms of storage space and processing while providing marginal benefits compared to static images when determining attentiveness (Dewan et al., 2019). Because of the marginal benefit, the 5-second interval was used. For the collection of eye data, it was determined that the collection of screenshots would also be necessary to give the eye data context. It was also decided that to accurately "chunk" the eye data and screenshots, the user input would need to be considered (If a student clicks on another tab in their browser, the context of the eye data would be lost). We grouped the eye data and screenshots in our database and chunked it based on user input to calculate the correlations between eye metrics and perceived facial attentiveness.

The individual components consist of our NiCATS student client C++ program, NodeJS Express web server, and Postgresql database. The client program is responsible for collecting and uploading the students' facial images, eye data, and screen captures collected during a classroom lecture. The web server is responsible for extracting eye metrics from raw eye data and using a convolutional neural network (CNN)

trained on attentiveness to predict the perceived facial attentiveness of the face image.

IV. Research Framework

Four field experiments were carefully planned and executed to collect data and validate aspects of the NiCATS systems' usage in different classroom settings. While the research questions investigated across the four experiments are the same, each experiment design evaluates those questions in different settings and builds on the earlier experimental results.

The independent variables are as follows.

- Eye Gaze Points: The eye gaze points (x-coordinate, y-coordinate, and timestamp for each gaze point) collected during the experiment varied for different subjects. These gaze points were used to calculate the eye metrics, which are analyzed to correlate with perceived facial attentiveness.

- Amount of screenshots: The number of screenshots varied depending on the user interaction during their recording session.

The dependent variable is as follows.

- Perceived attentiveness: The attentiveness scores were calculated that ranged between 0 (inattentive) and 1 (attentive), representing the level of perceived attentiveness of the student. For experiment 1, the attentiveness labels were produced by labelers following the BERI protocol. Experiment 1 was used to determine the feasibility of a correlation between perceived facial attentiveness (manually labeled) and eye metrics. Therefore, experiment 1 did not use a CNN model to produce an attentiveness label but instead used post-hoc manual labeling of collected images to create the attentive and inattentive dataset. The remaining experiments (experiments 2, 3, and 4) used the CNN model to produce attentiveness labels, and the model's performance was evaluated against manually labeled images.

The research questions for each respective experiment are as follows.

- RQ1 - Can CNN-based modeling of NiCATS 2.0 data (facial image alone) be used to understand and predict students' perceived attentiveness in various settings?
- RQ2 - Are there generalizable correlations between facial attentiveness and eye metrics that need to be considered for future research?

For all four experiments, the participants were instructed to sit in a computer classroom environment on computers equipped with a monitor-mounted 1080p webcam and a monitor-mounted Tobii Eye Tracker 4c. The webcam was mounted at the top center of the monitor, and the eye tracker was mounted at the bottom center of the monitor. The student computers contained an i7-3770 CPU and 16 GB of RAM, which is adequate for ensuring that all the data is captured during the NiCATS student client program execution. Figure 4 shows an example of the experiment setup.

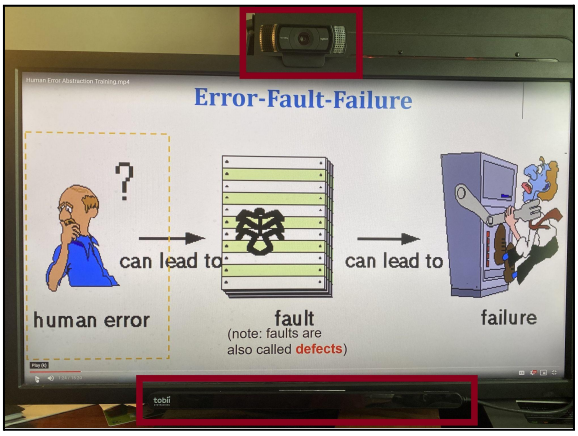

Figure 4. Webcam And Eye Tracker Mounted

In each experiment, the participants were instructed to calibrate their Tobii Eye Trackers and position the webcam to directly face the student before doing the assigned task. There were no common participants between experiments.

