# PhysioML: A Web-Based Tool for Machine Learning Education with Real-Time Physiological Data

Bryan Y. Hernández-Cuevas
bryan.hernandez5@upr.edu
University of Alabama
Tuscaloosa, AL

Myles Lewis
mlewis16@crimson.ua.edu
University of Alabama
Tuscaloosa, AL

Wesley Junkins
wcjunkins@crimson.ua.edu
University of Alabama
Tuscaloosa, AL

Chris S. Crawford
crawford@cs.ua.edu
University of Alabama
Tuscaloosa, AL

Andre Denham
adenham@ua.edu
University of Alabama
Tuscaloosa, AL

Feiya Luo
feiya.luo@ua.edu
University of Alabama
Tuscaloosa, AL

## Abstract

Artificial Intelligence and Machine Learning continue to increase in popularity. As a result, several new approaches to machine learning education have emerged in recent years. Many existing interactive techniques utilize text, image, and video data to engage students with machine learning. However, the use of physiological sensors for machine learning education activities is significantly unexplored. This paper presents findings from a study exploring students' experiences learning basic machine learning concepts while using physiological sensors to control an interactive game. In particular, the sensors measured electrical activity generated from students' arm muscles. Activities featuring physiological sensors produced similar outcomes when compared to exercises that leveraged image data. While students' machine learning self-efficacy increased in both conditions, students seemed more curious about machine learning after working with the physiological sensor. These results suggest that PhysioML may provide learning support similar to traditional ML education approaches while engaging students with novel interactive physiological sensors. We discuss these findings and reflect on ways physiological sensors may be used to augment traditional data types during classroom activities focused on machine learning.

## CCS Concepts

• **Applied computing** → **Interactive learning environments**; • **Computing methodologies** → *Machine learning*; • **Social and professional topics** → **Computing education**.

## Keywords

Physiological Computing, Computer Science Education, Machine Learning, Muscle Computer Interfaces, Electromyography (EMG)


*SIGCSE TS 2025, February 26-March 1, 2025, Pittsburgh, PA, USA*

ACM ISBN 979-8-4007-0531-1/25/02
https://doi.org/10.1145/3641554.3701815

**ACM Reference Format:**
Bryan Y. Hernández-Cuevas, Myles Lewis, Wesley Junkins, Chris S. Crawford, Andre Denham, and Feiya Luo. 2025. PhysioML: A Web-Based Tool for Machine Learning Education with Real-Time Physiological Data. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE TS 2025), February 26-March 1, 2025, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3641554.3701815

## 1 Introduction

Artificial Intelligence (AI) and Machine Learning (ML) are becoming increasingly prominent in everyday technologies. Users encounter increasingly intelligent software when interacting with social media, digital entertainment, games, online shopping, and more. However, few people understand the underlying processes used to develop ML applications. To address this gap, Touretzky et al. [42] introduced guidelines that discuss what learners should know about AI and machine learning. Researchers have also discussed various ways to conduct performance-based assessments for machine learning concepts [14]. Previous research that builds on these guidelines and frameworks is primarily dominated by audio, image, and text data [8]. This reality is likely due to the extensive availability of learning resources and datasets curated for these data types. However, emerging technologies such as physiological sensors offer the exploration of new data types that integrate our bodies and technology [23]. Furthermore, exploring such sensing approaches may encourage a type of self-discovery originally introduced by Papert [32] that is often overlooked [9]. To address this gap, we designed and evaluated PhysioML, a machine-learning education tool that leverages real-time physiological data (i.e., muscle activity). We explored learners' perceived understanding, performance, self-efficacy, and experience during a user study featuring 74 first- and second-year university students. In particular, we compared PhysioML with the popular Teachable Machine[1] tool. Our results suggest that PhysioML offers outcomes similar to those of Teachable Machine. However, learners expressed curiosity about machine learning more during the PhysioML condition.

## 2 Related Works

ML education research has continued to grow as AI/ML applications have become increasingly prominent in recent years. In particular, researchers have explored introducing ML through robotics [2, 20],

