Utilizing Convolutional Neural Networks and Eye-Tracking Data for Classroom Attention Tracking

Bradley Boswell
Augusta University
brboswell@augusta.edu

Andrew Sanders
Augusta University
asanders4@augusta.edu

Gursimran Singh Walia
Augusta University
gwalia@augusta.edu

Andrew Allen
Georgia Southern University
andrewallen@georgiasouthern.edu

Abstract

Instructors often use facial cues of their students as key indicators of student attention levels. However, this method can pose a problem in online and computer-based learning environments. While other research has shown computer vision and eye-tracking could be used with machine learning techniques to predict attentiveness, they have shown only moderate success in terms of accuracy. In this work, we improve upon existing techniques for student attention tracking. We employed our previously developed Non-Intrusive Classroom Attention Tracking System (NiCATS) to collect facial images and eye-tracking data of students during three controlled experiments that represent common academic scenarios. Our first contribution is using convolutional neural networks to predict student attentiveness with an F1-Score of 0.91. Our second contribution is the validation of using eye-tracking metrics in conjunction with machine learning models to predict the attentiveness of students with up to 0.78 F1-Score, which could be useful when webcam privacy is a concern.

Keywords: Eye-Tracking, Computer Vision, Education Technology, Machine Learning, Attention

1. Introduction

Understanding student attentiveness in classroom settings has always been a critical factor in improving the educational process. Traditionally, educators gauge student attention via direct observation or through feedback methods such as questionnaires and interviews. However, these methods can be time-consuming and disruptive. In traditional classrooms, it is easier for educators to detect inattention in students by observing their behavior and body language, but computer-based classrooms can present much more of a challenge. With the advent of online learning, students may be attending classes from various locations, which introduces more complex difficulties for instructors attempting to gauge the attention levels of their audience. Thus, as educational technologies continue to evolve, there is a growing interest in developing automated methods to monitor and assess student attentiveness. The urgency of this demand is further highlighted by studies demonstrating students' limited attention spans, with evidence suggesting that student attention levels drop only 10-30 minutes into a lecture (Young et al., 2009).

With the advent of artificial intelligence (AI) and machine learning (ML), many researchers have started to explore their potential in the context of education. In particular, convolutional neural networks (CNNs) have shown promise in a variety of image recognition tasks and, on the other hand, eye-tracking has been used to gain insight into cognitive processes such as attention and focus. Our research leverages these techniques to provide instructors with feedback about the perceived facial attentiveness of their students throughout the course of their lectures. By providing instructors with access to the attention levels of their students during their lecture, instructors can adjust the pace of their lecture according to these metrics and introduce interventions that may recapture the attention levels of their audience.

Our contributions are two-fold. First, we have improved upon our previously developed NiCATS pedagogical data collection tool by improving the performance of our CNN which predicts the perceived facial attentiveness of student face images which are captured via a webcam (Sanders et al., 2021). This approach provides an indicator of perceived facial attentiveness of students in a classroom, offering valuable insights for educators about student attentiveness during lectures. Second, we investigate the use of eye-tracking data to predict perceived facial attentiveness as an alternative approach when compared to using facial recognition. Due to the complexities that exist in the data collected by NiCATS, we evaluate the performance of four ML models in this context, including Random Forest, Logistic Regression, Artificial Neural Network, and K-Nearest Neighbors, to classify the eye metric data collected from our system (e.g., fixations, saccades). Our results indicate that the eye-metric models could serve as a suitable alternative to predicting student attention levels, which may be especially prevalent in settings where facial recognition technology may be a potential privacy concern for students in academia. These contributions demonstrate the value of NiCATS in supporting pedagogy in academic institutions.

The remainder of this paper is formatted as follows. The background literature is reviewed in Section 2. Section 3 provides an outline of the methodology. Section 4 presents the results and analysis of our findings. Section 5 provides a discussion of our results. Section 6 describes our conclusions and future work. Lastly, Section 7 provides our acknowledgments.

2. Background

The following section reviews the literature related to tracking attention using computer vision, eye-tracking, and other biometric data collection hardware.

2.1. Computer vision and attention tracking

Using webcams with computer vision has been previously used to classify attentiveness in classrooms.

