

# Privacy-Preserving and Scalable Affect Detection in Online Synchronous Learning

## Citation for published version (APA):

Böttger, F., Cetinkaya, U., Mitri, D. D., Gombert, S., Shingjergji, K., Iren, D., & Klemke, R. (2022). Privacy-Preserving and Scalable Affect Detection in Online Synchronous Learning. In I. Hiliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz, & T. Farrell (Eds.), *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption* (1 ed., pp. 45-58). Springer, Cham. Springer Lecture Notes in Computer Science (LNCS) Vol. 13450 [https://doi.org/10.1007/978-3-031-16290-9\\_4](https://doi.org/10.1007/978-3-031-16290-9_4)

## DOI:

[10.1007/978-3-031-16290-9\\_4](https://doi.org/10.1007/978-3-031-16290-9_4)

## Document status and date:

Published: 05/09/2022

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

## Take down policy

If you believe that this document breaches copyright please contact us at:

[pure-support@ou.nl](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 19 Nov. 2022

Open Universiteit  
[www.ou.nl](http://www.ou.nl)



# Privacy-preserving and scalable affect detection in online synchronous learning

Felix Böttger<sup>1</sup>, Ufuk Cetinkaya<sup>2</sup>, Daniele Di Mitri<sup>2</sup>,  
Sebastian Gombert<sup>2</sup>, Krist Shingjergji<sup>1</sup>, Deniz Iren<sup>1</sup>, Roland Klemke<sup>1</sup>

<sup>1</sup>Open University of The Netherlands, Heerlen, The Netherlands  
mail@felixboettger.de,

{krist.shingjergji, deniz.iren, roland.klemke}@ou.nl

<sup>2</sup>DIPF - Leibniz Institute for Research and Information in Education,  
Frankfurt, Germany

{cetinkaya, dimitri, gombert}@dipf.de

**Abstract.** The recent pandemic has forced most educational institutions to shift to distance learning. Teachers can perceive various non-verbal cues in face-to-face classrooms and thus notice when students are distracted, confused, or tired. However, the students' non-verbal cues are not observable in online classrooms. The lack of these cues poses a challenge for the teachers and hinders them in giving adequate, timely feedback in online educational settings. This can lead to learners not receiving proper guidance and may cause them to be demotivated. This paper proposes a pragmatic approach to detecting student affect in online synchronized learning classrooms. Our approach consists of a method and a privacy-preserving prototype that only collects data that is absolutely necessary to compute action units and is highly scalable by design to run on multiple devices without specialized hardware. We evaluated our prototype using a benchmark for the system performance. Our results confirm the feasibility and the applicability of the proposed approach.

**Keywords:** affect detection, action units, emotion recognition, privacy

## 1 Introduction

COVID-19 pandemic forced more than 1.6 billion learners out of school [31], becoming the most challenging disruption ever endured by the global education systems. In many countries, educational institutions were forced to move their regular activities online, relying on remote teaching to continue their education [16]. While the modality of education provision changed from physical to online presence the teaching methods in use remained essentially the same. For example, teachers often favored online synchronous classrooms (i.e., video conferencing tools) over asynchronous activities, discussion forums, or group work.

Physical distancing and learning in isolation posed severe challenges for learners worldwide by hindering their study success [24]. In this context, making education systems more resilient and less vulnerable to future disruptions became

a compelling need. In particular, we have to reconsider how digital technologies can support and better facilitate online and hybrid teaching. Digital education technologies such as *video conferencing tools* and *learning management systems* have made education more accessible and flexible. However, the modes of interaction respective systems implement remain unnatural for teachers and learners as it requires them to sit behind a computer screen for long hours. Furthermore, also communication in an online classroom has limitations. Teachers can perceive the students' affective states in a face-to-face classroom and notice when they are distracted, confused, or tired. This ability is somewhat hindered in online classrooms due to several limitations of the communication tools. For instance, video conferencing tools only show a limited number of participants on screen. Their images are displayed in small portions of the screen, leaving no space for showing body language. Thus, teachers using video conferencing tools cannot observe the non-verbal cues exhibited by the students. In addition, human communication is multimodal by nature [18], and students and teachers need to use a wide array of modes that go beyond the audio-visual support of the webcams and microphones to interact with each other. Such peripheral devices fall short in capturing and conveying non-verbal aspects of human communication such as body posture, facial expressions, prosody and intonation, and physical proximity. This poses a tremendous challenge for both teachers and learners and hinders the teachers' ability to give the classroom timely feedback. Thus, it potentially leads to learners lacking guidance and motivation.