For experiment 1, the participants watched a pre-recorded 15-minute lecture regarding Human Errors and their applications in everyday life. This lecture was selected because it was generic enough that prior knowledge of computer science would not heavily affect the measured attentiveness. Participants consisted of varied ages and backgrounds, with most having computer science experience.

For experiments 2 and 3, the participants reviewed Java source code samples that contained faults and identified the location of the faults and the reason they were faults. The code samples were different for experiment 2 and experiment 3. The code samples were displayed simultaneously in full-screen images that covered the participants' entire computer monitor. This was so the data collected could be consistent across participants. Experiments 2 and 3 are grouped because the type of environment is the same (Java code sample review), but the participants and code samples were different. Participants consisted of undergraduate computer science students enrolled in the second sequence of the introductory programming course (CS2).

For experiment 4, each participant took their beginner CS1 midterm exam. The content covered basic Java syntax and logic. The exam consisted of 15 multiple choice and true/false questions and an additional programming section that asked the students to write a simple program. The exam lasted for 75 minutes, and every participant stayed for the entire duration. Participants consisted of undergraduate computer science students enrolled in the first of a two sequence programming (CS1) course.

V. Evaluation Criterion

An overview of metrics and their calculations for all four experiments are presented in Table 3. The "evaluation criteria" briefly explain the analysis of individual data points that averaged for the entire experiment. For experiments that used the CNN model to evaluate attentiveness, Table 3 presents the

evaluation criteria. The evaluation criterion for independent variables (IV) 1-5 are listed below:

- IV1 - Fixation per second: Total numbers of fixations within a 5-second interval of a face image (i.e., all fixations between 2.5 seconds before the face image and 2.5 seconds after) divided by time interval (5 seconds).
- IV2 - Average fixation duration (ms): Average fixation duration (ms): All fixation durations within a 5-second window of a face image (between 2.5 seconds before and 2.5 seconds after) are summed up and divided by the number of fixations within the 5-second window.
- IV3 - Saccades per second: Total numbers of saccades within a 5-second face image (i.e., all fixations between 2.5 seconds before the face image and 2.5 seconds after) divided by time interval (5 seconds).
- IV4 - Average Saccade Duration (ms): All saccade durations within a 5-second window of a face image (between 2.5 seconds before and 2.5 after) are summed up and divided by the number of fixations within the 5-second window
- IV5 - Regression Rate: The duration of all saccades within a 5-second window of a face image (between 2.5 seconds before and 2.5 seconds after) whose directions are between 135° and 225° divided by the duration of all saccades within that window range

The evaluation criterion for the dependent variable is below:

- DV1 - Perceived Facial Attentiveness: For each student's face image collected during the entire session, the sum of all CNN-produced face image labels was divided by the total # of images (labeled attentive/inattentive).
- Performance of CNN models included:
    - Accuracy - All correct labels output by the CNN model divided by all labels;
    - Precision - All correct positive labels output by the CNN model divided by positive labels;
    - Recall: All correct positive labels output by the CNN model divided by all correct positive labels and all incorrect negative labels;
    - Specificity: All correct negative labels output by the CNN model divided by all correct negative labels and all incorrect positive labels.
    - F-1 Measure: (2*precision*recall)/(precision + recall).

VI. Results and Discussion

This section analyzes the data collected during each of the experiments. Experiment results are organized around research questions listed in Section 4.

A. Machine Learning Model

For each experiment (except for experiment 1), the CNN model was validated against the manually-labeled ground truth to determine the model's accuracy. Experiment 1 did not use the CNN model and was manually labeled to test for the ground truth of attentiveness. The purpose of experiment 1 was to test for the feasibility of finding correlations between perceived facial attentiveness and eye metrics. Once they were established from experiment 1, experiments 2, 3, and 4 used the CNN model to label images. To help readers understand the type of output produced, Figure 5 shows the single participant's attentiveness over time during the second experiment. As mentioned earlier, the CNN model outputs a value between 0 and 1 to indicate the level of perceived attentiveness. This varied for students and was used to evaluate the performance of NiCATS 2.0.

