

A multimodal dataset for understanding the impact of mobile phones on remote online virtual education

Roberto Daza^{1,*}, Alvaro Becerra², Ruth Cobos², Julian Fierrez¹, and Aythami Morales¹

¹Biometrics and Data Pattern Analytics Laboratory, Universidad Autonoma de Madrid, School of Engineering, Madrid, 28049, Spain

²Group for Advanced Interactive Tools, Universidad Autonoma de Madrid, School of Engineering, Madrid, 28049, Spain

*Corresponding author: Roberto Daza (roberto.daza@uam.es)

ABSTRACT

This work presents the IMPROVE dataset, a multimodal resource designed to evaluate the effects of mobile phone usage on learners during online education. It includes behavioral, biometric, physiological, and academic performance data collected from 120 learners divided into three groups with different levels of phone interaction, enabling the analysis of the impact of mobile phone usage and related phenomena such as nomophobia. A setup involving 16 synchronized sensors—including EEG, eye tracking, video cameras, smartwatches, and keystroke dynamics—was used to monitor learner activity during 30-minute sessions involving educational videos, document reading, and multiple-choice tests. Mobile phone usage events, including both controlled interventions and uncontrolled interactions, were labeled by supervisors and refined through a semi-supervised re-labeling process. Technical validation confirmed signal quality, and statistical analyses revealed biometric changes associated with phone usage. The dataset is publicly available for research through GitHub and Science Data Bank, with synchronized recordings from three platforms (edBB, edX, and LOGGE), provided in standard formats (.csv, .mp4, .wav, and .tsv), and accompanied by a detailed guide.

Background & Summary

Education is a fundamental pillar of society and has continually evolved throughout history. In recent years, it has been deeply influenced by the digital revolution, which has moved our society from the physical to the digital world. Evidence of this change in recent decades includes the growing significance of e-learning or online education, expected to grow exponentially over the next 20 years^{1,2}. Relevant institutions are adopting this model, especially those offering MOOCs (Massive Open Online Courses), due to their ability to reach a broader learner base³.

Learners have also been significantly impacted by the digital revolution, which has altered their habits to incorporate electronic devices into their daily routines such as mobile phones and tablets. Since 2012, studies^{4,5} have indicated that between 97% and 99% of learners own a mobile phone. Moreover, 95% of learners bring their phones to class every day, 92% use these devices to send text messages during class, and 10% even admit to texting during exams. On average, learners spend about five hours per day on their mobile phones, predominantly using messaging services (sending an average of 109 text messages per day) or social networks like WhatsApp and Instagram^{6,7}.

Therefore, the mobile phone raises numerous questions: Can it become a distraction to learners' attention? Or, conversely, could the absence of a mobile phone cause greater distraction or anxiety? Considering that attention is defined as the ability to concentrate, specifically, to exert conscious cognitive effort on a particular task or stimulus at any given moment⁸, and given that various studies suggest it is easy to lose focus due to short-term distractions such as messages, sounds, etc.^{9,10}, it is reasonable to consider how mobile phones could impact learner performance. This is not a new idea, and various studies have shown that mobile phone use negatively affects learners' attention, memory, and academic performance. This is mainly because it increases cognitive load by inducing learners to multitask^{5,10-12}. This effect becomes more noticeable as the class progresses or during long-duration activities¹¹. However, learners' performance can be influenced by their emotional state. For this reason, nomophobia, defined as the fear, stress, or anxiety caused by being unable to access or use a mobile phone, is an important factor to consider. Various studies have demonstrated how this condition affects memory and attention, ultimately impacting learners' academic performance^{11,13}.

Interestingly, although nomophobia can lead to reduced academic performance, the previously mentioned studies have shown that just the presence of a mobile phone also diminishes learners' attention levels, presenting a complex challenge that requires in-depth investigation.

