AI-Based Automatic Detection of Online Teamwork Engagement in Higher Education

Alejandra J. Magana, Syed Tanzim Mubarrat, Dominic Kao, and Bedrich Benes, *Senior Member, IEEE*

***Abstract*—Fostering productive engagement within teams has been found to improve student learning outcomes. Consequently, characterizing productive and unproductive time during teamwork sessions is a critical preliminary step to increase engagement in teamwork meetings. However, research from the cognitive sciences has mainly focused on characterizing levels of productive engagement. Thus, the theoretical contribution of this study focuses on characterizing active and passive forms of engagement, as well as negative and positive forms of engagement. In tandem, researchers have used computer-based methods to supplement quantitative and qualitative analyses to investigate teamwork engagement. Yet, these studies have been limited to information extracted primarily from one data stream. For instance, text data from discussion forums or video data from recordings. We developed an artificial intelligence (AI)-based automatic system that detects productive and unproductive engagement during live teamwork sessions. The technical contribution of this study focuses on the use of three data streams from an interactive session: audio, video, and text. We automatically analyze them and determine each team's level of engagement, such as productive engagement, unproductive engagement, disengagement, and idle. The AI-based system was validated based on hand-coded data. We used the system to characterize productive and unproductive engagement patterns in teams using deep learning methods. Results showed that there were $>91\%$ prediction accuracy and $<7\%$ mismatches between predictions for the three engagement detectors. Moreover, Pearson's r values between the predictions of the three detectors were >0.844. On a scale of -1 (unproductive engagement) to 1 (productive engagement), the scores for all teams were 0.94 ± 0.04, suggesting high productive engagement. In addition, teams tended to mostly be in productive engagement before transitioning to disengagement ($>90.34\%$ of the time) and to idle ($>93.69\%$ of the time). Before transitioning to productive engagement, we noticed almost equal fractions of teams being in idle and disengagement modes. These results show that the system effectively detects engagement and can be a viable tool for characterizing productive and unproductive engagement patterns in teamwork sessions.**

***Index Terms*—Artificial intelligence (AI), engagement, higher education, teamwork.**

I. Introduction

TEAMS are critical organizational structures for the success of any organization [1]. Consequently, higher education institutions are expected to deliver graduates capable of engaging in teams productively [2]. Although teamwork facilitation should be a necessary part of good undergraduate instruction [3], it is often difficult to implement teamwork practices in the context of large classes at the undergraduate level.

For instance, teamwork pedagogy has been implemented to orchestrate teamwork processes and to facilitate group work [4]. Specifically, in Science, Technology, Engineering, and Mathematics classrooms, teamwork pedagogy, such as cooperative learning [5], problem-based learning [6], and frame-of-reference training [7], among others, has been widely adopted for orchestrating team-based projects for an extended period of time, ranging from weeks to an entire semester. However, some research reporting on implementing teamwork pedagogies in large classes has reported a lack of success [8], [9]. This could be partly attributed to instructors having limited experience operating in teams or not knowing when students need support to overcome unproductive behaviors when engaged in teamwork interactions. Thus, instructors need training, guidance, and tools to help them teach and assess teamwork skills in their students effectively [10]. Furthermore, even when teamwork orchestration approaches are effectively implemented by instructors when providing upfront guidance in preparation for team-based projects, effective teamwork from a learner perspective requires team members to experience productive engagement during teamwork interactions [1], [11].

Research has determined that fostering productive engagement within teams is important as it can improve learning outcomes for students [12]. When students work in teams, they can share ideas and work collaboratively to solve problems as they learn from each other [13]. Similarly, research has identified that negative teamwork behaviors can be detrimental to overall team performance [14]. Thus, instructors need to know when students are disengaged or engaged in negative behaviors as they engage in teamwork interactions, so they can react and provide assistance or guidance accordingly. However, researchers have also identified the challenge of understanding when students are engaged in productive or unproductive behaviors [15]. Thus, research that characterizes productive and unproductive time during teamwork sessions is a critical preliminary step to increase

Received 22 September 2023; revised 19 February 2024 and 15 August 2024; accepted 3 September 2024. Date of publication 9 September 2024; date of current version 19 September 2024. This work was supported by the U.S. National Science Foundation under Award 2113991, Award 2219271, and Award 2417510. *(Corresponding author: Alejandra J. Magana.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the IRB-2021-1181.

Alejandra J. Magana is with the Department of Computer and Information Technology and the Department of Engineering Education, Purdue University, West Lafayette, IN 47907 USA (e-mail: admagana@purdue.edu).

Syed Tanzim Mubarrat and Dominic Kao are with the Department of Computer and Information Technology, Purdue University, West Lafayette, IN 47907 USA.

Bedrich Benes is with the Department of Computer Science, Purdue University, West Lafayette, IN 47907 USA.

Digital Object Identifier 10.1109/TLT.2024.3456447

engagement in teamwork meetings and productive time [16]. Furthermore, by characterizing the different forms of teamwork engagement, tools, such as dashboards [17], [18], [19], as well as mediating strategies [20], such as conflict resolution interventions, may provide instructors with means to support learners as they work in teams more productively. Such characterization may also provide instructors with a means to assess teamwork processes [21].

This study aims to characterize teamwork as productive and unproductive engagement during teamwork sessions among undergraduate students. Based on this characterization, we have developed an artificial intelligence (AI)-based automatic detection system. Our developed system can automatically detect teamwork engagement by analyzing audio, video, and text data streams from online teamwork sessions. Our approach uses the Russell diagram [22] that, in this context, relates the team members' valence and alertness, allowing us to characterize engagement. The AI-based system was validated on hand-coded qualitative data and allowed us to characterize teamwork engagement across multiple semesters without hand-coding.

The initial research question of our research is (RQ1) *"What are forms of productive and unproductive teamwork engagement in the context of an undergraduate course?"* Once characterizations of productive and unproductive engagement have been identified, the second goal of this study is to propose a computer-based approach to characterize forms of engagement during teamwork sessions. The second research question is (RQ2) *"What is the level of accuracy of a computer-based approach for characterizing productive and unproductive teamwork engagement?"* Finally, our proposed computer-based approach was applied to 85 additional teamwork sessions, allowing us to pursue our third goal of describing the forms of engagement enacted by such teams. Our third research question is (RQ3) *"What are the different ways in which teams of undergraduate students enacted productive and unproductive engagement during teamwork sessions?"*

Findings from the study include a detailed qualitative characterization of four different forms of teamwork engagement: 1) productive engagement; 2) unproductive engagement; 3) disengagement; and 4) idle. To determine if our AI-based automatic detection system could identify the four different forms of engagement, we used three streams of data from an interactive session (audio, video, and text). Results showed that there were $> 91\%$ prediction accuracy and $< 7\%$ mismatches between predictions for the three engagement detectors. Moreover, Pearson's r values between the predictions of the three detectors were > 0.844. On a scale of -1 (unproductive engagement) to 1 (productive engagement), the scores for all teams and teamwork sessions encompassing the full dataset were 0.94 ± 0.04, suggesting high productive engagement. In addition, teams tended to mostly be in *productive engagement* before transitioning to *disengagement* ($> 90.34\%$ of the time) and to *idle* ($> 93.69\%$ of the time). Before transitioning to *productive engagement*, we noticed almost equal fractions of teams being in *idle* and *disengagement*.

II. Related Work

Research on teamwork has been thoroughly investigated in the context of organizational psychology [23], as well as in collaborative learning [24]. Such research suggests that teamwork is a complex construct combining static and dynamic aspects of teams. For instance, a substantial body of work has focused on how static characteristics of teams, such as team members' backgrounds and expertise, elements of team formation, or developmental stages, among other features, affect team performance. Relatively less empirical work has examined the dynamic aspects of teams, such as those occurring during working sessions [25]. Thus, the scope of the study is centered on one of the dynamic aspects of teamwork, herein, teamwork engagement in academic settings. Research focused on dynamic aspects of teamwork has determined that productive teamwork engagement requires team members to experience behavioral synchronization processes during teamwork interactions [1], [11]. Behavioral processes consist of members' interdependent actions that translate efforts to outcomes via cognitive, verbal, and behavioral activities for applying knowledge and skills, organizing tasks toward achieving collective goals, and engaging in planning, goal setting, and coordinating processes [26]. In the context of higher education, these behavioral processes can be characterized as different forms of engagement. Engagement refers to the academic investment, motivation, and commitment students demonstrate within contexts associated with their educational formation [27]. Academic engagement has been linked to educational benefits, such as persistence, retention, and achievement [28].

A. Characterizing Teamwork Engagement

Educational psychologists, organizational psychologists, and education researchers have performed studies to characterize teamwork engagement during meetings. Findings from most of this body of work suggest that team engagement is a precursor of team performance [29]. These findings are supported by the idea that learning collaboratively can be more effective than learning individually [30], [31]. Thus, characterizing and measuring team engagement is an important area of research to support students effectively while they learn with their peers. Equally important is to understand and characterize unproductive engagement, as it negatively influences group performance [32]. Understanding when teamwork fails is equally critical [33], as negative experiences and unproductive behaviors can be detrimental to the learning process and cause unproductive conflict and social anxiety, among other things, negative behaviors.

However, a substantial body of work primarily focuses on generating quantitative team performance and engagement measures. One common metric is the number of artifacts or products teams generate [29]. Another common metric has focused on measures of team functioning [34], placing greater emphasis on the interaction processes than the outcomes [29].

Researchers have also switched methods for investigating teamwork engagement by implementing mixed-methods approaches. For instance, a study by Frank et al. [16] combining

surveys and qualitative observations characterized teamwork engagement in terms of conversational turn-taking, the topic of the discipline (e.g., architecture, engineering, etc.), the type of interaction (cocreation, presentation, negotiation, etc.), and the artifacts participants used or created during their meetings (e.g., sketches, schedules, models, etc.). Their findings indicated that social interaction during project meetings, including relationship building, is related to high levels of teamwork engagement. An important finding included using or generating shared artifacts, which supported collaboration.

B. Methods for Teamwork Engagement Detection

The use of computer-based methods to supplement quantitative and qualitative analyses in the context of teamwork research has recently been the focus of attention from researchers [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48]. A body of work has focused on teamwork interaction via discussion forums. For instance, a recent study from Zheng et al. [35] used computer-based methods to supplement qualitative and quantitative methods for analyzing discussion forum data to identify when learners produce off-topic information and variation in the level of engagement during online teamwork sessions. Wu [36] used machine learning (ML) methods to supplement qualitative methods to identify off-topic messaging. Another example is the study by Wong et al. [37], who characterized social interactions and contextual topics utilizing visualizations of online discussion forums to identify patterns in the data.

