

Article

Exploring an Affective and Responsive Virtual Environment to Improve Remote Learning

Jianing Qi ¹, Hao Tang ^{1,2,*}  and Zhigang Zhu ^{1,3} ¹ Graduate Center, City University of New York, New York, NY 10016, USA² BMCC, City University of New York, New York, NY 10007, USA³ City College, City University of New York, New York, NY 10031, USA

* Correspondence: htang@bmcc.cuny.edu

Abstract: Online classes are typically conducted by using video conferencing software such as Zoom, Microsoft Teams, and Google Meet. Research has identified drawbacks of online learning, such as “Zoom fatigue”, characterized by distractions and lack of engagement. This study presents the CUNY Affective and Responsive Virtual Environment (CARVE) Hub, a novel virtual reality hub that uses a facial emotion classification model to generate emojis for affective and informal responsive interaction in a 3D virtual classroom setting. A web-based machine learning model is employed for facial emotion classification, enabling students to communicate four basic emotions live through automated web camera capture in a virtual classroom without activating their cameras. The experiment is conducted in undergraduate classes on both Zoom and CARVE, and the results of a survey indicate that students have a positive perception of interactions in the proposed virtual classroom compared with Zoom. Correlations between automated emojis and interactions are also observed. This study discusses potential explanations for the improved interactions, including a decrease in pressure on students when they are not showing faces. In addition, video panels in traditional remote classrooms may be useful for communication but not for interaction. Students favor features in virtual reality, such as spatial audio and the ability to move around, with collaboration being identified as the most helpful feature.

Keywords: virtual reality; online education; emotion recognition



Citation: Qi, J.; Tang, H.; Zhu, Z. Exploring an Affective and Responsive Virtual Environment to Improve Remote Learning. *Virtual Worlds* **2023**, *2*, 53–74. <https://doi.org/10.3390/virtualworlds2010004>

Academic Editor: Christos J. Bouras

Received: 18 December 2022

Revised: 8 February 2023

Accepted: 17 February 2023

Published: 2 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The COVID-19 pandemic has led to a shift in the forms of education. Over the past two years, the widespread adoption of online classes has resulted in increased acceptance among both students and instructors/teachers. These classes are typically conducted through video conferencing software such as Zoom, Microsoft Teams, and Google Meet. Most students and teachers have experience with online classes. Even as in-person instruction resumes, some students and teachers continue to prefer the online format due to the convenience of saving commuting time and expenses, feeling more relaxed, and being without the requirement of an expensive physical space. It is likely that the trend toward online education will continue as we see more policy changes that support the growth of online classes and schools’ adoption of online and hybrid models.

Furthermore, remote learning has seen significant growth, increasing from 300,000 to 20 million students between 2011 and 2021. Additionally, venture capital funding in the education space also increased six-fold from 2017 to 2021 [1]. More universities began offering online education and degrees in the last decade. The market responded positively to remote education. The COVID-19 pandemic accelerated this trend even further. Augmented reality showed the potential for promoting collaborative and autonomous learning in higher education [2]. The recent interest in the “metaverse” also presents opportunities

for remote education. The Meta Corporation has proposed education scenarios and is working with universities, and it plans to continue investing in its virtual reality department in the future.

However, there are numerous complaints about online learning, specifically “Zoom fatigue”. Zoom fatigue refers to the exhaustion and stress that can result from excessive use of video conferencing technology, particularly Zoom, in remote learning environments when students must sit in front of and gaze at a screen for a very long time. Symptoms of Zoom fatigue can include difficulty learning online, body aches, and feelings of anxiety, among others [3]. According to a recent study [3], several issues contribute to this phenomenon. The main issues students are facing include distractions, difficulty focusing during Zoom sessions, and decreased learning effectiveness. Additionally, students often find themselves passively watching Zoom videos with little engagement in an online class, while several methods have been proposed to increase student engagement in class [4], such as encouraging active participation, the use of facial and body expressions, and the expression of students’ involvement, it is still difficult to implement these methods because students may not be willing to turn on their cameras for most of the time [5], due to privacy concerns, shyness or discomfort. This can decrease class engagement.

This paper addresses the challenges mentioned above through two approaches. The first approach is to use emojis for expressions, which can overcome the obstacle of students not turning on their cameras. The second approach is to create a 3D virtual classroom that increases student interactions using spatial audio and a 3D immersive virtual environment that simulates physical activities in a real classroom setting. The overall classroom experience is shown in Figure 1. It is a combination of virtual reality (VR) and affective communication. The virtual classroom has a resemblance to the real classroom experience: the teacher presents content while students listen to the teacher and interact with the teacher and other students.



Figure 1. A selection of the virtual classroom experience. (a) Left is the first-person perspective of a student with the slides in front of the view. (b) The middle is when students are sharing screens and discussing. (c) The right is when students use the pink drawing tool to interact with the slide.

The existing studies focus on testing the effectiveness of emotion classification in a VR environment using expensive and complicated headsets. On the contrary, this research presents a more cost-effective and user-friendly solution for communicating affective out the need for any specialized hardware. Students can participate in the emotion classification process using their existing computers, rather than expensive VR headsets. Additionally, the algorithm used in this research was selected based on real-time video feed data, rather than solely relying on results from datasets in a lab setting, ensuring its effectiveness in a real-world setting.

Overall, this paper aims to increase both the interaction between students and the responsive interaction between instructors and students, without the need for the video feed from participants. The paper also utilizes user surveys to investigate the effectiveness of the proposed approaches to increase interactions in online classes. The main contributions of the paper are as follows:

- The CARVE Hub enhances informal responsive interaction in a VR classroom setting by automatically analyzing users’ emotion data and generating corresponding emojis.

It is a new human–computer interface that is much cheaper, accessible, and ready to use compared to emotion data captured in VR headsets.

- The CARVE is powered by an emotion classification model through transfer learning in VR classrooms by addressing the limitations of current emotion datasets, such as small sample sizes and dissimilarity to classroom faces, to improve its real-world applicability.
- This study finds a strong correlation between the use of the automated emoji system and the increased interactions in a VR classroom. Additionally, the CRAVE Hub results in a statistically significant increase in student interactions compared with Zoom.

The paper is organized as follows: Section 2 discusses related work in remote learning using VR and AI technologies as well as studies of feedback from students. Section 4 explains the paper’s approaches and experimental design. Section 4 presents the experimental results. Section 5 describes the findings and some future work. Finally, the conclusions are provided in Section 6.

2. Related Work

Since the outbreak of COVID-19, there have been numerous studies on the effectiveness of teaching in remote classrooms. These studies can be broadly divided into two categories based on technological differences. The first category of studies uses video conferencing software and existing technology, such as Zoom, Google Meet, and Microsoft Teams. The second category focuses on the use of virtual reality (VR), including non-immersive VR, semi-immersive VR, and immersive VR. The majority of studies focus on immersive VR, which uses head-mounted displays (HMDs) [6]; while immersive VR provides a higher level of immersion, the cost of hardware and the discomfort of wearing HMDs are currently barriers for regular students to participate in VR education. There are a few studies that have compared the user experience results between HMD VR and desktop VR. One study in game design indicates that users’ performance is similar between HMD VR and desktop VR, and that users’ psychological needs are independent of the display modality [7]. Another study of remote education also found that, other than presence, desktop VR did not statistically differ from HMD VR in terms of the overall student experience [8].

2.1. VR in Education and Main Factors

Numerous studies try to identify key factors of remote education experiences, and here are some insights gathered from the results of current studies.

2.1.1. Interaction Format

The most common interaction software used in remote education is video conferencing, such as Zoom, Microsoft Teams, and Google Meet. These platforms offer various interaction formats, including audio, video, text box, danmaku, quizzes, and voting [9]. Several papers have explored the limitations of these video conferencing software for remote education [9,10]. These studies have noted that students are often reluctant to turn on their cameras and participate in video interactions. This reluctance is often attributed to privacy concerns, discomfort with displaying their appearance at home, and the desire to perform other activities during class, such as eating or taking notes [8–13]. Having a messy appearance, wearing pajamas, and eating food are top reasons for students not turning on their camera. This reluctance may prevent the full utilization of the video conferencing software and the video component. Additionally, students tend to prefer other formats, such as text box chat and danmaku, over video interactions [9]. Not being the focus of the classroom seems to make students more comfortable and easier to interact.

Most research conducted in VR classrooms is able to circumvent the video interaction format dilemma, as students do not need to turn on their cameras for more responsive and affective interactions. Results indicate that students have a high rating for not having to be seen or not having to use a webcam in VR classrooms [14]. However, most VR classrooms currently lack the capability for affective communication, as indicated [8]. In

order to address this issue, our research proposes the use of a deep emotion model to include emotional information without a video feed. This allows for the classified emotions to be transmitted, ensuring that students are not seen while still being able to interact and communicate emotionally with each other. By using this approach, students can still reap the benefits of VR classrooms while also being able to express their emotions in real time.