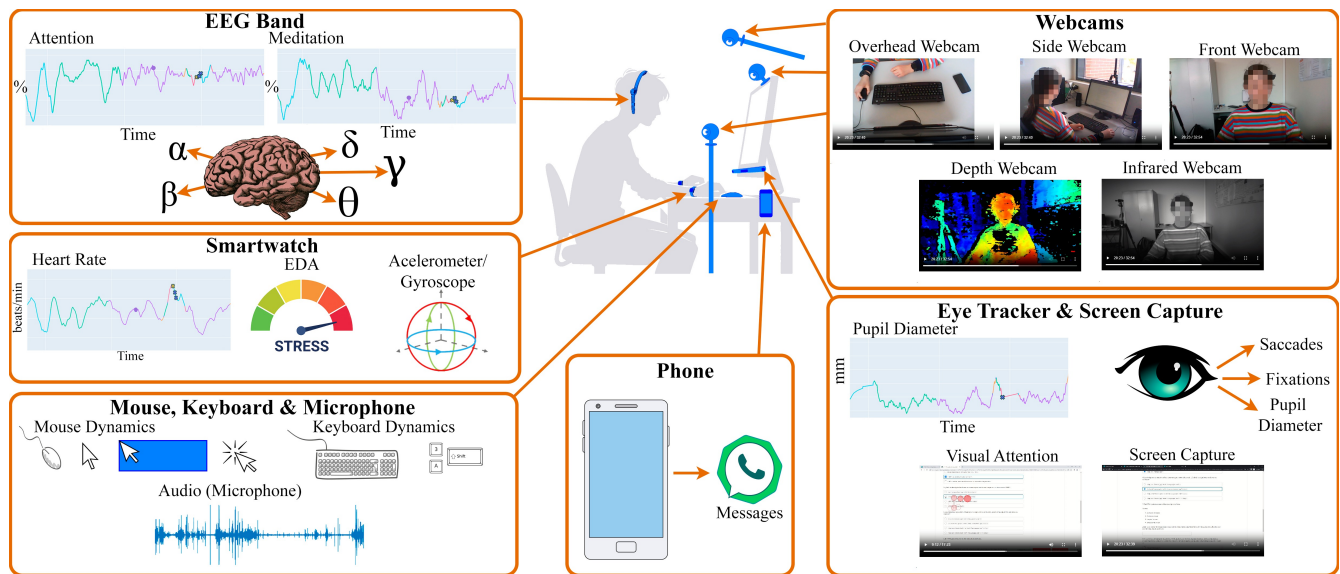


Figure 1. Acquisition setup used during data capture with the edBB^{1,14} platform, illustrating all sensors and devices employed, along with examples of the captured information.

The IMPROVE dataset was collected in a realistic online learning environment. A total of 120 learners were monitored, divided into three distinct groups to understand the effects of mobile phone usage. To our knowledge, IMPROVE dataset is the first to provide a comprehensive multimodal acquisition setup including behavioral^{1,14}, biometric, and learning analytics data from learners during an authentic online course that considers mobile phone usage. The dataset aims to facilitate understanding of the effects of mobile phone usage in online education. Additionally, it provides a robust dataset that can be utilized to develop new technologies to enhance learners' learning processes or ensure secure certification¹⁴⁻¹⁹.

We employed 16 sensors, recognized as effective indicators for understanding learner behavior^{8,20-22}, including an electrodermal activity (EDA) sensor, an electroencephalogram (EEG) band, two pulse meters, an accelerometer, a gyroscope, three RGB cameras, two Near-Infrared (NIR) cameras, a mouse, a keyboard, a microphone, a monitor, and an eye tracker (see Fig. 1). These sensors enabled us to monitor cognitive activity, biometric, and behavioral signals throughout 120 online learning sessions, each lasting approximately 30 minutes. During these sessions, learners interacted with educational content related to HTML, including video lectures, document reading, and test completion to assess their knowledge both before and after the session (with a total of 16 questions per session). This comprehensive setup allowed us to track not only how learners engaged with the material but also how external factors, such as mobile phone interruptions, affected their focus and learning outcomes.

In total, the dataset comprises 2.83 terabytes of data, making it one of the most extensive multimodal datasets in this domain. It includes 145 labeled mobile phone interruptions, providing crucial insights into how mobile phone usage affects learners.

Methods

This section outlines the experimental methods and materials used for data collection. The dataset was obtained in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Universidad Autónoma de Madrid (Approval No. CEI-130-2699, granted on 11/04/2023). Learners were informed about the experimental procedure and the data acquisition setup, and each participant signed a written consent form. The form clearly stated that the collected data would be used exclusively for research purposes and included in a publicly available dataset intended for the research community. Learners received an incentive for participating.

Dataset: IMPROVE

The dataset consists of 120 learners from the School of Engineering at the Universidad Autónoma de Madrid (UAM), who were monitored during a real online learning session focused on HTML, lasting between 20 and 35 minutes. The session was part of the MOOC titled "Introduction to Development of Web Applications" (WebApp MOOC for short) available on the edX platform (<https://www.edx.org>).

Category	Subcategory	Percentage	Number of Learners	Average Age
Degree Program	Computer Engineering	32.5%	39	21.77
	Telecommunications Engineering	30.0%	36	22.06
	Biomedical Engineering	25.0%	30	19.80
	Data Science and Engineering	3.3%	4	20.00
	Computer Engineering and Mathematics	9.2%	11	20.55
HTML Proficiency	None	42.5%	51	20.02
	Beginner	35.8%	43	21.95
	Intermediate	16.7%	20	21.80
	Advanced	5%	6	23.67
Medical Conditions	No issues	—	55	—
	Glasses	—	41	—
	Contact lenses	—	19	—
	Heart murmur	—	2	—
	Myopia without correction	—	4	—
Overall Averages	Overall Average Age	—	—	21.19
	Average Age of Male Learners	51.83%	61	21.64
	Average Age of Female Learners	49.16%	59	20.73

Table 1. Distribution of learners by degree program, HTML proficiency levels, medical conditions (percentages are not shown because learners can have more than one condition), and gender, along with their average ages.

Recruitment.

Learners from the School of Engineering at the Universidad Autónoma de Madrid (UAM) were recruited. The School of Engineering was selected because its five degree programs all require basic knowledge of computer science, which aligns well with the online course. An economic incentive was offered to encourage participation. A total of 170 learners signed up for the study, of whom 120 were selected. An initial pilot test was conducted with six learners to verify the experimental protocol and monitoring system. Adjustments were made after the initial test, including: requiring Group 1 learners to provide more detailed responses to messages (for more information about the groups formed, see the subsection Protocol), giving learners a more comprehensive explanation of the session beforehand, removing an initial form to avoid potential suggestion effects, and making technical adjustments to the EEG band and the eye tracker calibration. The six learners from the pilot test did not participate in the final data collection. Four additional learners were excluded due to issues during the session and replaced by other participants. The final 120 learners were divided into three groups of 40. The selection process considered several criteria to ensure balanced groups, including gender (61 male, 59 female), HTML proficiency (none, beginner, intermediate, advanced), degree program, among other factors. Learners ranged in age from 18 to 32 years, with an average age of 21 (see Table 1). All learners gave their informed consent for the publication of the dataset. Each learner was anonymized with an identifier that cannot be linked to their identity.