Another body of work has focused on detecting engagement from video data, either in a classroom or in an online environment. In the context of classroom environments, Xu et al. [38] utilized the interactive, constructive, active, passive, disengage (ICAPD) framework [49] and a simAM-YOLOv8n model to detect student cognitive engagement by combining visual behaviors with cognitive states. Pabba and Kumar [39] introduced a real-time system for monitoring student group engagement in large classrooms, which analyzed facial expressions to recognize academic affective states, such as boredom, focus, and frustration. Mehta et al. [40] proposed a 3-D DenseNet self-attention neural network for the automatic detection of student engagement in modern and traditional educational programs. Guhan et al. [41] proposed the Multimodal Perception of Engagement for Telehealth algorithm, a learning-based approach leveraging latent vectors corresponding to affective and cognitive features to estimate patient engagement levels during telehealth sessions. Zaletelj and Košir [50] employed ML algorithms to predict students' attention in classrooms using facial and body features, including gaze point and body posture. Mo et al. [46] proposed a multitask learning approach combining human pose estimation and object detection to recognize student behavior automatically. Hu et al. [47] constructed a noninvasive classroom learning engagement database using videos and designed a bimodal method based on ResNet50 and CoAtNet for recognizing learning engagement.

In the context of online learning environments, Levordashka et al. [43] developed a web-based application that uses real-time face tracking via a webcam to measure cognitive engagement in individuals. At the same time, they watched streamed theater content at home. Bosch and D'Mello [44] also used facial features to automatically detect mind wandering, where attention drifts from the current task to unrelated thoughts. In the laboratory, they analyzed face videos at six levels of granularity, including head pose, facial textures, and facial action units. Li et al. [45] employed supervised ML algorithms to predict students' cognitive engagement states based on their facial behaviors while solving clinical reasoning problems in an intelligent tutoring system. Bhardwaj et al. [48] employed deep transfer learning to predict student engagement from facial image data in the context of online courses.

Based on the previous work, it was identified that some of these studies have only focused on characterizing engagement using information extracted from discussion forum postings, while other studies focused on video data have developed methods for detecting a limited number of emotion ranges, such as only engagement and boredom. Moreover, some studies have focused on live classrooms while ignoring online team-based settings.

Based on findings from previous work, it was also identified that the majority of these aforementioned studies utilized ML and/or deep learning (DL) techniques for engagement detection, suggesting the high effectiveness of such methods. However, these studies mostly rely on a single modality of data (mainly facial images or facial expressions) for engagement detection. In a complex classroom environment (both online and offline), a student's face may not always be visible due to obstructions and/or video stream latency [47]. Moreover, video observation and face analysis usually require high-quality images and apply to single-person observations [50]. Therefore, focusing on only facial expression data limits the usability or reduces the accuracy and available complexity of engagement detection [50] in the classroom setting. Other modalities of data, such as text [51], [52], [53], [54], [55] and audio [56], [57], [58], have shown promise in sentiment analysis and emotion recognition. Consequently, there is a growing need to investigate student engagement using multiple modes (such as text transcriptions and audio conversations). Thus, other DL-based methods, combining video, text, and audio data, are needed to identify productive and unproductive engagement during live teamwork working sessions. This is the primary focus of our study.

C. DL Algorithms for Feature Extraction

Researchers have extensively utilized DL techniques to extract features from video, audio, and textual data [51], [55], [56], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70]. For instance, various DL models and algorithms have been used to analyze spatiotemporal features from video and motion data, such as 3-D convolutional neural networks (CNNs) [59], [60], convolutional long short-term memory networks [61], attentional convolutional network [62], multi-modal residual perceptron network (MRPN) [63], fully connected convolutional neural networks (FC-CNNs) [64], and residual neural networks [65]. For textual data streams, FastText [66], gate recurrent unit

model [56], graph convolutional networks [51], EmoTxt [67], Emotion-Sensitive TextRank [55], MRPN [63], and CNNs [68] have been utilized for feature extraction and label prediction. Finally, researchers have used DenseNet201 [69], bidirectional long short-term memory model [56], and OpenSMILE [70] for feature extraction and label prediction from audio data.

Among these studies, the approach taken by Zhou et al. [64] of using FC-CNN to develop a detector that provided a probability indicating the likelihood of an emotion being associated with one of the four quadrants in the Russell diagram from facial expressions informs our approach. In addition, DenseNet-201 has been extensively used for audio classification applications [71], [72], [73], [74], [75]. In a recent systematic literature review on DL-based audio classification techniques [74], the authors found that an approach based on DenseNets achieved the best performance [75]. Finally, the *fasttext* text classifier represents sentences as a bag of words and trains a simple linear classifier with rank constraints. To decrease the computational complexity and improve the real-time applicability, *fasttext* uses a hierarchical softmax [76] based on the Huffman coding tree [77]. This allows the *fasttext* text classifier to classify textual data very fast with minimal reduction in accuracy. As a result, we chose these three approaches for developing our DL-based engagement detectors.

III. Conceptual and Methodological Frameworks

This study used a multimethod approach to pursue the three defined research questions, and details of the methods are presented in the following section. The combination of qualitative, quantitative, and computational approaches was deemed adequate as integrative research has concluded that combining multiple methods can dramatically increase the accuracy and quality of any research's analysis and conclusions [78]. Specifically, to answer our first research question, we implemented a qualitative approach, whose findings were then quantized to then implement computational approaches to respond to the second and third research questions.

For the qualitative component of the study, we implemented content analysis for analyzing the data as it is a structured yet flexible approach [79]. Content analysis is appropriate when the aim of the research is to interpret meaning from the content of the data, thus uncovering characteristics inherent to the phenomenon under investigation, in this case, teamwork engagement. Although research in cognitive science has identified different levels of student engagement during learning [49], recent empirical work in the context of classrooms and within online team interactions [38], [80], [81] has also identified disengagement and unproductive engagement behaviors. Thus, we needed a conceptual framework that guided us to better uncover and characterize productive and unproductive forms of engagement or disengagement. For this, we used the Russell diagram (see Fig. 1) to guide our exploration. The Russell diagram [22] is often used to map emotional states into a 2-D plane according to two principal dimensions: 1) a dimension related to valence following a negative $\leftrightarrow$ positive continuum on the x-axis and 2) a dimension related to alertness following a deactivated $\leftrightarrow$ activated continuum on the y-axis (see Fig. 1). Various emotions can be situated in the 2-D plane due to a linear combination of the two variables [82]. Each emotion can then be mapped onto the outer edge of the Russell diagram, e.g., satisfaction (see the lower right quadrant in Fig. 1) can be classified as positive and deactivated, whereas frustration (upper left quadrant) as negative and activated.

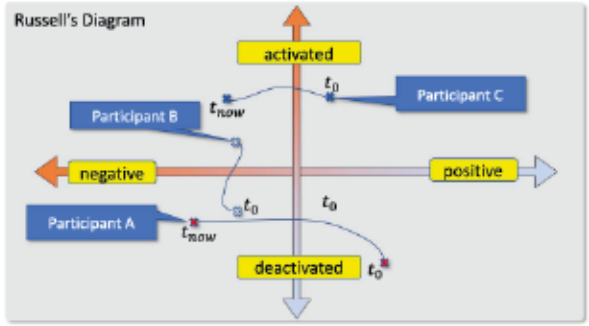

Fig. 1. Russell diagram maps an engagement state to a point in a 2-D plane according to activation and type. Three participants are displayed, and the changes in their engagement state are visualized as trajectories over time from t_0 to t_{now}.

In the context of technology-enhanced learning environments, the Russell diagram has been used to detect how learners' emotions evolve during learning while interacting with technology and to provide feedback to improve the learning experience [83]. Applications have also been used in intelligent tutoring systems, virtual agents [84], affective computing, and AI in education to elicit positive emotions [85].

Research on emotions has been associated with student engagement. For instance, positive emotions during the learning process have been directly linked to student engagement [86], [87]. Given that 1) research on emotions has focused on activated and deactivated, as well as positive and negative aspects of emotional states and 2) recent empirical research on engagement has gone beyond productive forms of engagement [49] to also characterize unproductive forms of engagement [38], [80], we deemed necessary to contextualize our research in a broader view of teamwork engagement. Specifically, it was deemed necessary to characterize engagement considering different levels of being activated or deactivated, aligning with the activation dimension of the Russell diagram, as well as positive and negative forms of productive and unproductive engagement, aligning with the valence dimension of the Russell diagram. Thus, this study used Russell diagram dimensions of valence and activation as guidance to characterize different forms of student engagement and disengagement.

Once we characterized different forms of productive and unproductive teamwork engagement, we used our findings to develop a system that uses DL-based automatic engagement classification. Specifically, we detected the engagement state S_A of a participant A as the two extreme values on both axes $S_A = \{[negative, deactivated],\ [negative, activated],\ [positive, activated], [positive, deactivated]\}$. This quantized the 2-D plane of Russell diagram into four values corresponding to each quadrant, which was deemed adequate for this study.

For this, we used DL algorithms to automatically extract the engagement state of the team from three data streams: audio (voice), video (facial expression), and text (video transcripts). Once the system was developed and validated, we used it to expand our analysis of teamwork engagement to a larger sample of data, thus allowing us to identify patterns of interaction.

IV. Methods

A. Context and Participants

The context of the study was a second-year systems analysis and design course offered at a public university in the USA. The course was required for all undergraduate students pursuing computer and information technology majors. The course aimed to introduce students to the practice of documenting and modeling requirements of a system to be developed and then constructing and modeling and corresponding functional prototypes of the proposed system. These skills were practiced through a team-based semester-long project, coordinated following principles of cooperative learning [88], [89], [90]. There are five principles, which were implemented as follows.

1) The principle of individual and group accountability was achieved by having a portion of the semester-long project be submitted and graded individually.
2) The principle of interpersonal and small group skills was achieved by intentionally discussing and reflecting on effective teamwork skills and facilitating conflict management and conflict resolution training.
3) The principle of face-to-face promotive interaction was facilitated by devoting lecture time for teams to work on their projects.
4) The principle of positive interdependence was established by setting clear goals for each project milestone and assigning roles.
5) Finally, the principle of group processing was facilitated by having students reflect on their team processes and submit these team-based reflections along with each project milestone submission.

The project was organized into four major milestones for preparing the system documentation. The prototype was implemented in five iterations called sprints. The course was delivered twice a week (Tuesday and Thursday), and each class period lasted 75 min. The Tuesday lecture was often devoted to introducing new concepts and skills along with practice activities. In contrast, the Thursday lecture was often devoted to teamwork, where students applied the learned skills to their semester-long projects. The course has a regular enrollment of 120–150 students every semester. According to institutional data, in every given semester, the course had a population of ≈24% female students and ≈76% male students. Students were organized into teams of four or five students to work on the semester-long project.

B. Procedures and Data Collection Method

The Microsoft Teams platform facilitated the teamwork dynamics during and outside lecture time. Each team was provided with a private channel to interact and share files in the same location. On some occasions, fully online teamwork sessions were scheduled throughout the semester. Specifically, five online teamwork sessions were scheduled in weeks 3, 4, 7, 8, and 10 of the semester. The purpose of the first session was for students to get used to the Microsoft Teams platform. In that first session, students joined their respective team channels, met their team members, and discussed and signed their team contracts. The team contracts provided students with guidelines on engaging in coordination processes (e.g., internal deadlines, meeting times outside of class, preferred methods of communication, and preliminary role assignments). The online teamwork sessions were scheduled right before a major project deadline. Also, the last two weeks of the semester were facilitated entirely online, as students would mainly work on completing their final project submissions during that time.