2.1.2. Student Feedback

An important factor for effective teaching in the classroom is students' feedback. Social, body gesture, and facial expression feedback are important for teaching and learning. However, video conferencing software typically lacks this type of feedback for both teachers and students [3]. Teachers tend to adjust their teaching based on students' feedback during in-person teaching, but they are not able to make similar adjustments during online lectures [10] if students turn off their cameras most of the time. Instructors have difficulty understanding whether the teaching information is understood by students or not, and they cannot tell whether students are engaged or not [3,10].

An experiment was performed using face icons to show students' emotions and mental states instead of their real faces, and it showed that the abstracted data are sufficient for teachers to receive students' feedback [15]. Sometimes, teachers only need to have aggregated responses to understand the entire classroom, and they do not need to look at each face individually [10,13,15,16]. Studies show that students are willing to provide feedback, such as their emotion data, if it can help the class [10]. Our research has decided to follow the usage of emotional state with icon representations as in the study of [10], but instead of using colors, our research uses emojis. To be specific, this research proposes to use emojis in real time, as they are a widely accepted format for communicating emotions. This allows students to communicate their emotions through emojis without sharing any other personal data. This approach ensures that students can interact and communicate emotionally without compromising their privacy.

2.1.3. Informal Social Interactions

Social interactions are one of the key components in remote education, and the main advantage of an in-person classroom compared to a video conference class is social interaction. Students' feedback in various studies highlights the importance of casual and responsive interactions between people, as it is essential for creating a sense of community and connections [6,9,10].

One study indicates that when compared with video interactions, students find that text chats and danmaku can significantly increase informal interactions [9]. Another study shows that the informal aspect increases social interactions, and the analysis is that text chats would create less social anxiety and fewer interruptions [10]. Some papers also propose using web cameras to capture data non-intrusively [17], which can be a good alternative solution to sending video data.

Compared to video conference software with broadcasting audio, a VR classroom study that tested spatial audio supports informal audio interaction. Students rated voice chat as the most helpful communication method in that study [8]. Other experiments have text chat as the most favored by students [9,10]. By allowing informal interactions, audio chat surpasses text chat as the most helpful feature in that experiment.

Students like to talk to people near them spatially during in-person classes compared to online classes [9]. Through casual talks, students tend to have more self-disclosure during group discussions and form more friendships [10]. Ad hoc and private conversations have high approval rates from users, even in virtual conferences [11]. In addition, different communication methods such as discussion forums, bulletin boards, and concept mapping software can support social interaction and encourage discussion [18]. These informal communication channels can facilitate more interactions, promote a sense of community, and create connections among students.

In conclusion, it is essential for remote education to incorporate informal and responsive interactions, as it can help to create a sense of community and connections among students, which can greatly enhance the overall learning experience. This can be achieved through the use of various communication methods such as text chats, danmaku, spatial audio, and non-intrusive web camera capture. Furthermore, virtual reality classrooms can provide an immersive and interactive environment that can support informal interactions, and provide a more engaging learning experience for students. Our research focuses on utilizing the informal interaction format in VR classrooms based on the findings of [10]. This includes incorporating spatial audio to allow for natural group discussions among students, as well as promoting features such as drawing tools, screen sharing, and the ability to place virtual models within the virtual environment. This approach aims to enhance the overall learning experience and foster a more collaborative and interactive atmosphere for students rather than the more lecture-style VR class in [8].

2.2. Emotion Detection

Emotion detection by analyzing facial images or videos is a widely studied problem in computer vision. Two popular facial emotion detection datasets are FER2013 [19] and CK+ [20]. Both datasets have their own advantages. FER2013 has a larger sample size, with 28,709 examples, while CK+ has 593 examples. However, the quality of the FER2013 dataset is not ideal, as there are mislabeled images, anime characters, and hard-to-determine emotions. Additionally, the dataset is not representative of what students' faces look like when captured by web cameras during online classes. On the other hand, CK+ data have higher quality and more closely resemble the faces captured by web cameras during online classes. As this research focuses on detecting emotions in a college class setting, the intended detection subjects are adults. In order to ensure that the model is as accurate as possible for this specific population, the CK+ dataset [20] was used. This dataset includes images of adults in a variety of emotional states and is, therefore, more closely aligned with the intended detection subjects than other datasets. By using the CK+ dataset, the model is able to better understand the nuances of adult emotions and provide more accurate results. However, it is limited by its small sample size of 593 examples.

Currently, machine learning models can achieve an accuracy of 90.49% for FER data [21] and 99.8% for CK+ data [20]. However, when these models are applied to real-world scenarios and faces captured by web cameras, their performance tends to be low. In order to address this issue, our research proposes the use of transfer learning. By first training the model on the FER2013 dataset and then fine-tuning it on the CK+ dataset, the model can learn the features of emotions and better adapt to the situation of web camera-captured faces. This approach is expected to improve the model's performance and accuracy in real-world scenarios.

3. Methods and Experiment

3.1. Overview

The main research questions of this paper are as follows:

1. How does informal social interaction (responsive interaction) influence individual learning experiences in remote classrooms?
2. How does the use of emotion data for affective interaction compare to the use of video feeds in terms of improving remote classroom experiences?
3. How does a 3D virtual environment via desktop VR compare to video conferencing software in terms of influencing learning experiences in remote classrooms?

To answer these questions, a survey containing multiple sub-questions was designed and administered to 9 undergraduate students in a remote laboratory class. The experiment was conducted during the Fall 2022 semester and compared the use of Zoom and the proposed solution, the CUNY Affective and Responsive Virtual Environment (CARVE) Hub, based on Mozilla Hubs. The students were given a two-hour class through Zoom for the first session, followed by the survey. For the second session, the students were given

the same class content, but through CARVE Hub, and then completed the survey again. The class environment is shown in Figure 2.



Figure 2. A Class in the CUNY Affective and Responsive Virtual Environment (CARVE) Hub. We can see students focusing on the PowerPoint, drawings on the ground, and a captured smile emoji.

3.2. The CARVE Hub Setup

This study employs Mozilla Hub to develop the CARVE Hub virtual classroom environment. This decision was made after comparing various products and determining that Mozilla Hub possesses advantageous existing features and ease of distribution. In CARVE Hub, adjustments were made to Mozilla Hub's spatial audio, and modifications were made to room and audio zones. The decay rate of spatial audio was increased to reduce the amount of noise heard by students when they are positioned far apart. Additionally, a platform for broadcasting voices was created to ensure that all students could hear the teacher's voice. Furthermore, modifications were made to the Mozilla Hub client, which will be described in further detail in a later section. The modified Hub was deployed on AWS, allowing students and teachers to access it through a single link.

The main features of CARVE Hub include avatar movement within a virtual space, screen sharing, communication through AI-based emojis, drawing, and voice chat. All participants accessed the virtual classroom using their personal computers and did not require any specialized software. The Mozilla Hub-based CARVE Hub operates on browsers, making it accessible by simply clicking the provided link. The idea is to simulate the real classroom experience. The flow chart of the entire CARVE Hub is in Figure 3. Students can move in a 3D virtual environment, where their emotions are captured through webcams and communicated through emojis. They can walk around with their avatars, talk to other avatars close to them, perceive class content, and have interactions similar to a real classroom setting. This allows for a more natural and immersive experience in a virtual setting, where affective communication is made possible through the use of emotion classification models and virtual avatars. The use of webcams and avatars allows for a more realistic and interactive experience.

The main libraries and stacks of the experiment include TensorFlow.js, Mozilla Hub, and AWS.