Researchers have used images of student faces, collected from videos recorded during a “Cognitive Skills Training” experiment, to train a machine learning model to predict whether or not a face appeared to be engaged (Whitehill et al., 2014). Using this model, they found it had a similar binary classification performance to that of human classifiers. They also found that the engagement labels produced by both humans and the machine learning model had a moderate correlation with task performance.

Computer vision algorithms can detect head movements and posture to determine if a student is facing the screen or if they are distracted, such as looking down at their phone or talking to a classmate (Feng et al., 2021). Furthermore, facial recognition algorithms can identify whether a student is expressing emotions such as confusion or boredom, providing valuable information that could be used by teachers to adjust their teaching methods in real-time (Tabassum et al., 2020). By integrating computer vision technology into classroom management systems, instructors could have the ability to receive notifications or alerts when students are distracted or disengaged, allowing them to intervene quickly and keep students on track. The use of computer vision in automated attention tracking offers a non-intrusive and scalable solution to improving the learning experience of students in computer classrooms.

2.2. Eye-tracking and attention tracking

Another promising approach to automated attention tracking is through the use of eye-tracking technologies. By tracking eye movements, it is possible to determine whether a student is looking at the screen, as well as where they are looking on the screen. Prior research has shown that these data points can be invaluable when predicting student attention levels (Veliyath et al., 2019), causes for inattention (Rosengrant et al., 2011), and detecting cognitive interference (Rizzo et al., 2022). For example, if a student’s gaze wanders frequently or fixates on non-class-related content, this could indicate that they are not engaged in the lecture. Conversely, if a student’s gaze remains fixed on the lecture materials, this could suggest that the student is paying attention.

Research that uses eye-trackers commonly uses the following eye-tracking terms. *Gaze points* are the immediate direction of a person’s eyes at a given point in time and are commonly represented in (X,Y) coordinates with respect to the dimensions of their computer monitor. *Fixations* are the stabilization of the eye on part of the stimulus for a short period of time, and that usually last between 200-300ms (Sharafi et al., 2015). *Saccades* are the quick movements between fixations that usually last around 50ms (Sharafi et al., 2015).

2.3. Other attention tracking methods

Researchers have used the Kinect One full-body motion sensor to build a feature set characterizing the facial and body properties of students to build a machine learning model that can predict attentiveness with moderate success (Zaletelj and Košir, 2017). Researchers have used smartwatches to track hand motions and heart activity and built a high-accuracy machine learning model for predicting attentiveness (Zhu et al., 2017).

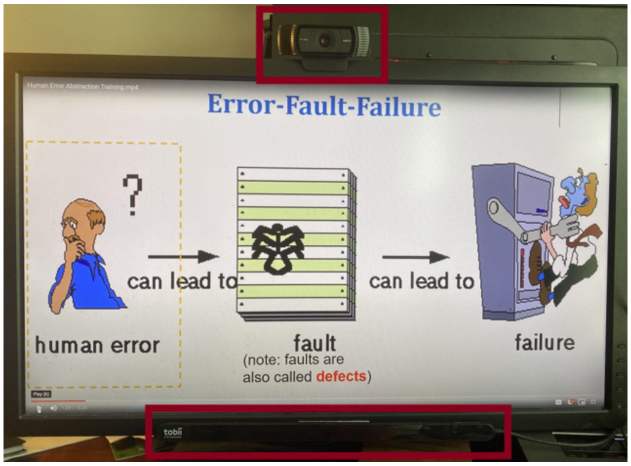

Figure 1. Example setup configuration

3. Methodology

The following section presents our methodology for data collection, pre-processing, and training of our classification models.

3.1. Non-intrusive classroom attention tracking system

An author-developed data collection tool, Non-Intrusive Classroom Attention Tracking System (NiCATS) (Sanders et al., 2021), was developed to collect data from students in computer-based classrooms by using an eye tracker and a webcam. In this work, each participating student was instructed to sit at a computer that was equipped with a Tobii Eye Tracker 4c, and a Logitech C920 webcam. An example setup configuration is shown in Figure 1.