Protocol.

The objective was to examine the effects of mobile phone presence or absence on learner behavior and academic performance. To this end, three distinct groups were formed:

1. **Group 1:** Mobile phone use and possession were allowed. The device was placed on the learner's desk, visible to the learner, with both sound and vibration activated.
2. **Group 2:** Mobile phone possession was allowed, while use was prohibited. The device was placed on the learner's desk with the screen facing down, and sound and vibration were activated.
3. **Group 3:** The mobile phone was confiscated for the entire duration of the learning session.

To ensure consistency across all groups, all learners underwent the same learning session, delivered in a fixed sequence without enforced time limits between tasks. Learners were otherwise free to behave as they typically would in an e-learning context, with the only variation stemming from the mobile phone usage condition assigned to each group. No explicit instructions were given to avoid distractions unrelated to mobile phone use. However, two general restrictions applied: (i) learners could not return to previous tasks, and (ii) they were not allowed to control video playback (i.e., pause, rewind, or skip).

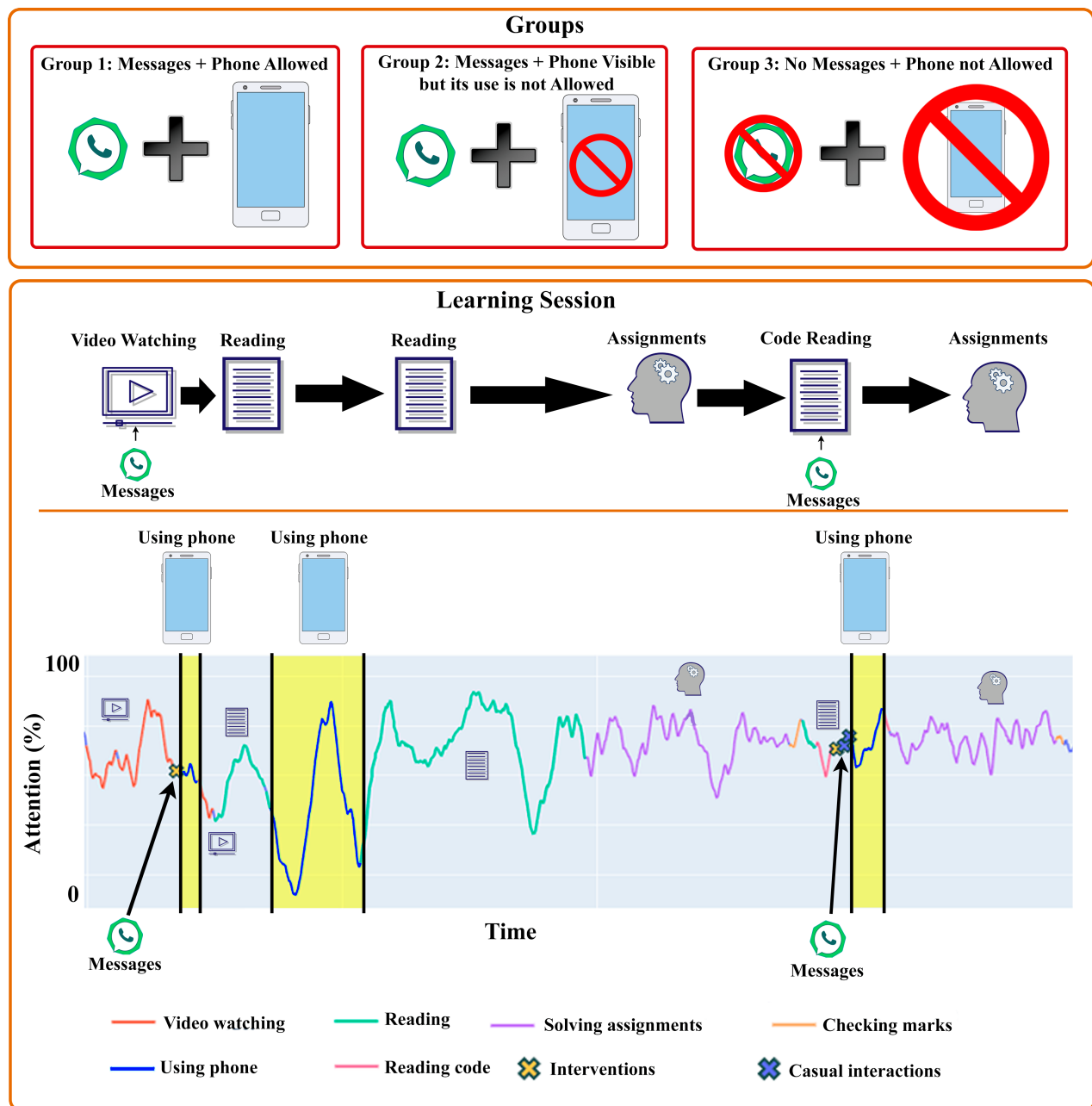


Figure 2. Protocol followed during the learning session to evaluate the impact of mobile phone usage on learners' academic performance. The top section presents the experimental design, illustrating the division of learners into three groups based on the permitted level of mobile phone interaction: unrestricted use, visible but prohibited use, and complete removal. The bottom section details the structure of the learning session, including the sequence of activities (e.g., video watching, reading, and completing assignments). It also shows a sample graph of a learner's attention signal, synchronized with the session timeline. Time intervals in the signal are color-coded to reflect each activity, each accompanied by its corresponding icon, while windows of phone usage are highlighted by vertical boundary lines, with contextual icons indicating the type and timing of interruptions. This provides a visual representation of when phone-related interruptions occurred during the session.