[1]https://teachablemachine.withgoogle.com/

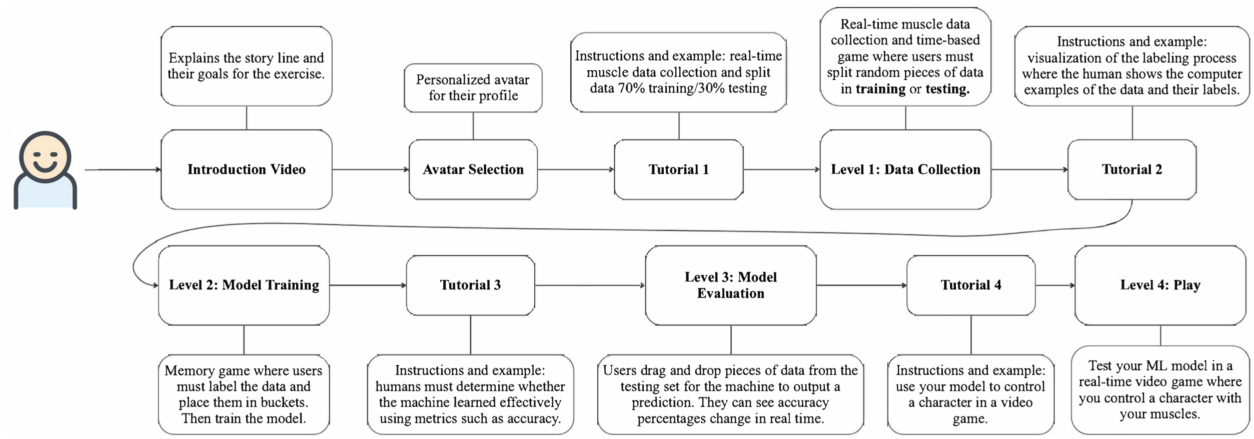



**Figure 1: Outline of activities during PhysioML condition.**

body motions [17, 47, 22], conversational agents [43, 21], image data [36, 44], text data [34], microcontrollers [45, 40], simulations [6], public datasets [28, 39], drawing [41], and gaming [35].

Several previous studies have expressed the importance of enabling students to sample personalized data to build datasets [18]. This is typically done through the collection of sound, image, or video data [8]. Recent rapid changes in our technological infrastructure have led to new opportunities to explore connections between learners' bodies and sensing technologies [23]. However, limited work investigates students using their physiological data during ML learning activities. Physiological sensing involves the measurement of bio-electrical signals from the body [33, 11]. These signals are often used to measure activity associated with the brain (Electroencephalogram - EEG), heart (Electrocardiogram - ECG/EKG), and muscles (Electromyogram - EMG) [12].

Over the past two decades, education-focused studies featuring physiological sensors have primarily used physiological data to evaluate students' learning experiences. For example, Yuan et al. investigated the feasibility of using EEG to evaluate students' reading comprehension [46]. Researchers have also presented work that focuses on using physiological sensors to measure learners' attention levels [25, 7], engagement [24], mental effort [19], anxiety [38, 13], cognitive load [3, 31], and programming [5].

Other researchers have proposed a different use case scenario that involves the construction of interactive applications that respond to changing physiological activity. This approach differs from the evaluation approach in that the physiological sensors are not intended to measure the learners' user state for evaluation purposes. Instead, the devices are used to drive interactive applications similar to peripheral devices such as mice and keyboards.

EEG sensors, which measure electrical activity in the brain, have been used previously to engage learners in STEM activities [16]. However, recent observations suggest that muscle sensors may be more intuitive for novice learners [27]. This paper seeks to contribute additional knowledge regarding how physiological sensors can be leveraged to introduce basic machine learning concepts. To our knowledge, this is the first study exploring physiological sensors for ML education. We argue that exploring novel physiological sensor types aligns with 2 core "Big Ideas" in AI [42]: (1) Computers perceive the world using sensors, and (2) Computers can learn from data. This paper contributes knowledge related to these ideas through a user study comparing educational activities leveraging image and physiological data.

## 3 PhysioML Tool Design

We developed Physio-ML to explore students' experiences learning ML concepts with real-time physiological data. PhysioML consists of 3 core components: **Real-Time Muscle Data**, **Data Processing**, and the **Web Application**. The following sections discuss each core component.

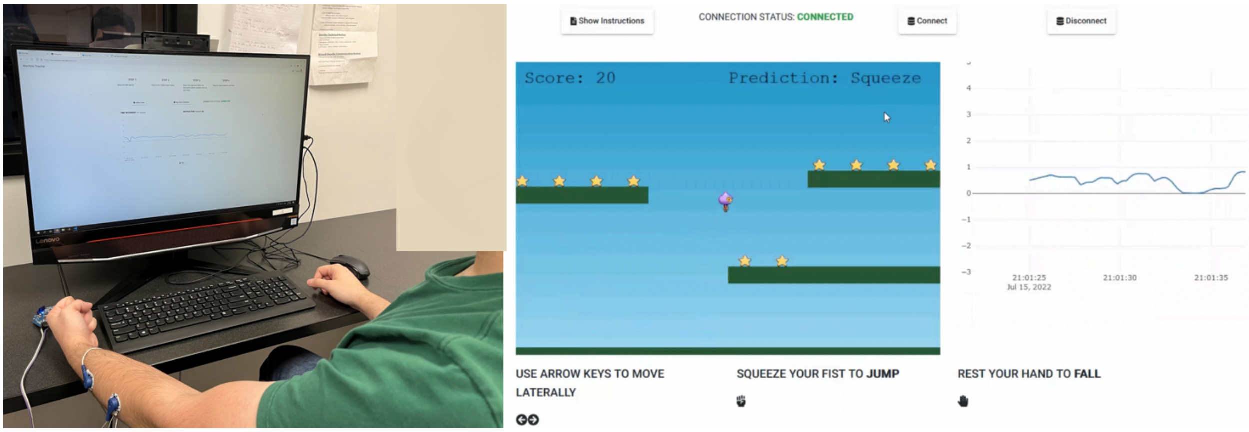
### 3.1 Real-Time Muscle Data (Electrical Activity)

PhysioML's significant feature is its support for real-time muscle data collection via OpenBCI's Ganglion device. The Ganglion is a low-cost, open-source biosensing wearable that allows students to measure and record electrical signals produced by their muscles in real time (e.g., Electromyography—EMG). It features 4 channels that sample data at 200Hz per channel. We leveraged Ganglion's support for Bluetooth connections to stream real-time muscle data to the computer.