In addition to setting up our hardware, we preinstalled the Tobii Eye Tracker 4c drivers as well as the NiCATS client software. NiCATS was used to gather the subject's facial images and eye movements as students reviewed a pre-recorded lecture on their computers. Following the data collection session, we evaluated the system to 1) verify that the system was capable of accurately capturing face images and eye metrics and 2) identify strong correlations between metrics that could potentially be used to predict student attentiveness in the future. This work was followed by additional data collection experiments from various learning environments where the face image data was used to train a Convolutional Neural Network (CNN) to predict perceived student attentiveness with 77% accuracy (Sanders et al., 2022). Furthermore, a correlation analysis between the attentiveness predictions made by our CNN and various

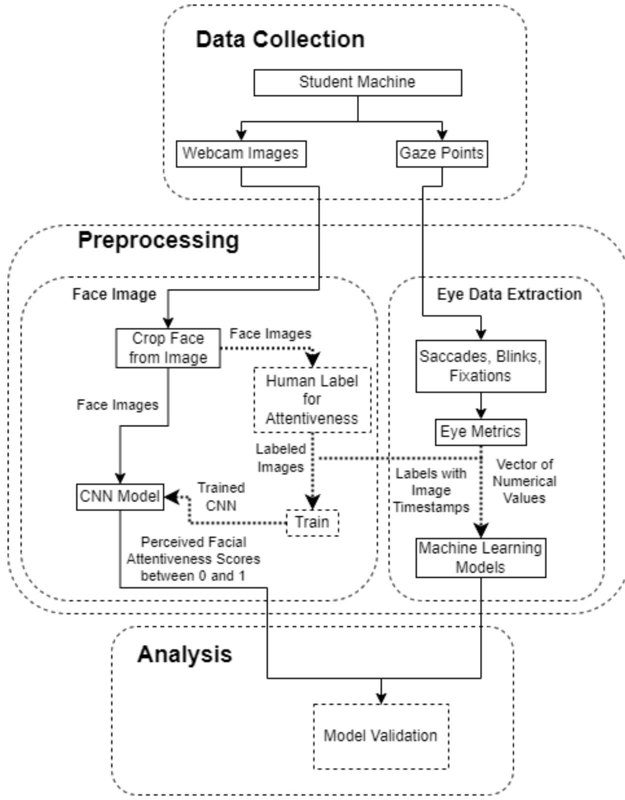

Figure 2. High-level architecture of the proposed data pipeline

eye metrics was conducted which provided a strong foundation for our research questions in this work. An overview of the proposed NiCATS architecture, for the purposes of answering the research questions in Section 3.2, is shown in Figure 2.

3.2. Research questions

The research questions we explore in this work are as follows:

- *RQ1: Can a convolutional neural network be trained to predict perceived facial attentiveness from face images with high accuracy?*

As we, and others, had previously studied, machine learning can be used to predict perceived facial attentiveness from face images (Sanders et al., 2022, Whitehill et al., 2014) with moderate accuracy (77% accuracy from (Sanders et al., 2022), 0.72 Cohen's κ from (Whitehill et al., 2014). To improve upon previous work, we utilized class weights to limit the biasing effect of class imbalance on our convolutional neural network model. To determine the effectiveness of our model

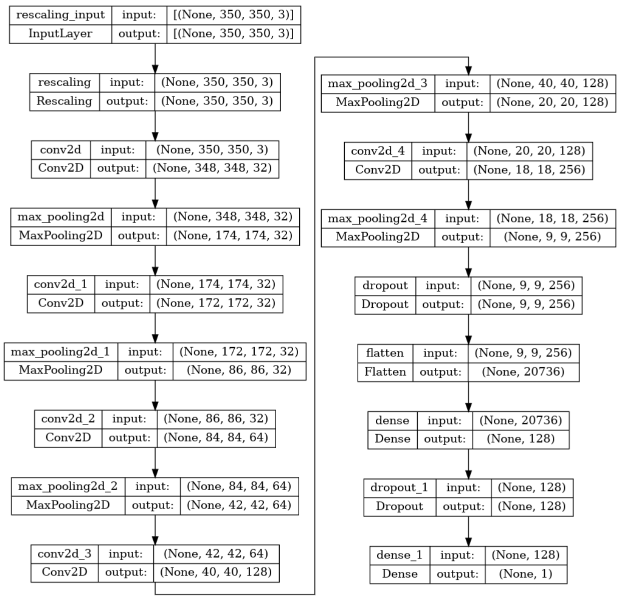

Figure 3. Overview of NiCATS CNN model (Sanders et al., 2021)

we evaluate this in terms of accuracy, recall, precision, AUC, AUPRC, and F1-Score.