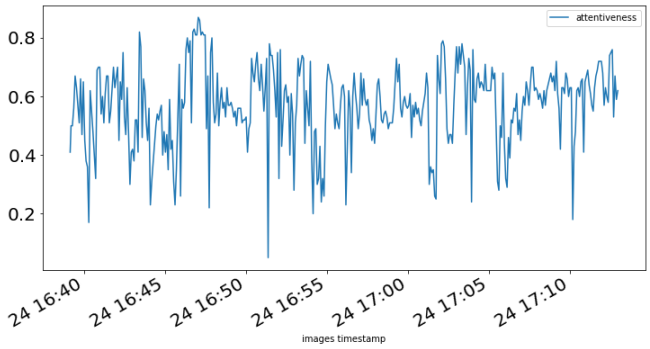

Figure 5. A Single Participant's Attentiveness Score During Experiment 2

B. CNN Model Evaluation Results

A classification report was generated for each experiment that used the CNN model to determine the effectiveness of the CNN model for predicting student attentiveness. The results are organized around each experiment.

Experiment 2: The accuracy of the machine learning model for experiment 2 is 77%. This means that the model correctly predicts the label of the image 77% of the time. The precision for attentiveness is 85%, the recall is 88%, and the specificity is 14%. These results match the evaluation of the CNN model during training and testing. In general, the model is better at predicting truly attentive images as being attentive than it predicts truly inattentive images as being inattentive. This experiment, in particular, has higher accuracy than what was initially indicated during the training and testing. This performance can be attributed to the settings where it was a highly-controlled experiment environment as students were asked to review source code (for short intervals) in short intervals. As is shown, 1832 images collected were labeled attentive, which translates to higher precision and recall (since the model is shown to identify attentive students better).

Experiment 3: The accuracy of the machine learning model for experiment 3 is 67%. This means that the model correctly predicts the label of the image 67% of the time. The precision for attentiveness is 85%, the recall is 88%, and the

specificity is 23%. Like with the training and testing set and the results from experiment 2, the model seems to be better at predicting the label of images with a ground truth of attentive rather than inattentive. This experiment has higher accuracy than what was produced in the training and testing of the model. Like with experiment 2, this can be attributed to the experiment environment (source code review in short time intervals) contributing to a better performance in labeling.

Experiment 4: The accuracy of the machine learning model for experiment 4 is 50%. This means that the model correctly predicts the label of the image 50% of the time. The precision for attentiveness is 85%, the recall is 51%, and the specificity is 51%. As is the same with the previous experiments, the model is better at predicting attentive labels than inattentive labels. This classification report differs from the previous experiments in that the accuracy is closer to 50% than the accuracy of the previous experiments. This can be interpreted as exams generally producing more ambiguous images that are harder to accurately label and may be due to writing code vs. attempting multiple-choice questions, scrolling/moving between questions, and working with multiple windows during the exam. While the accuracy in this experiment was below expectation, some useful insights were gained with respect to the eye-tracking data that can be used to determine the feasibility of using NiCATS for attentiveness prediction in exams.

C. *Relationship between Eye Metrics and Facial Attentiveness*

The results from the above analysis are based on the CNN model that only considered face images. A large part of the NiCATS ability allows researchers to understand how the eye metrics (e.g., fixations and saccades) are related to facial attentiveness. Understanding the correlation can help researchers use multiple factors (facial attentiveness, gaze-based attention) to predict student attentiveness in real-time. The NiCATS 2.0 can capture each participant's fixation and saccade patterns (in a quantifiable manner) that can be correlated with students' facial attentiveness post-hoc. As an example, Figure 6 shows the duration of average fixation and the average saccades of a single participant during experiment 4 (a timed mid-terms exam). Regression analysis was conducted for IV 1-5 (fixation per duration, average fixation duration, saccades per second, average saccade duration and regression rate) vs. DV1 (facial attentiveness).