### 3.2 Data Processing

Data extracted from the Ganglion device was captured by using the BrainFlow[2] library. PhysioML's underlying processing pipeline is informed by previous literature on muscle-computer interfaces [37]. Once captured, the data was rectified and filtered using the Scipy[3] library. To attenuate high-frequency noise and extract the low-frequency components relevant to muscle activity, a 4th-order Butterworth low-pass filter with a cutoff frequency of 3 Hz was applied to the acquired biosignal. Widely used Python libraries such as NumPy, Pandas, and Scikit-learn were used to handle model training and prediction. Linear Discriminant Analysis (LDA) was

[2]https://brainflow.org/
[3]https://scipy.org/




A) B)

**Figure 2: An illustration of a participant using PhysioML to (A) collect/label their muscle data and (B) Use the trained model that predicts muscle activations to control a game.**

leveraged as the default model during training tasks. Communication between the Python environment and web application were implemented using Flask[4] and SocketIO[5].

## 3.3 Web Application

PhysioML engages students through an interactive web application featuring a gamified muscle-computer interface. At its core, PhysioML is a small web-based game where students control an on-screen character using their muscles and keyboard input. By squeezing and resting their fist, students can cause the captured muscle activity data to change in real time. This feature aligns well with previous interactive machine learning approaches discussed in previous literature [48]. In particular, PhysioML steps students through the process of collecting and labeling muscle data into two classes: (1) **squeezing** and (2) **resting** fist. Afterward, the students use the PhysioML interface shown in Figure 3 to train and test a model that predicts the state of the student's muscle. During Level 4, predictions are mapped to the game character through **falling** and **jumping** actions. As shown in Figure 1, a total of 4 tutorials and challenge activities were featured in our user study. Students were required to complete each level before moving forward.

# 4 Methods

We conducted a user study to explore the students' experience while engaging with PhysioML and the Teachable Machine software. With this study, our team aims to address the following research questions: **RQ1:** To what extent do physiological signals affect students' **perceived understanding** of ML and **performance-based outcomes** during ML activities?, **RQ2:** To what extent does working with physiological sensors during ML activities impact students' **self-efficacy** compared to traditional methods (e.g., images)? and **RQ3:** How does using physiological sensors affect students' **user experience** during ML activities compared to traditional methods (e.g., images)?

[4]https://flask.palletsprojects.com/
[5]https://python-socketio.readthedocs.io/

## 4.1 Participants

The University of Alabama's Institutional Review Board (IRB) approved this study before recruitment. A total of 74 first- and second-year university students were recruited. Most were between 18 and 21 (78%), with age ranges of 22-25 and 31-35 as a far second (8% and 7% respectively). Around 62% of participants identified as male, 31% as female, 4% as non-binary, and 1% as gender-fluid. Student majors included Computer Science (62%), Computer Engineering (7%), Educational Psychology (7%), Nursing (4%), Mechanical Engineering (3%), Mathematics (3%) and more.

To understand the participants' backgrounds in programming and Machine Learning, they were asked about their skill levels and basic definitions for each category. Most identified as either novice (32%) or competent programmers (41%). However, 23% of participants reported no experience at all in programming, and only three participants self-reported as experts (4%). Around 25% of participants said they were not aware of what machine learning is, 16% were unsure, and 41% said they know what the technology means. Only 18% of participants reported a definite understanding of what machine learning entails.

## 4.2 Study Procedures

After an introductory explanation and ensuring their comfort, participants completed a pre-survey which provided non-identifiable demographic data. The pre-survey was followed by a between-subjects design with two conditions: **image data** and **physiological data (muscle activity)**. There was a total of 37 participants per condition. The key difference between each condition is the use of physiological sensors or image data. Teachable Machine, a state-of-the-art web-based ML educational tool developed by Google, was used to support educational ML activities in the image data condition. During the physiological data condition, students used our PhysioML software shown in Figure 2 and 3. PhysioML tasks were similar to tasks completed during the image data condition. Each session lasted approximately 60 minutes. Participants were guided through each web app to collect data, label data, and test a