- *RQ2: Can eye tracking data be used to predict perceived facial attentiveness?*

Previous work has shown that gaze data from eye trackers could be used to predict self-reported attentiveness levels by students with moderate accuracy (Veliyath et al., 2019). To work towards predicting perceived facial attentiveness using eye tracking data, we utilize fixation and saccade information to target human-labeled perceived facial attentiveness. To determine the effectiveness of our models we evaluate them using accuracy, recall, precision, AUC, AUPRC, and F1-Score.

3.3. Experiments

To answer our research questions, we used our previously developed NiCATS data collection tool to collect facial images and eye-tracking data from students in a variety of controlled experiments. Each controlled experiment was designed to emulate scenarios that are commonly experienced by students, allowing for data that is representative of real-world settings.

Experiment 1 - pre-recorded lecture: Our first experiment involved asking students to watch a fifteen-minute pre-recorded lecture about software faults. Then they were asked to take a pre-test and post-test with questions related to the contents presented in the lecture. The content of this lecture was deliberately chosen so it would be unfamiliar to the students such that data about the learning process could be accurately measured for our analysis through completing a pre-test, containing questions about the unseen lecture contents, and an identical post-test, which students completed after watching the lecture. The NiCATS client software was used to record the student data during the lecture so it could be used later for post-hoc analysis. The purpose of this experiment was to collect student information during the learning process, emulating the common lecture scenario that students are required to do.

Experiment 2 - code review: Our second experiment involved asking students in an introductory programming course to review fault-seeded Java code for four minutes and locate, in an open-response question, where the faults lie. We chose Java as the programming language for this experiment as this was the language that was familiar to all participants. Each student reviewed four code samples of fault-seeded Java code. Each code sample was seeded with common faults experienced by students in the target classroom. While students were reviewing the code sample, we used our NiCATS client software to collect student data during the code review so it could be used later for our post-hoc analysis. The purpose of this experiment was to collect student information during the application of comprehension in a coding environment. This is a common scenario for computer science students as they are expected to be able to write and debug code, especially for code that they did not write.

Experiment 3 - CS1 exam: Our third experiment involved students taking a CS1 midterm exam while the NiCATS client software collected their face images and eye-tracking data. The goal of this experiment was to collect student data during the application of comprehension in an examination environment. The exam included multiple-choice questions, open-answer questions, and a coding portion where the students were asked to create a program to complete a task. This setting is common for computer science students as they are asked to display their knowledge of what they have learned in both question-answering and code writing.

3.4. Convolutional neural network model

Convolutional Neural Networks (CNN) are a type of neural network model that is commonly used for

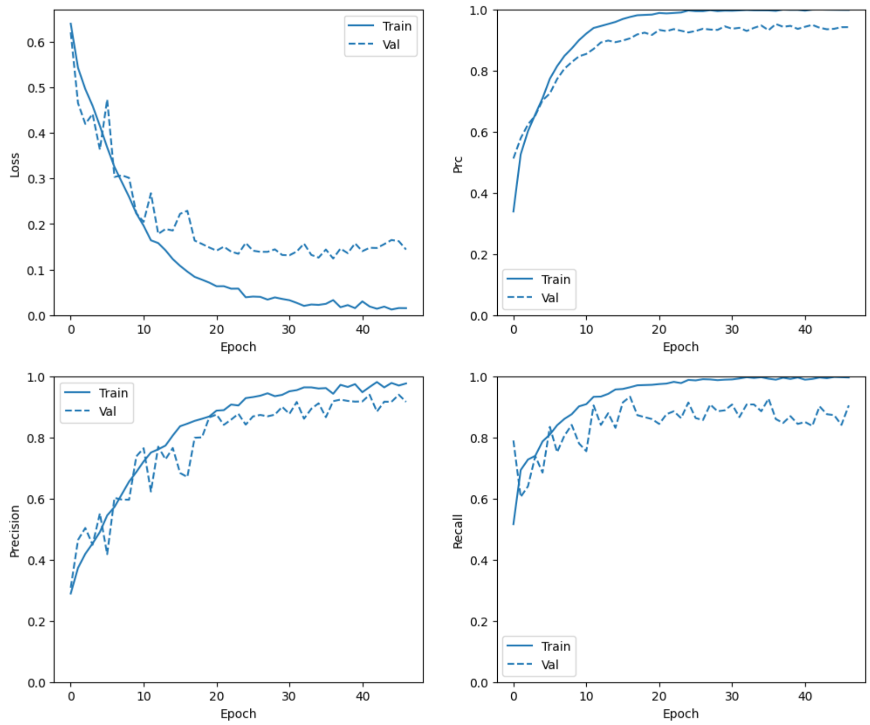

Figure 4. CNN model training history

image classification tasks. They are generally used due to their superior performance in image feature learning and classification as compared to other machine learning models (Kumar and Rao, 2018). Face image attentiveness classification provides a natural fit for the use of convolutional neural networks.