Students were required to record two of the five online teamwork sessions, which were used as the data collection method for this study. Specifically, students were asked to record the second teamwork session (week 4), where they worked on preparing the submission of the first project milestone. Students were also asked to record the fifth online teamwork session (week 10), where they worked on the third project milestone. Each of the recordings had an average duration of 40 min. For these two sessions, students collaborated online and self-recorded their teamwork sessions using the features provided by the collaboration platform. Students were then asked to download and submit the video recordings to the course instructor via a secure file-sharing system. Each video recording consisted of screen captures of each team member's camera from the Microsoft Teams platform (focused on their facial expressions) as well as audio of the conversation between team members. All the video files were stored in a secure cloud content management and file-sharing system, with only the investigators accessing the files. The Institutional Review Board deemed the study exempted with the protocol number IRB-2021-1181. The data were collected in the Fall of 2021 and 2022 and the Spring of 2022. The total video recordings consisted of 144 files. The recordings had a duration between 15 and 90 min long. The total number of teamwork sessions used for hand-coding the data was 59, the total number used for training the AI-based automatic detection algorithm was 30 (out of the original 59), and the final total number of additional teamwork sessions analyzed was 85.

C. Data Analysis Methods

The data analysis was approached as three separate studies (see Fig. 2), each based on the previous study's findings. Study 1 was a qualitative study in which human raters applied content analysis to characterize forms of teamwork engagement from the video recordings. Study 1 approached (RQ1) *"What are forms of productive and unproductive teamwork engagement in the context of an undergraduate course?"* Then, Study 2 utilized the findings from Study 1 to develop and validate a collection of computer-based algorithms that utilized video, text, and audio data to characterize students' levels of engagement. Study 2 approached (RQ2) *"What is the level of accuracy of*

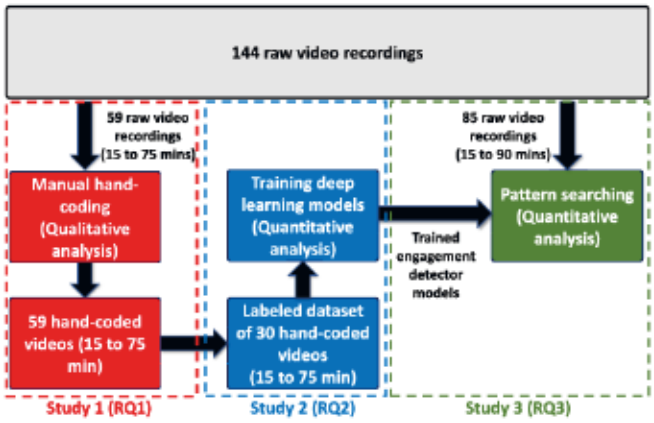

Fig. 2. Our approach is a sequence of three studies, each corresponding to one research question.

a computer-based approach for characterizing productive and unproductive teamwork engagement?" Finally, Study 3 aimed at identifying overall patterns of students' enacted levels of engagement on the entire dataset of 144 files. Study 3 approached (RQ3) *"What are the different ways in which teams of undergraduate students enacted productive and unproductive engagement during teamwork sessions?"* The specifics of how the data analysis was approached and the corresponding findings are detailed in the following sections.

1) Qualitative Human-Coded Data Analysis Method (RQ1): Study 1 (see Fig. 2) approached the first research question regarding the characterization of productive and unproductive teamwork engagement in the context of an undergraduate course. A qualitative study was performed to respond to this research question.

The analysis initiated following a content analysis [79] to answer the first research question, RQ1, regarding the characterization of engagement productivity. For this, the initial qualitative analysis was performed with a subsample of 59 video recordings from the Fall 2021 and the Spring 2022 cohorts (15–75 min). Students' names were omitted, and each team member was identified with a number. The analysis consisted of a cyclical three-step procedure of chunking the data, annotating the data, and categorizing the data. Chunking the data involved defining the unit of analysis. Thus, the analysis was performed at the student level and in 2-min intervals. Annotating the data involved the generation of a code with the identified level of engagement and a qualitative description of the code. Finally, when categorizing the data, the codes were distributed across the four quadrants of the Russell diagram.

Since some of the codes and categorizations of the data were debatable, the initial codebook was developed and validated among three raters, all with expertise in qualitative research to verify the trustworthiness of the codebook. For this, the first step of the qualitative analysis was ensured by involving the three raters in coding 20% of the data together. In this stage, discrepancies in the coding process were discussed and agreed upon with specific arguments and justifications. For example, the code "ignoring others' comments or suggestions" could have been categorized as disengaged or idle as a given team member was unresponsive. However, the categorization was decided as "unproductive engagement" since the resulting action within that same 2-min interval was other team members needing a response in order to continue working. Similarly, if a student's behavior was coded as "being late for the group meeting," it might still contribute to some "productive engagement" later in the meeting; however, in that 2-min interval, other team members had to wait for that team member in order to initiate a task, thus resulting in unproductive behavior.

The raters underwent three analysis cycles to ensure inter-rater reliability was strong for the first 20% of the data. In the first round, all raters analyzed five videos individually, after which they came together to discuss similarities, differences, and areas of improvement for the coded data. In the second round of analysis, the raters jointly recoded the same data to ensure that everyone agreed with the comprehension of the coding process. In the third and final round of analysis, the raters coded the data individually again and arrived at an inter-rater reliability score of 88%. The other 80% left of the data was coded single-handedly by two raters. Once the initial codebook was validated, two trained raters individually coded an additional set of 29 videos. The coding of this second set of videos ensured the trustworthiness of our approach.

The coding process was initiated by watching each video and identifying codes of behaviors describing students' actions. Examples of coded actions included behaviors directly related to learning, such as discussing or brainstorming ideas, methods, or approaches to solving a problem, asking questions in one's own words, taking notes, verbally comparing information discussed in the group, and drawing analogies from learning materials. Other types of actions involved teamwork processes, such as communication of information like discussing deadlines, coordination processes involving task planning and division of labor, leadership behaviors, socialization processes, giving or seeking feedback, and helping or volunteering to do work, among others. We also categorized behaviors that involved passive actions such as listening or negative actions such as disruptive behavior such as overpowering conversations, yelling at or ignoring others, not paying attention, or being distracted with other tasks. Students' initial actions and behaviors were then categorized into four major forms of engagement: productive engagement, unproductive engagement, disengagement, and idle.

Study 2 (see Fig. 2) approached the study's second research question aimed at identifying the level of accuracy of a computer-based approach for characterizing productive and unproductive teamwork engagement. For this, a set of algorithms was implemented and used to detect teamwork engagement. The findings from Study 1 were used to guide and validate Study 2.

2) DL-Based Data Analysis Method (RQ2): The input to the DL-based data analysis was the video, and the output was the set of three emotional states of the team extracted from the video stream S^V, the audio stream S^A, and the text stream S^T (see Fig. 3).

The datasets for training and testing the video, audio, and text stream engagement detectors (described below) were extracted from a subsample of 30 video recordings out of the 59 videos

Fig. 3. Input to the DL-based data analysis was the video, and the output was the set of three emotional states of the team extracted from the video S^V, audio S^A, and text S^T streams.

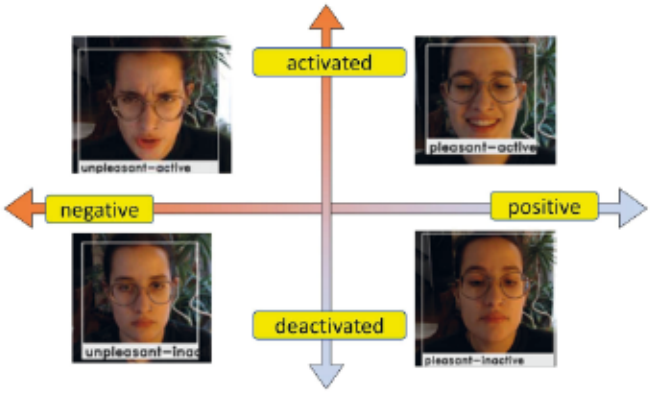

Fig. 4. Output of the DL emotion detector from video and its mapping to different quadrants of Russell diagram.

hand-coded in Study 1 (only the video recordings from the Fall 2021 cohort were used). In Study 1, students in each team in each video were classified into four different forms of engagement every 2-min interval. We extended this into a collective classification scheme for the whole team by considering the engagement category with the highest number of students in a 2-min interval as the category for the whole team. Using this classification scheme, we labeled the datasets into four classes, namely, *productive engagement*, *unproductive engagement*, *idle*, and *disengagement*. In a tie between multiple classes, the team classification was considered to be *neutral*. We omitted them from the datasets as only 0.44% of the total number of video frames were in the *neutral* category.

The *video stream* engagement detector uses the work of Zhou et al. [64], a series of FC-CNNs trained on 853 624 manually annotated images from several databases [see Fig. 7(a)]. To increase the performance, we created our own dataset by extracting faces from each data frame from all video files using the *Face Recognition* [91] library and labeling them according to the hand-coded data [see Fig. 7(a)]. The dataset consisted of all detected faces in 30 video recordings from the Fall 2021 cohort. We considered all detected faces in a classified 2-min interval to belong to the same class. The implementation works in real time on a desktop computer, and an example of the classification is shown in Fig. 4.

For the *text stream* engagement detector, we first transcribed all 30 videos using *whisper* [92] by OpenAI to generate subtitle files for each video. Using the subtitle files, we created a text dataset containing a sentence per line and the labels based on hand-coded data [see Fig. 7(b)], similar to the video dataset. This dataset consisted of all transcribed sentences extracted from the

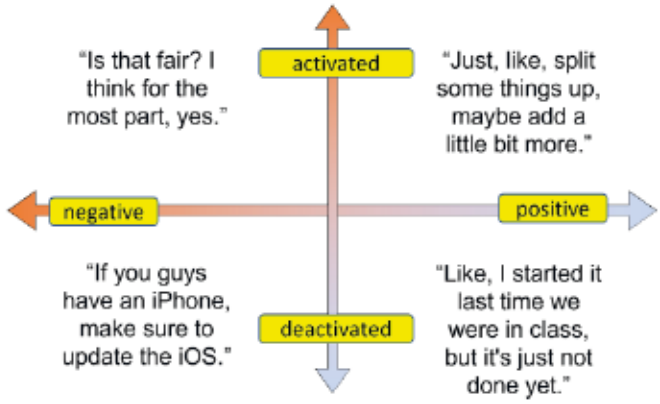

Fig. 5. Output of the DL emotion detector from transcript text and its mapping to different quadrants of Russell diagram.

30 videos from the Fall 2021 cohort. As there was a single stream of transcribed text for a single video, we considered all sentences in a classified 2-min interval to belong to the same class. Finally, we trained a supervised text classifier using the *fasttext* library [66] on our custom dataset with stochastic gradient descent and a linearly decaying learning rate. Similar to the video and audio stream engagement detector, this implementation works in real time on a desktop computer (see Fig. 5).

We developed the *audio stream* engagement detector based on the DenseNet-201 architecture [69], which is a 201-layer deep CNN with all layers connected directly with each other to ensure maximum information flow between layers in the network. First, we used *moviepy* [93] to extract audio files from the videos and *pydub* [94] to denoise and isolate the sentence utterances from the audio files (based on the timestamps on the subtitle files) [see Fig. 7(c)]. Then, we used *librosa* [95] to extract Mel spectrogram of each sentence utterance [see Fig. 7(c)]. We used these Mel spectrogram images to create an audio dataset with labels based on the hand-coded data from Study 1 [see Fig. 7(c)]. This dataset consisted of Mel spectrogram images of all sentence utterances from all 30 videos from the Fall 2021 cohort. The classification scheme for the audio dataset was the same as the text dataset. Finally, we trained a pretrained DenseNet-201 network with two additional fully connected (FC) layers on our custom audio dataset to output labels denoting the engagement levels on Russell diagram in real time (see Fig. 6).

