Leveraging Artificial Intelligence (AI) to Enhance Computer Science Instruction

HyeonJin Yoon
Nebraska Center for Research on Children, Youth, Families, and Schools
University of Nebraska-Lincoln
Lincoln, USA
hyoon5@unl.edu

Xin Zhong
Department of Computer Science
University of Nebraska Omaha
Omaha, USA
xzhong@unomaha.edu

Agnibh Dasgupta
Department of Computer Science
University of Nebraska Omaha
Omaha, USA
adasgupta@unomaha.edu

Gwen Nugent
Nebraska Center for Research on Children, Youth, Families, and Schools
University of Nebraska-Lincoln
Lincoln, USA
gnugent1@unl.edu

Guy Trainin
Teaching, Learning and Teacher Education
University of Nebraska-Lincoln
Lincoln, USA
gtrainin2@unl.edu

***Abstract*— Measuring and understanding instructional processes in classrooms is essential for enhancing teaching and learning outcomes. Classroom observations provide educators with integral opportunity for professional development, fostering reflective practice and continuous improvement. However, traditional observation and coaching procedures are time-consuming, cumbersome, and expensive, requiring travel or video coding. Artificial intelligence (AI) technologies present a promising solution, enabling streamlined, efficient, and nuanced analysis of classroom observation data. This study aims to develop AI-based computer vision and multi-modal models to measure computer science (CS) instructional processes in video data and validate their accuracy. Using two K-8 CS classroom instruction videos, we tested two deep-learning approaches: 1) Convolutional Neural Network (CNN) for analyzing visual imagery, and 2) a multimodal deep learning approach that analyzed both visual and auditory information. Results showed that the multimodal approach achieved the highest accuracy. Our findings support the broader application of these methods for analyzing more intricate teaching and learning processes. We discuss the implications of the results for CS education practice and research.**

Keywords—computer science education, classroom observation, instructional behaviors, deep learning, multi-modal analysis

I. Introduction

Understanding and reflecting on instructional processes in classrooms is crucial for enhancing teaching and learning. Systematic classroom observations offer educators direct evidence and feedback on instructional quality leading to better student learning. Such observations provide educators with the most tangible insights into the nature and quality of classroom instruction, facilitating reflection on teaching practices and student learning. Additionally, classroom observations can capture and describe the great variability across different classrooms, schools, and districts [1].

Despite its advantages, conducting and analyzing classroom observation is a laborious and time-consuming task, requiring educators, administrators, or researchers to physically visit local sites or thoroughly code from video. Observational coding can often take up to twice as long as the actual classroom session. Moreover, in scenarios where peer observation and coaching aren't feasible, particularly in under-resourced schools with limited teachers or professional development staff, the process becomes even more challenging [2].

AI technologies have demonstrated their utility in analyzing audio and text-based source material [3] and offer tremendous potential for streamlining and analyzing classroom observations [4]. The purpose of this work-in-progress (WIP) paper is to develop and validate advanced AI-based approaches for measuring instructional processes captured in K-12 computer science (CS) classroom videos. By leveraging advanced AI technology in analyzing CS instructional processes, this study provides educators and researchers with a means to conduct more refined analysis of teaching and learning processes.

II. Theroretical Framework

A. Impact of Classroom Observation on Teaching and Learning

Using observational data to provide performance feedback has been linked to improvements in teaching and learning (e.g., instructional practices and student engagement [5], [6]. Pianta, Hamre, and Downer [7] emphasize how classroom observations provide insights into teaching effectiveness, facilitating targeted feedback for professional growth. In an empirical study, Rauch [8] found that pre-service teachers perceived feedback from peer observations as improving the quality of their teaching and valued it as a helpful experience. Furthermore, training and coaching with performance feedback was found to be associated with promoting the use of evidence-based instructional practices and improving teacher performance in preschool settings [9]. Dragon and his colleagues [10] also found that observational data can yield valuable insights into student behavior and emotional states, which can in turn inform teaching practices.