To answer research question 1, we improved upon our previously developed CNN model (Sanders et al., 2022) by using class weights. The purpose of class weights is to reduce the biasing effect of class imbalance. The total number of trainable parameters in our model is 3,052,129. Our total dataset was 8858 face images, of which 7,296 were labeled attentive, and 1,562 were labeled inattentive. The training-test split was 80-20. Due to the class imbalance, class weights of 0.61 for attentive and 2.84 for inattentive were used.

The images were human labeled and all images were labeled according to the Behavioral Engagement Related to Instruction protocol (BERI) (Lane and Harris, 2015) on if they appeared "attentive" or "inattentive". For purposes of being included in the dataset, only images that were fully labeled attentive or inattentive by all three labelers were included. The labelers showed moderate agreement with a Krippendorff Alpha of 0.48 in determining if an image appeared attentive or not. We used a batch size of 32, with Adam as the optimizer. To prevent overfitting, we utilized early stopping to cut off the training of the model when the loss of the validation set did not improve over ten epochs. Our model trained for 46 epochs before stopping early. Our model architecture is shown in Figure 3. The training and testing loss, precision, recall, and AUPRC are shown in Figure 4. The training set confusion matrix is shown in Figure 5.

3.5. Eye tracking models

To answer our second research question, RQ2, we compare the performance metrics of four machine learning models, each with their own strengths with respect to classifying student attentiveness based on eye metric data: Random Forest, Logistic Regression, Artificial Neural Networks (ANN), and K-Nearest Neighbors (KNN).

To process the raw gaze data collected by our eye tracking hardware, the Identification by Dispersion-Threshold algorithm (IDT) algorithm is applied (Salvucci and Goldberg, 2000). This is a popular algorithm for identifying fixations and saccades from raw gaze point data streams. The algorithm identifies a threshold distance value that separates fixations from saccades based on the dispersion of gaze points within a specific time window. These metrics are further processed to calculate features such as:

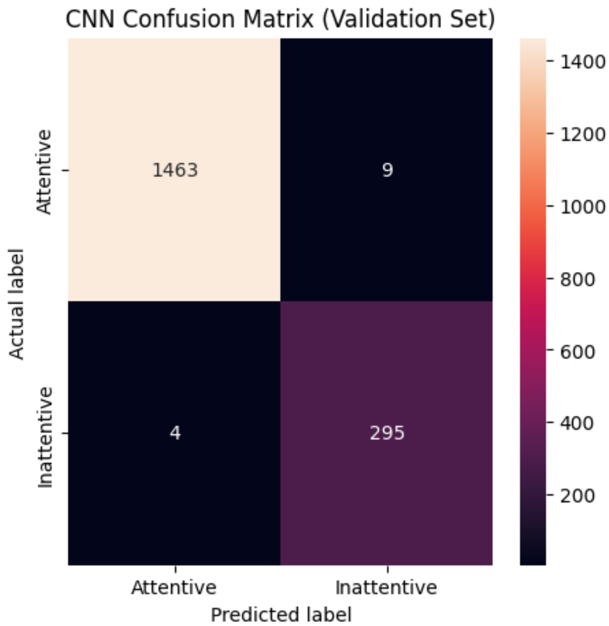

Figure 5. CNN model confusion matrix (validation set)

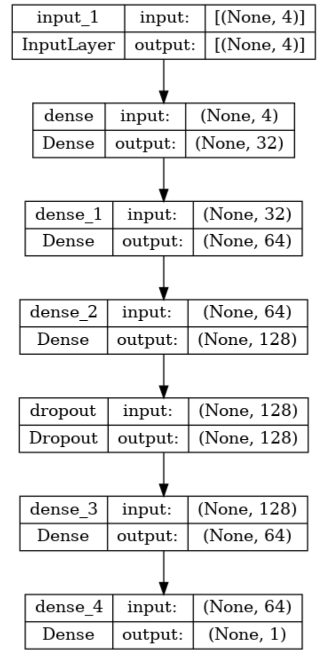

Figure 6. Overview of ANN model

- Average Occurrence Duration - The average duration that a fixation/saccade occurred during a 5-second time window period.