Three supervisors were present throughout the session. The first supervisor was responsible for explaining the session in detail to the learner, describing the operation of the sensors, placing and calibrating them, and ensuring that the platforms functioned correctly. The second supervisor used the LOGGE tool²³ (for more details on this tool, see the "Data, platforms and sensors" subsection) to label the moments when the learner interacted with the mobile phone or when it rang, as well as to label other relevant information for the session. The third supervisor was responsible for sending two messages to the

learners at key moments during the learning session for both Group 1 and Group 2, reviewing the information recorded by the second supervisor, noting any possible incidents during the session, and assisting the learner in case of doubts. We will differentiate between two different types of interruptions: interventions (the two controlled messages sent by the supervisor) and casual interactions (uncontrolled messages/calls received by the learner that were not related to the interventions). Therefore, interventions are considered controlled events, while interactions are classified as uncontrolled. The interventions were always delivered at the same points during the session. The first intervention occurred during an explanatory video of the lesson, and the second during the reading of HTML code, which was part of the learner evaluation (see Fig. 2). Learners in Group 1 were required to respond with a detailed reply. The messages consisted of simple questions such as “What do you think of the cafeteria of our School of Engineering?”, or an image from a popular movie like Harry Potter, accompanied by the question: “What do you think of this film series? Would you recommend it?” On average, learners took about 26 seconds to respond to the first message and about 36 seconds to respond to the second message. Learners in Group 2 were not allowed to use their mobile phones, so their phones simply rang and vibrated during the interruptions.

Additionally, learners in Group 1 were given the freedom to respond/send messages or calls to their contacts. Each of these interactions was recorded. On average, the learners in Group 1 chose to interact with their mobile phones approximately 5 times (including controlled messages and uncontrolled messages or calls), while the mobile phone in Group 2 rang/vibrated approximately 7 times.

Tasks

Each learner completed a single learning session focused on HTML, taken from the WebApp MOOC. The session lasted an average of 30 minutes, with most learners completing it in 20 to 35 minutes, and a maximum duration of 40 minutes. The tasks performed during the session were as follows:

Before starting the HTML unit:

- **Calibration and Learner Information:** Learners were briefed again on the session procedures. Sensors, such as the eye tracker and EEG band, were fitted and calibrated. Learners were asked to provide information that could affect the sensors, such as the use of glasses/contact lenses, cardiac conditions, handedness (right or left-handed), etc. Specific medical conditions reported by the learners, including their prevalence, are summarized in Table 1.
- **EDA Test:** Before starting the session, an initial EDA test was conducted to understand the learners’ stress levels. Learners were also asked to self-report their current state as “Very stressed, stressed, normal, relaxed, extremely relaxed” to obtain their feedback.
- **Form and Pretest:** Additionally, they filled out a form with the following two questions, i) What is your current level of stress/anxiety? and ii) What is your current level of distraction? This information was collected to gather learner feedback prior to the session and to validate it against sensor data. Learners also completed a pretest consisting of eight multiple-choice questions of low to intermediate difficulty on HTML, with two objectives: *i)* to assess the learners’ current level of HTML knowledge and *ii)* to balance the three groups.

HTML Unit: The HTML unit consisted of a theoretical part where learners were introduced to HTML concepts in web design through explanatory videos and reading web documents, and a second part focused on assessing their understanding. This evaluation involved eight multiple-choice questions. After completing the test, learners were able to review their incorrect answers. At no point were learners allowed to return to previous activities, pause the videos, or skip forward/backward in the video. The tasks were ordered as follows (see Fig. 2):

1. **Videos:** Two introductory videos on interactive web and basic HTML concepts. The first video lasted 20 seconds, while the second lasted 1 minute and 48 seconds. During the second video, a message was sent to the learners in Groups 1 and 2 after 1 minute, coinciding with the explanation of HTML concepts.
2. **Document Reading:** Two documents further explaining the videos, covering concepts such as HTML structure, different tags, etc.
3. **Test:** This test was designed to be compared with the pretest, in order to evaluate whether the students had learned during the learning session compared to their initial knowledge. The questions were similar to those in the pretest. Six multiple-choice questions followed by a phase of reading HTML code to answer two more multiple-choice questions. During the HTML code reading, the second message was sent to learners at the beginning of the reading when comprehension of the code required greater attention. However, learners could resume reading the code after answering the message, unlike the video where they could not rewind.
4. **Review Phase:** Finally, learners accessed a review phase where they could check and understand the mistakes they made.

Learner Feedback	Group 1			Group 2			Group 3		
	Low	Medium	High	Low	Medium	High	Low	Medium	High
Anxiety at Start	52.63	39.47	7.89	51.43	40.0	8.57	43.24	51.35	5.41
Distraction at Start	47.37	52.63	0.0	51.43	34.29	14.29	67.57	21.62	10.81
Anxiety During Session	60.53	28.95	10.53	60.0	34.29	5.71	64.86	35.14	0.0
Distraction During Session	44.74	39.47	15.79	42.86	45.71	11.43	72.97	21.62	5.41
Session Difficulty	37.5	55.00	7.5	40.0	42.5	17.5	45.0	47.5	7.5
Performance Level	7.89	50.0	42.11	8.57	60.0	31.43	10.81	43.24	45.95
Phone Distraction	26.32	50.0	23.68	40.0	50.0	10.0	-	-	-
Stress Due to Phone Inaccessibility	-	-	-	73.68	26.32	0.0	95.0	5.0	0.0

Table 2. Learner feedback from forms, categorized into three levels (low, medium, and high) and expressed as percentages. The responses cover aspects such as anxiety, distraction, perceived difficulty, and self-assessed performance, and are presented separately for each experimental group.