Previous studies have indicated the potential impact of the use of classroom observation data on student learning. Providing performance feedback to teachers on observation data has been shown to improve class engagement, reduce problem behaviors [5], [7], and enhance social-emotional outcomes in preschool children [11].

B. Artificial Intelligence (AI)-based Analysis of Classroom Observations

Over the past decade, there has been remarkable growth in automated human behavior analysis in classroom settings

using artificial intelligence [4]. AI enables automated analysis of large volumes of data, saving significant time and resources compared to human observation. AI-based video analysis offers more refined insights into student/teacher interactions, while enhancing objectivity and reliability and mitigating biases. Crucially, providing real-time results and feedback based on AI-driven analysis enables prompt adjustments in teaching strategies. This allows teachers to refine lesson planning and adapt instructional strategies to meet diverse student needs.

A range of studies have demonstrated the potential of artificial intelligence (AI)-based observation automated systems, such as machine learning (ML) and deep learning, in analyzing classroom observation data [12]-[14]. These AI based automated analysis approaches were designed to detect and categorize instructor/teacher behaviors [15], assess student engagement and attention during class [16], student-teacher interaction [17], and classroom climate [18]. The primary goal was to enhance teaching effectiveness and offer educators actionable feedback for instructional improvement. More recently, multi-modal recognition method emerged to improve the accuracy of classroom behavior analysis [19], [14], [18], incorporating a variety of mode of the data extracted from the video (e.g., image, audio) in analyses. For instance, Ramakrishnan and colleagues [18] developed a multi-modal machine learning-based system to automatically detect positive classroom climate and negative classroom climate through classroom observation. This system utilized image data (e.g., facial expression) and audio data (e.g., warm voice, yelling) with high accuracy.

These studies highlight the transformative potential of AI in analyzing classroom observations, offering scalable, objective, and data-driven approaches to enhance teaching quality and student learning outcomes. Despite rapid advancements in AI, the utilization of more sophisticated AI techniques for analyzing classroom observations remains relatively underexplored, particularly in the context of computer science education within K-12 classroom settings. Given the promising utility of AI-driven analysis of classroom observation data for improving teaching and learning in K-12 CS classrooms, our ultimate project goal is to develop and validate advanced multi-modal deep learning techniques for analyzing computer science (CS) classroom instruction observation video data.

The purpose of this WIP study is to develop AI-based computer vision and multi-modal models for measuring instructional processes (i.e., instructional format; teaching whole class vs. teaching individual students) in video data and to validate their accuracy. The research question guiding this study is as follows:

Research Question. How can we accurately measure instructional format (teaching whole class vs. teaching individual students) from K-12 CS classroom videos using advanced computer vision and multi-modal-based algorithms?

III. METHOD

A. Data Source and Labeling

Two K-8 computer science (CS) classroom instruction observation videos were utilized for this study. These videos were collected under a National Science Foundation (NSF)-funded project. Each video came from a CS class in urban Grade 3 and Grade 8 classrooms. The Grade 3 class was taught by a male teacher, and the Grade 8 class was taught by a female teacher. Class time for both videos was 40 minutes in length, with approximately 17 students included in each video.

In order to define the ground truth, each video was initially coded by an educational research expert, one of the authors of this study, for instructional format and corresponding timestamps. The expert speifically categorized a series of frames as either “teaching whole class” or “teaching individual students.” This labeled video was then analyzed with AI computer vision and multi-modal models for training and validation. Each frame of the video is labeled to indicate the instructional format (i.e., “teaching whole class” or “teaching individual students”). Some examples are visualized in Figure 1.