- Number of Occurrences per Second - The number of fixations/saccades per second that occurred over a 5-second time window at 90Hz.

After processing the raw gaze data into these metrics, the resulting data set is used to train our four candidate machine learning models.

Random forest: The Random Forest model, an ensemble learning method that leverages the use of multiple decision trees (Ho, 1995), provides a promising solution for handling complex interactions between different eye metrics. The model is resilient to outliers (Gunduz and Fokoue, 2015) and can effectively address the issue related to class imbalances (Khoshgoftaar et al., 2015). Importantly, the Random Forest model offers feature importance scores, granting us insight into which eye metrics significantly influence the prediction of student attentiveness. For our hyper-parameters, we set the number of trees to 100, which generally offers a good balance between model performance and computational efficiency, and the maximum depth of these trees was limited to 5.

Logistic regression: Logistic Regression, though a simpler model, can be a highly effective solution for binary classification tasks, aligning well with our research goal of differentiating between attentive and inattentive students. The model offers probabilities as outputs, providing an easy-to-interpret measurement of certainty in its predictions. Moreover, logistic regression is less prone to overfitting, ensuring the model generalizes well to unseen data.

Artificial neural network: Artificial Neural Networks, due to their ability to capture non-linear relationships within data, often cater to more complex classification tasks. In cases where relationships between eye metrics and attentiveness are non-linear or involve high-dimensional interactions, ANNs could outperform linear models. However, the inherent complexity of ANNs may render the decision-making process less interpretable.

Our model is shown in Figure 6. It has 18913 trainable parameters, uses Adam as the optimizer, and utilizes early stopping to prevent overfitting. To limit the biasing effect of imbalanced classes, we used class weights of 0.814 for the majority class and 1.296 for the minority class.

K-nearest neighbors: Lastly, the K-Nearest Neighbors (KNN) model classifies new instances based on their similarity to other instances within the

Table 1. CNN model performance comparison

Model	Accuracy	Precision	Recall	F1
CNN 1.0	0.77	0.85	0.88	0.86
CNN 2.0	0.96	0.92	0.90	0.91

training dataset. This non-parametric model can be particularly effective when irregular decision boundaries exist. The flexibility of KNN allows us to adjust the model's complexity and decision boundary by altering the k parameter, enabling an effective balance of bias and variance. For this work, we set this parameter to 30. This means the model considers the 30 closest samples to make a final prediction. This number was chosen to ensure that the model is not overly sensitive to noise in the data, which can be a problem when using a smaller value for k.

4. Results

The following subsections provide an analysis of the results from our experiment in the context of our two research questions.

4.1. RQ1: CNN for predicting student attentiveness

To evaluate the effectiveness of utilizing our proposed Convolutional Neural Network (CNN) model to predict students' attentiveness based on their facial cues during classroom sessions, we conducted a comprehensive evaluation of the model's performance. The evaluation was carried out on labeled validation data, and the metrics employed were accuracy, precision, recall, AUC, AUPRC, and F1-score. Although our preceding research yielded encouraging outcomes for this task (Sanders et al., 2022), further improvements were required to make this a feasible approach for our target application domain. To address our prior limitations, we introduced a weighting scheme to our samples before initiating the training process. The implementation of this method was aimed at circumventing possible biases within our model, wherein the class appearing more frequently could skew the target predictions due to its over-representation. A comparison of the performance metrics of our initial model (termed 'CNN 1.0') and the improved model (designated 'CNN 2.0') is provided in Table 1.

Our new model, CNN 2.0, exhibits an accuracy of 96%, indicating that it accurately predicted the attentiveness levels of students in 96% of our validation samples. The model also shows a precision of 0.92, meaning that when it predicts a student as attentive, it is correct 92% of the time. With a recall of 0.90, our model correctly identifies 90% of all truly attentive students. The F1-score, which is the harmonic mean of precision and recall, stands at 0.91, indicating a balanced performance between precision and recall. When compared to our previous model, CNN 1.0, these metrics illustrate a notable performance improvement that underscores the significant advancements made and reinforces the practicality and effectiveness of employing CNN models for predicting student attentiveness in educational settings.