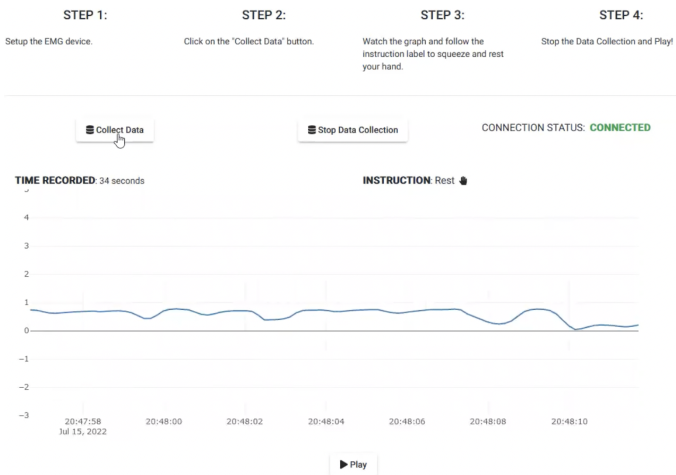



**Figure 3: User interface used to guide students through each level.**

model (using either images or physiological data). Screen recordings were captured to facilitate comprehensive data collection. The study concluded with a post-survey and a short semi-structured interview.

## 4.3 Data Collection

This study investigated students' understanding of machine learning concepts, self-efficacy, and user experience after each condition. The data collection methods used to aid our analysis are discussed in the following sections.

*4.3.1 Performance-Based Rubric & Perceived Understanding.* Each session was screen-recorded for post-analysis of students' understanding of machine learning concepts. We utilized a Performance-Based Rubric [14] to assess participants during each condition. Furthermore, we surveyed students' perceived understanding of concepts such as data collection/processing, model training, classification/inference, and evaluation.

*4.3.2 Self-Efficacy.* Self-efficacy scores were collected through the pre-post surveys, using a previously published questionnaire as a reference [4]. The modifications made the survey questions specific to machine learning education applications (ten-point Likert scale). The self-efficacy scores have a maximum of 100.

*4.3.3 Stress.* The study explored the stress-related feelings induced by each condition using the Short Stress State Questionnaire (SSSQ) [15], a modified version of the Dundee Stress State Questionnaire [30, 29]. The SSSQ comprises three main categories: distress, engagement, and worry. The survey was condensed from 24 to 11 items to prevent survey fatigue. The item selection process captured each of the core three components.

*4.3.4 Usability.* The study assessed usability in each condition using the System Usability Scale (SUS) [1], a reliable and valid standardized scale. Participants completed the SUS scale post-survey following the recommended format and scoring procedures.

*4.3.5 Interviews.* A short semi-structured interview session was held after participants had finished their assigned tasks and the post-survey. The goal was to identify detailed information about their experiences. The questions were kept broad so as not to introduce any bias. Participants were asked about their general thoughts on the system, what they learned, and what they thought about the data type. Through these interviews, participants provided in-depth insights into their self-efficacy, knowledge gains or barriers, and system usability.

## 4.4 Data Analysis

*4.4.1 Quantitative Data.* We used R to analyze the quantitative data. Paired t-tests were used to analyze pre-post self-efficacy survey responses. Data normality was evaluated through visual inspection and the Shapiro-Wilk Normality test. Skewness and kurtosis tests were also used to understand data distribution. Independent t-tests were used to analyze differences across conditions. In particular, we utilized post-survey responses to examine differences in performance, perceived understanding, self-efficacy, and stress across conditions.

Screen recordings were utilized during data collection, training, and classification lessons to analyze participants' performance. A rubric presented by Von Wangenheim et al. [14] was utilized to evaluate participants' performance. The rubric was used to score artifacts created by participants and was tailored for image classification tasks and physiological data tasks.

*4.4.2 Qualitative Data.* The process of analyzing the qualitative data began with the transcription of all audio recordings from the interviews. Once the interviews were converted into text format, qualitative analysis software (specifically, Atlas.ti) was used for coding. A content analysis approach was employed, drawing upon established methodologies [10, 26]. This involved thoroughly reviewing the dataset and assigning labels to recurring topics or themes. The coding process followed a deductive approach to minimize bias. The resulting codes were then abstracted into a list of concepts and categories, referred to as "themes," which captured participant experiences and trends. Additionally, notable quotes were identified and marked for later reference to highlight significant remarks.

# 5 Results

## 5.1 Survey Results

*5.1.1 Perceived Understanding.* We found no difference between the PhysioML and Teachable Machine conditions when evaluating students' perceived understanding. Students perceived understanding of data collection and cleaning ("The system helped me better understand how a model is trained for a machine learning system.") which were higher in the PhysioML (M=4.29, SD=0.90) in comparison to the Teachable Machine (M=3.91, SD=1.21) condition but not significantly ($p>0.05$). When analyzing students' perceived understanding of model training ("The system helped me better understand how a model is trained for a machine learning system."), we observed higher responses in the Teachable Machine (M=4.13, SD =1.00) condition than in the PhysioML (M=4.05, SD =1.10) condition. However, this difference was not significant ($p>0.05$). Perceived