In the last decade, the technological leaps in artificial intelligence have paved the way for novel human-computer interaction methods. State-of-the-art affective computing technologies can automatically recognize non-verbal cues such as gestures and body posture [15], facial expressions [20], and speech intonation [3]. Such technologies can alleviate the challenges of online education by analyzing and aggregating many signals from the microphones and webcams of learners, narrowing the communication modality gap between video conferencing and face-to-face communication. Teachers who are equipped with such information can alter their teaching strategy when needed, such as taking a break or changing the course of the learning activities. Moreover, they can adapt their teaching styles and course structures based on data.

Despite apparent benefits, affective computing systems are not without any risks. Debatably, the most critical threat is the invasion of learners' privacy[6]. Therefore, it is imperative to design such systems in a way that ensures the protection of the same [9]. The designs must adhere to privacy and data protection regulations and must employ privacy-by-design principles [23]. These principles include practices such as purposeful data collection (e.g., collecting and sharing only the data relevant to the teacher), clearly informing the subjects of the method, asking for consent, and using anonymization and aggregation to avoid tracing the data back to individuals.

To address these challenges, we seek to answer the following research question.

*How can we enable teachers to sense the affective states of the classroom in online synchronized learning environments in a privacy-preserving way?*

This paper addresses these challenges by proposing a pragmatic approach to detecting student affect in online synchronized learning classrooms in a privacy-preserving and highly scalable manner. We present *Sense the Classroom - Live (STC-Live)*, a research prototype that addresses these challenges and can run on many different end-user platforms, thus not requiring costly specialized equipment. Moreover, we evaluate the prototype’s performance.

The remaining of this paper is structured as follows. First, Section 2 presents the background information on emotions, emotion recognition, and privacy-preserving design in the context of learning. Then, in section 3, we describe the details of STC-Live and the evaluation procedure. Next, section 4 presents the results of the system evaluation. Finally, in section 5, we discuss the results, reflect on them, and conclude our paper.

## 2 Background

### 2.1 Emotions and emotion recognition in learning

Emotions are complex reaction patterns involving experiential, behavioral, and physiological elements by which humans attempt to cope with a matter or event [1]. Ekman defined a set of ‘basic emotions’ [11] as anger, disgust, sadness, happiness, fear, surprise, and neutrality. The primary emotions are universal in how they are expressed and perceived. More complex emotions are nuances or combinations of the basic emotions. A similar term, affective state, refers to longer-lasting emotions and moods. Several studies exist that define affective states in the context of educational sciences [27]. Some of the affective states relevant to educational sciences are engagement, concentration, boredom, anxiety, confusion, frustration, and happiness [8]. Students’ emotional states affect their learning experience by influencing their motivation to learn, engagement, and self-regulation [25]. Many studies report pieces of evidence of a relationship between emotional states and learning experience. For example, it is shown that enjoyment and pride positively predicted academic achievement, while the opposite holds for emotions like anger, anxiety, shame, boredom, and hopelessness [28]. The affective states can be perceived by observing nonverbal cues, e.g., gestures, body posture, micro-expressions, and activities such as not actively listening or looking away. Therefore, in recent years, affective computing in education has received widespread attention from researchers [32].





















There are many methods and tools to measure emotions in online learning environments [17] that can be categorized into three different areas: psychological, physiological, and behavioral [13]. The psychological measurement methods are based on the self-reporting of emotions, e.g., questionnaires such as the Academic Emotions Questionnaire (AEQ) by Pekrun et al. [26], and self-report systems such as *emot-control* [14]. The physiological measurement methods use

sensors to collect signals from the skin, heart, etc. This method requires specific instruments and sensors, making it challenging to use in an online setting [17]. Lastly, the behavioral measurement tools use behavioral expressions to measure emotions in, for example, natural language [10] and facial expressions. Examples in the literature include a system that detects boredom and lack of interest using eye and head movement [19] and a method that uses eyeball movement and head gestures observed from the real-time feed of the students’ web cameras to estimate the corresponding concentration levels[30].