3) Training and Validation of DL Models: As the datasets for the video stream, the audio stream, and the text stream engagement detectors were unbalanced (>90%) samples in the category of *productive engagement*), we used different data augmentation techniques to balance the datasets before training (see Fig. 7). For the video dataset, we used the *imgaug* library [96] to augment the face images, resulting in 3 551 666 samples. In addition, we used *librosa* [95], *pydub* [94], and *colorednoise* [97], [98] libraries to augment audio files by speeding them up, slowing them down, modifying the pitch, and adding various noise before extracting Mel spectrogram from them. In total, the audio dataset contained 17 157 samples. Similarly, for augmenting the sentence samples in the text dataset, we used the *textaugment* [99] and the *nlpaug* [100] libraries, resulting in 17 676 samples. Then,

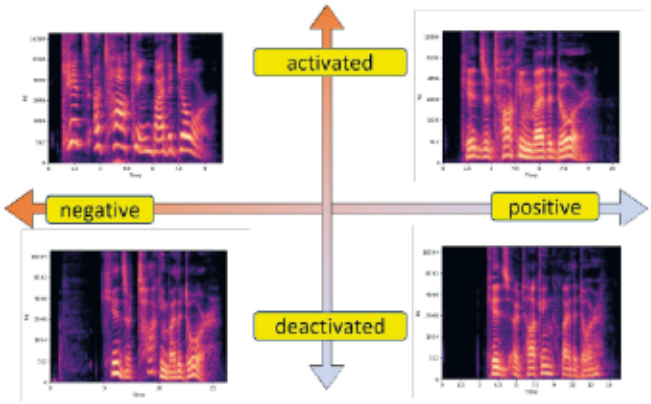

Fig. 6. Output of the DL emotion detector from audio and its mapping to different quadrants of Russell diagram.

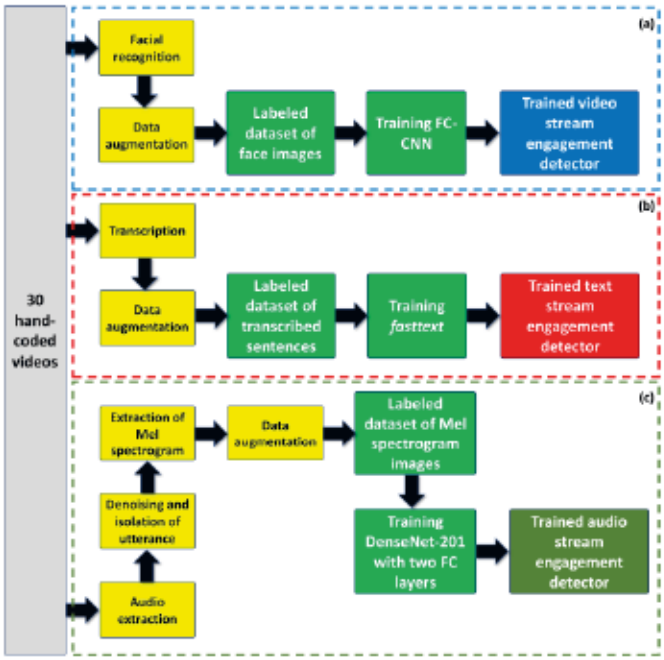

Fig. 7. Outline of the preprocessing steps, methodology, and data used for training (a) the video stream, (b) the text stream, and (c) the audio stream engagement detector.

we trained the three engagement detectors on their respective datasets with an 80:20 split (80% of the data used for training and the rest for testing). The training phase was carried out once for each engagement detector. We ran the trained models through the testing phase 50× for the video stream, 100× for the audio stream, and 100× for engagement detectors, respectively. The variation of the results was, on average, under 1%.

After the training and testing phase, we used the video stream, the audio stream, and the text stream engagement detectors to develop software using Python that parsed through a video file frame by frame, predicted the team engagement from the three streams of data, and displayed them on Russell diagram in real time (see Fig. 8). As there were multiple streams of video data for a single frame (i.e., multiple team members), the location of

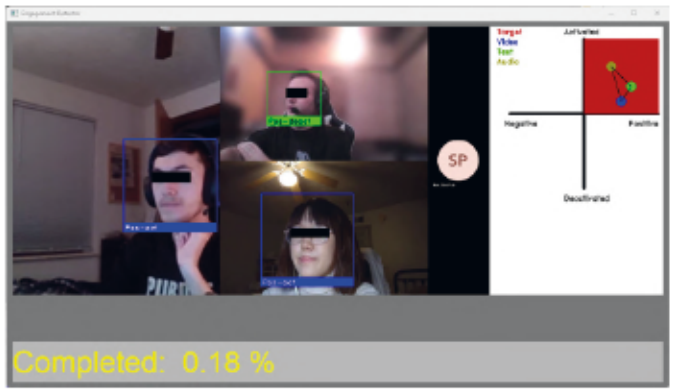

Fig. 8. Engagement detector software incorporates the detection of engagement from video, audio, and text. The detected engagements are displayed on Russell diagram in real time for each frame of data. Blue, yellow, and green dots denote the predicted engagements from video, audio, and text streams, respectively. The red square denotes the team engagement based on the hand-coded data. This figure denotes the predicted engagement for a 2-s validity interval.

the team engagement on the Russell diagram (the blue dot in Fig. 8) was calculated by averaging the predicted engagement categories of all team members detected in that particular frame. Inspection of the dataset revealed that audio clips and textual transcription data consisted of only 36% of the total number of video frames. Therefore, we employed the use of validity intervals for the audio stream and the textual interaction stream engagement detectors in the developed software. This process ensured that the engagement class predicted from a single audio clip or a single string of textual transcription data remained in the memory for a specific interval, i.e., the validity interval. We used validity intervals from 0 to 10 s with 1-s increments. As 99.46% of the total number of video frames from the dataset contained at least one recognizable face, we did not use any validity intervals for the video stream engagement detector. We used the developed software to process all 30 videos from the Fall 2021 cohorts and recorded the prediction accuracies by comparing the predictions to the actual team engagement based on the hand-coded data (i.e., the red square in Fig. 8). We determined the most suitable validity range based on prediction accuracies and mismatches between the three engagement detectors.

Study 3 approached the third research question aimed at identifying the different ways in which teams of undergraduate students enacted productive and unproductive engagement during teamwork sessions. For this, the software developed and validated in Study 2 was applied to the overall dataset to characterize the overall patterns of students' levels of engagement.

4) N-GAGE Software-Based Analysis (RQ3): We modified the software developed in Study 2 to identify levels and patterns of team engagement on videos that do not have hand-coded classification by removing the red square denoting team engagement based on hand-coded data, resulting in the N-GAGE (Processing) software (see Fig. 9). Furthermore, we added an "engagement score" bar that displays the aggregated team engagement score for each frame of data based on the predictions of the three engagement detectors (see Fig. 9). This software processed a user-selected video frame by frame, displayed the

Fig. 9. N-GAGE (Processing) software incorporates the detection of engagement from three separate data streams (video, audio, and text). The detected engagements are displayed on the Russell diagram in real time for each frame of data. Blue, yellow, and green dots denote the predicted engagements from video, audio, and text streams, respectively. The "engagement score" bar denotes the level of team engagement for each frame of the video based on the predictions of the three engagement detectors, with scores of 1, 0.5, −0.5, and −1 for *productive engagement*, *idle*, *disengagement*, and *unproductive engagement*, respectively.

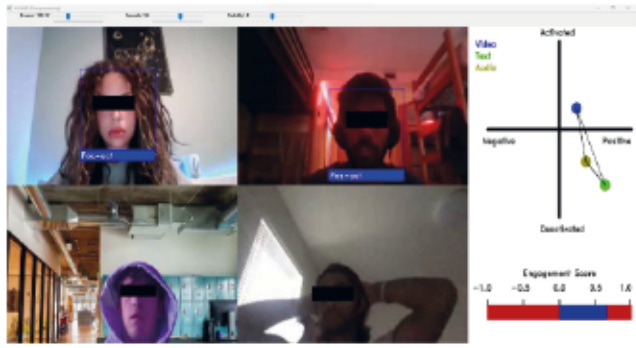

Fig. 10. Postprocessing software, with video-seeking functionality and smoothing windows of variable sizes. The smoothing is carried out using the moving average method with a user-defined window size. The team engagement from three different data streams is displayed on the Russell diagram for each data frame. Blue, yellow, and green dots denote the predicted engagements from video, audio, and text streams, respectively. The red square denotes the actual team engagement based on the hand-coded data.

results in real time, and stored the results in a data file. In addition, to assist with identifying patterns of engagement, this software also incorporated the use of validity intervals for the audio stream and the textual interaction stream engagement detectors, similar to Study 2.

For developing the "engagement score" bar, we first assigned scores for the four different engagement categories as follows: *productive engagement* = 1, *idle* = 0.5, *disengagement* = −0.5, and *unproductive engagement* = −1. Then, we assigned weights to the predictions of the video stream, the audio stream, and the text stream engagement detectors. Finally, we calculated the engagement score of a single frame of data using the following:

$$sc_i = sc_{A_i} \times w_A + sc_{T_i} \times w_T + sc_{V_i} \times w_V \tag{1}$$

where sc_i, sc_{A_i}, sc_{T_i}, and sc_{V_i} are the team engagement score, the score from audio prediction, the score from text prediction, and the score from video prediction for the ith frame of data, respectively; w_A, w_T, and w_V are the weights assigned to the audio stream, the text stream, and the video stream predictions, respectively.

As the three engagement detectors demonstrated similar levels of performance from Study 2, we assigned them equal weights (i.e., $w_A = w_T = w_V = 0.33$). Moreover, after processing all frames of a video file, we calculated the normalized team engagement (NTE) score for the team using

$$sc_{\text{norm}} = \frac{\sum_{i=1}^{F_{\text{tot}}} sc_i}{F_{\text{tot}}} \tag{2}$$

where sc_{norm} denotes the "NTE score," sc_i denotes the team engagement score for the ith frame of data, and F_{tot} is the total number of scoreable frames (frames containing data from at least one of the three data streams) in the video file.

In addition to the N-GAGE (Processing) software, we also developed software to investigate the processed videos and identify the different ways in which teams of undergraduate students enacted productive and unproductive engagement during teamwork sessions, i.e., the N-GAGE (Postprocessing) software (see Fig. 10). This software contains video-seeking functionality (using the "Frames" slider in Fig. 10) and allows the user to display results for different validity intervals (0–10 s range) (using the "Validity" slider in Fig. 10). In addition, it also implements smoothing of the change of predicted classes from the three engagement detectors using the moving average method with user-defined window sizes (using the "Smooth" slider in Fig. 10). The moving averages are calculated as follows:

$$vid_{x_i} = \sum_{j=i-k/2}^{i+k/2} (vid_{x_j}), \quad \left[i > \frac{k}{2}\right] \tag{3}$$

$$vid_{y_i} = \sum_{j=i-k/2}^{i+k/2} (vid_{y_j}), \quad \left[i > \frac{k}{2}\right] \tag{4}$$

where k is the size of the windows, vid_{x_i} and vid_{y_i} are, respectively, the x-coordinate and the y-coordinate of the predicted engagement from the video stream in the Russell diagram (i.e., the blue dot in Fig. 10). We used equations similar to (3) and (4) to smooth the predictions from the audio stream and the text stream engagement detectors.