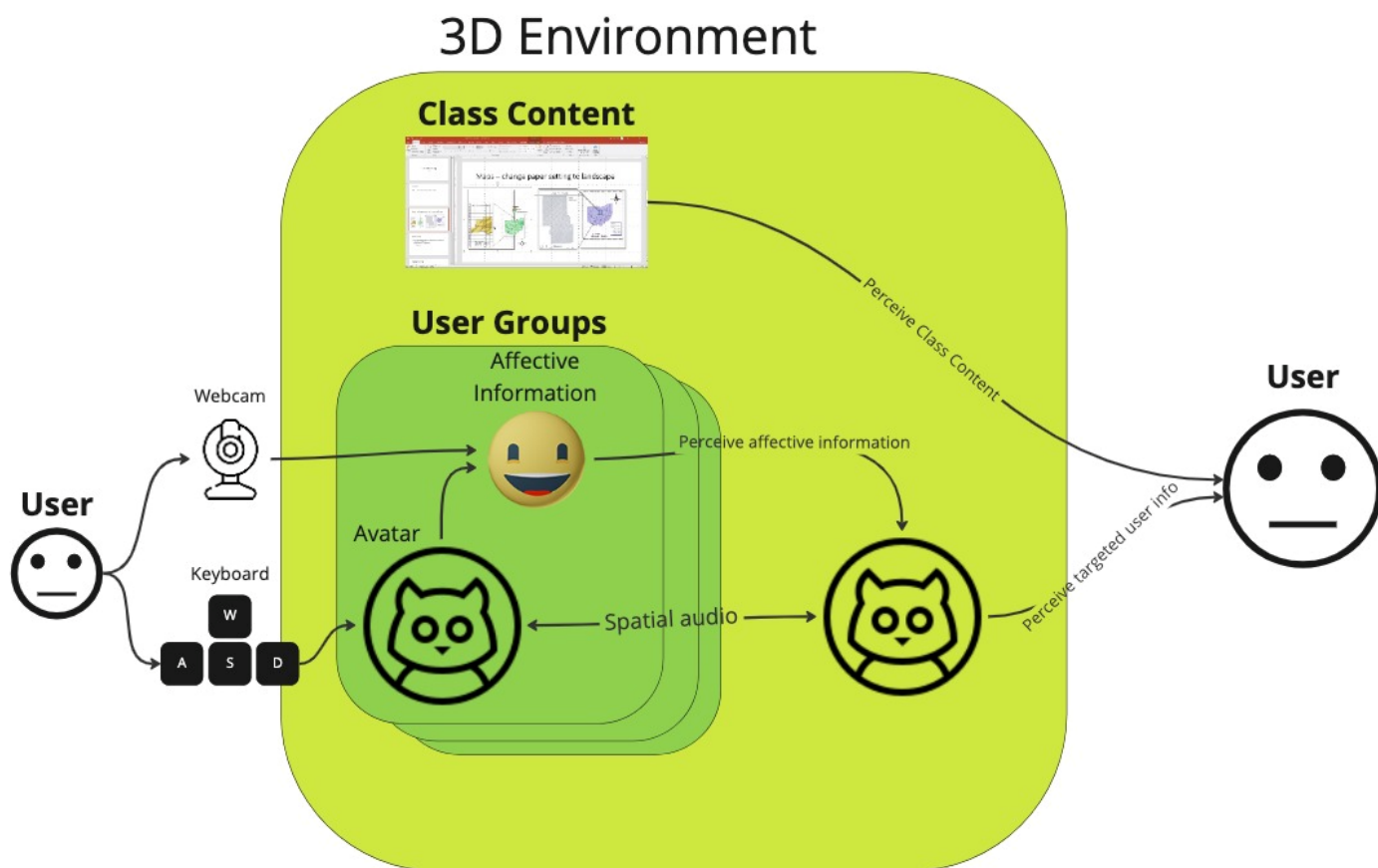


Figure 3. The flow chart of CARVE Hub design.

3.3. Deep Facial Emotion Recognition Model Setup

A deep emotion classification model was added to capture students' emotions through web cameras. Students can communicate emotions with each other without video panels, but through automated emojis instead. Facial expressions are captured through the web camera and classified through an emotion recognition model to send emojis. Furthermore, the emotion recognition model is deployed and run on students' browsers, and it processes locally on each student's desktop; thus, no video is needed to be transmitted out of their local machines.

A deep emotion classification model was trained through transfer learning from MobileNet [22]. The MobileNet pre-trained with ImageNet [23] was fine-tuned, and a similar architecture as Delgado [24] was used. The architecture added more layers from the original MobileNet, including one global average pooling 2D, one dense layer with 128 neurons with ReLU activation, and a 4-neuron classification with softmax activation. Augmentation and rescaling layers were also added to the model. The model architecture can be seen in Figure 4.

The aim is to improve the performance of a facial expression recognition model by utilizing the advantages of two different datasets, FER2013 [19] and CK+ [20]. It is found that while both datasets yield promising results during testing, they do not perform well when applied to real-life scenarios using faces captured from web cameras. To address this, the approach is to combine the strengths of both datasets and fine-tune the model using a pre-trained MobileNet. Furthermore, it is observed that certain emotions are not relevant in classroom settings, and that the FER2013 dataset is imbalanced, particularly with regard to the Disgust label. In the following, we will discuss details of emotion class selection and transfer learning implementation.

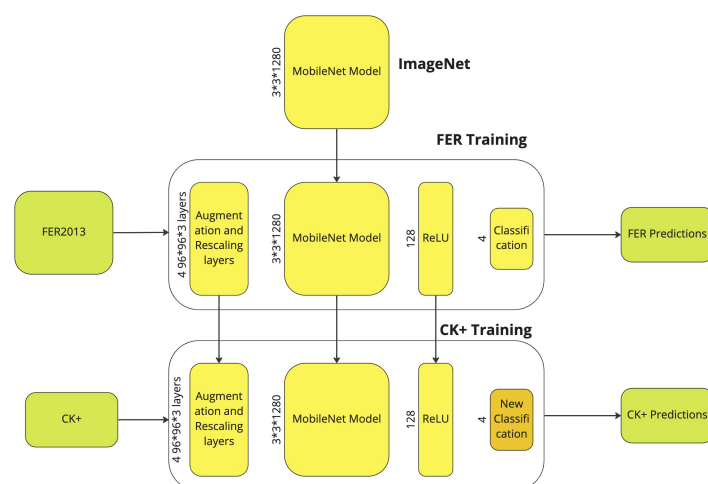


Figure 4. Architecture of deep facial emotion recognition model. For each task, the model is trained with frozen weights with new head first and then fine-tuned.

3.3.1. Class Label Selection for Real Classroom Settings

The design of the classification system uses four categories: happy, sad, surprise, and neutral. The other three categories, including anger, disgust, and fear, were removed for several reasons. First, the emotions of anger, disgust, and fear are less relevant to the classroom setting, as students do not frequently display these negative emotions during class. The main goal is to capture whether students are interested in the class, do not like the class, or are surprised by the content, as this is more helpful for teachers to receive emotional feedback. Second, the FER2013 dataset used to train the emotion classification system had a very imbalanced label distribution for disgust, which is a magnitude less compared to the other labels. This would affect the training results and lead to a decrease in accuracy during testing with real web camera footage. Even though the testing accuracy with FER2013 was high, the preliminary models did not perform well on real faces of real classroom settings, especially with biased labels.

In conclusion, the decision to remove the three labels of anger, disgust, and fear was made to improve the accuracy of the emotion classification system and to provide more relevant and useful feedback for teachers. The focus is on capturing the emotions of interest, disinterest, and surprise, which are more common and useful in classroom settings. The imbalanced dataset and poor performance on real faces were also considered important factors in this decision. The mapped emotion of the remaining 4 classes to emojis can be found in Figure 5.

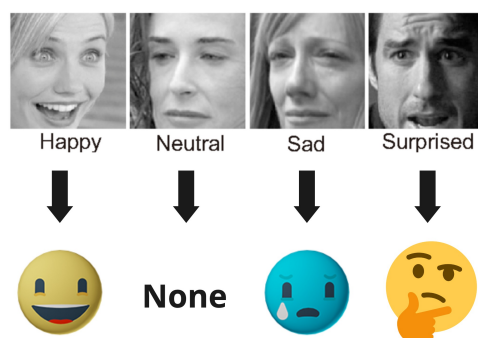


Figure 5. Mapped emojis for each emotion label.

3.3.2. Transfer Learning: Implementation Details

Transfer learning from MobileNet was selected as the emotion model due to its superior performance compared to other models. This is evident from the real web camera

testing results. Traditional models that were trained from scratch using the CK+ and FER2013 datasets had high accuracy results during testing, but their performance was poor when applied to real faces captured from web cameras. The poor performance was due to the fact that the features extracted from these models were not robust enough to handle the variations in real faces. On the other hand, transfer learning from MobileNet utilizes pre-trained layers that have better features, making the model more robust. As a result, live faces captured from web cameras have more accurate classifications, and the results are more consistent.

FER2013 [19] dataset was utilized to fine-tune the model, followed by fine-tuning it again with CK+ [20] data. The weights of a pre-trained MobileNet were frozen, and the model was trained with FER2013 first; then, the weights were unfrozen and the model was fine-tuned. The process was repeated with CK+. An accuracy of 87.5% was achieved for 4-class classification, which is comparable to the benchmark set by the FER2013 dataset. Accuracy refers to the ability to correctly classify 4-class emotions into each label. An accuracy of 87.5% means the model is able to identify most of the emotions correctly. Additionally, the model is able to provide acceptable results with real faces captured from web cameras. It does a better performance from researchers' observation with live web camera captured faces.

The model was implemented with TensorFlow.js [25], allowing it to run on web browsers on the client side, which helps to protect the privacy of students by not transmitting their videos.

The model was integrated as a real-time capture app within the CARVE Hub, where each student can see corresponding emojis pop up for their emotions, as shown in Figure 6. The model automatically ran when participants joined the Hub.



Figure 6. Student's happy emotion captured and sent as an emoji automatically.

3.4. System Design

The design of the CARVE Hub is aimed at making it scalable, efficient, and robust. By combining deep learning with VR, the system can transmit affective information in real time without placing a significant burden on the server. This allows for a more seamless and immersive virtual classroom experience. Additionally, by utilizing the latest deep learning techniques and models, the system is able to accurately classify emotions in real time, providing valuable feedback to teachers and allowing for a more personalized learning experience for students. The design of the system aims to enhance the VR classroom experience and make it more engaging and effective for both students and teachers.