## 2.2 Facial expressions and action units

Facial expression is one of the most effective channels humans use to communicate their emotions [20]. Many studies have documented that basic human emotions are expressed and recognized universally across cultures [21]. Emotions are expressed in the face by combining multiple muscle movements and contractions, i.e., action units (AU). Researchers have developed systematic approaches to categorize and decode action units [12], and such practices have formed a solid basis for automated facial emotion recognition [20].

Table 1: The 20 AUs as classified by the AU detection step of STC-Live

AU1  Inner Brow Raiser	AU2  Outer Brow Raiser	AU4  Brow Lowerer	AU5  Upper Lid Raiser	AU6  Cheek Raiser
AU7  Lid Tightener	AU9  Nose Wrinkler	AU10  Upper Lip Raiser	AU11  Nasolabial Deepener	AU12  Lip Corner Puller
AU14  Dimpler	AU15  Lip Corner Depressor	AU17  Chin Raiser	AU20  Lip Stretcher	AU23  Lip Tightener
AU24  Lip Pressor	AU25  Lips Part	AU26  Jaw Drop	AU28  Lip Suck	AU43  Eyes Closed

## 2.3 Privacy in learner emotion detection

Scheffel et al. [29] identified data privacy as the most critical factor for users’ trust in systems processing learner data. According to Drachslar & Greller [9],

“there are hesitations regarding, among other things, [...] violation of personal privacy rights; [...] intransparency of the learning analytics systems; [...] the impossibility to fully anonymize data; safeguard access to data; and, the reuse of data for non-intended purposes.” For this reason, they conclude, among other aspects, that learner data needs to be “anonymize[ed] as far as possible”.

Research on achieving privacy for the specific use case of emotion detection is sparse. Past publications mainly focused on achieving privacy at the machine learning stage by minimizing the possibility of extracting sensitive information from neural networks while maximizing their ability to recognize human emotions [22]. The vector representations produced by these networks are aimed to be sent over the network for downstream classification.

It is debatable what exact types of vector representations are appropriate for preserving privacy in online learner emotion detection, as many representations allow at least for linking attacks. Nonetheless, acquiring vector representations which contain only the data which is absolutely necessary for detecting affect on the client-side and then transferring these to a server for downstream classification reduces sensitivity of the stored data by a large degree. This contributes to preserving the privacy of the classified individuals.

### 3 Method

In this study, we designed and developed a software prototype that detects the students’ affective states in online synchronized learning environments. This section details the proposed system architecture, the collection, storage, and processing of the data, including the action unit detection method based on machine learning. Finally, we report the evaluation of the proposed system.

#### 3.1 System architecture

STC-Live is a web-based affective learning analytics platform. It uses machine learning models embedded inside the web browser to extract data from the user’s webcam without transmitting or storing any video data. Only the outcomes of the machine learning process (i.e., numerical representations of the facial expressions) are transferred to the server, stored inside a database, and displayed to the teacher in an aggregated manner. Additionally, the platform offers a dashboard that visualizes the collected data in real-time. As an open-source project, it can be used as a starting point for similar study designs and adapted for specific requirements.

#### 3.2 System overview

The system comprises three main components: a) the student-side component that runs on student computers for data collection, b) the server back-end component that receives, stores, and forwards the data to the teacher, and c)

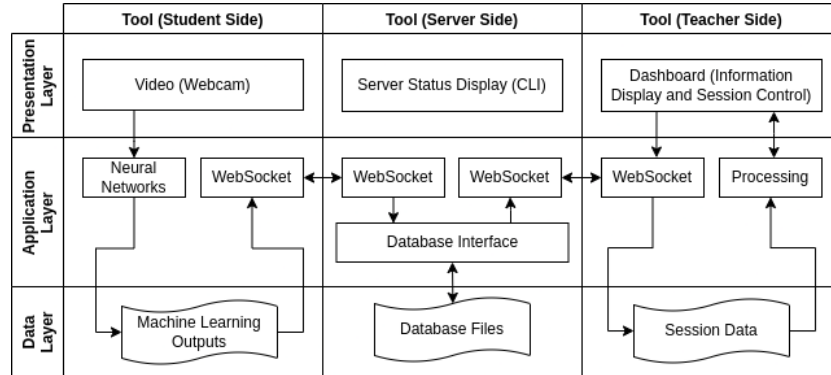


Fig. 1: The system architecture of STC-Live.

the teacher-side component that allows session handling and access to the session data (see fig. 1). The teacher- and student-side components are accessible through a website hosted on the server back-end. This approach ensures multi-platform compatibility without the need to develop and maintain separate code bases for different platforms. From the user’s perspective, web-based programs are also more trusted than their native counterparts, as browsers limit the capabilities of web-based programs (e.g., restricted file access, asking users to allow camera/microphone access).