4.2. RQ2: Eye-tracking for predicting student attentiveness

The performance of our four candidate models — Random Forest (RF), Logistic Regression (LOG), Artificial Neural Network (ANN), and K-Nearest Neighbors (KNN) — evaluated on eye-tracking data to predict perceived facial attentiveness is displayed in Table 2. The findings provide affirmative evidence to our research question, substantiating the premise that eye-tracking data can indeed be used to predict perceived facial attentiveness. Firstly, the accuracy of all four models, ranging from 0.68 to 0.69, indicates that a significant portion of the predictions made by these models are correct.

These results show a strong connection between the eye-tracking data and perceived facial attentiveness. This connection is further emphasized by the AUC values, ranging from 0.70 to 0.71, which denote that all four models have a good performance at distinguishing between attentive and inattentive states. The precision scores, which reflect the proportion of true positive predictions among all positive predictions, range from 0.67 to 0.71. This relatively high precision across all models indicates a low false-positive rate, implying that when the models predict a student to be attentive, they are usually correct. The recall scores, ranging from 0.85 to 0.93, demonstrate that a high proportion of truly attentive instances were correctly identified by the models, further supporting the viability of eye-tracking data for this prediction task. The F1-scores, which serve as a balanced measure of precision and recall, reinforce the balanced performance of the models, showing that they are not biased towards either precision or recall. The AUPRC values, ranging from 0.75 to 0.77, demonstrate the effectiveness of the models at differentiating between attentive and inattentive students, especially in contexts where the classes may be imbalanced.

Table 2. Performance comparison of eye-metric classifiers

Model	Accuracy	Precision	Recall	AUC	AUPRC	F1
RF	0.69	0.69	0.87	0.71	0.77	0.77
LOG	0.69	0.69	0.88	0.71	0.76	0.78
ANN	0.68	0.67	0.93	0.70	0.76	0.76
KNN	0.69	0.71	0.85	0.70	0.75	0.77

5. Discussion of results

This paper presents a comprehensive exploration of the applicability of computer vision and eye-tracking technology in predicting student attention levels in a classroom setting. We sought to address two primary research questions. Firstly, we investigated the feasibility of training a Convolutional Neural Network (CNN) to predict perceived facial attentiveness from webcam-captured images with high accuracy. We found that CNNs, due to their capability to hierarchically extract and learn features from image data, are indeed proficient at this task. This contributes significantly to the educational landscape by providing educators with a reliable indicator of student attentiveness. By monitoring and adjusting teaching strategies based on these attentiveness indicators, educators can foster an enhanced learning environment, ensuring their content resonates effectively with their audience.

Secondly, we delved into the potential of eye-tracking data as a means of predicting perceived facial attentiveness. To evaluate this, we employed a variety of machine learning models including Random Forest, Logistic Regression, Artificial Neural Networks, and K-Nearest Neighbors. The results from each model were largely promising and displayed comparable performance metrics, thus affirming our research proposition of the effectiveness of eye-tracking data in predicting attentiveness. However, when selecting a model for practical application, one should consider the specific requirements of their educational setting. Each model we investigated displayed subtle differences in precision, recall, F1-score, and area under the curve (AUC) values, suggesting that the "best" model may vary depending on whether the priority lies with precision, recall, or balanced consideration of both alongside other metrics such as AUC.

6. Conclusion and future work

In conclusion, this research validates the practicality of using both CNNs for image-based attentiveness prediction, and machine learning models for predicting attentiveness using eye-tracking data. We not only improved the performance of the CNN classifier originally used in NiCATS, but we also introduced a methodology for predicting student attention levels using only eye metrics with 69% accuracy.

Future work could progress in several directions. While convolutional neural networks are great for image classification, and artificial neural networks have shown promise in eye-tracking data classification, they each lack the ability to analyze intricate temporal patterns, such as how a person's facial expression evolves over the course of a lecture. We recommend looking into attention-based models or transformers (Vaswani et al., 2023), which have become more popular machine learning methods for learning nuances and applying contextual information in sequential data.

Designing interventions based on attentiveness feedback and assessing their impact on the learning outcomes and overall student experience is another logical step in this research area. One potential way to accomplish this is to investigate the impact of additional eye-tracking data on the predictive performance of our approach, such as gaze paths and blink frequency. We plan to apply dynamic time-warping clustering analysis on students' gaze paths to make statements about how high performers optimize their scan path as compared to low performers.