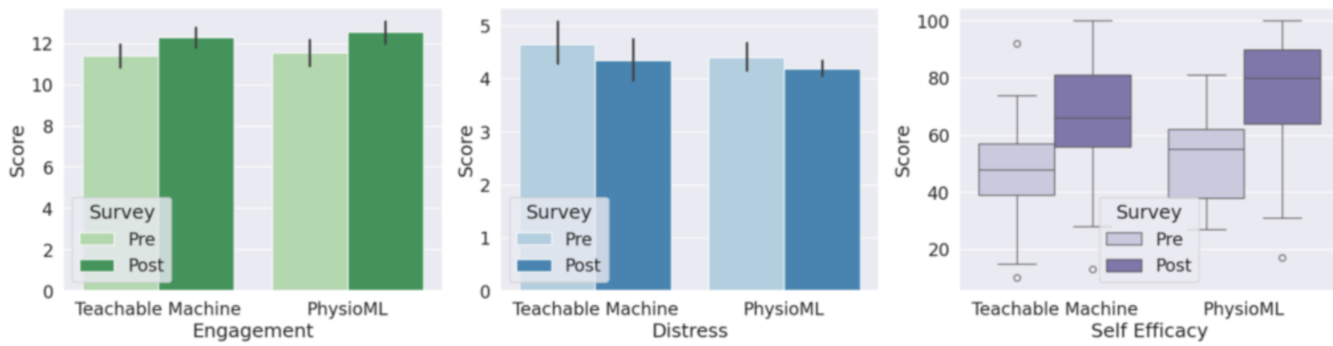



**Figure 4: Pre-Post analysis of participants' engagement, distress, and self-efficacy responses.**

understanding of classification ("The system helped me better understand classification in a machine learning system.") during the PhysioML (M=4.02, SD =0.95) and Teachable Machine (M=4.0, SD =1.15) condition was also similar (p>0.05). Although insignificant (p>0.05), we observed higher responses related to students' perceived understanding of model evaluation ("The system helped me better understand the evaluation process of a machine learning system.") during the Teachable Machine (M=4.21, SD=0.75) condition compared to PhysioML (M=3.94, SD=0.99). We observed higher responses to students' overall understanding of the ML process ("The system helped me better understand the creation process of a machine learning system.") during the PhysioML (M=4.16, SD=1.09) condition compared to Teachable Machine (M=3.91, SD=1.03). However, this difference was not significant (p>0.05).

*5.1.2 Self-Efficacy.* Pre-post surveys were conducted to assess participants' confidence during each condition. Participants initially reported low mean self-efficacy scores (PhysioML: 47.30, Teachable Machine: 52.41) but showed significant improvement after using both applications (PhysioML: 66.24, Teachable Machine: 74.19, **p < 0.001**) as shown in Figure 4. There was no significant difference (p>0.05) found between the post scores of PhysioML and Teachable Machine.

*5.1.3 Stress.* Results of the modified SSSQ were used to measure students' experience during both conditions. An independent t-test of post-survey responses did not show differences (p>0.05) across each condition in **engagement** (PhysioML: 12.5, Teachable Machine: 12.3) or **distress** (PhysioML: 4.19, Teachable Machine: 4.35). **Worry** scores were significantly lower during the PhysioML (PhysioML: 3.35, Teachable Machine: 4.11) condition. However, this may be explained by significantly lower worry scores captured during the PhysioML (M=3.03) and Teachable Machine (M=3.81) pre-surveys.

*5.1.4 Usability.* The post-survey results indicated that both systems exhibited excellent usability. Teachable Machine achieved an average scaled score of 84.60, while PhysioML obtained an average scaled score of 82.50. According to literature, usability scores above 68 are considered above average. Therefore, both systems received significantly higher ratings than the average usability score.

## 5.2 Assessment Results

We leveraged a performance-based assessment method proposed by Wangenheim et. al. [14] to evaluate screen recordings of each session. In the Teachable Machine condition, participants achieved an average score of 14.78 out of 22.0 (67.18%). For the PhysioML condition, the participants' average score was 14.27 (64.90%). An independent t-test of scores did not indicate a significant difference across conditions **(p>0.05)**. However, PhysioML resulted in higher average scores (p<0.05) for concepts related to *"distribution of the dataset"* (PhysioML: 1.51, Teachable Machine: 1.14), *"labeling"* (PhysioML: 1.69, Teachable Machine: 1.33), and *"training"* (PhysioML: 1.6, Teachable Machine: 1.03). Teachable Machine resulted in higher average scores **(p<0.05)** linked to *"Test with new objects"* (PhysioML: 0.00, Teachable Machine: 0.58) and *"Data Cleaning"* (PhysioML: 0.94, Teachable Machine: 1.75).