### 3.3 Student-side component: data collection

The student-side component is a JavaScript program that runs inside the web browser. It periodically takes an image from the webcam’s video feed, which is then used as input for the machine learning pipeline. The machine learning pipeline transforms the images into numerical values representing the facial action units. Consecutively, the numerical values are converted into JSON (Javascript Object Notation) format that contains the following information for each time interval; the prominent emotion detected, timestamp, a list of the spatial coordinates of the 68 facial landmarks, and a list of 5408 Histogram of Oriented Gradient (HOG) values. This JSON object is sent to the server-side tool (back-end) via a WebSocket connection. The images themselves are neither stored nor transferred, therefore avoiding any risk of a privacy breach. The frequency of data collection is configured on the server-side, taking into account the time required to generate a JSON data point. Our recommendation to ensure reliable data collection is for the worst-performing student computer to be used as a baseline for this interval. We evaluate the performance of the data collection tool on different sets of hardware (Section 3.6).

**Machine learning pipeline:** The student-side component incorporates a machine learning pipeline (see fig. 2) that consists of three different neural networks

provided by FaceAPI, a commonly used computer vision library for face detection and emotion recognition. Specifically, the pipeline comprises the steps of i) face detection, ii) landmark identification and facial emotion recognition, and iii) AU classification. The first two steps use the following models provided by the FaceAPI; *ssdMobilenetv1*, *faceLandmark68Net*, and *faceExpressionNet*, and the third step uses the *Py-Feat AU classification* model [4].

The face detection step uses *ssdMobilenetv1*, which was trained on the WIDER-FACE - dataset [33], and is used to detect the faces on the given image. The model calculates the location of every face and returns a bounding box for each face and a confidence probability associated with the bounding box.

The landmark identification and facial emotion recognition step use *faceLandmark68Net* and *faceExpressionNet* simultaneously. The *faceLandmark68Net* is a lightweight landmark detection network that identifies the location of prominent facial features, i.e., landmarks. It has been trained on approximately 35.000 face images, and it recognizes 68 unique facial landmarks on a given image of a face. In contrast, the *faceExpressionNet* constitutes a Convolutional Neural Network (CNN) that takes an image as an input and returns the predicted emotion.

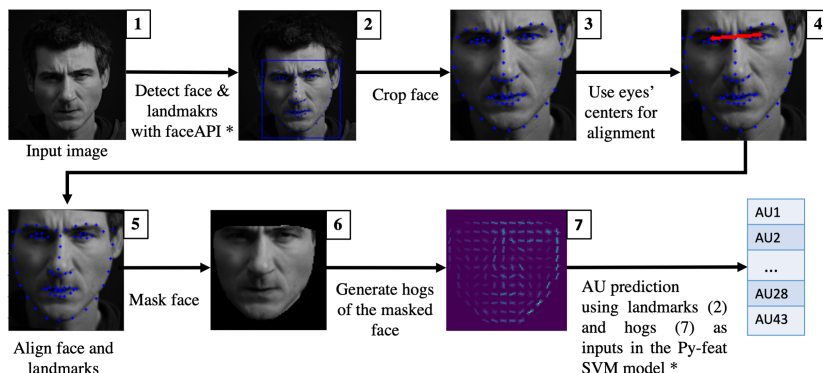


Fig. 2: The pipeline of AU detection

The steps with an asterisk (\*) are non-deterministic methods of machine learning algorithms with different performance accuracy measures.