End of the Session:

- A final **EDA Test** was conducted to detect post-session stress levels, and learners' feedback was collected again.
- Learners completed a post-session **Form**, which asked about their levels of stress, distraction, perceived difficulty, and performance during the session. Additionally, learners in Groups 1 and 2 were asked about phone distraction, while Groups 2 and 3 were asked about stress from phone deprivation. Table 2 presents the responses collected from the self-reported forms.
- Learners were asked to fill out a form with the following six questions:
 - "What was your level of stress/anxiety during the learning session?"
 - "What was your level of distraction during the learning session?"
 - "How would you evaluate the difficulty of the learning session you just completed?"
 - "How would you rate your performance during the learning session?"
 - "Observations: Please add any comments you wish to make after the data collection."

The final question varied depending on the group assigned to each learner:

- For Groups 1 and 2: "How distracted were you by your mobile phone?"
- For Groups 2 and 3: "How stressed did you feel due to not being allowed to use your mobile phone?"

Data, Platforms and Sensors

Two different platforms were used to monitor the learners: the edX and edBB platforms^{1,14}. The edX platform is an online MOOC platform used to conduct the learning sessions. It also allows for the capture of session metadata from learners, such as test answers, task start and end times, and more. The edBB platform was used to monitor behavioral and biometric signals that have been shown to contain relevant information for understanding learner engagement and performance^{1,8,14,20,22,24,25}. See Daza et al.¹ for a video demonstration (<https://www.youtube.com/watch?v=JbcL2N4YcDM>). A multimodal acquisition framework was designed, incorporating 16 sources of information/sensors (see Fig. 1 and Table 3):

- **Cameras:** Two Logitech C170 cameras (side and overhead) operating at 20 Hz with a resolution of 640×480 , and one front-facing Intel RealSense camera were used. The RealSense²⁶ camera includes one RGB camera and two NIR cameras, with dimensions of 90 mm length $\times$ 25 mm (depth) $\times$ 25 mm (height). The NIR cameras are monochrome and sensitive to both the visible spectrum and NIR, following the sensitivity curve of the CMOS sensors. The three Intel RealSense cameras were configured to operate at 30 Hz and a resolution of 1280×720 , and depth images were obtained by combining their three image channels. See Fig. 1 for more details on the position of each camera.

Information Type	Sensors or Platforms	Sampling Rate	Features
Video	3 RGB cameras 2 Infrared cameras 1 Depth camera	20 Hz - 30 Hz	MP4 files with codec H264
Desktop Video	Screen	1 Hz	MP4 file with codec H264
Audio	Microphone	8000 Hz	Uncompressed WAV files
Keystroke	Keyboard	12 Hz	Keypress and key release events
Mouse Dynamics	Mouse	895 Hz	Mouse events: Move, press/release, drag and drop and mouse wheel spin
EEG	Band	1 Hz	Power Spectrum Density of five frequency bands. Level of attention (from 0 to 100) and eyeblink strength
Pulse, Stress, temperature and Inertial	2 SmartWatch: Huawei Watch 2 and FitBit Sense with PPG, Gyroscope, Accelerometer, Electrodermal Activity	100 Hz	Timestamps and data from the pulse, EDA sensor (skin conductance), body temperature and the inertial sensors (accelerometer, gyroscope, magnetometer)
Visual attention, Gaze, and Eyeblink	2 Eye Tracking Cameras	120 Hz	Gaze origin and point, saccades, fixations, gaze event duration, pupil diameter, eyeblink, data quality
Face position, Head Pose and Facial Landmarks	edBB platform	30Hz	Facial bounding box information, 468 facial landmarks, Euler angles
Context Data	edBB platform, LOGGE Data-M2LADS, edX, learner, Forms	NA	Learner Information (gender, HTML proficiency level, field of study, health issues, etc.), mobile phone usage events, computer details including name, private IP, public IP, MAC address, operating system (OS), architecture, keyboard language, screen resolution, available memory, total memory, start and finish times of each task session, and test answers, etc.

Table 3. Summary of sensors and data collection from the IMPROVE dataset.

- Smartwatches: Two different smartwatches, the Huawei Watch 2 and the FitBit Sense, were used to monitor heart rate^{14,27,28}, stress levels via the electrodermal activity (EDA) sensor^{27,29,30}, and inertial data using gyroscopes and accelerometers³¹. Two different smartwatch models were used, each worn on a different wrist, so that the heart rate monitored by the devices could be compared. The heart rate measurement was set to 1 Hz for the Huawei Watch 2 and 0.2 Hz for the FitBit Sense. The FitBit Sense has one sensors that are not available in the Huawei Watch 2: the EDA sensor. The Huawei Watch 2 was worn on the dominant hand of the learner to detect the movements produced by mouse usage.
- EEG: A NeuroSky EEG headset, was used to measure the power spectral density across five electroencephalographic frequency bands ($\alpha, \beta, \gamma, \delta, \theta$). Through the preprocessing of these EEG channels, estimates of attention and meditation levels, as well as the occurrence of eyeblinks, are obtained. EEG captures voltage signals typically generated by synaptic excitations of the dendrites in the pyramidal cells located in the top layer of the brain cortex³². These signals are primarily produced by the synchronous firing of numerous neurons and fibers³³. EEG is an important, effective, and objective measure for estimating cognitive load and learner states^{34,35} because these signals are sensitive to the mental effort/cognitive works and mental states such as learning, lying, perception and stress. Extensive research in neurology and learning fields has explored the relationship between EEG signals and mental activities^{24,33,36–39}, demonstrating the potential of the five electroencephalographic channels and the correlation between eyeblinks and cognitive load^{22,24,40,41}. In the context of this dataset, EEG data were included to provide a direct and continuous measurement of learners’