On the technical aspect, the CARVE Hub system comprises two main components. The first component is the CARVE Hub client, which is based on a modified version of the Mozilla Hub and is capable of sending emojis based on emotion classification to the

CARVE interface on AWS. The second component is the emotion classification system, which utilizes Tensorflow.js and runs on each client's browser.

The CARVE Hub client injects JavaScript into the Mozilla Hub and runs the emotion classification model upon the user's joining a room. The system then utilizes an emoji system to send corresponding emojis based on the emotion classification results. The CARVE Hub's own Mozilla Hub server is deployed on AWS, allowing students to join and run the system on their web browsers via a provided link.

The emotion classification model is hosted on AWS S3. Once users join a room, their browsers download the machine learning model and utilize Tensorflow.js to run the emotion classification model locally. All classification tasks and results are computed on each user's browser and computer. The main networked components for users are their avatars and emojis.

The architecture of the system is depicted in Figure 7. The client web browsers consist of two main components: Tensorflow.js and the modified Mozilla Hub client. Tensorflow.js utilizes the emotion classification model downloaded from S3 storage to run. The modified Mozilla Hub client connects with the AWS backend server to the network and communicates with other users.

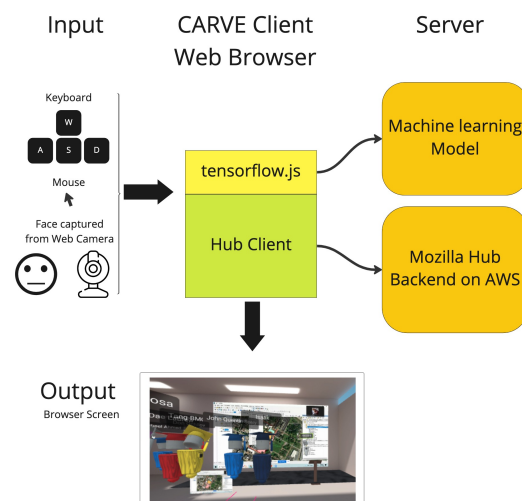


Figure 7. System architecture.

Overall, the deployment of the emotion classification system on the user's web browser allows for a more scalable, efficient, and robust system while also protecting users' privacy:

- Users do not need to transmit video data, which protects the privacy of users' data. Instead, only the classified emotions are sent, which saves bandwidth and reduces network traffic. This also makes the system more efficient and secure.
- Users have the ability to process emotions locally on their machine, allowing for the system to be scalable to a larger population. Currently, due to CARVE Hub VR restrictions, there can be a maximum of 20 people in a class, but the emotion classification system has the potential to scale to a much larger number of users. This makes it more accessible and efficient for a larger group of students to use in a virtual classroom setting.
- The classification model is stored online on AWS S3, which allows the users' client to only download the model parameters. This means that the emotion classification model can be updated in real time, and users do not have to manage anything. Everything is taken care of by the researchers. This approach of storing the model on a cloud server makes it easily accessible for users and also eliminates the need for managing and updating the model on each user's device. Additionally, it also allows for quick integration of any new models that may be developed in the future.

- The full system is completely web browser-based, which means there is no installation required for users. This enhances the overall user experience as it makes it easier for users to join the VR classroom and send their emotions. Not having to go through a complicated installation process also makes it more accessible to a wider range of users. Additionally, the web browser-based system allows for real-time updates and easy management for researchers, as they can add more useful models such as attention and engagement in the future.

In summary, by processing the emotion locally on the user's machine, the system is able to scale to a larger population. The emotion classification system is deployed on the user's web browser, making it more scalable and efficient for a larger population. This also eliminates the need for any installations, making it more user-friendly and easy to join the CARVE Hub. Overall, the design of the system focuses on making emotion classification in VR scalable, efficient, and robust.

3.5. Content and Structure of the Lessons

The experiment was conducted in laboratory classes of Introduction to Geospatial Information Systems (GIS), which introduces various methods for interpreting and analyzing spatial data. The experiments were conducted in two laboratory classes, and students used software called ArcGIS to complete GIS lab exercises. The first lab was to georeference a historical map through Zoom. Students needed to select a scanned historical map image of New York City created 200 years ago, identify landmarks on the historical map and select the corresponding points in a GIS layer (Openstreet Map is used in the lab). Then ArcGIS can compute the coordinate transformation between the historical map image and the Openstreet map based on the selected corresponding landmarks. The transformation aligns the historical map image with the real world. The challenge of the lab was to understand the georeference concepts. The historical map image was not very accurate; hence, it was important to identify more appropriate landmarks on both the image and the map to ensure the quality of the georeference.

The second lab class taught students to create a geospatial layer through the proposed approach taught in the class. Students needed to create a university campus GIS layer. Points, lines, and polygons were drawn on a GIS base layer in ArcGIS to represent single real-world locations, roads, and buildings or parking lots, respectively. The challenges of the lab were to understand the concept of vector data, and to create accurate vector data. For example, polygons showing property boundaries must be closed, undershoots or overshoots of the border lines were not allowed, and contour lines in a vector line layer must not intersect.

In both lab classes, the instructor used PowerPoint slides to review the basic GIS concepts learned in previous lectures, introduced the features of the GIS software, and showed a brief demonstration and emphasized the challenges of the lab exercises. Students were allowed to ask questions or talk to their classmates.

Breakout rooms allow students to easily collaborate in small teams, and they are often used when teamwork or group discussion is needed. In the GIS classes, students were required to complete individual lab exercises; therefore, group discussion was not needed in the lab classes; hence, we did not use a breakout room with fixed team members in any lab classes. However, students often collaborate with each other in an in-person class. For example, a student sometimes got stuck when they had difficulty localizing a function or feature in ArcGIS or did not set up the function parameters correctly. Therefore, an online classroom with easy collaboration among students can improve their learning experiences and effectiveness, which will be tested in our future experiments.

3.6. Survey and Experiment Setup

The survey responses in this study utilize a 7-point rating scale. Two surveys were performed, one for the Zoom session and one for the CARVE session, and they contained

the same 18 questions to facilitate comparison. The survey questions can be found in Appendix A.

The survey was designed to gather feedback from students on two different virtual classroom platforms, Zoom and CARVE Hub. The survey was divided into three main groups to test three hypotheses: social interaction and connections, classroom experience, and product features. For each hypothesis, a set of questions were proposed for students to rate on a scale of 1 (strongly disagree) to 7 (strongly agree).

The first hypothesis is that there is a significant difference in students' social interactions and connections between Zoom and CARVE Hub. To test this hypothesis, we designed 8 questions related to social interactions and connections. These questions aimed to gather feedback on how students felt about their interactions and connections with other students and teachers in the virtual classroom.

The second hypothesis is that there is a significant difference in students' overall classroom experience between Zoom and CARVE Hub. To test this hypothesis, we designed 4 questions related to the classroom experience. These questions aimed to gather feedback on how satisfied students were with the format, how much content they learned, how engaged they were, and how comfortable they felt in the virtual classroom.

The third hypothesis is that there is a significant difference in students' perceptions of the product features between Zoom and CARVE Hub. To test this hypothesis, there were 6 questions related to product features. These questions aimed to gather feedback on how useful and helpful the various features of the virtual classroom are for communication and interaction.

In addition to the quantitative data, there were also 5 open-ended questions for qualitative analysis. These questions allowed students to leave comments and provide more detailed feedback on their experience with the CARVE Hub. This qualitative data can provide more in-depth insights into students' perceptions and opinions of the virtual classroom.

The survey collected both quantitative data from the rating scales and qualitative data from open-ended questions. The goal of the survey was to compare the ratings from students using Zoom and CARVE Hub in order to validate the hypotheses. The results of the survey can be used to understand students' remote preferences for each learning software. Overall, the survey design aimed to gather comprehensive data from students on their experiences with two remote learning tools, Zoom and CARVE, and Zoom was used as a benchmark for comparison.

On the day of the experiment, a link was sent to all students and the teacher to join the CARVE session. Before the class began, 15 min were dedicated to explaining the basic features and movements to the students. The lecture then proceeded, with content similar to that of the Zoom class. Students were taught and asked to operate software during the class. After the class concluded, students were asked to complete a survey. The two classes, the CARVE and Zoom class, offered different content but were comparable in terms of teaching and learning materials. The overall process is depicted in Figure 2.

4. Experimental Results

The evaluation of this research is focused on understanding the effectiveness of the emotion classification system in VR classrooms. To achieve this, a set of metrics were used to analyze the data collected from students' feedback. These metrics include mean, median, variance, statistical comparison using Wilcoxon Signed Rank tests, T-test, and correlation measures. This is similar to the evaluation methods used in other papers testing surveys in VR classes, such as [8]. Additionally, this research adds the use of variance and T-test to provide a more comprehensive understanding of the results.