The AU classification step uses the pre-trained Support Vector Machine (SVM) model provided by the Py-Feat [4]. The model receives two vectors as input: the facial landmarks, a  $(68 \times 2)$  vector of the landmark locations, and the HOGs, a vector of  $(5408 \times 1)$  features that describe an image as a distribution of orientations [7]. The model's output is a list of the AUs classified as present among the 20 possible AUs (Table ??). Pre-processing the image is required for alignment with the input format used for training the classifier [2]. The pre-processing, in summary, consists of the following steps: cropping, resizing, alignment, and masking. In the initial stages, the detected face is cropped



from the initial image and resized <sup>1</sup>. Respectively, the detected landmarks are projected in the new image. In the following steps, the cropped face is aligned using the positions of the two eyes and rotating the image so that the line that connects them is horizontal. Similarly, as in the previous step, the detected landmarks are rotated respectively. Lastly, the face is masked using the positions of the landmarks. The vector of the HOG values of the pre-processed image is calculated using eight orientations,  $8 \times 8$  pixels per cell and  $2 \times 2$  pixels per block. STC-Live saves the vectors of the HOG values and the landmarks of the pre-processed image, but not the camera image itself, reducing the amount of possibly sensitive data. The data can be used as inputs to the SVM classifier to detect the AUs.

### 3.4 Server back-end component: data storage and transfer

The server-side tool is a Node.js program that functions as a back-end for the distributed system. It receives periodic updates from the student-side component and stores the contained data in a MongoDB database. The current status of all participants of an individual session is bundled and periodically sent to the teacher-side component for visualization. The back-end can run multiple sessions simultaneously, making it possible to have a shared instance. When a new session is created using the web interface, the back-end creates a 12-digit session key, which the students use to enter a session. Access to the session data is only granted to the creator of the session, i.e., the host. The server can be configured to either automatically delete all session data shortly after a session has ended or keep the data in the database for the after-the-fact review. The resource-intensive computation through neural networks is done solely on the students' machines, so the system is highly scalable. It can handle several hundreds of participants in multiple sessions, even on weaker server hardware.

### 3.5 Teacher-side component: session management

The teacher-side tool is a JavaScript program that runs inside the teacher's web browser. It connects to the backend via a WebSocket connection used to control the session and receive periodic updates from the back-end. Users can create sessions through a web interface. The session host is granted access to a web-based dashboard that contains real-time information about the current participants' states, such as the detected affective states, as well as the control elements to invite new participants, download all corresponding data, or close the session. Sessions without active participants are automatically closed after a configurable delay.

### 3.6 System evaluation

To evaluate the actual performance of our prototype, we created a benchmark scenario that uses the same machine-learning pipeline to extract data from the

<sup>1</sup> The size used is  $112 \times 112$

webcam video feed but does not transmit the extracted data to the server. We decided this to ensure that the performance measurement is accurate and not influenced by the stability or speed of the connection to the server. As the machine learning process is by far the most resource-demanding task for the prototype, the results should indicate the overall system performance. The benchmark scenario consisted of 1000 executions of the pipeline, with a new image being passed to the pipeline every second. We measured the time it takes to process the facial data, emotion, and landmark recognition and generate the HOG features, but not the AU detection from these data points as the latter is performed on the server-side. We recorded a video clip of a face moving around to create challenging - but not impossible - situations for face detection. We then tested the actual performance of the system using this pre-recorded video clip<sup>2</sup> on different computers with varying hardware, operating systems, and browsers. We tested the platform on all hardware configurations that were available to us. We have shown that it's feasible to run our platform on lower-end hardware with a status interval of one second, the status interval can be shorter on higher-end hardware.

## 4 Results

While a correlation between the response time and the systems clock rate and memory size can be shown, performance depends on additional factors such as L1, L2, L3 cache, thermal design and processor architecture. We therefore also report the performance testing results on real hardware configurations. Figure 3 shows a violin plot of the benchmark results, i.e., the distribution of the time required for each pipeline iteration on each computer. The specifications of the computers are listed below the device names. Each graph displays a different number of clusters indicating the concentration of the measurements within that range. The density of the charts indicates a low variance in execution time, suggesting consistent performance. The ThinkPad Yoga 370 and HP Envy x360 15 show occasional spikes of about 900 ms per run. The weakest performer among the tested devices was the ThinkPad T420 running Ubuntu 21.10, with an average run time of 853 ms and occasional spikes to over 1 second.

For most computers, the average data processing time was below the 400 ms mark, except for the ThinkPad T420, with an average time duration of 866 ms. The time needed to initiate the data processing was left out for calculating the average time. While the initialization may take some time, this can be easily compensated for by starting the prototype before the actual teaching session.