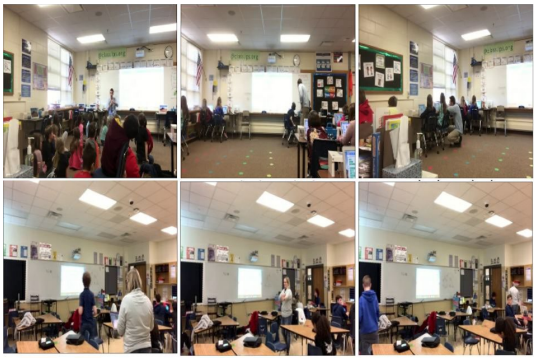

Fig. 1. Sample frames from both datasets. The first row contains random frames from the first dataset, and the second row contains from the second dataset.

The challenge of measuring instructional format is essentially recognizing the primary in class activity for each frame of the classroom videos. Specifically, in this paper, we model this as a classification problem where each frame is categorized based on whether the teacher is instructing the whole class or individual students, as labeled.

B. Methodological Design

1) Convolutional Neural Network (CNN)

To address the challenge of classifying instructional formats, we propose to test deep learning-based approaches, particularly convolutional neural networks (CNN), due to their effectiveness in learning semantic features for classification tasks. CNNs are well-suited for our task as they can analyze each frame of the classroom videos, identifying essential features at various levels of granularity. The architecture of the model we used, ResNet50 [20], is illustrated in Figure 2. This model takes an image of shape 224 x 224 x 3 as input and classifies it into one of the two categories.

ResNet50’s architecture comprises several components: a zero-padding layer to ensure that the input dimensions are correctly aligned with subsequent layers, a 7x7 convolutional layer with 64 filters, followed by a max-pooling layer with a 3x3 filter, and 16 residual blocks separated into four stages. Each residual block consists of a 1x1 convolutional layer that reduces the number of filters, followed by a 3x3 convolutional layer, and another 1x1 convolutional layer that restores the number of filters. The uniqueness of ResNet50 lies in its use of skip connections, or shortcuts, that bypass

one or more layers in the residual blocks. These connections add the input of a block to its output, allowing gradients to flow through the network more effectively during backpropagation, which facilitates the training of much deeper networks.

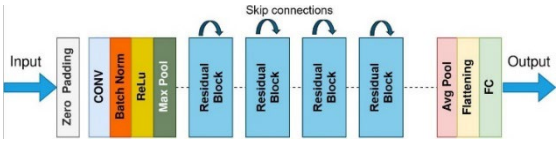

Fig. 2. ResNet50 architecture.

2) Mutlimodal Network

To enhance instructional format classification, we propose a multimodal approach that leverages CNNs for spatial information and incorporates temporal and audio features from videos. Videos, with their additional temporal axis, provide richer data than images, offering context through motion, actions, and events. Audio adds complementary information such as dialogue, background noise, and music, crucial for emotion recognition and event detection. Combining video and audio data in a multimodal deep learning approach allows for learning more detailed information for complex tasks. Figure 3 provides an overview of our multimodal model used for classification.

The model takes as input a small video clip comprising of 10 frames, hence the shape 10 x 224 x 224 x 3 and the corresponding audio clip, producing one of two classes. The architecture of the video portion of the model is very similar to ResNet50, except using (2+1)D convolutional layers instead of 2D layers to accommodate the video input. The audio feature extractor uses Mel-frequency cepstral coefficients (MFCC) to capture the essential characteristics of audio signals in a compact form. The features from each case are learned and fused together by concatenating along the channel axis of the projected features to produce a class for that segment.

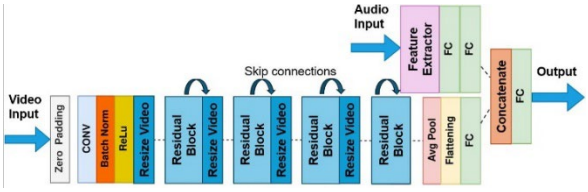

Fig. 3. Multimodal architecture

C. Loss

For both CNN and multimodal methods, we use categorical cross-entropy loss to measure the difference between the true label distribution and the predicted probability distribution. The equation below mathematically describes the loss calculation.

$$Loss = -\frac{1}{N}\Sigma_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)]$$

Where y_i is the true label (0 or 1), p_i is the predicted probability for class 1, and N is the number of samples. The goal is to minimize this loss, which means improving the model's predictions to be as close to the actual labels as possible.