We used the N-GAGE (Processing) software on 85 videos from the Spring 2022 and Fall 2022 cohorts and recorded the NTE scores, level of mismatches between the three engagement detectors, and patterns of changes in engagement states.

V. Results

The findings of the three studies are presented in the following sections, each responding to each of the research questions.

A. Results Characterizing Productive and Unproductive Teamwork Engagement (RQ1)

Four different forms of engagement were identified during the qualitative process, which were mapped to the Russell diagram, as shown in Fig. 11.

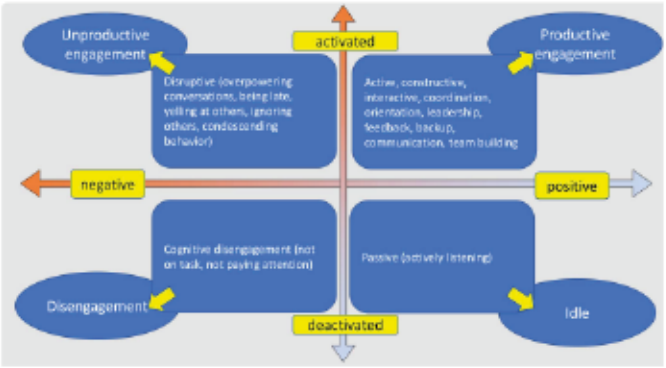

Fig. 11. Four different forms of engagement that resulted from Study 1 and their mapping to the Russell diagram.

TABLE I
PRECISION, RECALL, AND *F1*-SCORES OF THE AUDIO STREAM, THE TEXT STREAM, AND THE VIDEO STREAM ENGAGEMENT DETECTORS DURING THE TRAINING PHASE

	Accuracy	Precision	Recall	*F1*-score
Text	0.943	0.955	0.946	0.950
Audio	0.965	0.965	0.965	0.965
Video	0.968	0.906	0.981	0.904

1) *Productive engagement:* It consisted of activated and positive behaviors where students: a) worked together in creating new knowledge such as brainstorming or generating ideas; b) worked individually to produce an additional externalized output or idea like creating a diagram; and c) performed a manipulation of an artifact like taking notes or received information without overtly acting but were listening or being attentive. This form of engagement also consisted of communication and coordination processes such as planning, helping or reassuring group members, socializing, deciding roles and tasks, giving or seeking feedback, volunteering or helping each other, and deciding on meeting times.
2) *Unproductive engagement:* It consisted of activated or negative disruptive behaviors such as overpowering conversations, being late for the group meeting, yelling at others, ignoring others' comments or suggestions, and demonstrating condescending behavior.
3) *Disengagement:* It consisted of deactivated and negative behaviors such as not being on task or not paying attention.
4) *Idle:* It consisted of positive deactivated behaviors such as being passive and actively listening simultaneously.

B. Results Evaluating a Computer-Based Approach for Characterizing Engagement (RQ2)

From the training phase, the accuracies were found to be 96.81%, 96.47%, and 94.27% for the video stream, the audio stream, and the text stream engagement detectors, respectively. In addition, the precision, recall, and *F1*-scores for the three streams are shown in Table I. We ran the testing phase 100× for the video stream, 100× the audio stream, and 100× for the text stream engagement detectors and found minimal variance (under 1%) (see Table II).

TABLE II
PRECISION, RECALL, AND *F1*-SCORES OF THE AUDIO STREAM, THE TEXT STREAM, AND THE VIDEO STREAM ENGAGEMENT DETECTORS DURING THE TESTING PHASES (MEAN ± STANDARD DEVIATION), RAN FOR 100, 100, AND 50 TIMES, RESPECTIVELY

	Accuracy	Precision	Recall	*F1*-score
Text	0.931 ± 0.003	0.930 ± 0.003	0.918 ± 0.003	0.924 ± 0.003
Audio	0.874 ± 0.003	0.874 ± 0.003	0.874 ± 0.003	0.873 ± 0.003
Video	0.911 ± 0.0002	0.736 ± 0.003	0.907 ± 0.0001	0.840 ± 0.002

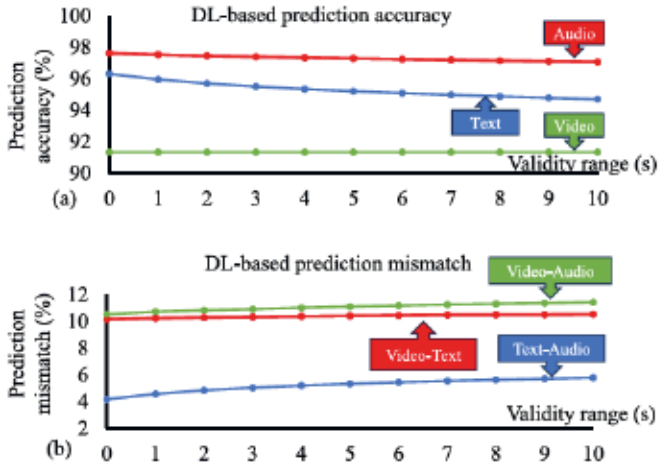

Fig. 12. (a) Prediction accuracy (%) of the audio, text, and video stream engagement detectors (top). (b) Percentage mismatch (%) between them for different validity intervals (0–10 s), averaged across the 30 videos from the Fall 2021 cohort.

TABLE III
PREDICTION ACCURACY (MEAN ± STANDARD DEVIATION) OF THE AUDIO, TEXT, AND VIDEO STREAM ENGAGEMENT DETECTORS, AVERAGED ACROSS ALL 30 VIDEOS FROM THE FALL 2021 COHORT

	Prediction accuracy Mean (%)	Std Dev
Text	96.30	4.96
Audio	97.61	4.12
Video	91.32	9.25

As 0-s validity interval showed the highest prediction accuracy for all three engagement detectors and the lowest percentage mismatches between all three engagement detectors (see Fig. 12), we used this for the remaining analysis.

Evaluation of the engagement detector software showed high prediction accuracies for all three engagement detectors (see Table III). In addition, the Pearson's *r* values between the predictions of video and audio, video and text, and text and audio stream engagement detectors were 0.844, 0.946, and 0.956, respectively. Out of all videos from the Fall 2021 cohort, 56.67%, 93.33%, and 90.00% of videos showed greater than 90% prediction accuracy for the video stream, the audio stream, and the text stream engagement detectors, respectively [see Fig. 13(a)]. Similarly, 80.00%, 93.33%, and 76.67% of videos demonstrated less than 10% mismatches between video and text, text and audio, and video and audio, respectively [see Fig. 13(b)]. Moreover, the comparison of the trend of categorized

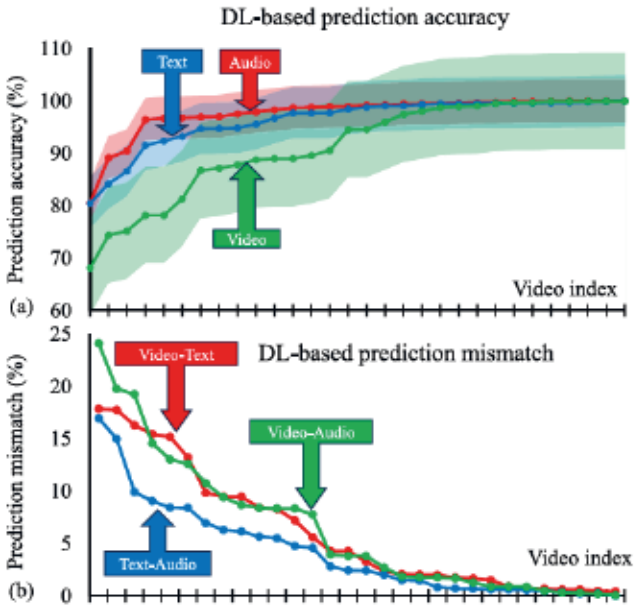

Fig. 13. (a) Prediction accuracy (%) trends of the audio, text, and video stream engagement detectors. (b) Percentage mismatch (%) trends between them across the 30 videos from the Fall 2021 cohort, ordered from low to high. The shaded areas denote the standard deviation, calculated across the 30 videos.

TABLE IV
COMPARISON OF THE TREND OF CATEGORIZED VIDEO FRAMES BETWEEN THE HAND-CODED DATA AND THE PREDICTIONS FROM THE VIDEO STREAM ENGAGEMENT DETECTOR

	Hand-coded		DL-based	
	Count	[%]	Count	[%]
Frames with face	738 704	100	738 704	100
Positive	732 134	99.11	728 776	98.66
Negative	4 901	0.66	874	0.12
Active	671 084	90.85	721 609	97.69
Inactive	65 951	8.93	8 041	1.09
Productive engagement	671 084	90.85	721 609	97.69
Unproductive engagement	0	0.00	0	0.00
Idle	61 050	8.26	7 167	0.97
Disengagement	4 901	0.66	874	0.12

TABLE V
COMPARISON OF THE TREND OF CATEGORIZED VIDEO FRAMES BETWEEN THE HAND-CODED DATA AND THE PREDICTIONS FROM THE AUDIO STREAM ENGAGEMENT DETECTOR

	Hand-coded		DL-based	
	Count	[%]	Count	[%]
Frames with audio	360,275	100	360,275	100
Positive	347,682	96.50	347,564	96.47
Negative	11,476	3.19	12,711	3.53
Active	329,241	91.39	334,732	92.91
Inactive	29,917	8.30	25,543	7.09
Productive engagement	329,241	91.39	333,703	92.62
Unproductive engagement	0	0.00	1,029	0.29
Idle	18,441	5.12	13,861	3.85
Disengagement	11,476	3.19	11,682	3.24

video frames between the hand-coded data and the predictions demonstrated high similarity for all three engagement detectors (see Tables IV–VI).