The research also focuses on comparing the responses for each question under the three main hypotheses: on social interaction, on classroom experience, and on product features. The social interaction hypothesis has eight questions (marked by the red bar in Figure 8), the classroom experience hypothesis has four questions (marked by the green

bar in Figure 8), and the product feature hypothesis has six questions (marked by the blue bar in Figure 8). By comparing the mean, median, and p values of the Wilcoxon test and t -test for each hypothesis in Table 1, the research aims to understand the effectiveness of the emotion classification system in facilitating social interaction, improving classroom experience, and enhancing product features.

The survey results of each question are visualized in Figure 8. The purple line shows the median for each response. The results show higher mean and median for questions in social interactions, including Students Interactions, Talk to Teacher, Teacher Interaction, Students Connection, and Teacher Connection.

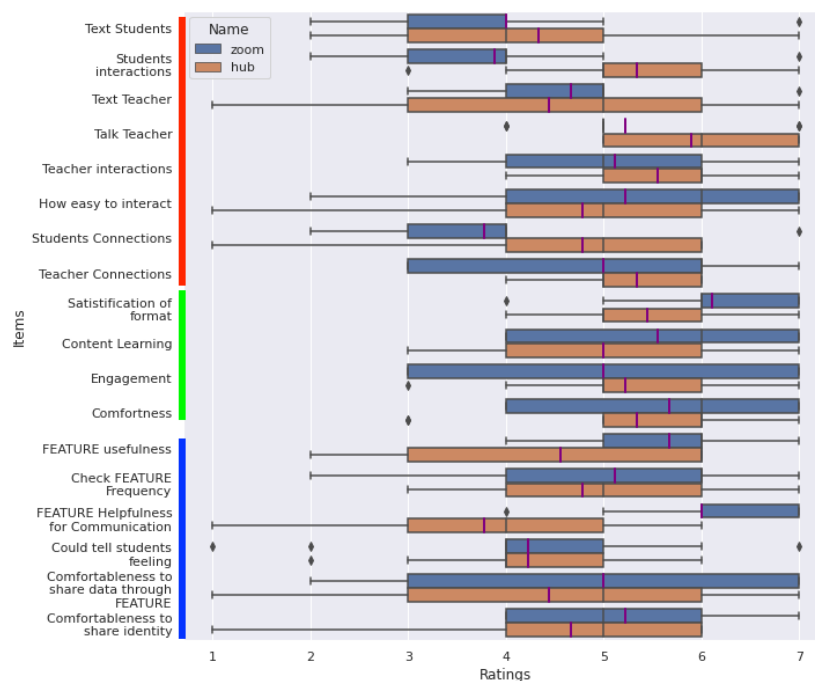


Figure 8. Survey result: Purple line is median. Red section questions are social interactions, green section questions are experience-related, and blue section questions are automated emoji features vs. camera features. Student rating scale of each item from 1 to 7 (strongly disagree to strongly agree).

As a result of this study, the following advantages of the CARVE Hub were observed by comparing the aggregated mean rating of all questions in each of three categories: Interaction, Experience, and Feature. The results are shown in Table 1. The CARVE Hub was found to have higher mean scores for social interactions when compared to Zoom responses. In addition to the above metrics, this research also includes the calculation of p -values using the Wilcoxon Signed Rank test to compare the scores between the CARVE Hub and Zoom. The results of this test show that there is a significant difference in scores between the two platforms, with the CARVE Hub scoring higher than Zoom in terms of social interactions. This indicates that the CARVE Hub is a more effective platform for facilitating student interactions in a remote classroom setting. No significant differences were found in the Experience category. However, when it comes to feature preference, Zoom received a higher score than the CARVE Hub. This is expected, as the two products are significantly different in terms of their features and capabilities, and students might prefer Zoom since they have used it for more than 2 years for online education.

When breaking down the significant differences for each question, two questions stood out with significant p values. The first was related to student interactions, and the second was related to the helpfulness of communication features, specifically the video panel in Zoom and the emotion communication feature in the CARVE Hub.

Some more details on the three hypotheses will be discussed in the following sections.

Table 1. Aggregated response result. Each category is the aggregated mean score for all questions in that section. Interactions are informal social interaction, Experience is learning experience, and Feature is Hub vs. Zoom features. Please see Appendix A. The higher values are in bold. Significant p values are also in bold.

Category	Interactions	Experience	Feature
Zoom Mean	4.6	5.6	5.2
Hub Mean	5.1	5.3	4.4
Zoom Median	4.0	6.0	5.5
Hub Median	5.0	5.0	5.0
Wilcoxon Tests p Value	0.03	0.20	0.01
t -Tests p Value	0.08	0.29	0.01

4.1. Social Interactions and Connections

The following survey questions were found to have significantly higher ratings in the CARVE Hub compared to Zoom: “student interactions among themselves” and “students’ feeling of connections among themselves”. Despite the small sample size, a significant difference was found in the student interaction question, with a p -value of less than 5% when comparing students’ interactions in Zoom and in CARVE Hub. Additionally, the students reported feeling more connected to the teacher in the CARVE Hub, and the standard deviation was smaller, indicating that students had more interactions and connections with the teacher and that there was less disagreement among students in their responses.

The correlation among questions provides interesting insights. The CARVE Hub and Zoom question correlation matrices can be found in Appendix B. By analyzing the correlations among factors, it was found that “students’ connections” in the CARVE Hub have high correlations with “how identity is shared”, “sharing data through emojis”, and “automated emojis”. “students’ connections” in Zoom were found to have high correlations with “texting with students”, “texting with the teacher”, and “student interactions”. Similarly, “students’ interactions” in the CARVE Hub were found to be highly correlated with “how their data is shared” and “using automated emojis”, while “students’ interactions” in Zoom were found to be highly correlated with “texting”.

4.2. Classroom Experience

When examining the overall experience for the class, we focused on “satisfaction with the format”, “content learning”, “engagement”, and “comfort”. The results show that students generally agreed to have better satisfaction with the Zoom format. It seems that students had better content learning during the Zoom session. For engagement and comfort, students had divided opinions for the Zoom class, but their opinions were more consistent for the CARVE Hub class. The CARVE Hub’s engagement was found to be higher than Zoom, and the upper interquartile range was greater than the lower interquartile range when compared to Zoom. This suggests that students were more consistent in agreeing that the CARVE Hub had better engagement, and that there were divided opinions on engagement when rating the Zoom class.

For the qualitative responses, the preference for Zoom and CARVE Hub’s classes is divided. Four out of seven responses preferred the CARVE Hub’s format, while the remaining three preferred the Zoom format. However, out of the three responses preferring Zoom, one reported disliking “too many distracting elements to learn”, and the other one reported disliking “only because we were unable to finish the lab guide in time”. For those who favored the CARVE Hub over Zoom, interaction was the common topic they liked about the CARVE Hub.

Additionally, when asked about the most helpful feature, most of the students liked the ability to share multiple screens and to walk around in the virtual environment. This

suggests that interactions were a significant factor that students favored in the CARVE Hub class.

When examining the correlation between overall classroom experience and the survey data, the following high correlations were found for the CARVE Hub class: *“using emojis with other students”*, *“talking to students and teachers”*, *“spatial audio”*, and *“interacting with other students”*. More interestingly, it was found that students who preferred using cameras over emojis had the highest negative correlations with classroom experience in the CARVE Hub class. Those who preferred using cameras were not satisfied with the CARVE Hub class. One qualitative response suggests that a neurological disorder or simulator sickness might prevent learning. For the Zoom classes, it seems that *“teacher interaction”*, *“using cameras”*, and *“easiness to interact”* were highly positively correlated with *“satisfaction”*, *“content learning”*, *“engagement”*, and *“comfort”*.

Furthermore, for both Zoom and CARVE Hub, engagement was found to have a high correlation with the number of content students learned from the class. For engagement in the CARVE Hub, it seems that students were engaged in class by talking, using emojis, and interacting with other students. In contrast, students were engaged in Zoom by interacting with the teacher, talking to the teacher, and checking video panels. This suggests that students are more engaged through peer interactions when given the opportunity, but Zoom does not support this factor as much as the CARVE Hub.

4.3. Emotion Data: CARVE’s New Feature

The results of this study indicate that there are no advantages to abstracting emotion data into emojis when compared to Zoom. There were also no negative effects reported by students due to automated emoji bugs. The automated emoji system was found to be less useful than the cameras, and Zoom’s video feed was found to have a significant advantage for communication with a p -value of less than 5%. Other ratings for both the Zoom and CARVE Hub experiments were found to be similar. Overall, it appears that students have more acceptance of the camera compared to the automated emojis.

An interesting result is that students’ interactions were found to be statistically significantly higher in the CARVE Hub, and they were highly correlated with sharing their data through emojis and with communication through automated emojis. This high correlation was not observed among *“interaction”*, *“how comfortable data is shared through video”*, and *“video helpfulness for communication”* in the Zoom class. However, the automated emoji system was found to be significantly less useful for communication when compared to video feeds. The CARVE Hub had higher interactions than Zoom, but the feature of automated emojis was found to be less useful for communication. It is possible that there are other factors contributing to interactions other than communication, and that automated emojis may be helpful for interactions but not communication. The emojis might provide more non-verbal and non-video interactions.