## 5.3 Themes from Post-Experience Interviews

Codes were developed using the process discussed earlier to represent the themes that participants repeatedly brought up during interviews. The following results focus on the final common themes based on these codes. The most mentioned theme was **"ease of use"** with 52 occurrences. This includes any mention of how simple, fun, or enjoyable their experience was. Another category comprises their **"curiosity"** about the physiological signals with 30 occurrences, which includes mentions of the visualizations and the usefulness of using human data in real-time. One participant stated: *"I was curious on how that works (the physiological device), like how it collects the data..."*

The experiences appeared to stimulate participants' critical thinking regarding the data and its potential influence on the performance of ML models. For instance, participants contemplated strategies to enhance the accuracy of their artifacts. Regarding PhysioML, users engaged in discussions about the impact of body differences, filters, data volume, and arm positions. The third most prevalent theme, occurring in 25 instances, was the **"utilization of data for making adjustments or improvements"** in the final results of their application.

The fourth prevalent theme (24 occurrences) involved participants expressing their **"acquisition of knowledge"** regarding the overall process of creating a complete ML application. They discussed the three main stages and occasionally provided specific details about each stage. During an interview one participant stated: *"I thought that was so much fun, and it's a great opportunity to learn more about machine learning and kind of like, really how it works and the process behind it."*

The fifth prevailing theme (24 occurrences) revolved around participants expressing their **"enjoyment"** of the gamified nature of PhysioML. They appreciated the structured and engaging aspects of the application, as well as the helpful visualizations that facilitated the connection between their data and the interactive elements. Participants also found the game to be beneficial in enhancing their understanding of the modeling processes, particularly feature extraction. One participant stated: *"I really like the process of actually getting to see how the model is working. When you get to play the games, you just see the data collection cards and training, whatever, but then you got to see it in action, and you got to be involved with it, and I thought that was really cool."*

Participants expressed a desire for **"more technical details"** regarding the ML systems (21 occurrences). They found their assigned system to be a good introduction to ML but expressed increased curiosity about the underlying backend processes. They specifically mentioned their interest in specific algorithms and visualizations for the training sections.

## 6 Discussion

Observations from this study suggest that physiological sensors, such as EMG, can provide positive hands-on experiences during educational activities featuring machine learning. Students' perceived understanding of machine learning and performance was not significantly different between the physiological sensor and image conditions (**RQ1**). While image data is commonly used for ML education, this finding suggests that physiological sensors could provide similar support for students' perceived understanding of ML. Findings related to students' perceived understanding were also aligned with self-efficacy responses (**RQ2**). In particular, pre-post measures of self-efficacy during the PhysioML and Teachable machine condition also significantly increase. These results support the view that physiological sensors could promote positive attitudes towards ML. Additionally, both PhysioML and Teachable Machine received higher-than-average usability scores (**RQ3**).

Overall, students had similar performances across the PhysioML and Teachable Machine conditions. However, users in both conditions performed very poorly on *"Testing with new objects"* and *"Adjustments/improvements"*. As an independent task, Teachable Machine did not motivate students to continue trying new classes for improved accuracy. PhysioML had two issues: testing with new objects was not included as a feature, and the leveling system might have proven to be a bit long for those looking to improve an ML model. Though students regularly tried the evaluation mini-game (level four), they did not start the experience again due to the required time. Future work might explore improving these motivational factors to allow students to improve their models since this skill is necessary for ML and physiological computing. Students also struggled with *data cleaning* with the PhysioML condition. This is likely due to concepts such as digital signal processing (DSP) being new to students. Additional research on ways to integrate DSP and ML in educational settings is needed to address this challenge in the future.

## 7 Limitations

This study was designed to compare the performance of groups assigned to two conditions. The approach discussed in this paper limited the results to provide an overall comparison between participants in their respective conditions regarding perceived understanding, performance, self-efficacy, stress, and usability. Although the study's goal by design was to gather an overview of the participants' performances, we recognize the possible threats to validity that could arise from this methodology.

PhysioML allows novices to begin working with physiological-based ML applications. Participants showed positive learning outcomes; however, most participants wanted to delve further into the topic. Interviews with participants revealed that they were intrigued by the technologies and would have liked to learn more about the model training aspects (21 occurrences). PhysioML is currently tailored to those with less background knowledge about ML and physiological computing. As a result, most of the lower-level technical components are abstracted. Another finding was the oversimplification of the entire process, which limited users' ability to experiment freely. The structure of PhysioML helps novices but simultaneously constrains more experienced students. Therefore, the lack of complexity and strict structure can be addressed to improve the system. Future iterations could explore the additions of detailed training algorithms, provide options for the user to tune parameters, and manually manage the labels. Overall, PhysioML provided learning support similar to traditional image-based ML education.

## 8 Conclusion

This paper discusses a study exploring the use of physiological sensors (i.e., EMG - muscle activity) during a machine learning education activity. In particular, the physiological sensor approach was compared to students' experiences using image data to learn concepts relevant to machine learning. Results from this study suggest that physiological sensors may support ML education in a manner similar to image-based approaches. Furthermore, integrating physiological sensors during ML education activities may encourage students to be more curious about machine learning.

## Acknowledgments

This work was supported in part by a National Science Foundation (NSF) grant (#2045561).