D. Dataset Preparation and Preprocessing

For our experiments, we used the labels generated earlier in section III. A. and empirically categorized two classroom teaching videos into two classes. The classes were labeled as 'Ind,' indicating the teacher helping individual students, and 'Teach,' indicating the teacher addressing the whole class. These two labels function as the ground truth for the dataset. The videos were 2,367 seconds and 2,203 seconds long, respectively, with a resolution of 1280 x 720 pixels. With 29.97 frames per second for each video, the number of labelled frames were approximately 70,938 and 66,023, respectively. Each video was resized to 224 x 224 pixels and then either split into individual frames for our CNN model or into video and audio segments for our multimodal DL model. We refer to these videos as dataset 1 and dataset 2, respectively.

For our CNN model, we split each video into frames and selected every 5th frame to avoid redundancy. For the multimodal DL approach, we divided each video into 6-second segments, creating mp4 files for video and wav files for audio. From each video segment, we selected 10 frames, skipping 15 frames in between to avoid redundancy, thus constructing the final video data. Each frame and segment also had an associated label as described above. Finally, each dataset (video) was split into training and validation sets. We randomly used 90% of the shuffled data to train our model and used the remaining 10% as the validation set to evaluate the model after training.

IV. Results

A. Spatial Analysis

The CNN model classified individual frames into two classes: whole class instruction and individual instruction. For dataset 1, we achieved a validation accuracy of 96.9%, and for dataset 2, an accuracy of 97.7%. A heatmap, generated using Grad-CAM and the output of the CNN's convolutional layer, highlighted the model's areas of attention. Warmer colors (red) indicate high-interest areas, while colder colors (blue) indicate low-interest areas. Figure 4 shows the heatmap results. In dataset 1, the model focuses on the teacher for 'Ind' and on the students for 'Teach'. For dataset 2, the focus areas are less clear, but the accuracy remains high.

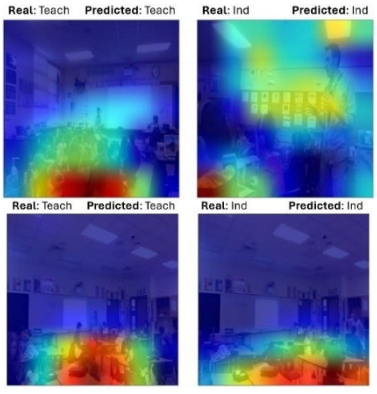

Fig. 4. Heat map showcasing the regions of focus by the model.

B. Temporal Analysis

To further evaluate the model's classification capabilities, we extended the input dimensionality to include video sequences instead of individual frames and adjusted the model architecture accordingly. Specifically, we applied the main pipeline (shown in Figure 3) to process video inputs while

excluding audio data. This adaptation allowed the model to capture temporal dynamics within the video sequences.

We applied this temporal approach to dataset 1 and achieved a validation accuracy of 95.4%. For dataset 2, the model attained a validation accuracy of 90.6%. These results indicate that incorporating temporal information from video sequences lead to a dip in performance compared to the spatial results. We conjecture that the performance decline observed in the temporal model is due to the relatively low complexity of the classification task, which involves only two classes. This simplicity does not fully leverage the increased dimensionality of the input data and the advanced capabilities of the model.

C. Multimodal Analysis

To further enhance the model's performance, we integrated audio data into the temporal model, creating a multimodal approach as depicted in Figure 3. This architecture allows the model to utilize both visual and auditory information from the video segments, providing a more comprehensive analysis of the instructional formats. With the inclusion of audio data, the model achieved significant improvements in accuracy relative to the temporal model and provides some seemingly data dependent improvement compared to the spatial model. For dataset 1, we obtained a validation accuracy of 98.3%, while for dataset 2, we achieved a validation accuracy of 96.6%.