5. Discussions

5.1. Interaction

Our correlation study suggests that emojis may help with interactions. Zoom’s video feature is ideal for communication but may not be sufficient for interactions. Video may not be the most effective form of interaction.

The scores for interactions were found to have a high correlation with how data are shared with automated emojis. From the qualitative responses, it was observed that the CARVE Hub has advantages in interactions when compared to Zoom, but Zoom is much better with communication. We suspect this is due to the limitations of how the question was asked, as interactions might capture a broader range of activities than communication, and Zoom is better suited for clearly talking and transmitting information.

Automated emojis were found to have a high correlation with students’ interactions. Students were more active in using emojis, and they even used emojis outside of the four automated emoji categories. Students frequently used clapping, love, hand raising, and

other emojis. Since they had to manually click on those emojis in order to use them, students actively used the emoji system. Emojis may be a good way for students to interact with each other, as they encourage active participation.

Furthermore, the highest correlated factor for interactions from the Zoom survey was text, not video. Even though students did not turn on cameras during the CARVE Hub class, their interaction score was still higher than in the Zoom class. This suggests that video may not be the most effective method for interactions. When comparing the CARVE Hub's automated emojis and Zoom's video feed, students did not favor one over the other, as shown in Figure 9.

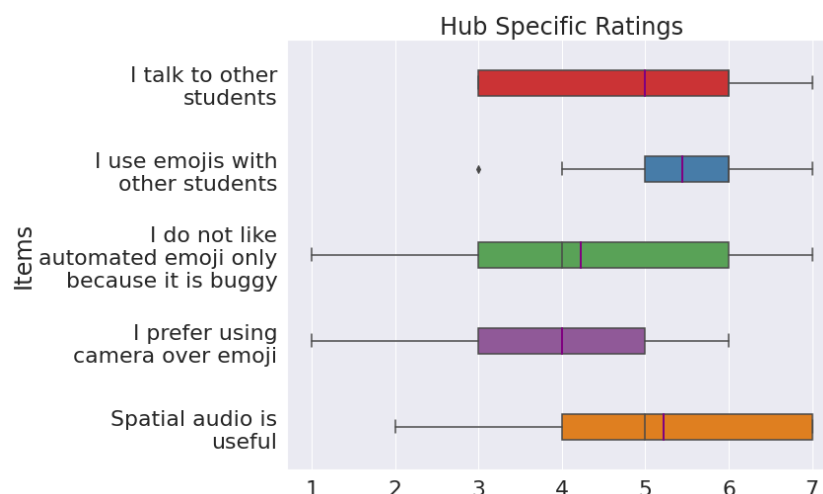


Figure 9. Hub Specific Ratings: student rating scale from 1 to 7 (strongly disagree to strongly agree).

Considering that the highest correlation for interactions in the CARVE Hub is related to how comfortable students were with sharing data, we speculate that sharing data without showing faces can increase interactions. Students might feel less comfortable interacting when they are in the spotlight and showing their faces. Being hidden behind an avatar or a name might make it easier to interact. A private and informal method of communication might be able to facilitate more interactions.

We believe that avoiding forcing students to be under a spotlight might help increase their interactions with each other. If students are forced to show their faces while everyone can see them all the time, they might feel anxious. Both emojis and text are more private and informal channels. Students are able to hide behind an identity, and they are not under the constant pressure of being watched by others. This could be the reason why there are higher correlations for those methods.

5.2. Collaboration

One recurring theme mentioned in the qualitative responses and observed during the experiment is collaboration. Share screen and collaboration were the most frequently mentioned feedback from students, and it was observed that students actively drew, talked, and checked other students' screens. The virtual classroom environment provided by the CARVE Hub may facilitate a better collaborative environment than traditional video conferencing, with the added convenience of collaboration tools. We observed that collaboration also increases interactions between students and the teacher, and it is more similar to in-person classes where the teacher can walk around and assist students with their problems.

Collaborative learning can lead to the development of higher-level thinking, oral communication, self-management, and leadership skills. It also promotes student–faculty interaction, increases student retention and self-esteem, and exposes students to diverse perspectives. Additionally, it helps students have higher achievement in university classes [26].

In the Zoom class, only one user is allowed to share their screen. Hence, when students need help in lab activities, the instructor asks one student to share their screen. Other students have to wait until the previous student is finished, while in the CARVE Hub, students are allowed to share their screens anytime and anywhere; some students project their screens on virtual walls, and others project their screens on virtual tables. Any two students can work together as long as they are next to each other.

Since students contribute to the success of the CARVE Hub through collaboration, it is suggested that a more comprehensive study of collaboration in a virtual classroom should be a priority for future research.

5.3. Weakness of CARVE Hub

During the experiment, two students reported a dislike for the CARVE Hub and rated most of the questions far below the average score. Our analysis found that those who preferred not to use cameras had a negative correlation with the interaction rating. It is possible that these students are shy and prefer to learn through listening rather than interacting. One student also mentioned the issue of simulator sickness, which may have affected their ability to learn effectively in the 3D environment of the CARVE Hub. A possible solution to this issue is to create a focus mode that only displays the teacher's slides and audio, rather than the other students and the room environment, as suggested by a previous study [9]. This may help students focus on the lecture and improve their learning experience.

5.4. Future Directions

The students who participated in this study had a positive reaction when they first joined the CARVE Hub. Many of them were curious about the platform and asked the researchers how they could access it in the future. When asked about their preferences, some students said they preferred the CARVE Hub class "because it was new and engaging for someone like me who mainly on the computer". One student specifically mentioned that they enjoyed the overall concept of the Hub and that it could be useful in various classroom settings, such as science classes and research projects. Another student commented that "it could be useful in science classrooms (physics, general high school sciences, nursing programs, and so on), as well as research projects (social experiments and so on)". Overall, the students were excited about the potential of the CARVE Hub in education.

Even though the sample size of our experiment is small, there are statistically significant comparisons. Some students also mentioned that it may be more effective with larger class size. A study with more participants would be beneficial in order to further validate the results.

We also observed that some qualitative responses were not reflected in the quantitative data. Collaboration was a key focus for students in the qualitative responses, and students reported that features such as improved drawing tools, better screen sharing, and better spatial audio would improve their remote classroom experience.

Multiple students mentioned spatial audio as helpful in qualitative responses. Spatial audio is a technology that allows for the creation of a more immersive audio experience in virtual environments. It enables users to hear sounds as if they were coming from different locations, making it feel as though they are in the same room as the source of the sound. This technology is particularly useful in virtual classrooms, as it allows students to join and switch group discussions more naturally. They can discuss while listening to the lecture, without having to switch into different breakout rooms, and have easier collaboration. It is one of the highlights of students' responses.

The CARVE Hub has similar settings as the real classroom for spontaneous discourses. Spontaneous discourses are beneficial for both students and teachers. They provide opportunities for students to engage in active learning, where they can apply and practice the knowledge and skills they have acquired. Some researches show a significant increase in students' performance with spontaneous group discussion [27].

When there are no emotion data used, the research intends to use spatial locations, avatar movement activities, and communication activities. The idea is to use different features to identify affective interactions, such as attention, engagement, etc.

This work focuses on capturing emotions that are related to the class and not the full range of emotions. In future research, attention and engagement will be studied in addition to emotions. This allows for a more comprehensive understanding of student engagement and participation in virtual classrooms. Additionally, the use of spatial locations, avatar movement activities, and communication activities can also provide valuable insights into affective interactions in VR environments.

Based on the results of this study, we propose the following features for future remote classroom platforms:

1. A more natural interaction method that does not force participants to be in the spotlight. The automated emoji system was highly correlated with increased interactions, and there may be other solutions that can be considered.
2. Improved user collaboration methods. Students favored collaboration features, and it may help to create a better classroom environment.
3. A simplified control interface. Students reported that the complexity of the CARVE Hub's control interface was a challenge, and they spent a lot of time learning how to navigate the platform.

6. Conclusions

In conclusion, the CARVE Hub system increased interactions among students compared with a classic Zoom class. The top three features related to informal interactions were: the usage of automated emojis, talking with each other, and the spatial audio in the VR classroom. Qualitative responses also reflected similar feedback, such as "increased interactions", "better screen sharing capabilities", and "the ability to walk around in the virtual environment". The Hub's automated emoji system, although not yet effective for natural communication, did increase private interactions among students.

This study also found that students did not have a clear preference for using emojis or video when comparing the two features directly, but video feeds may not be as crucial for facilitating interactions. Instead, collaboration was identified as the most important feature for students and should be prioritized in future remote classroom design. Furthermore, this study suggests that alternatives to current video feeds, where everyone shows their faces at all times, should be considered in remote classrooms. The use of emojis without video feeds was found to have no negative effect on interactions, even though this is apparently not sufficient for effective communication. Interactions in the Hub are statistically significantly higher than in Zoom. Better design choices may exist, and they can perform better than video feeds. Additionally, students' feedback highlighted the importance of collaboration, which should be a focus in future remote classroom design.