Furthermore, we derived regression plots of the average time duration for each device. The average time illustrates the dependent variable, whereas the RAM and clock rate are the independent variables. As figure 4 shows, the amount of RAM and the time needed for one pipeline iteration are negatively correlated. With an increase in RAM, we observe a decrease in time duration, which improves the device's overall performance. The regression between the CPU clock

---

<sup>2</sup> We used a virtual webcam for this purpose

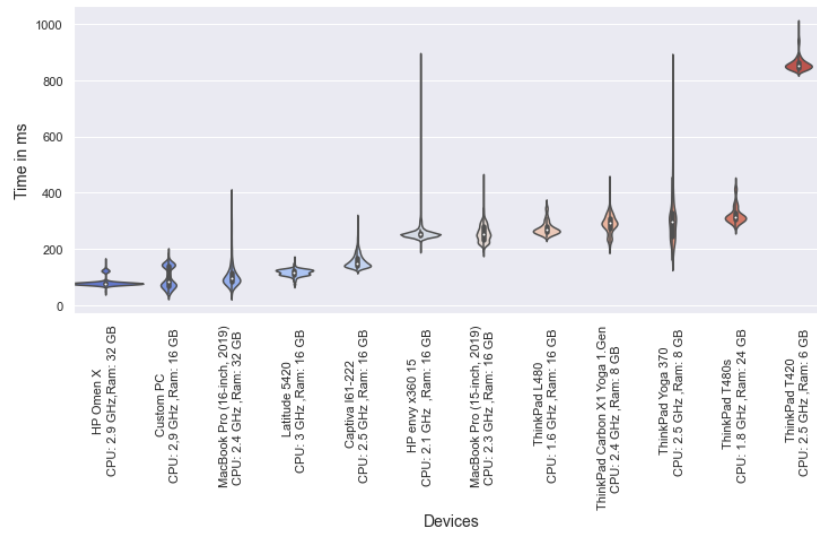


Fig. 3: Violin plot for device performance

and the same execution time also shows a slight negative correlation, as shown in figure 4. Unsurprisingly, the results show that better hardware leads to increased performance and, therefore, a decrease in the time needed to run the pipeline on a picture. The most important observation is that, with the scarce exception of a small number of iterations on ThinkPad T420, all iterations finished under a second, which successfully demonstrates the real-time operation capability of the prototype.

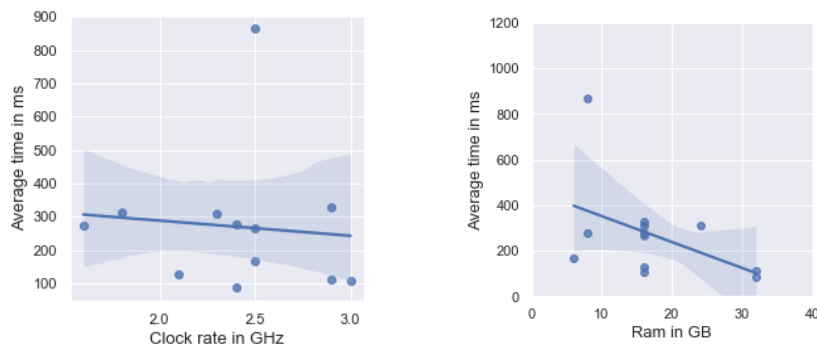


Fig. 4: Average response time vs. hardware specifications

## 5 Discussion and conclusion

The forced shift to hybrid learning in most educational institutions during the recent pandemic has affected the majority of learners and teachers throughout the globe. In this study, we aimed to explore the ways to alleviate the challenges posed by non-verbal communication limitations of synchronized online learning. Specifically, we designed STC-Live to automatically detect the learners' affective states and communicate this information to the teachers so that they can sense the overall affective status in the classroom and adapt their teaching style to improve the students' learning experience potentially. Furthermore, we implemented a machine learning pipeline that processes the webcam feed of the students to detect and extract facial expressions without the need to transfer the images to a remote server, thus, preserving the privacy of the student by design. The performance evaluation of the student component of the prototype indicates that it can run on most modern computers without causing resource bottlenecks. Moreover, the distributed architecture of STC-Live makes it highly scalable.

With the continuous advances in machine learning and affective computing, we envision many more automated methods being developed and used in practice soon. However, to reap the benefits of these technologies while avoiding the potential risks, researchers must study the underlying concepts from theoretical and practical perspectives.