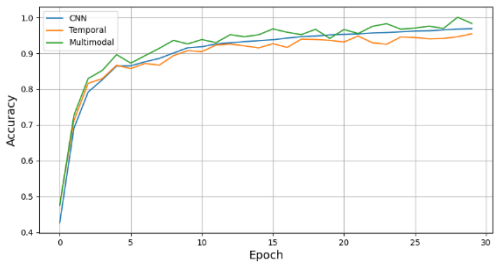

Fig. 5. Validation acuracy graph for dataset 1

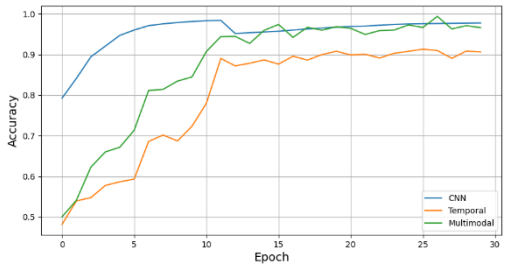

Fig. 6. Vadliation accuracy graph for dataset 2

Figures 5 and 6 illustrate the combined validation accuracy plots from all three experiments for each dataset. The X-axis represents the number of epochs, which is the number of times the model was trained on the entire dataset, while the Y-axis indicates the mean accuracy of all predictions made on the validation set after each epoch of training. The plots show that the highest or near-highest accuracy was consistently observed when both audio and video data were utilized together.

V. DISCUSSION

In this WIP study, we developed advanced AI computer vision and multi-modal algorithms to analyze classroom observation data from K-8 computer science classrooms. We examined the accuracy of these algorithms' predictions regarding instructional formats using a validation dataset.

Our results showed modeling both audio and video data achieved the highest accuracy, compared to using only image frames or standalone video data. We conjecture that the high starting accuracy of the CNN model for dataset 2, as seen in Figure 6, may be attributed to easily learnable spatial features and a static camera, unlike dataset 1. This likely explains why the CNN model outperforms the multimodal approach for this dataset. Further testing on a variety of datasets will illuminate the best applications of this approach. Combining audio and video data leverages additional information that can be more informative in complex classification tasks which may lead to discrepancy in results based on dataset characteristics (e.g., student collaboration and engagement in learning). Nevertheless, using complementary modalities shows potential to improve the classification of instructional formats compared to using only image frames or standalone video data. This underscores the importance of multimodal data for capturing classroom dynamics. Our findings validate the proposed framework and highlight the effectiveness of AI-based classification in educational settings.

The pilot results highlight the potential of using AI to analyze real-time CS classroom instruction. Incorporating AI can enhance the depth and accuracy of learning analytics, providing educators with actionable insights that are directly relevant to their specific classroom environments. Real-time, automated analysis can help teachers understand nuanced information about the teaching and learning process and provide a foundation for reflecting on effective instructional decisions tailored to their specific class. Ultimately, this approach is expected to improve the quality of CS instruction, thereby helping all students effectively achieve their learning goals.

In particular, the automated classroom observation results can significantly benefit CS teachers with limited personal experience and training, especially those working in high-needs environments such as rural, low-income, and diverse schools, where quality CS education and resources are limited. For example, CS teachers in rural school districts with only one CS teacher, where peer observation and feedback are not possible, will greatly benefit from the results of the automated observation analysis. Similarly, CS teachers in diverse schools can use the analysis results to identify notable individual differences in learning CS among different student subgroups and adjust their teaching strategies to meet the unique needs and learning behaviors of their students.

Finally, we expect that AI-driven analysis of classroom observation video data opens new horizons for educational research. Researchers can identify patterns of effective or ineffective teaching and learning strategies, and explore nuanced relationships between teaching practices, student learning behavior, and academic outcomes captured from the massive amount of classroom observation data. This streamlined, efficient, and nuanced analysis has the potential to advance data-driven pedagogy by enabling the continuous examination and refinement of pedagogical theories through real-time and contextually informed analytics.