Furthermore, our research intends to test more affective information in the future, such as attention and engagement, to enhance communication in a virtual classroom. We plan to use more advanced machine learning models that can detect these additional emotional states and provide a more comprehensive understanding of students' engagement levels. This will allow us to improve the overall communication and learning experience in a virtual classroom. Additionally, it will also provide insights on which features are more important for effective virtual communication and learning. This will help us develop better models and improve the overall system performance.

The code implementation can be found at (accessed on 1 March 2023) <https://github.com/user074/hubs/tree/hubs-cloud>.

Author Contributions: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, visualization: J.Q.; conceptualization, writing—review and editing, supervision, project administration, funding acquisition: H.T. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The research is supported by the National Science Foundation (NSF) through Awards #2131186 (CISE-MSI), #2202043 (ATE), #1827505 (PFI), and #1737533 (S&CC). The work is also supported by the US Air Force Office of Scientific Research (AFOSR) via Award #FA9550-21-1-0082 and the ODNI Intelligence Community Center for Academic Excellence (IC CAE) at Rutgers University (#HHM402-19-1-0003 and #HHM402-18-1-0007).

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the research, conducted in established or commonly accepted educational settings, that specifically involves normal educational practices that are not likely to adversely impact students' opportunity to learn required educational content or the assessment of educators who provide instruction. This includes most research on regular and special education instructional strategies, and research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Survey Questions

Informal social interactions

- I chat with other students through text
- I talk to other students (chat, asking questions, discussing problems, etc) (*Hub only*)
- I use emojis with other students (*Hub only*)
- I interact with other students
- I chat with teacher through text
- I talk to teacher (asking questions, discussing problems, etc)
- I interact with teacher
- I find it easy to interact with other people
- I feel connected with other students
- I feel connected with teacher

Learning experiences

- I am satisfied with overall format of the course
- The platform helps me to learn a lot of content from the class
- I am engaged in the class
- I feel comfortable during the class

Hub Features vs. Zoom Features

- FEATURE(Camera or Emoji) is useful
- I check FEATURE(Camera video stream or Emoji) frequently
- FEATURE(Camera or Emoji) is helpful for my communication
- I do not like automated emoji only because it is buggy(*Hub only*)
- I could tell how other students feel
- I feel comfortable to share my data through FEATURE(Camera or Emoji)
- I prefer using camera over emoji (*Hub only*)
- I feel comfortable regarding how my identity is shared on this platform
- Spatial audio is useful (*Hub only*)

Open ended short answers (*Hub only*)

- Do you prefer last week (Zoom) or this week's class (Hub)? Why?
- What is the most helpful feature and why do you like it?
- What frustrations do you have for remote class?
- What would you suggest to improve?
- Anything else you want to tell us? Feel free to leave anything here

Appendix B. Correlation Matrix

The list of questions for both Zoom and Hub correlation matrices are:

- I chat with other students through text
- I interact with other students
- I chat with teacher through text
- I talk to teacher (asking questions, discussing problems, etc)
- I interact with teacher
- I find it easy to interact with other people
- I feel connected with other students
- I feel connected with teacher
- I am satisfied with overall format of the course
- The platform helps me to learn a lot of content from the class
- I am engaged in the class
- I feel comfortable during the class
- FEATURE(Camera or Emoji) is useful
- I check FEATURE(Camera video stream or Emoji) frequently
- FEATURE(Camera or Emoji) is helpful for my communication
- I could tell how other students feel
- I feel comfortable to share my data through FEATURE(Camera or Emoji)
- I feel comfortable regarding how my identity is shared on this platform

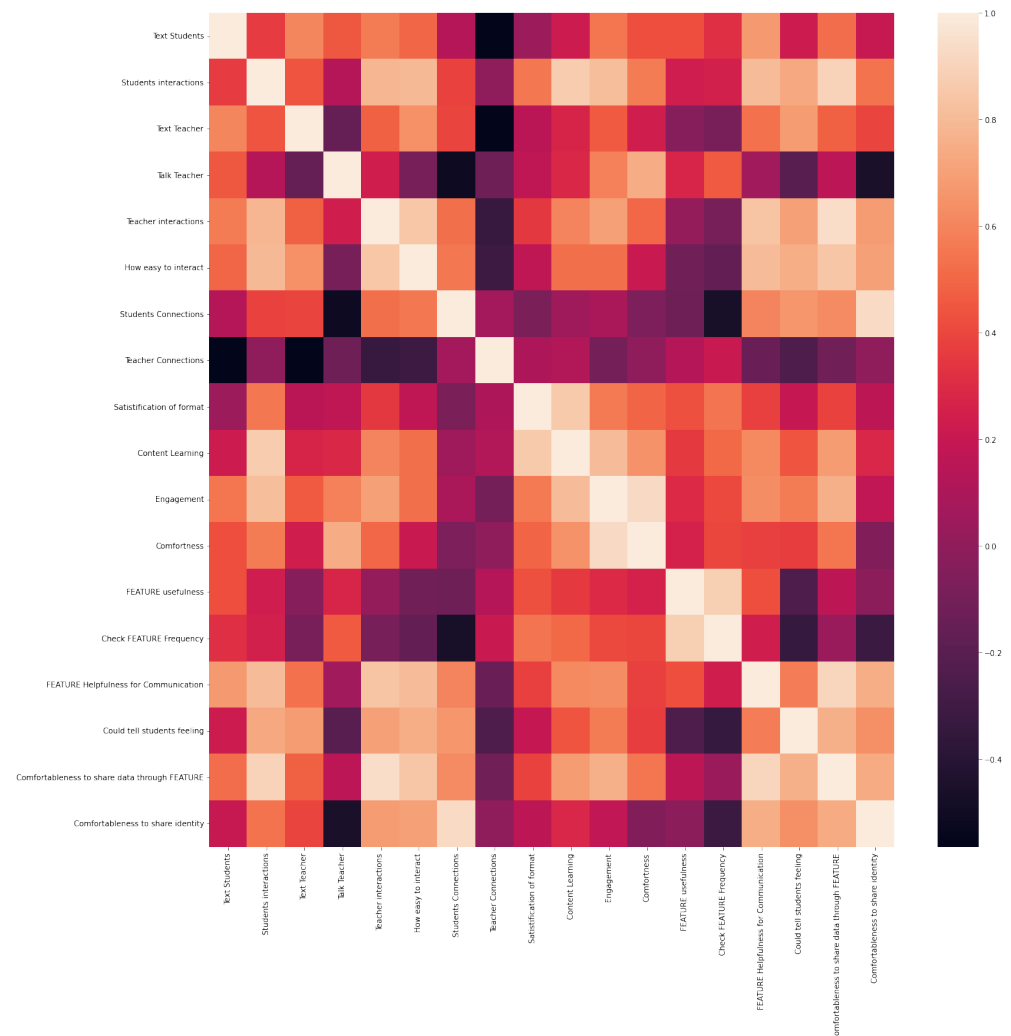


Figure A1. Hub correlation.