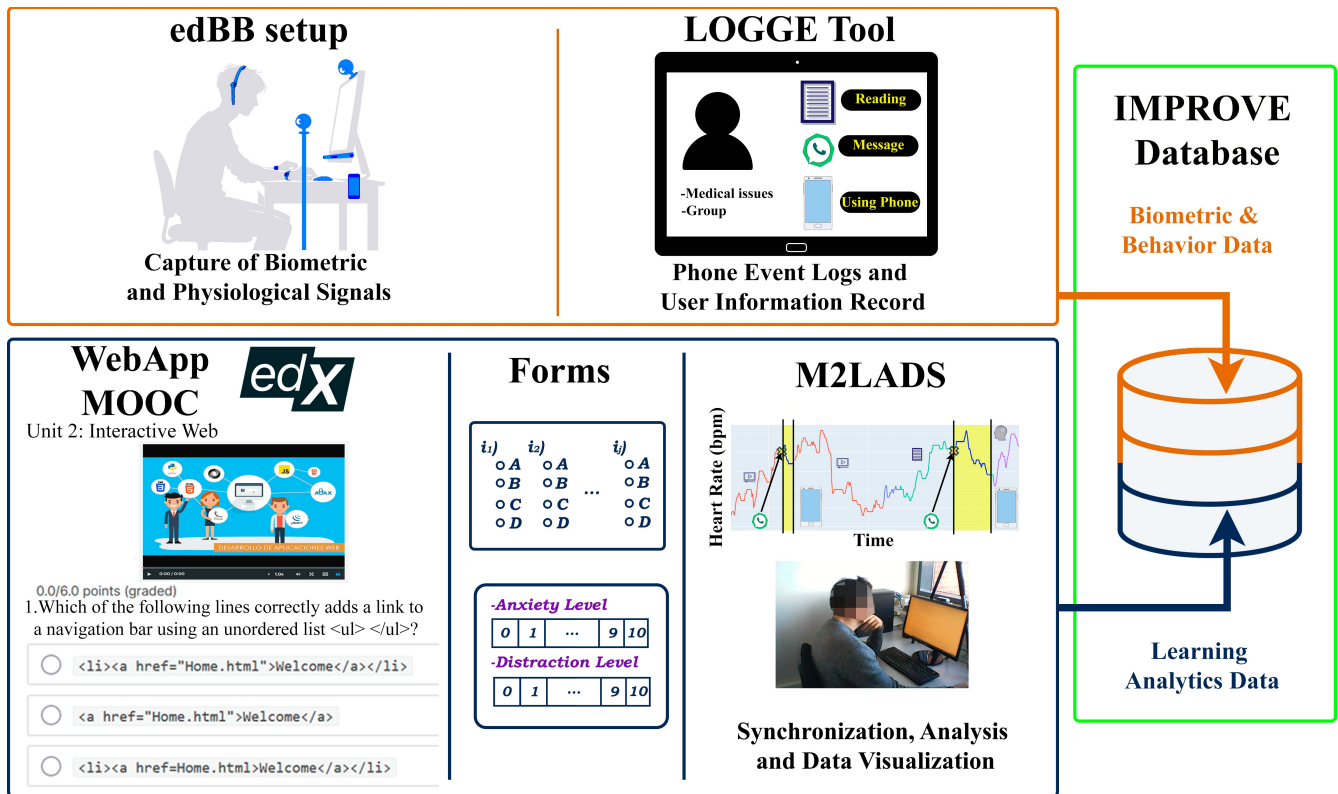


Figure 3. Overview of the platforms and tools used to capture the IMPROVE dataset. The edBB setup was used to monitor the learner’s biometric and behavioral signals during online sessions. The edX platform delivered the course content and collected session metadata. The LOGGE Tool recorded additional information, such as medical issues and group assignments, and mobile phone usage. The data were aggregated into the IMPROVE dataset, which could be analyzed using the M2LADS platform to visualize and synchronize the captured signals and events.

cognitive states throughout the session. This information constitutes a key component for evaluating the impact of mobile phone usage on cognitive load, particularly with regard to fluctuations in attention levels and potential effects associated with nomophobia or mobile-induced distractions. This is especially relevant given that attention is defined as the cognitive effort directed toward a task and plays a pivotal role in ensuring accurate comprehension during learning.

- Visual attention: A Tobii Pro Fusion eye tracker equipped with two high-speed infrared cameras configured at 120 Hz for eye tracking. This device estimated the following data: gaze origin and point, pupil diameter, eye movement type (fixation, saccade, unclassified, eyes not found), event duration, data quality, eyeblink, and more; allowing us to measure visual attention. Previous studies have demonstrated that visual attention is a key factor in learning, providing a deeper understanding of learner performance^{42,43}. There is a strong correlation between learner performance and visual attention, with features such as fixations, pupil diameter, eyeblink and saccades being closely associated with learning outcomes⁴², fatigue detection⁴⁴, and task prediction⁴⁵.
- Computer Information: Relevant information for the session were obtained from various sources on the learner’s computer. i) Microphone Information: Audio was recorded from the computer’s microphone at a sampling rate of 8000 Hz. The audio was recorded throughout the session, and all learners pronounce an initial phrase, "Sound test for the edBB platform," which can be used as a baseline for future research. ii) Keyboard and Mouse Activity: Keystrokes, mouse position, inter-keystroke timing, mouse wheel movements, etc., are monitored. Keystrokes and mouse dynamics have proven to be effective mechanisms for understanding learner sessions and providing secure certification^{46,47}. iii) Screen Capture: The monitor screen was recorded at a frequency of 1 Hz. iv) Session metadata: Information such as keyboard type, logging data, IP and MAC addresses, operating system, etc., was collected.