TABLE VI
COMPARISON OF THE TREND OF CATEGORIZED VIDEO FRAMES BETWEEN THE HAND-CODED DATA AND THE PREDICTIONS FROM THE TEXT STREAM ENGAGEMENT DETECTOR

	Hand-coded		DL-based	
	Count	[%]	Count	[%]
Frames with transcripts	360,275	100	360,275	100
Positive	347,682	96.50	349,252	96.97
Negative	11,476	3.19	11,023	3.06
Active	329,241	91.39	337,936	93.80
Inactive	29,917	8.30	22,339	6.20
Productive engagement	329,241	91.39	337,936	93.80
Unproductive engagement	0	0.00	0	0.00
Idle	18,441	5.12	11,316	3.14
Disengagement	11,476	3.19	11,023	3.06

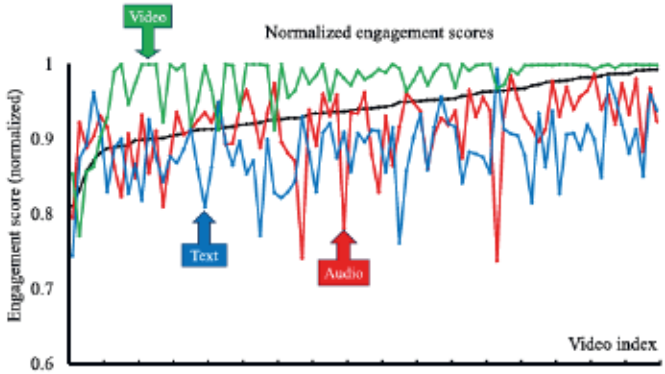

Fig. 14. Team engagement scores (normalized with respect to the total number of frames in each video) for 85 videos from the Spring 2022 and Fall 2022 cohorts (the black line), ordered from low to high. The corresponding NTE scores computed from the predictions of the audio stream, the text stream, and the video stream engagement detectors for the videos are also shown.

TABLE VII
LEVEL OF MISMATCHES BETWEEN THE PREDICTIONS FROM VIDEO, AUDIO, AND TEXT STREAMS FOR ALL 85 VIDEOS FROM THE SPRING 2022 AND FALL 2022 COHORTS

	Fraction of videos with mismatches (%)		
Mismatch level (%)	**Video–Text**	**Text–Audio**	**Video–Audio**
>0 and ≤ 10	31.76	5.88	50.59
>10 and ≤ 20	58.82	60.00	37.65
>20 and ≤ 30	5.88	28.24	7.06
>40 and ≤ 50	3.53	3.53	2.35
>50 and ≤ 60	0.00	2.35	2.35
>60	0.00	0.00	0.00

C. Results Identifying the Patterns of Student Engagement During Teamwork Sessions (RQ3)

On a scale of -1 (*unproductive engagement*) to 1 (*productive engagement*), the mean NTE scores across all teams were 0.94 ± 0.04, suggesting high productive engagement (see Fig. 14). In addition, the mean NTE scores computed from the three data streams separately were 0.98 ± 0.04, 0.91 ± 0.05, and 0.88 ± 0.05 for video, audio, and text streams, respectively (see Fig. 14). Moreover, 90.59%, 65.88%, and 88.24% of teams demonstrated less than 20% mismatches between the predictions from video and text, text and audio, and video and audio streams, respectively (see Table VII).

We identified some consistent patterns regarding changes in engagement states from frame to frame for all teams and for all three engagement detectors. For example, teams tended

to be mostly in *productive engagement* before transitioning to *disengagement* (> 90.34% of the time) and *idle* (> 93.69% of the time). However, we also identified some irregular patterns. For instance, although teams tended to transition to *productive engagement* mostly from *idle* (90.79% of the time) for the video stream, the patterns were slightly different for the audio (transitioned to *productive engagement* 51.15% and 48.81% of the time from *idle* and *disengagement*, respectively) and the text streams (transitioned to *productive engagement* 37.38% and 62.62% of the time from *idle* and *disengagement*, respectively).

VI. Discussion and Implications

This study approached three research questions aimed at identifying the level of engagement students enacted as members of teams as they collaborated through online teamwork sessions. One of the primary contributions of the study is that it provided qualitative characterizations of different forms of teamwork engagement, including productive engagement, unproductive engagement, disengagement, and idle. This is an important contribution in itself, as a substantial body of research informing our understanding of teamwork performance has focused on investigating static characteristics of teams, such as team members' backgrounds and expertise, elements of team formation, or developmental stages [25]. Instead, our study examined dynamic aspects of teams, such as those occurring during working sessions. A second contribution of the study focused on identifying overall patterns of engagement during online teamwork sessions in the context of an undergraduate course. Overall, it was observed that students enacted forms of productive engagement, including instances of students actively contributing to the completion of a task where they worked together in brainstorming or generating ideas, producing additional externalized outputs, manipulating an artifact like taking notes or receiving information without overtly performing an action but were listening or being attentive. Productive engagement also consisted of team members communicating and coordinating processes such as planning, helping, or reassuring group members, socializing, deciding roles and tasks, giving or seeking feedback, volunteering or helping each other, and deciding on meeting times. These behaviors are significant as they represent interdependent team processes and dynamics that helped teams achieve their collective goals [101], therefore improving teams' ability to function effectively [102]. The different forms of productive engagement combined aspects of team members contributing with their knowledge and expertise [103] and coordination and communication processes [104].

The study also contributed to technological advances in learning technologies for monitoring teamwork engagement. Specifically, we have deployed three modalities (video, audio, and text) of real-time AI-based classifiers, and we have validated them on human-coded data showing high reliability of results. The classifiers work with a wide variety of learners with different accents and arriving from different cultures. Moreover, the video classifier is robust in poor lighting conditions and works for the detection of people with different clothing and glasses (see Figs. 9 and 10).

The first implication of the study relates to the importance of instructors' role in planning and coordinating teamwork projects [2]. Research has identified that assuming that participants will socially interact in an online environment merely because the environment makes it possible may not result in engaged teams [105]. While team training interventions are a viable approach for enhancing teamwork performance in industry [106], teamwork pedagogy can be equally effective in education environments [2]. In this study, cooperative learning was used as the pedagogical approach to facilitate teamwork pedagogy. The coordination and orchestration of the course, along with the timing of teamwork sessions and project deadlines, may have promoted the different forms of engagement or disengagement during the working sessions.

A second implication of this study relates to the potential of our AI-based detection system to identify forms of engagement in a timely manner. Although our implementation of the system, along with its validation, was performed on historical data, the validity of the approach opens new venues for automatic detection of engagement in working classrooms. Specifically, in the context of higher education, implementing teamwork practices is hard, and it is even harder to monitor whether students are benefiting from the learning experience [107]. Specifically, in large-class settings, where there may be over 120 students per class, resulting in over 30 teams, effective collaborative behaviors can make a difference in students' social interaction [81]. Such timely information can be delivered to instructors in the form of digital dashboards [17], [18], [19] that can notify instructors of unproductive or negative behaviors. By providing instructors with information about teamwork engagement, they could facilitate better social interaction through mediation [108]. Instructor mediation can include eliciting ideas, managing conflict, providing immediate feedback, and weaving ideas together, among others [105]. Dashboards may also provide instructors with a means to assess teamwork processes [10].

VII. Conclusion, Limitations, and Future Work

The growing complexity of modern work environments and the increase in distributed organizations have made virtual teamwork essential to an organization's success. Thus, higher education institutions should cultivate students who will be effective collaborators in teams, both in-person and virtually. Specifically, in the context of teams working virtually, it is necessary to identify strategies that maximize student engagement. To do this effectively, it is first necessary to characterize forms of engagement, detect when unproductive engagement or disengagement is happening, and provide adequate training, pedagogical support, and mediation if needed. The study contributes in part toward this goal by identifying forms of online teamwork engagement and detecting those automatically.

However, our approach also has several limitations and scope for future work. One of the primary limitations is that the engagement detection was done a posteriori; thus, no immediate feedback was provided to the instructor. Thus, our future work could include the implementation of a dashboard that can track teamwork engagement in real time. A second limitation was

that we focused solely on one aspect of teamwork, specifically cognitive engagement. However, in academic settings, engagement is also operationalized in terms of behavioral engagement, including aspects such as effort and participation, and affective engagement, including demonstrating a positive attitude toward learning, trust, and a sense of belonging [109]. Therefore, future work could focus on characterizing behavioral and attitudinal aspects of engagement and static aspects of teamwork, such as team members' backgrounds, expertise, and personality, among many other characteristics.

A third limitation focused on the AI-based approach. The deep neural classifiers respond immediately and do not consider a more prolonged time context, causing sudden spikes in responses. We mediate for this by using the floating average, but a more advanced method that considers extended context, such as the Transformer [110], could provide more stable results. Also, the wealth of responses varies significantly. Students, at times, did not say anything but responded continuously with facial expressions. A better version of the classifier would weigh the responses accordingly, but the actual weight of each response is currently unknown. In addition, for training the video stream engagement detector, we considered behaviors from all detected participants in a classified 2-min interval to belong to the same class as the hand-coding was carried out at 2-min intervals for each video. This may have caused the video-based prediction to have the lowest accuracy among the three engagement detectors, as all team members in a 2-min interval may not belong to the same class. If we had chosen a smaller interval, the accuracy of video-based prediction may have improved at the expense of the hand-coding process becoming very lengthy and tedious.

Future studies can focus on creating a more refined dataset using smaller intervals. Moreover, the datasets used for training the engagement detectors were heavily imbalanced toward *productive engagement*. Although we applied various data augmentation techniques to balance the datasets, future studies should consider using a more balanced dataset including more *unproductive engagement* samples. Finally, the three engagement detectors produce three different predictions in our developed system. In the future, it may be interesting to investigate data fusion techniques, such as feature-level or decision-level fusion, to obtain a single prediction from the three engagement detectors. Moreover, the developed detectors could be used individually in other contexts, such as measuring the level of engagement in chats, dialogues, or video classes. We attempted to use as many unintrusive input channels as possible. It would be interesting to enhance the data with additional biometric inputs, such as heart rate, temperature, etc. However, this requires access to the data and their labeling with respect to engagement level.

Despite these limitations, the contributions of the study are significant as they go beyond the qualities of team members and the products they generate to focus on interactive aspects of teamwork mediated by technology.

ACKNOWLEDGMENT

The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. National Science Foundation or the U.S. Government.

REFERENCES

[1] L. A. DeChurch and J. R. Mesmer-Magnus, "The cognitive underpinnings of effective teamwork: A meta-analysis," *J. Appl. Psychol.*, vol. 95, no. 1, pp. 32–53, 2010.

[2] L. Riebe, A. Girardi, and C. Whitsed, "A systematic literature review of teamwork pedagogy in higher education," *Small Group Res.*, vol. 47, no. 6, pp. 619–664, 2016.

[3] D. Reed, "How productive meetings can support teamwork," *Dent. Nurs.*, vol. 8, no. 9, pp. 588–591, 2012.

[4] J. Chen, A. Kolmos, and X. Du, "Forms of implementation and challenges of PBL in engineering education: A review of literature," *Eur. J. Eng. Educ.*, vol. 46, no. 1, pp. 90–115, 2021.

[5] D. W. Johnson, R. T. Johnson, and K. A. Smith, "Cooperative learning: Improving university instruction by basing practice on validated theory," *J. Excellence Univ. Teach.*, vol. 25, no. 4, pp. 1–26, 2014.

[6] S. Baharom and B. Palaniandy, "Problem-based learning: A process for the acquisition of learning and generic skills," *PBL Across Cultures*, vol. 47, pp. 47–55, 2013.

[7] S. Kim, L. Copeland, E. Cohen, J. Galt, C. A. Terregino, and A. Pradhan, "Frame-of-reference training for students: Promoting a shared mental model for clerkship performance with an online, interactive training module," *J. Gen. Intern. Med.*, vol. 37, no. 6, pp. 1575–1577, 2022.

[8] P. M. Alexander, "Virtual teamwork in very large undergraduate classes," *Comput. Educ.*, vol. 47, no. 2, pp. 127–147, 2006.