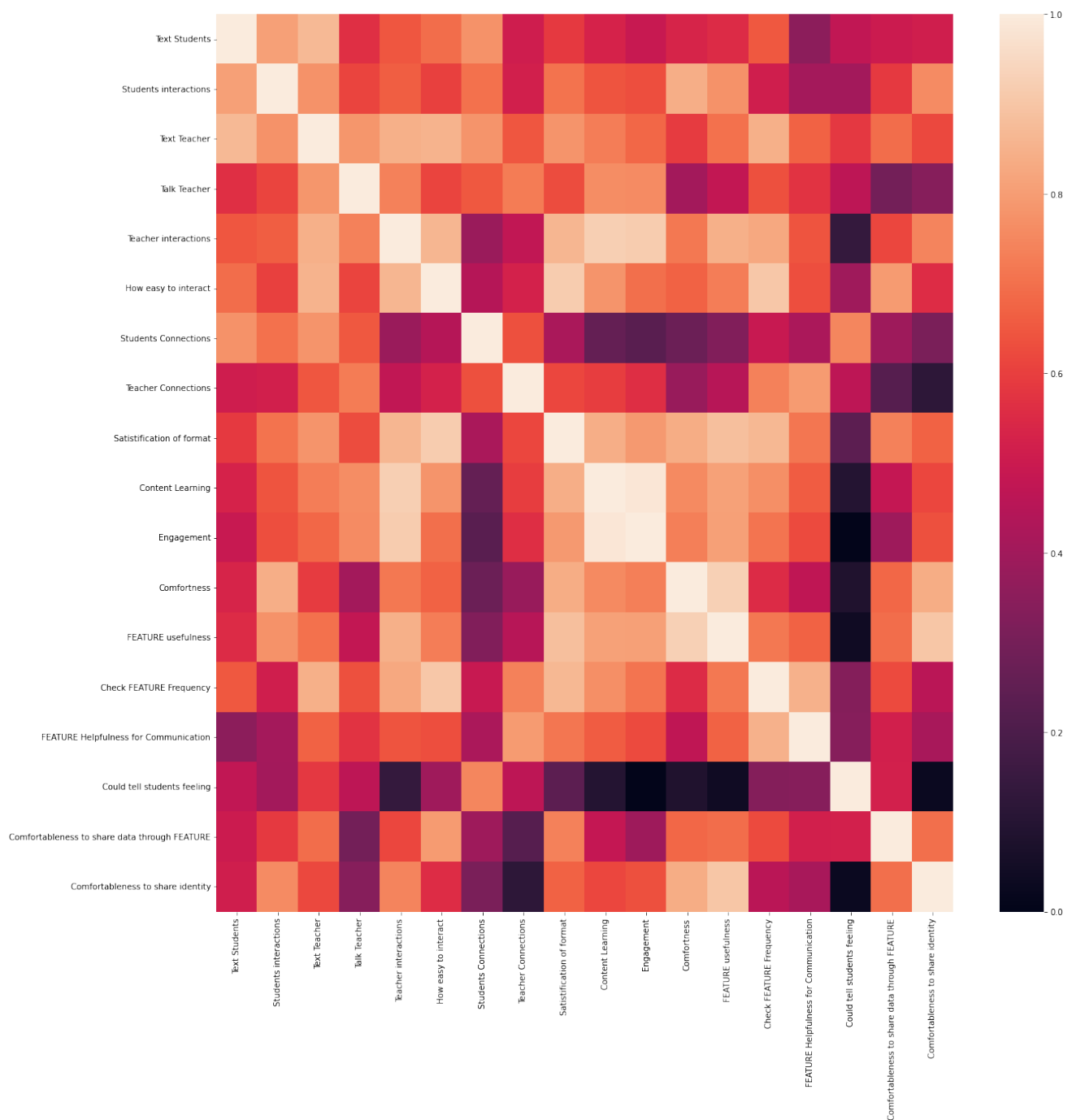


Figure A2. Zoom correlation.

References

1. Demand for Online Education is Growing. Are Providers Ready? Available online: <https://www.mckinsey.com/industries/education/our-insights/demand-for-online-education-is-growing-are-providers-ready> (accessed on 18 November 2022).
2. Martín-Gutiérrez, J.; Fabiani, P.N.; Benesova, W.; Meneses, M.D.; Mora, C.E. Augmented Reality to Promote Collaborative and Autonomous Learning in Higher Education. *Comput. Hum. Behav.* **2015**, *51*, 752–761. [CrossRef]
3. Peper, E.; Wilson, V.; Martin, M.; Rosegard, E.; Harvey, R. Avoid Zoom Fatigue, Be Present and Learn. *NeuroRegulation* **2021**, *8*, 47–56. [CrossRef]
4. Peper, E.; Yang, A. Beyond Zoom Fatigue: Re-energize Yourself and Improve Learning. *Acad. Psychiatry* **2021**. [CrossRef]
5. Castelli, F.; Sarvary, M. Why students do not turn on their video cameras during online classes and an equitable and inclusive plan to encourage them to do so. *Ecol. Evol.* **2021**, *11*, 3565–3576. [CrossRef] [PubMed]

6. Jin, Q.; Liu, Y.; Yarosh, S.; Han, B.; Qian, F. How Will VR Enter University Classrooms? Multi-Stakeholders Investigation of VR in Higher Education. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans LA USA, 29 April–5 May 2022; pp. 1–17. [\[CrossRef\]](#)
7. Pallavicini, F.; Pepe, A. Comparing Player Experience in Video Games Played in Virtual Reality or on Desktop Displays: Immersion, Flow, and Positive Emotions. In Proceedings of the Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, Barcelona Spain, 22–25 October 2019; pp. 195–210. [\[CrossRef\]](#)
8. Yoshimura, A.; Borst, C.W. Evaluation and Comparison of Desktop Viewing and Headset Viewing of Remote Lectures in VR with Mozilla Hubs. In Proceedings of the ICAT-EGVE 2020—International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, Virtual, 2–4 December 2020; pp. 51–59. [\[CrossRef\]](#)
9. Chen, Z.; Cao, H.; Deng, Y.; Gao, X.; Piao, J.; Xu, F.; Zhang, Y.; Li, Y. Learning from Home: A Mixed-Methods Analysis of Live Streaming Based Remote Education Experience in Chinese Colleges during the COVID-19 Pandemic. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, ACM, Yokohama, Japan, 8–13 May 2021; pp. 1–16. [\[CrossRef\]](#)
10. Yarmand, M.; Solyst, J.; Klemmer, S.; Weibel, N. “It Feels Like I Am Talking into a Void”: Understanding Interaction Gaps in Synchronous Online Classrooms. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI ’21, Yokohama, Japan, 8–13 May 2021; pp. 1–9. [\[CrossRef\]](#)
11. Song, J.; Riedl, C.; Malone, T.W. Online Mingling: Supporting Ad Hoc, Private Conversations at Virtual Conferences. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI’21, Yokohama, Japan, 8–13 May 2021; pp. 1–10. [\[CrossRef\]](#)
12. Yao, N.; Brewer, J.; D’Angelo, S.; Horn, M.; Gergle, D. Visualizing Gaze Information from Multiple Students to Support Remote Instruction. In Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, CHI EA ’18, Montreal, QC Canada, 21–26 April 2018; pp. 1–6. [\[CrossRef\]](#)
13. Wang, Q.; Jing, S.; Joyner, D.; Wilcox, L.; Li, H.; Plötz, T.; Disalvo, B. Sensing Affect to Empower Students: Learner Perspectives on Affect-Sensitive Technology in Large Educational Contexts. In Proceedings of the Seventh ACM Conference on Learning @ Scale, Virtual, 12–14 August 2020; pp. 63–76. [\[CrossRef\]](#)
14. Yoshimura, A.; Borst, C.W. Remote Instruction in Virtual Reality: A Study of Students Attending Class Remotely from Home with VR Headsets. In Proceedings of the Mensch und Computer 2020—Workshopband, Magdeburg, Germany, 6–9 September 2020. [\[CrossRef\]](#)
15. Broussard, D.M.; Rahman, Y.; Kulshreshtha, A.K.; Borst, C.W. An Interface for Enhanced Teacher Awareness of Student Actions and Attention in a VR Classroom. In Proceedings of the 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Lisbon, Portugal, 27 March–1 April 2021; pp. 284–290. [\[CrossRef\]](#)
16. Chen, M. Visualizing the Pulse of a Classroom. In Proceedings of the Eleventh ACM International Conference on Multimedia, Berkeley, CA, USA, 2–8 November 2003; pp. 555–561. <https://dl.acm.org/doi/10.1145/957013.957130>.
17. Asteriadis, S.; Tzouveli, P.; Karpouzis, K.; Kollias, S. Estimation of Behavioral User State Based on Eye Gaze and Head Pose—Application in an e-Learning Environment. *Multimedia Tools Appl.* **2009**, *41*, 469–493. [\[CrossRef\]](#)
18. De Freitas, S. *Learning in Immersive Worlds: A Review of Game-Based Learning*; JISC: Bristol, UK, 2006.
19. Goodfellow, I.J.; Erhan, D.; Carrier, P.L.; Courville, A.; Mirza, M.; Hamner, B.; Cukierski, W.; Tang, Y.; Thaler, D.; Lee, D.H.; et al. Challenges in Representation Learning: A report on three machine learning contests. *arXiv* **2013**, arXiv:1307.0414. [\[CrossRef\]](#)
20. Kanade, T.; Cohn, J.; Tian, Y. Comprehensive database for facial expression analysis. In Proceedings of the Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), Grenoble, France, 28–30 March 2000; pp. 46–53. [\[CrossRef\]](#)
21. Li, H.; Wang, N.; Ding, X.; Yang, X.; Gao, X. Adaptively Learning Facial Expression Representation via C-F Labels and Distillation. *IEEE Trans. Image Process.* **2021**, *30*, 2016–2028. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861. [\[CrossRef\]](#)
23. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [\[CrossRef\]](#)
24. Delgado, K.; Origgi, J.M.; Hasanpoor, T.; Yu, H.; Alessio, D.; Arroyo, I.; Lee, W.; Betke, M.; Woolf, B.; Bargal, S.A. Student Engagement Dataset. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 3621–3629. [\[CrossRef\]](#)
25. Smilkov, D.; Thorat, N.; Assogba, Y.; Yuan, A.; Kreeger, N.; Yu, P.; Zhang, K.; Cai, S.; Nielsen, E.; Soergel, D.; et al. TensorFlow.js: Machine Learning for the Web and Beyond. *arXiv* **2019**, arXiv:1901.05350. [\[CrossRef\]](#)
26. Chandra, R. Collaborative Learning for Educational Achievement. *Int. J. Res. Method Educ.* **2015**, *5*, 4–7.
27. Lestari, R. The use of spontaneous group discussion (SGD) to teach writing descriptive text: A case of the eighth grade students of SMP 1. *Eternal Engl. Teach. J.* **2018**, *5*, 2. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.