## References

[1] John Brooke. 1996. Sus: a "quick and dirty" usability. *Usability evaluation in industry*, 189, 3.
[2] Harald Burgsteiner, Martin Kandlhofer, and Gerald Steinbauer. 2016. Irobot: teaching the basics of artificial intelligence in high schools. In *Proceedings of the AAAI conference on artificial intelligence* number 1. Vol. 30.
[3] Chih-Ming Chen and Chung-Hsin Wu. 2015. Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, 80, 108–121.
[4] Deborah R Compeau and Christopher A Higgins. 1995. Computer self-efficacy: development of a measure and initial test. *MIS quarterly*, 189–211.

[5] Igor Crk, Timothy Kluthe, and Andreas Stefik. 2015. Understanding programming expertise: an empirical study of phasic brain wave changes. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23, 1, 1–29.

[6] Adrian A de Freitas and Troy B Weingart. 2021. I'm going to learn what?!? teaching artificial intelligence to freshmen in an introductory computer science course. In *Proceedings of the 52nd ACM technical symposium on computer science education*, 198–204.

[7] Spyridon Doukakis, Mary Angela Papalaskari, Panayiotis Vlamos, Antonia Plerou, and Panagiota Giannopoulou. 2018. Assessing attention in visual and textual programming using neuroeducation approaches. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, 392–392.

[8] Stefania Druga, Nancy Otero, and Amy J Ko. 2022. The landscape of teaching resources for ai education. In *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1*, 96–102.

[9] Michael Eisenberg, Sherry Hsi, and HyunJoo Oh. 2017. Machines and minds: the new cognitive science, and the potential evolution of children's intuitions about thinking. *International journal of child-computer interaction*, 14, 1–4.

[10] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *Journal of advanced nursing*, 62, 1, 107–115.

[11] Stephen H Fairclough. 2009. Fundamentals of physiological computing. *Interacting with computers*, 21, 1-2, 133–145.

[12] Stephen H Fairclough. 2010. Physiological computing: interfacing with the human nervous system. In *Sensing emotions*. Springer, 1–20.

[13] Jamie Gorson, Kathryn Cunningham, Marcelo Worsley, and Eleanor O'Rourke. 2022. Using electrodermal activity measurements to understand student emotions while programming. In *Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1*, 105–119.

[14] Christiane Gresse Von Wangenheim, Nathalia da Cruz ALVES, Marcelo F Rauber, Jean CR Hauck, and Ibrahim H Yeter. 2022. A proposal for performance-based assessment of the learning of machine learning concepts and practices in k-12. *Informatics in Education*, 21, 3, 479–500.

[15] William S Helton. 2004. Validation of a short stress state questionnaire. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* number 11. Vol. 48. Sage Publications Sage CA: Los Angeles, CA, 1238–1242.

[16] Bryan Hernandez-Cuevas, William Egbert, Andre Denham, Ajay Mehul, and Chris S Crawford. 2020. Changing minds: exploring brain-computer interface experiences with high school students. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–10.

[17] Tom Hitron, Yoav Orlev, Iddo Wald, Ariel Shamir, Hadas Erel, and Oren Zuckerman. 2019. Can children understand machine learning concepts? the effect of uncovering black boxes. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, 1–11.

[18] Tom Hitron, Iddo Wald, Hadas Erel, and Oren Zuckerman. 2018. Introducing children to machine learning concepts through hands-on experience. In *Proceedings of the 17th ACM conference on interaction design and children*, 563–568.

[19] Yerika Jimenez. 2017. Towards understanding student's mental effort in block-based programming environments using electroencephalogram (eeg). In *Proceedings of the 2017 ACM Conference on International Computing Education Research*, 285–286.

[20] Martin Kandlhofer, Gerald Steinbauer, Julia P Laßnig, Wilfried Baumann, Sandra Plomer, Aron Ballagi, and Istvan Alfoldi. 2019. Enabling the creation of intelligent things: bringing artificial intelligence and robotics to schools. In *2019 IEEE frontiers in education conference (FIE)*. IEEE, 1–5.

[21] Gloria Ashiya Katuka et al. 2023. A summer camp experience to engage middle school learners in ai through conversational app development. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, 813–819.

[22] Vishesh Kumar and Marcelo Worsley. 2023. Scratch for sports: athletic drills as a platform for experiencing, understanding, and developing ai-driven apps. In *Proceedings of the AAAI Conference on Artificial Intelligence* number 13. Vol. 37, 16011–16016.

[23] Victor Lee. 2014. *Learning technologies and the body*. Taylor & Francis.

[24] Victor R Lee. 2023. Electrodermal activity wearables and wearable cameras as unobtrusive observation devices in makerspaces. In *Unobtrusive Observations of Learning in Digital Environments: Examining Behavior, Cognition, Emotion, Metacognition and Social Processes Using Learning Analytics*. Springer, 217–235.

[25] Yong-Teng Lin and Chih-Ming Chen. 2019. Improving effectiveness of learners' review of video lectures by using an attention-based video lecture review mechanism based on brainwave signals. *Interactive Learning Environments*, 27, 1, 86–102.