[9] C. C. Danko and A. A. Duarte, "The Challenge of implementing a student-centred learning approach in large engineering classes," 2009.

[10] R. Lingard and S. Barkataki, "Teaching teamwork in engineering and computer science," in *Proc. Front. Educ. Conf.*, 2011, pp. F1C-1–F1C-5.

[11] R. A. Sottilare, C. S. Burke, E. Salas, A. M. Sinatra, J. H. Johnston, and S. B. Gilbert, "Designing adaptive instruction for teams: A meta-analysis," *Int. J. Artif. Intell. Educ.*, vol. 28, pp. 225–264, 2018.

[12] M. Bond and S. Bedenlier, "Facilitating student engagement through educational technology: Towards a conceptual framework," *J. Interact. Media Educ.*, vol. 2019, no. 1, 2019, Art. no. 11.

[13] F. Pattanpichet, "The effects of using collaborative learning to enhance students English speaking achievement," *J. College Teach. Learn.*, vol. 8, no. 11, pp. 1–10, 2011.

[14] D. R. Seibold, P. Kang, and B. M. Gailliard, "Communication that damages teamwork: The dark side of teams," in *Destructive Organizational Communication*. Evanston, IL, USA: Routledge, 2010, pp. 283–306.

[15] M. Tissenbaum, "I see what you did there! Divergent collaboration and learner transitions from unproductive to productive states in open-ended inquiry," *Comput. Educ.*, vol. 145, 2020, Art. no. 103739.

[16] M. Frank, R. Fruchter, and M. Leinikka, "Global teamwork: Components of engaging and productive meetings," in *Proc. Int. Conf. Comput. Civil Building Eng.*, 2016, pp. 1933–1941.

[17] J. Zheng, L. Huang, S. Li, S. P. Lajoie, Y. Chen, and C. E. Hmelo-Silver, "Self-regulation and emotion matter: A case study of instructor interactions with a learning analytics dashboard," *Comput. Educ.*, vol. 161, 2021, Art. no. 104061.

[18] I. Amarasinghe, D. Hernández-Leo, K. Michos, and M. Vujovic, "An actionable orchestration dashboard to enhance collaboration in the classroom," *IEEE Trans. Learn. Technol.*, vol. 13, no. 4, pp. 662–675, Oct.–Dec. 2020.

[19] G. M. Fernandez-Nieto et al., "Storytelling with learner data: Guiding student reflection on multimodal team data," *IEEE Trans. Learn. Technol.*, vol. 14, no. 5, pp. 695–708, Oct. 2021.

[20] K. J. Behfar, R. S. Peterson, E. A. Mannix, and W. M. Trochim, "The critical role of conflict resolution in teams: A close look at the links between conflict type, conflict management strategies, and team outcomes," *J. Appl. Psychol.*, vol. 93, no. 1, 2008, Art. no. 170.

[21] B. B. Tomić, A. D. Kijevčanin, Z. V. Ševarac, and J. M. Jovanović, "An AI-based approach for grading students' collaboration," *IEEE Trans. Learn. Technol.*, vol. 16, no. 3, pp. 292–305, Jun. 2023.

[22] J. A. Russell, "Core affect and the psychological construction of emotion," *Psychol. Rev.*, vol. 110, no. 1, 2003, Art. no. 145.

[23] J. E. Mathieu, P. T. Gallagher, M. A. Domingo, and E. A. Klock, "Embracing complexity: Reviewing the past decade of team effectiveness research," *Annu. Rev. Org. Psychol. Org. Behav.*, vol. 6, pp. 17–46, 2019.

[24] J.-W. Strijbos, "Assessment of collaborative learning," in *Handbook of Human and Social Conditions in Assessment*. Evanston, IL, USA: Routledge, 2016, pp. 302–318.

[25] F. Delice, M. Rousseau, and J. Feitosa, "Advancing teams research: What, when, and how to measure team dynamics over time," *Front. Psychol.*, vol. 10, 2019, Art. no. 1324.

[26] N. J. Cooke, J. C. Gorman, C. W. Myers, and J. L. Duran, "Interactive team cognition," *Cogn. Sci.*, vol. 37, no. 2, pp. 255–285, 2013.

[27] B. London, G. Downey, and S. Mace, "Psychological theories of educational engagement: A multi-method approach to studying individual engagement and institutional change," *Vanderbit Law Rev.*, vol. 60, 2007, Art. no. 455.

[28] M. Bond, S. Bedenlier, K. Buntins, M. Kerres, and O. Zawacki-Richter, "Facilitating student engagement in higher education through educational technology: A narrative systematic review in the field of education," *Contemporary Issues Technol. Teacher Educ.*, vol. 20, no. 2, pp. 315–368, 2020.

[29] M. Cater and K. Y. Jones, "Measuring perceptions of engagement in teamwork in youth development programs," *J. Exp. Educ.*, vol. 37, no. 2, pp. 176–186, 2014.

[30] S. J. Hazel, N. Heberle, M.-M. McEwen, and K. Adams, "Team-based learning increases active engagement and enhances development of teamwork and communication skills in a first-year course for veterinary and animal science undergraduates," *J. Vet. Med. Educ.*, vol. 40, no. 4, pp. 333–341, 2013.

[31] B. Rogoff, "Cognition as a collaborative process," in *Handbook of Child Psychology*. Hoboken, NJ, USA: Wiley, 1998.

[32] M. D. Cannon and A. C. Edmondson, "Confronting failure: Antecedents and consequences of shared beliefs about failure in organizational work groups," *J. Org. Behav.: Int. J. Ind., Occup. Org. Psychol. Behav.*, vol. 22, no. 2, pp. 161–177, 2001.

[33] N. Frey, D. Fisher, and S. Everlove, *Productive Group Work: How to Engage Students, Build Teamwork, and Promote Understanding*. Alexandria, VA, USA: ASCD, 2009.

[34] B. Senior and S. Swailes, "Inside management teams: Developing a teamwork survey instrument," *Brit. J. Manage.*, vol. 18, no. 2, pp. 138–153, 2007.

[35] L. Zheng, L. Zhong, and Y. Fan, "An immediate analysis of the interaction topic approach to promoting group performance, knowledge convergence, cognitive engagement, and coregulation in online collaborative learning," *Educ. Inf. Technol.*, vol. 28, pp. 9913–9934, 2023.

[36] S.-Y. Wu, "Construction and evaluation of an online environment to reduce off-topic messaging," *Interact. Learn. Environ.*, vol. 30, no. 3, pp. 455–469, 2022.

[37] G. K. Wong, Y. K. Li, and X. Lai, "Visualizing the learning patterns of topic-based social interaction in online discussion forums: An exploratory study," *Educ. Technol. Res. Develop.*, vol. 69, no. 5, pp. 2813–2843, 2021.

[38] Q. Xu, Y. Wei, J. Gao, H. Yao, and Q. Liu, "ICAPD framework and simAM-YOLOv8n for student cognitive engagement detection in classroom," *IEEE Access*, vol. 11, pp. 136063–136076, 2023.

[39] C. Pabba and P. Kumar, "An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition," *Expert Syst.*, vol. 39, no. 1, 2022, Art. no. e12839.

[40] N. K. Mehta, S. S. Prasad, S. Saurav, R. Saini, and S. Singh, "Three-dimensional DenseNet self-attention neural network for automatic detection of student's engagement," *Appl. Intell.*, vol. 52, no. 12, pp. 13803–13823, 2022.

[41] P. Guhan et al., "Developing an effective and automated patient engagement estimator for telehealth: A machine learning approach," 2020, *arXiv:2011.08690*.

[42] L. Shan, "Measuring cognitive engagement: An overview of measurement instruments and techniques," *Int. J. Psychol. Educ. Stud.*, vol. 8, no. 3, pp. 63–76, 2021.

[43] A. Levordashka, D. S. Fraser, and I. D. Gilchrist, "Measuring real-time cognitive engagement in remote audiences," *Sci. Rep.*, vol. 13, no. 1, 2023, Art. no. 10516.

[44] N. Bosch and S. K. D'mello, "Automatic detection of mind wandering from video in the lab and in the classroom," *IEEE Trans. Affect. Comput.*, vol. 12, no. 4, pp. 974–988, Oct.–Dec. 2021.

[45] S. Li, S. P. Lajoie, J. Zheng, H. Wu, and H. Cheng, "Automated detection of cognitive engagement to inform the art of staying engaged in problem-solving," *Comput. Educ.*, vol. 163, 2021, Art. no. 104114.

[46] J. Mo, R. Zhu, H. Yuan, Z. Shou, and L. Chen, "Student behavior recognition based on multitask learning," *Multimedia Tools Appl.*, vol. 82, no. 12, pp. 19091–19108, 2023.

[47] M. Hu et al., "Bimodal learning engagement recognition from videos in the classroom," *Sensors*, vol. 22, no. 16, 2022, Art. no. 5932.

[48] P. Bhardwaj, P. Gupta, H. Panwar, M. K. Siddiqui, R. Morales-Menendez, and A. Bhaik, "Application of deep learning on student engagement in e-learning environments," *Comput. Elect. Eng.*, vol. 93, 2021, Art. no. 107277.

[49] M. T. Chi and R. Wylie, "The ICAP framework: Linking cognitive engagement to active learning outcomes," *Educ. Psychol.*, vol. 49, no. 4, pp. 219–243, 2014.

[50] J. Zaletelj and A. Košir, "Predicting students' attention in the classroom from kinect facial and body features," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, pp. 1–12, 2017.

[51] M. Marreddy, S. R. Oota, L. S. Vakada, V. C. Chinni, and R. Mamidi, "Multi-task text classification using graph convolutional networks for large-scale low resource language," in *Proc. Int. Joint Conf. Neural Netw.*, 2022, pp. 1–8.

[52] R. Tokuhisa, K. Inui, and Y. Matsumoto, "Emotion classification using massive examples extracted from the web," in *Proc. 22nd Int. Conf. Comput. Linguistics*, 2008, pp. 881–888.

[53] M. A. Mageed and L. Ungar, "EmoNet: Fine-grained emotion detection with gated recurrent neural networks," in *Proc. 55th Annu. Meeting Assoc. Comput. linguistics*, 2017, pp. 718–728.

[54] Z. Chen, N. Ma, and B. Liu, "Lifelong learning for sentiment classification," 2018, *arXiv:1801.02808*.

[55] S. D. Gollapalli, P. Rozenshtein, and S. K. Ng, "ESTer: Combining word co-occurrences and word associations for unsupervised emotion detection," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2020, pp. 1043–1056.

[56] Y. Shen, H. Yang, and L. Lin, "Automatic depression detection: An emotional audio-textual corpus and a GRU/BILSTM-based model," in *Proc. IEEE Int. Conf. Acoust.*, 2022, pp. 6247–6251.

[57] S. Yoon, S. Dey, H. Lee, and K. Jung, "Attentive modality hopping mechanism for speech emotion recognition," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 3362–3366.

[58] K. Han, D. Yu, and I. Tashev, "Speech emotion recognition using deep neural network and extreme learning machine," in *Proc. Interspeech Conf.*, 2014, pp. 223–227.

[59] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4489–4497.

[60] S. T. Mubarrat, S. K. Chowdhury, and A. D. Nimbarte, "Predicting shoulder joint reaction forces from 3D body kinematics: A convolutional neural network approach," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 64, 2020, pp. 939–941.