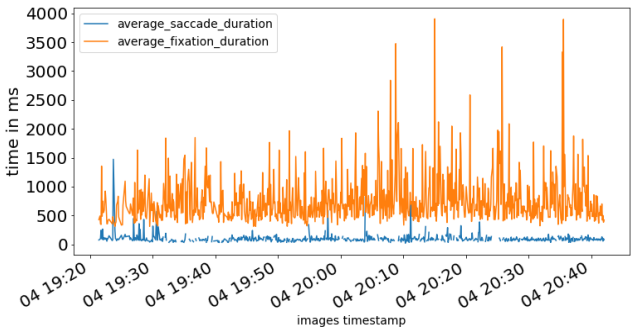

Figure 6. A Single Participant's Average Fixation and Saccade Patterns During Experiment 4

The results from the regression analysis are discussed below and are organized for each experiment. Only the significant correlations are reported below:

Experiment 1: Fixations per second were positively correlated with perceived facial attentiveness (p-value of 0.003). Saccades per second had a negative correlation with perceived facial attentiveness (p-value of 0.002). Average saccade duration also had a negative correlation with perceived facial attentiveness (p-value of <0.001).

Experiment 2: Saccades per second had a negative correlation with perceived facial attentiveness (p-value <0.001).

Experiment 3: Fixations per second were positively correlated with perceived facial attentiveness (p-value <0.001). Average fixation duration was positively correlated with perceived facial attentiveness (p-value = 0.01). Saccades per second had a **negative** correlation with perceived facial attentiveness (p-value <0.001). Average saccade duration also had a **negative** correlation with perceived facial attentiveness (p-value of <0.001).

Experiment 4: Fixations per second were positively correlated with perceived facial attentiveness (p-value <0.001). Average fixation duration was positively correlated with perceived facial attentiveness (p-value = 0.01). Saccades per second had a **negative** correlation with perceived facial attentiveness (p-value <0.001). Average saccade duration also had a **negative** correlation with perceived facial attentiveness (p-value of <0.001).

The commonality of results across experiments: Fixations per second was always a positive correlation with perceived facial attentiveness but was only statistically significant in experiments 1, 3, and 4. Saccades per second had a weak negative statistically significant correlation in all four experiments and were the only correlation to be statically significant in all four experiments. In all cases, it seems that saccades per second as an eye metric can be used to increase the accuracy of a predictive model.

Based on these results, combining the facial attentiveness CNN model's output with one or more of the stronger correlations could make a more accurate predictive model. Fixations per second, average fixation duration, saccades per second, and average saccade duration all have consistent enough results to be considered good candidates for creating a better predictive model. Based on the results from all four experiments, the regression rate seems to not be consistent enough for consideration.

VII. Limitations

Our results contain a number of limitations that should be considered before constructing and using a similar data collection and attentiveness-predicting system.

The use of NiCATS in a real-time setting could prove unrealistic. A common lecture scenario is a professor

presenting PowerPoint slides to the class. Finding time to check a "professor dashboard" to monitor average attentiveness levels in the class could be unreasonable. A small "alert" could be used instead, which could give a pop-up to the instructor if attentiveness levels have been low for a long period of time.

The upfront cost of eye-tracking and face image capturing equipment could be deemed an inefficient use of resources without further results. While the face image classifier CNN model in this paper reached 77% accuracy, more data will need to be collected to evaluate the practicality and accuracy of using it in real-time (in combination with the eye-tracking data). In addition, data will need to be collected on the effectiveness of altering lecture content based on the attentiveness labels produced by the system.

VIII. Conclusion and Future Work

This paper presents the investigation, design, analysis, and results of exploratory work to identify and automate significant indicators of students' attentiveness and the relationship between attentiveness and eye metrics. This paper iterates on our previous NiCATS paper by being a use-case study and collecting data in different settings. The experimental designs collected students' eye data and facial images as input for analysis.