[26] Britt-Marie Lindgren, Berit Lundman, and Ulla H Graneheim. 2020. Abstraction and interpretation during the qualitative content analysis process. *International journal of nursing studies*, 108, 103632.

[27] Feiya Luo, Ruohan Liu, Idowu David Awoyemi, Chris Crawford, and Fatema Nasrin. 2024. Novel insights into elementary girls' experiences in physiological computing. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 764–770.

[28] Ruizhe Ma, Ismaila Temitayo Sanusi, Vaishali Mahipal, Joseph E Gonzales, and Fred G Martin. 2023. Developing machine learning algorithm literacy with novel plugged and unplugged approaches. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, 298–304.

[29] Gerald Matthews, Sian E Campbell, Shona Falconer, Lucy A Joyner, Jane Huggins, Kirby Gilliland, Rebecca Grier, and Joel S Warm. 2002. Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry. *Emotion*, 2, 4, 315.

[30] Gerald Matthews, Lucy Joyner, Kirby Gilliland, Sian Campbell, Shona Falconer, and Jane Huggins. 1999. Validation of a comprehensive stress state questionnaire: towards a state big three. *Personality psychology in Europe*, 7, 335–350.

[31] Caitlin Mills, Igor Fridman, Walid Soussou, Disha Waghray, Andrew M Olney, and Sidney K D'Mello. 2017. Put your thinking cap on: detecting cognitive load using eeg during learning. In *Proceedings of the seventh international learning analytics & knowledge conference*, 80–89.

[32] Seymour A Papert. 2020. *Mindstorms: Children, computers, and powerful ideas*. Basic books.

[33] Rosalind W Picard. 2000. *Affective computing*. MIT press.

[34] Patricio Quiroz and Francisco J Gutierrez. 2024. Scratch-nb: a scratch extension for introducing k-12 learners to supervised machine learning. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 1077–1083.

[35] Prapa Rattadilok, Chris Roadknight, and Li Li. 2018. Teaching students about machine learning through a gamified approach. In *2018 IEEE international conference on teaching, assessment, and learning for engineering (TALE)*. IEEE, 1011–1015.

[36] Alpay Sabuncuoglu. 2020. Designing one year curriculum to teach artificial intelligence for middle school. In *Proceedings of the 2020 ACM conference on innovation and technology in computer science education*, 96–102.

[37] T Scott Saponas, Desney S Tan, Dan Morris, and Ravin Balakrishnan. 2008. Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 515–524.

[38] Natalia Silvis-Cividjian, Joshua Kenyon, Elina Nazarian, Stijn Sluis, and Martin Gevonden. 2024. On using physiological sensors and ai to monitor emotions in a bug-hunting game. In *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1*, 429–435.

[39] James Skripchuk, Yang Shi, and Thomas Price. 2022. Identifying common errors in open-ended machine learning projects. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education-Volume 1*, 216–222.

[40] Hussel Suriyaarachchi, Paul Denny, and Suranga Nanayakkara. 2022. Scratch and sense: using real-time sensor data to motivate students learning scratch. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education-Volume 1*, 983–989.

[41] Tapani Toivonen, Ilkka Jormanainen, Matti Tedre, Radu Mariescu-Istodor, Teemu Valtonen, Henriikka Vartiainen, and Juho Kahila. 2022. Interacting by drawing: introducing machine learning ideas to children at a k–9 science fair. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–5.

[42] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning ai for k-12: what should every child know about ai? In *Proceedings of the AAAI conference on artificial intelligence* number 01. Vol. 33, 9795–9799.

[43] Jessica Van Brummelen, Tommy Heng, and Viktoriya Tabunshchyk. 2021. Teaching tech to talk: k-12 conversational artificial intelligence literacy curriculum and development tools. In *Proceedings of the AAAI Conference on Artificial Intelligence* number 17. Vol. 35, 15655–15663.

[44] Henriikka Vartiainen, Matti Tedre, and Teemu Valtonen. 2020. Learning machine learning with very young children: who is teaching whom? *International journal of child-computer interaction*, 25, 100182.

[45] Randi Williams, Michał Moskal, and Peli De Halleux. 2022. Ml blocks: a block-based, graphical user interface for creating tinyml models. In *2022 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 1–5.

[46] Yueran Yuan, Kai-min Chang, Jessica Nelson Taylor, and Jack Mostow. 2014. Toward unobtrusive measurement of reading comprehension using low-cost eeg. In *Proceedings of the Fourth International Conference on learning analytics and knowledge*. ACM, 54–58.

[47] Abigail Zimmermann-Niefield, Shawn Polson, Celeste Moreno, and R Benjamin Shapiro. 2020. Youth making machine learning models for gesture-controlled interactive media. In *Proceedings of the interaction design and children conference*, 63–74.

[48] Abigail Zimmermann-Niefield, Makenna Turner, Bridget Murphy, Shaun K Kane, and R Benjamin Shapiro. 2019. Youth learning machine learning through building models of athletic moves. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 121–132.