Given the encouraging results with eye-tracking data, future work could be done to investigate the application of similar methods to other physiological data that can be non-intrusively collected. These data sources have the potential to provide additional or complementary information to eye-tracking and image-based data, enriching the models' capability to predict attentiveness. Integrating multiple data sources could pave the way for a comprehensive attentiveness prediction system that dynamically adapts to individual students' responses, further optimizing the learning experience.

7. Acknowledgements

We acknowledge the many students that volunteered for our experiments and the Georgia Southern University Institutional Review Board for their support and approval of our research protocols.

References

Feng, Y., Yu, S., Peng, H., Li, Y., & Zhang, J. (2021). Detect faces efficiently: A survey and evaluations. *CoRR*, *abs/2112.01787*. https://arxiv.org/abs/2112.01787

Gunduz, N., & Fokoue, E. P. (2015). Robust classification of high dimension low sample size data. *arXiv: Applications*.

Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, *1*, 278–282 vol.1. https://doi.org/10.1109/ICDAR.1995.598994

Khoshgoftaar, T. M., Fazelpour, A., Dittman, D. J., & Napolitano, A. (2015). Alterations to the bootstrapping process within random forest: A case study on imbalanced bioinformatics data. *2015 IEEE International Conference on Information Reuse and Integration*, 342–348.

Kumar, K., & Rao, A. C. S. (2018). Breast cancer classification of image using convolutional neural network. *2018 4th International Conference on Recent Advances in Information Technology (RAIT)*, 1–6.

Lane, E. S., & Harris, S. (2015). Research and teaching: A new tool for measuring student behavioral engagement in large university classes. *The journal of college science teaching*, *44*.

Rizzo, A., Ermini, S., Zanca, D., Bernabini, D., & Rossi, A. (2022). A machine learning approach for detecting cognitive interference based on eye-tracking data. *Frontiers in Human Neuroscience*, *16*. https://doi.org/10.3389/fnhum.2022.806330

Rosengrant, D., Hearrington, D., Alvarado, K., & Keeble, D. (2011). Following student gaze patterns in physical science lectures. *Physics Education Research Conference 2011*, *1413*, 323–326.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 71–78. https://doi.org/10.1145/355017.355028

Sanders, A., Boswell, B., Allen, A., Singh Walia, G., & Shakil Hossain, M. (2022). Development and field-testing of a non-intrusive classroom attention tracking system (nicats) for tracking student attention in cs classrooms. *2022 IEEE Frontiers in Education Conference (FIE)*, 1–9. https://doi.org/10.1109/FIE56618.2022.9962447

Sanders, A., Boswell, B., Walia, G. S., & Allen, A. (2021). Non-intrusive classroom attention tracking system (nicats). *2021 IEEE Frontiers in Education Conference (FIE)*, 1–9. https://doi.org/10.1109/FIE49875.2021.9637411

Sharafi, Z., Shaffer, T., Sharif, B., & Guéhéneuc, Y.-G. (2015). Eye-tracking metrics in software engineering. *2015 Asia-Pacific Software Engineering Conference (APSEC)*, 96–103. https://doi.org/10.1109/APSEC.2015.53

Tabassum, T., Allen, A. A., & De, P. (2020). Non-intrusive identification of student attentiveness and finding their correlation with detectable facial emotions. *Proceedings of the 2020 ACM Southeast Conference*, 127–134. https://doi.org/10.1145/3374135.3385263

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). Attention is all you need.

Veliyath, N., De, P., Allen, A. A., Hodges, C. B., & Mitra, A. (2019). Modeling students' attention in the classroom using eyetrackers. *Proceedings of the 2019 ACM Southeast Conference*, 2–9. https://doi.org/10.1145/3299815.3314424

Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing*, *5*(1), 86–98. https://doi.org/10.1109/TAFFC.2014.2316163

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

Zaletelj, J., & Košir, A. (2017). Predicting students' attention in the classroom from Kinect facial and body features. *EURASIP Journal on Image and Video Processing*, *2017*(1), 80. https://doi.org/10.1186/s13640-017-0228-8
75 citations (Crossref) [2023-06-14]

Zhu, Z., Ober, S., & Jafari, R. (2017). Modeling and detecting student attention and interest level using wearable computers. *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 13–18. https://doi.org/10.1109/BSN.2017.7935996