The findings from the current study provide direction for the next steps of our project. The immediate next step will be to validate the proposed solution using a larger dataset with a more diverse range of classroom settings and instructional practices. We will also continue to analyze more complex constructs, such as student engagement and teacher instructional strategies.

Our long-term project goal is to develop a comprehensive AI-based classroom observation system that captures and analyzes video recordings of diverse interactions in K-12 CS classrooms. The system processes video data to generate the reports on teacher instruction (e.g., feedback quality, explanations, questioning strategies), and student learning processes and behaviors (e.g., seeking help, collaboration, engagement). It then provides real-time, personalized instructional recommendations to teachers via a teacher dashboard. Ultimately, this tool aims to enhance students' proficiency in computer science while improving instructional capacity for K-12 CS teachers, particularly in high-needs schools, including rural and racially diverse areas.

REFERENCES

[1] F. Martinez, S. Taut, & K. Schaaf, “Classroom observation for evaluating and improving teaching: An international perspective,” *Studies in Educational Evaluation*, vol. 49, pp. 15-29, June 2016.

[2] D. Gitomer, C. Bell, Y. Qi, D. Mccaffrey, B. K. Hamre, & R. C> Pianta, “The Instructional Challenge in Improving Teaching Quality: Lessons from a Classroom Observation Protocol,” *Teachers College Record*, Vol. 116(6), pp. 1-32, June 2014.

[3] A. Moreno and T. Redondo, "Text analytics: the convergence of big data and artificial intelligence," *IJIMAI*, vol. 3, no. 6, pp. 57-64, Dec 2016.

[4] Z. S. Syed, F. K. Shaikh, M. S. S. Syed, and A. Syed, "Visual Analytics for Automated Behavior Understanding in Learning Environments: A Review of Opportunities, Emerging Methods, and Challenges," in *Intelligent Image and Video Analytics*, El-Alfy, El-Sayed M, Bebis, George, Zhou, Mengchu, Ed. Boca Raton: CRC Press, 2023, pp. 221-250, 2023.

[5] G. Colvin, K. Flannery, G. M. Sugai, and J. Monegan, "Using Observational Data to Provide Performance Feedback to Teachers: A High School Case Study," *Preventing School Failure: Alternative Education for Children and Youth*, vol. 53, no. 2, pp. 104-115, Apr. 2009.

[6] M. M. MacKinnon, "Using observational feedback to promote academic development," *International Journal for Academic Development,* vol. 6, no. 1, pp. 21-28, Jan. 2001.

[7] P. M. Pianta, B. K. Hamre, and J. T. Downer, "How does classroom quality relate to program and teacher quality in sustaining effects over time?," *Early Childhood Research Quarterly*, vol. 36, no. 4, pp. 239-248, Oct. 2016.

[8] K. L. Rauch and C. R. Whittaker, "Observation and Feedback during Student Teaching: Learning from Peers," *Action in Teacher Education*, vol. 21, no. 3, pp. 67-78, Sep. 1999.

[9] M. L. Hemmeter, J. K. Hardy, A. G. Schnitz, J. M. Adams, and K. A. Kinder, "Effects of training and coaching with performance feedback on teachers’ use of Pyramid Model practices," *Topics in Early Childhood Special Education*, vol. 35, no. 3, pp. 144-156, Sep. 2015.

[10] T. Dragon, I. Arroyo, B. P. Woolf, W. Burleson, R. El Kaliouby, and H. Eydgahi, "Viewing student affect and learning through classroom observation and physical sensors," in *Proc. 9th Int. Conf. Intelligent Tutoring Systems (ITS 2008), June 23-27, 2008, Montreal, Canada* [Online]. Available: IEEE Xplore,. https://link.springer.com/chapter/10.1007/978-3-540-69132-7_8.