An essential concern regarding the use of affective machine learning technologies is the the user's privacy. From a student's perspective, there are several concerns. Emotions are highly personal. Therefore, recording and disclosing of emotions can lead towards student profiling and eventually constitute a privacy threat. Educational providers that consider using the proposed technology must inform students and teachers regarding any attempt to analyze emotions automatically, and they must seek students' informed consent to carry out the analysis. From a teacher's perspective, such a data-intensive approach for measuring of the classroom's affective might backfire, as it could be used as an indicator to monitor teachers' performance and undermine their independence. Therefore, we caution against the use of aggregate affective measurements as performance goals and highlight the importance of using such information only for decision support to improve students' learning experience.

This study has implications for both research and practice. We described a method and the implementation details of a prototype that can detect students' affective states in an online classroom. Our method and the open-source prototype can enable educational scientists to study the effect of affective states in synchronized online education. The machine learning pipeline that we propose comprises a novel way of affective state recognition, which practitioners can tailor to fit specific purposes. In practice, such a prototype can be used by teachers in online courses that may alleviate the hardships posed by the lack of non-verbal communication between the teachers and the students, potentially improving the learning experience.

Despite the aforementioned contributions, this study is not without limitations. The first limitation relates to the accuracy of the system. STC-Live incorporates a series of underlying machine learning models which can limit its performance. Additionally, the privacy-preserving design of STC-Live makes it challenging to measure the system’s accuracy as a whole. One possible way to overcome this challenge is to conduct a separate experiment in which the participants’ video data can be recorded and manually annotated by the researchers. Only then can practitioners compare the system’s output against the ground truth annotations created by the researchers. Additionally, the role of affective states in learning must be explored by additional research. For instance, which affective states are relevant, and how can we define them in terms of observable non-verbal cues? The answer to these questions will help us improve the system and communicate the information with the teachers in an optimal way.

Another limitation relates to the privacy of the system. The contribution lies in the possibility to detect action units of students without ever collecting any imagery of them. While not collecting any images of participants certainly improves the privacy aspect of the system, the collected data (HOG values and landmarks) can still be considered sensitive data. Furthermore, linking attacks [5] could allow to identify participants using the stored data. To further improve the privacy of the system, we plan create a model for action unit detection that can be run in the browser, thus eliminating the need to send HOG values and landmarks to the server.

In the future, we will continue our research in affective state detection in learning. Specifically, we will examine how the affective states manifest as non-verbal cues in online education settings. We will study how teachers and students perceive the system, focusing on their preferences and concerns. Finally, a relevant milestone for the proposed system is to evaluate its effect in multiple courses.

## References

1. Apa dictionary of psychology, emotions. <https://dictionary.apa.org/emotions>, accessed: 2022-04-19
2. Baltruaitis, T., Mahmoud, M.M., Robinson, P.: Cross-dataset learning and person-specific normalisation for automatic action unit detection. 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) **06**, 1–6 (2015)
3. Bromuri, S., Henkel, A.P., Iren, D., Urovi, V.: Using ai to predict service agent stress from emotion patterns in service interactions. *Journal of Service Management* (2020)
4. Cheong, J.H., Xie, T., Byrne, S., Chang, L.J.: Py-feat: Python facial expression analysis toolbox. ArXiv **abs/2104.03509** (2021)
5. Chiu, W.C., Fritz, M.: See the difference: Direct pre-image reconstruction and pose estimation by differentiating hog (2015). <https://doi.org/10.48550/ARXIV.1505.00663>, <https://arxiv.org/abs/1505.00663>
6. Correia, A.P., Liu, C., Xu, F.: Evaluating videoconferencing systems for the quality of the educational experience. *Distance Education* **41**(4), 429–452 (2020)