The edBB platform facilitated optimal synchronization across all sensors, recording the captured samples in both UNIX time and local time, and sending the information to a common server for easy access. By using a single platform to control all

sensors, it was possible to synchronize the activation and termination times of each sensor. This configuration also enabled the acquisition of precise information, such as the exact moment a learner pressed a key and the corresponding video frame, without requiring additional post-processing.

In addition to using these platforms, a logging tool called LOGGE²³ was developed with the primary goal of labeling the mobile phone events for Group 1 and Group 2, which consisted of learner interactions with their mobile phones or when their phones rang. Additionally, this tool collected learner-related information (e.g., use of glasses/contact lenses, cardiac conditions, handedness, etc.) and enhanced the edX log traces with details about the activities performed during the learning sessions, such as task start and end times, error logs, and more.

Finally, a web-based system called M2LADS^{23,48} was designed to visualize the information captured from each monitored learner. This system enabled the following: *i*) Validation of correct monitoring and synchronization: visualizing the information from the captured signals (webcam videos, screen recordings, heart rate, attention level, etc.) and ensuring their correct synchronization. *ii*) Statistical analysis of the data: comparing learner performance (pretest vs. posttest scores), analyzing signals across activities (e.g., identifying which activities had higher attention levels or heart rates), and correlating biometric signals. The overall architecture, including the platforms and tools used to gather IMPROVE dataset, is summarized in Fig. 3.

Data Processing

Facial Video Processing. The videos from the RealSense RGB camera were processed to extract the learner's facial position, facial landmarks, and 3D head pose for each frame. Three state-of-the-art methods based on Convolutional Neural Networks (CNNs) were employed:

- **Facial Detection:** The facial detection module employed MediaPipe's BlazeFace⁴⁹ model (full-range version) for 2D facial image detection. This module is based on CNNs specifically designed to significantly reduce computational cost and inference time without compromising accuracy. BlazeFace integrates a lightweight feature extraction architecture inspired by MobileNetV1/V2⁵⁰, leveraging their efficiency and speed. Specifically, the model's architecture applies advanced convolution techniques, similar to those employed in networks like CenterNet⁵¹, with a customized encoder. The detector was trained on large private datasets containing geographically diverse facial detection samples.
- **Head Pose Detection:** A real-time head pose detector based on the WHENet⁵² architecture was employed. The Euler angles —pitch (vertical), yaw (horizontal), and roll (longitudinal)— were estimated from 2D head images, allowing the 3D orientation of the head to be deduced from 2D facial images. This architecture is based on fully connected CNNs for angle classification. Specifically, EfficientNet⁵³ was used as the backbone for feature extraction. The architecture consists of several layers, starting with convolutional layers that capture spatial features of the image. These convolutional layers were followed by pooling layers. A series of fully connected dense layers interprets the extracted features and performs the classification and regression of the Euler angles. In the classification process, each of the angles (yaw, pitch, and roll) was classified into bins, while regression was used to accurately predict the Euler angle values within those bins. The head pose detector was trained on the 300W-LP⁵⁴ and CMU⁵⁵.
- **Facial Landmark Detection:** The MediaPipe library was employed to estimate 478 3D facial landmarks in real time from facial images. This enables the extraction of relevant information from facial regions such as the eyes, forehead, mouth, and more. The model is based on a CNN, similar to MobileNetV2⁵⁰, but with custom blocks designed for real-time performance. The model was evaluated on a private dataset comprising 1,700 examples uniformly distributed across 17 geographical subregions. It achieved a normalized mean error of 2.62%, calculated relative to the inter-ocular distance, with similar error rates observed for each subregion. Consequently, each video is accompanied by the 478 3D facial landmarks for every frame, along with detailed information about the estimation of each landmark, including its visibility and presence in the image.

Eye Tracker Signal Processing. The data obtained from the eye tracker was processed to extract relevant features such as fixations and saccades. On average, Tobii Pro Fusion captures one sample with gaze coordinates every 8.3 ms. These samples can be processed over the entire video duration to obtain relevant visual attention features, such as fixations and saccades. For this purpose, one of the built-in filters provided by the Tobii Pro Lab software was selected, specifically the I-VT (Fixation) filter. We selected this filter because it is suitable for our study, where head movements are minimal and the predominant eye movements are fixations and saccades. The filter classifies gaze points as either fixations or saccades based on their velocity: points with speeds below 30 degrees per second are grouped into a single fixation. Some characteristics of this filter include: *i*) Noise reduction: A moving average is applied every 3 samples. *ii*) Velocity calculation: Velocity is calculated in 20 ms windows. *iii*) Minimum fixation duration: Fixations shorter than 60 ms are discarded. *iv*) Merging adjacent fixations: Adjacent fixations are merged if the interval between them is 75 ms or less and the angular distance is 0.5 degrees or less.