[11] D. M. Florescu and L. E. Ciolan, "The impact of classroom observation in early education on children’s outcomes," *Journal of Pedagogy*, vol. 1, pp. 217-236, Jan. 2023.

[12] M. Korban, S. Singh, P. Youngs, G. S. Watson, and S. T. Acton, "AI-assisted activity detection in K-6 classroom environments: A preliminary framework to assist in pedagogical performance evaluation," in *Proc. 2021 55th Asilomar Conf. Signals, Systems, and Computers, October 31-November 3, 2021, Pacific Grove, CA, USA* [Online]. Available: IEEE Xplore, https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9723283&casa_token=LJx1Rona1ykAAAAA:pIYzbeNFcw8keQDp1kVdt-9EzcSQM5BDbAqVS046PnXpU-bClMaNNARPmQsrpZaQsKK-0w4oRQ.

[13] S. Shapsough and I. Zualkernan, "Using machine learning to automate classroom observation for low-resource environments," in *Proc. 2018 IEEE Global Humanitarian Technology Conference (GHTC), October 18-21, 2018, San Jose, CA, USA* [Online]. Available: IEEE Xplore, https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8601902&casa_token=zqVShEOU4MgAAAAA:a4mQRMu2aAr_OulEV6OHR-cYVlAERqaCcG6FOc0m_T3ZJyA_X-IQqYZRambOLun-phOt97xFWQ.

[14] N. Daksith, S. Wanaguru, U. Kolonne, A. Dolawatta, L. Abeywardhana, and D. Kasthurirathna, "Multi Model Approach to Evaluate and Enhance Student-Teacher Interactions," in *Proc. 2023 5th International Conference on Advancements in Computing (ICAC), December 7-9, 2023, Battaramulla, Sri Lanka* [Online]. Available: Xplore, https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10417599&casa_token=jzgnxVw6CYEAAAAA:xkkckzaGWFkhY06KgVaM74kBqLwNI1FExUIPoC7lcdLVNClJBfTjVfnVZ0iGSNaclYEiLtZYiw.

[15] Y. Hu, R. F. Mello, and D. Gašević, "Automatic analysis of cognitive presence in online discussions: An approach using deep learning and explainable artificial intelligence," *Computers and Education: Artificial Intelligence*, vol. 2, p. 100037, May 2021.

[16] Z. Trabelsi, F. Alnajjar, M. M. A. Parambil, M. Gochoo, and L. Ali, "Real-time attention monitoring system for classroom: A deep learning approach for student’s behavior recognition," *Big Data and Cognitive Computing*, vol. 7, no. 1, pp. 48, Jan. 2023.

[17] J. Yu, Z. Li, Z. Liu, M. Tian, and Y. Lu, "A Student-Teacher Multimodal Interaction Analysis System for Classroom Observation," in *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky, AIED 2023, July 3-7, 2023, Tokyo, Japan*, N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, and O. C. Santos, Eds. Sringer: Cham, 20023. pp. 193-199.

[18] A. Ramakrishnan, B. Zylich, E. Ottmar, J. LoCasale-Crouch, and J. Whitehill, "Toward automated classroom observation: Multimodal machine learning to estimate class positive climate and negative climate," *IEEE Trans. Affect. Comput*., vol. 14, no. 1, pp. 664-679, Jan. 2021.

[19] M. Li, M. Liu, Z. Jiang, Z. Zhao, J. Zhang, M. Ge, H. Duan, and Y. Wang, "Multimodal Emotion Recognition and State Analysis of Classroom Video and Audio Based on Deep Neural Network," *J. Interconnect. Networks*, vol. 22, pp. 2146011:1-2146011:20, Mar. 2022.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*., June 26-July 1, 2016, *Las Vegas, USA* [Online]. Available: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf.

[21] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis*., October 22-29, 2017, Venice, Italy [Online]. Available: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://openaccess.thecvf.com/content_ICCV_2017/papers/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.pdf.