[61] S. Mubarrat and S. Chowdhury, "Convolutional lstm: A deep learning approach to predict shoulder joint reaction forces," *Comput. Methods Biomech. Biomed. Eng.*, vol. 26, no. 1, pp. 65–77, 2023.

[62] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," *Sensors*, vol. 21, no. 9, 2021, Art. no. 3046.

[63] X. Chang and W. Skarbek, "Multi-modal residual perceptron network for audio–video emotion recognition," *Sensors*, vol. 21, no. 16, 2021, Art. no. 5452.

[64] W. Zhou, J. Cheng, X. Lei, B. Benes, and N. Adamo, "Deep learning-based emotion recognition from real-time videos," in *Human-Computer Interaction. Multimodal and Natural Interaction*. Cham, Switzerland: Springer, 2020, pp. 321–332.

[65] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[66] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, Apr. 2017, pp. 427–431.

[67] F. Calefato, F. Lanubile, and N. Novielli, "EmoTxt: A toolkit for emotion recognition from text," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. Workshops Demos*, 2017, pp. 79–80.

[68] D. Hazarika, S. Poria, R. Mihalcea, E. Cambria, and R. Zimmermann, "Icon: Interactive conversational memory network for multimodal emotion detection," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 2594–2604.

[69] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[70] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE: The Munich versatile and fast open-source audio feature extractor," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 1459–1462.

[71] R. Dhiman, G. S. Kang, and V. Gupta, "Modified dense convolutional networks based emotion detection from speech using its paralinguistic features," *Multimedia Tools Appl.*, vol. 80, no. 21-23, pp. 32041–32069, 2021.
[72] W. Bian, J. Wang, B. Zhuang, J. Yang, S. Wang, and J. Xiao, "Audio-based music classification with DenseNet and data augmentation," in *16th Pacific Rim Int. Conf. Artif. Intell.*, 2019, pp. 56–65.
[73] O. O. Abayomi-Alli, R. Damaševičius, A. Qazi, M. Adedoyin-Olowe, and S. Misra, "Data augmentation and deep learning methods in sound classification: A systematic review," *Electronics*, vol. 11, no. 22, 2022, Art. no. 3795.
[74] A. F. R. Nogueira, H. S. Oliveira, J. J. Machado, and J. M. R. Tavares, "Sound classification and processing of urban environments: A systematic literature review," *Sensors*, vol. 22, no. 22, 2022, Art. no. 8608.
[75] Z. Mushtaq and S.-F. Su, "Efficient classification of environmental sounds through multiple features aggregation and data enhancement techniques for spectrogram images," *Symmetry*, vol. 12, no. 11, 2020, Art. no. 1822.
[76] J. Goodman, "Classes for fast maximum entropy training," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Proc.*, 2001, vol. 1, pp. 561–564.
[77] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.
[78] R. Vivek and Y. Nanthagopan, "Review and comparison of multi-method and mixed method application in research studies," *Eur. J. Manage. Issues*, vol. 29, no. 4, pp. 200–208, 2021.
[79] H.-F. Hsieh and S. E. Shannon, "Three approaches to qualitative content analysis," *Qualitative Health Res.*, vol. 15, no. 9, pp. 1277–1288, 2005.
[80] A. J. Magana, A. Jaiswal, T. L. Amuah, M. Z. Bula, M. S. U. Duha, and J. C. Richardson, "Characterizing team cognition within software engineering teams in an undergraduate course," *IEEE Trans. Educ.*, vol. 67, no. 1, pp. 87–99, Feb. 2024.
[81] A. J. Magana, T. Amuah, S. Aggrawal, and D. A. Patel, "Teamwork dynamics in the context of large-size software development courses," *Int. J. STEM Educ.*, vol. 10, no. 1, 2023, Art. no. 57.
[82] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Develop. Psychopathol.*, vol. 17, no. 3, pp. 715–734, 2005.
[83] L. Shen, M. Wang, and R. Shen, "Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment," *J. Educ. Technol. Soc.*, vol. 12, no. 2, pp. 176–189, 2009.
[84] A. P. Lawson, R. E. Mayer, N. Adamo-Villani, B. Benes, X. Lei, and J. Cheng, "Do learners recognize and relate to the emotions displayed by virtual instructors?," *Int. J. Artif. Intell. Educ.*, vol. 114, pp. 1560–4306, 2021.
[85] J. M. Harley, S. P. Lajoie, C. Frasson, and N. C. Hall, "Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features," *Int. J. Artif. Intell. Educ.*, vol. 27, pp. 268–297, 2017.
[86] T. Sandanayake, A. Madurapperuma, and D. Dias, "Affective E learning model for recognising learner emotions," *Int. J. Inf. Educ. Technol.*, vol. 1, no. 4, pp. 315–320, 2011.
[87] A. C. Strain and S. K. D'Mello, "Affect regulation during learning: The enhancing effect of cognitive reappraisal," *Appl. Cogn. Psychol.*, vol. 29, no. 1, pp. 1–19, 2015.
[88] A. J. Magana, T. Karabiyik, P. Thomas, A. Jaiswal, V. Perera, and J. Dworkin, "Teamwork facilitation and conflict resolution training in a HyFlex course during the COVID-19 pandemic," *J. Eng. Educ.*, vol. 111, no. 2, pp. 446–473, 2022.
[89] A. Jaiswal, T. Karabiyik, P. Thomas, and A. J. Magana, "Characterizing team orientations and academic performance in cooperative project-based learning environments," *Educ. Sci.*, vol. 11, 2021, Art. no. 520.
[90] A. J. Magana, Y. Y. Seah, and P. Thomas, "Fostering cooperative learning with scrum in a semi-capstone systems analysis and design course," *J. Inf. Syst. Educ.*, vol. 29, no. 2, pp. 75–92, 2018.
[91] A. Geitgey, "Face recognition: 1.2.2," 2018. [Online]. Available: https://github.com/ageitgey/face_recognition
[92] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 28492–28518.
[93] Zulko, "Moviepy: 1.0.3," 2020. [Online]. Available: https://github.com/Zulko/moviepy
[94] J. Robert, "Pydub: 0.25.1," 2021. [Online]. Available: https://github.com/jiaaro/pydub
[95] B. McFee et al., "Librosa/librosa: 0.10.0," Feb. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7657336
[96] A. B. Jung et al., "Imgaug," 2020. Accessed: Feb. 1, 2020. [Online]. Available: https://github.com/aleju/imgaug
[97] J. Timmer and M. Koenig, "On generating power law noise," *Astron. Astrophys.*, vol. 300, 1995, Art. no. 707.
[98] F. Patzelt, A. Spaeth, and O. Eberhard, "Colorednoise: 2.0.0," 2022. [Online]. Available: https://github.com/felixpatzelt/colorednoise
[99] V. Marivate and T. Sefara, "Improving short text classification through global augmentation methods," in *Machine Learning and Knowledge Extraction*. Cham, Switzerland: Springer, 2020, pp. 385–399.
[100] E. Ma, "Nlpaug: 1.1.11," 2019. [Online]. Available: https://github.com/makcedward/nlpaug
[101] M. A. Marks, J. E. Mathieu, and S. J. Zaccaro, "A temporally based framework and taxonomy of team processes," *Acad. Manage. Rev.*, vol. 26, no. 3, pp. 356–376, 2001.
[102] K. W. Buffinton, K. W. Jablokow, and K. A. Martin, "Project team dynamics and cognitive style," *Eng. Manage. J.*, vol. 14, no. 3, pp. 25–33, 2002.
[103] N. J. Cooke, "Team cognition as interaction," *Curr. Directions Psychol. Sci.*, vol. 24, no. 6, pp. 415–419, 2015.
[104] C. G. Collins, C. B. Gibson, N. R. Quigley, and S. K. Parker, "Unpacking team dynamics with growth modeling: An approach to test, refine, and integrate theory," *Org. Psychol. Rev.*, vol. 6, no. 1, pp. 63–91, 2016.
[105] K. Kreijns, P. A. Kirschner, and W. Jochems, "Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research," *Comput. Hum. Behav.*, vol. 19, no. 3, pp. 335–353, 2003.
[106] D. McEwan, G. R. Ruissen, M. A. Eys, B. D. Zumbo, and M. R. Beauchamp, "The effectiveness of teamwork training on teamwork behaviors and team performance: A systematic review and meta-analysis of controlled interventions," *PLoS one*, vol. 12, no. 1, 2017, Art. no. e0169604.
[107] D. Allen and K. Tanner, "Infusing active learning into the large-enrollment biology class: Seven strategies, from the simple to complex," *Cell Biol. Educ.*, vol. 4, no. 4, pp. 262–268, 2005.
[108] C. N. Gunawardena, "Social presence theory and implications for interaction and collaborative learning in computer conferences," *Int. J. Educ. Telecommun.*, vol. 1, pp. 147–166, 1995.
[109] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," *Rev. Educ. Res.*, vol. 74, no. 1, pp. 59–109, 2004.
[110] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

Alejandra J. Magana received the Ph.D. degree in engineering education from Purdue University, West Lafayette, IN, USA, in 2009.

She is currently the W.C. Furnas Professor in Enterprise Excellence with the Department of Computer and Information Technology and a Professor with the School of Engineering Education, Purdue University. Her research program investigates how model-based cognition in Science, Technology, Engineering, and Mathematics can be better supported using expert tools and disciplinary practices, such as data science computation, modeling, and simulation.

Dr. Magana is a Fellow of the American Society for Engineering Education.

Syed Tanzim Mubarrat received the B.S. degree in electrical and electronics engineering from the Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 2019, and the M.S. degree in industrial engineering from Texas Tech University, Lubbock, TX, USA, in 2021. He is currently working toward the Ph.D. degree with the Department of Computer and Information Technology, Purdue University, West Lafayette, IN, USA.

His research focuses on virtual reality, human–computer interaction, haptics, deep learning, and game research. His research interests include leveraging artificial-intelligence-based tools to increase the realism and effectiveness of virtual worlds, particularly gamified learning and training environments.

Dominic Kao received the B.S. degree in computer engineering from the University of Alberta, Edmonton, AB, Canada, in 2009, the M.S.E. degree in computer science from Princeton University, Princeton, NJ, USA, in 2012, and the Ph.D. degree in computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2018.

He is currently an Assistant Professor with Purdue University, West Lafayette, IN, USA, where he is also the Director of the Virtual Futures Lab. He researches virtual worlds, games, and education. His research interests include building virtual worlds and leveraging existing ones, primarily in understanding how virtual worlds influence people and developing best practices for developing games and learning environments.

Dr. Kao received the National Science Foundation CAREER Award in 2024.

Bedrich Benes (Senior Member, IEEE) received the M.S. and Ph.D. degrees in computer science from the Czech Technical University in Prague, in 1991 and 1999, respectively.

He is a Professor and Associate Head of Computer Science with Purdue University, West Lafayette, IN, USA. He works in generative methods for geometry synthesis and deep learning. He has authored or coauthored more than 200 research papers. His research interests include procedural and inverse procedural modeling, simulation of natural phenomena, and additive manufacturing.

Dr. Benes is a Fellow of the European Association for Computer Graphics and a Senior Member of the Association for Computing Machinery.