7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) **1**, 886–893 vol. 1 (2005)
8. D’Mello, S.: A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology* **105**(4), 1082 (2013)
9. Drachler, H., Greller, W.: Privacy and analytics: It’s a delicate issue a checklist for trusted learning analytics. In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. p. 89–98. LAK '16, Association for Computing Machinery, New York, NY, USA (2016)
10. D’Mello, S., Graesser, A.: Affect detection from human-computer dialogue with an intelligent tutoring system. In: *International Workshop on Intelligent Virtual Agents*. pp. 54–67. Springer (2006)
11. Ekman, P.: An argument for basic emotions. *Cognition & emotion* **6**(3-4), 169–200 (1992)
12. Ekman, P., Friesen, W.V.: *Facial Action Coding System*. No. v. 1, Consulting Psychologists Press (1978)
13. Feidakis, M., Daradoumis, T., Caballé, S.: Emotion measurement in intelligent tutoring systems: what, when and how to measure. In: *2011 Third International Conference on Intelligent Networking and Collaborative Systems*. pp. 807–812. IEEE (2011)
14. Feidakis, M., Daradoumis, T., Caballé, S., Conesa, J.: Measuring the impact of emotion awareness on e-learning situations. In: *2013 Seventh international conference on complex, intelligent, and software intensive systems*. pp. 391–396. IEEE (2013)
15. Ghaleb, E., Mertens, A., Asteriadis, S., Weiss, G.: Skeleton-based explainable bodily expressed emotion recognition through graph convolutional networks. In: *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*. pp. 1–8. IEEE (2021)
16. Hodges, C.B., Moore, S., Lockee, B.B., Trust, T., Bond, M.A.: *The Difference Between Emergency Remote Teaching and Online Learning* (2020), <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>
17. Imani, M., Montazer, G.A.: A survey of emotion recognition methods with emphasis on e-learning environments. *Journal of Network and Computer Applications* **147**, 102423 (2019)
18. Kress, G.: *Multimodality: A social semiotic approach to contemporary communication*. Routledge (2009)
19. L.B, K., Gg, L.P.: Student emotion recognition system (sers) for e-learning improvement based on learner concentration metric. *Procedia Computer Science* **85**, 767–776 (2016)
20. Li, S., Deng, W.: Deep facial expression recognition: A survey. *IEEE transactions on affective computing* (2020)
21. Matsumoto, D., Hwang, H.S.C.: Culture, emotion, and expression. *Cross-Cultural Psychology: Contemporary Themes and Perspectives* pp. 501–515 (2019)
22. Narula, V., Feng, K., Chaspari, T.: Preserving Privacy in Image-Based Emotion Recognition through User Anonymization, p. 452–460. Association for Computing Machinery, New York, NY, USA (2020)
23. Newlands, G., Lutz, C., Tamò-Larrieux, A., Villaronga, E.F., Harasgama, R., Scheitlin, G.: Innovation under pressure: implications for data privacy during the covid-19 pandemic. *Big Data & Society* **7**(2), 2053951720976680 (2020)

24. Onyema, E.M., Eucheria, N.C., Obafemi, F.A., Sen, S., Atonye, F.G., Sharma, A., Alsayed, A.O.: Impact of coronavirus pandemic on education. *Journal of Education and Practice* **11**(13), 108–121 (2020)
25. Pekrun, R.: Emotions and learning. *Educational practices series* **24**(1), 1–31 (2014)
26. Pekrun, R., Goetz, T., Frenzel, A.C., Barchfeld, P., Perry, R.P.: Measuring emotions in students' learning and performance: The achievement emotions questionnaire (aeq). *Contemporary educational psychology* **36**(1), 36–48 (2011)
27. Pekrun, R., Goetz, T., Titz, W., Perry, R.P.: Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist* **37**(2), 91–105 (2002)
28. Pekrun, R., Lichtenfeld, S., Marsh, H.W., Murayama, K., Goetz, T.: Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child development* **88**(5), 1653–1670 (2017)
29. Scheffel, M., Drachler, H., Stoyanov, S., Specht, M.: Quality indicators for learning analytics. *Journal of Educational Technology & Society* **17**(4), 117–132 (2014)
30. Sharma, P., Joshi, S., Gautam, S., Maharjan, S., Filipe, V., Reis, M.J.: Student engagement detection using emotion analysis, eye tracking and head movement with machine learning. *arXiv preprint arXiv:1909.12913* (2019)
31. UNESCO: UN Secretary-General warns of education catastrophe, pointing to UNESCO estimate of 24 million learners at risk of dropping out (Aug 2020), <https://en.unesco.org/news/secretary-general-warns-education-catastrophe-pointing-unesco-estimate-24-million-learners-0>
32. Yadegaridehkordi, E., Noor, N.F.B.M., Ayub, M.N.B., Affal, H.B., Hussin, N.B.: Affective computing in education: A systematic review and future research. *Computers & Education* **142**, 103649 (2019)
33. Yang, S., Luo, P., Loy, C.C., Tang, X.: Wider face: A face detection benchmark. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016)