The novel contribution of this paper is the creation of the NiCATS data collection system and the establishment of the link between eye metrics and perceived facial attentiveness. This research added to the body of knowledge regarding attention tracking in a classroom setting and sought to validate previous research on the topic. This research could prove invaluable in shaping how classroom lectures and materials are structured and improving institutions' and professors' feedback regarding their teaching methods.

Based on the findings from the research, a link between perceived facial attentiveness and the amount per second and duration of both fixations and saccades was identified. It suggests that measuring attentiveness could be combined with eye-tracking with facial expressions to create a more accurate predictive machine learning model than a machine learning model without this information. There is no definitive type of attentiveness-predicting eye metric when referring to "per second" and "average duration" metrics. Notably, saccades per second are the most reliable eye metric regarding statistical significance and effect on attentiveness. Additional research is needed to more confidently state the potential of using eye-tracking to measure attentiveness, along with which eye metrics are most relevant in the prediction of attentiveness.

References

[1] B. Anderson, "Stop paying attention to 'attention,'" *WIREs Cognitive Science*, 2021.

[2] W. James, "The principles of psychology / by William James.," 1910.

[3] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," Review of Educational Research, vol. 74, no. 1, pp. 59–109, 2004.

[4] M. S. Young, S. Robinson, and P. Alberts, "Students pay attention!," Active Learning in Higher Education, vol. 10, no. 1, pp. 41–55, 2009.

[5] M. A. Dewan, M. Murshed, and F. Lin, "Engagement detection in online learning: A Review," Smart Learning Environments, vol. 6, no. 1, 2019.

[6] E. Lane and S. Harris, "Research and teaching: A new tool for measuring student behavioral engagement in large university classes," Journal of College Science Teaching, vol. 044, no. 06, pp. 83–91, Jul. 2015.

[7] T. Tabassum, A. A. Allen, and P. De, "Non-intrusive identification of student attentiveness and finding their correlation with detectable facial emotions," Proceedings of the 2020 ACM Southeast Conference, 2020.

[8] A. Sanders, B. Boswell, G. S. Walia, and A. Allen, "Non-intrusive classroom attention tracking system (nicats)," 2021 IEEE Frontiers in Education Conference (FIE), 2021.

[9] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagemen tfrom facial expressions," IEEE Transactions on Affective Computing, vol. 5, no. 1, pp. 86–98, 2014.

[10] W.-H. Yun, D. Lee, C. Park, J. Kim, and J. Kim, "Automatic recognition of children engagement from facial video using Convolutional Neural Networks," IEEE Transactions on Affective Computing, vol. 11, no. 4, pp. 696–707, 2020.

[11] B. T. Carter and S. G. Luke, "Best practices in eye tracking research," International Journal of Psychophysiology, vol. 155, pp. 49–62, 2020.

[12] Z. Sharafi, T. Shaffer, B. Sharif, and Y.-G. Gueheneuc, "Eye-tracking metrics in software engineering," 2015 Asia-Pacific Software Engineering Conference (APSEC), 2015.

[13] N. Veliyath, P. De, A. A. Allen, C. B. Hodges, and A. Mitra, "Modeling students' attention in the classroom using Eyetrackers," Proceedings of the 2019 ACM Southeast Conference, 2019.

[14] D. Rosengrant, D. Hearrington, K. Alvarado, D. Keeble, N. S. Rebello, P. V. Engelhardt, and C. Singh, "Following student gaze patterns in physical science lectures," AIP Conference Proceedings, 2012.

[15] J. Zaletelj and A. Košir, "Predicting students' attention in the classroom from Kinect facial and body features," EURASIP Journal on Image and Video Processing, vol. 2017, no. 1, 2017.

[16] Z. Zhu, S. Ober, and R. Jafari, "Modeling and detecting student attention and interest level using wearable computers," 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN), 2017.

[17] Y. Feng, S. Yu, H. Peng, Y.-R. Li, and J. Zhang, "Detect faces efficiently: A survey and evaluations," IEEE

Transactions on Biometrics, Behavior, and Identity Science, vol. 4, no. 1, pp. 1–18, 2022.