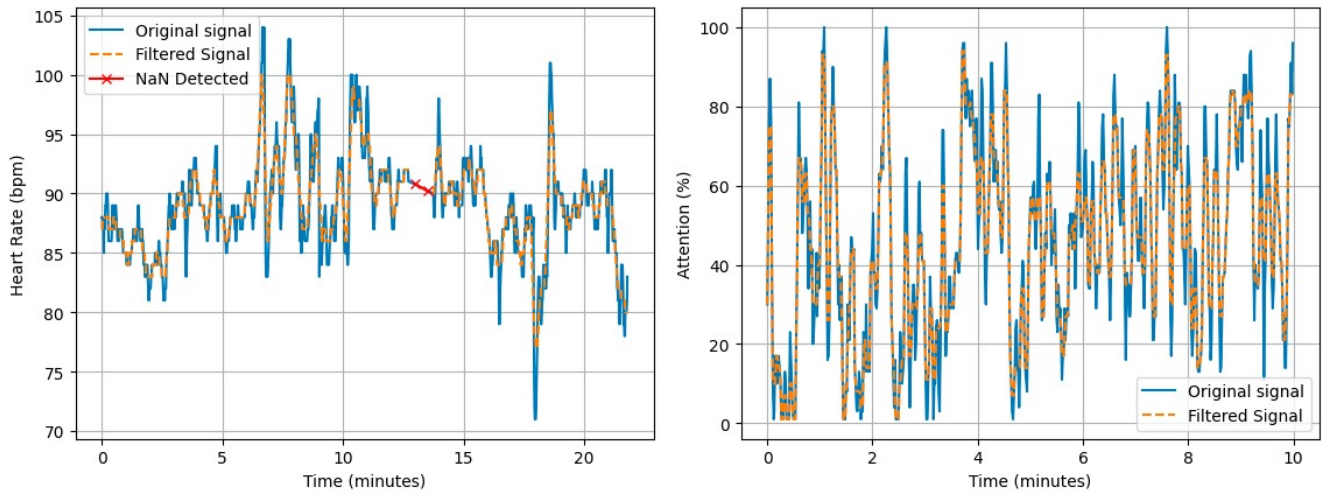


Figure 4. Example of filtered signals in IMPROVE: The left graph displays the raw heart rate signal and the same signal filtered using a moving average filter (with a 15-second window). The right graph shows the raw attention signal derived from the EEG and filtered using a median filter (with a 5-second window). In both graphs, the original signals are represented by a solid blue line, while the filtered signals are shown with a solid yellow line.

EEG Signal Processing. The data obtained from the EEG headband were filtered to mitigate potential errors, such as signal fluctuations caused by sudden participant movements, ensuring more accurate EEG recordings. The signals were processed using a median filter with a window size of 5, effectively reducing noise from artifacts, such as participant movements, eyeblinks, and other sources, without distorting the peaks or extreme values in the signal. Additionally, interpolation was applied to data losses shorter than 5 seconds, while longer data gaps were marked as missing. Both the raw and filtered signals are available in the dataset.

Smartwatch Signal Processing. The heart rate data were also filtered to remove minor fluctuations and smooth the signal. Specifically, a moving average filter with a 15-second window was applied, as the pulse signal in this context is expected to change slowly, with beat-to-beat variations typically not being abrupt at the sampling rate of smartwatches. The signal obtained from the smartwatch’s accelerometer was filtered using a fourth-order Butterworth low-pass filter with a cutoff frequency of 15 Hz. This approach was used to eliminate high-frequency noise, such as vibrations or unintentional movements, while preserving the rapid and relevant movements during the learner’s interactions with the mouse. For the gyroscope signal, a similar Butterworth low-pass filter was applied, but with a cutoff frequency of 10 Hz, which allowed for the capture of rapid wrist movements while eliminating potential high-frequency artifacts. Both the raw and filtered data are available in the dataset. Fig. 4 presents examples of filtered signals from the IMPROVE dataset.

Data Records

The IMPROVE dataset is available from the Science Data Bank⁵⁶ (<https://doi.org/10.57760/sciencedb.14565>) and GitHub (<https://github.com/BiDALab/IMPROVE>). The dataset comprises 2.83 terabytes of information, including biometric data, physiological signals, videos, supervisor-labeled events related to mobile phone usage, and more. This data was collected from the edBB and edX platforms and the LOGGE tool, and includes both raw and processed signals. The dataset is organized into three main folders, each corresponding to a different group based on mobile phone usage (see Subsection Protocol for more details). Within each of these folders, there are 40 main subfolders, each labeled with the learner’s ID, containing the monitored data for each learner during the learning session. Each of these subfolders is further divided into four categories: ‘edBB’, ‘edBB Processed Data’, ‘edX,’ and ‘LOGGE,’ corresponding to the information collected from each platform and tool. The structure and contents of the dataset are summarised in the following tables: Table 4, Table 5, and Table 6. The processed signals and videos obtained through the edBB platform are stored in the ‘edBB Processed Data’ folder and are described in Table 4. Table 5 lists the files and descriptions from the edX platform and the LOGGE tool. Additionally, all files from the edBB platform and their descriptions are summarized in Table 6.

Folder Name	Files	Description	Format	File Contents
Video	box head_pose landmarks	Processed video data: Bounding boxes, facial landmarks, and head pose of all users in the RealSense video	.csv .csv .csv	Bounding box is formatted as [x, y, width, height], 478 landmarks per detected face, Euler angles (yaw, pitch, and roll) in degrees, frames
MindWave	filter_ATT filter_MED filter_BANDPOWER	EEG band signals filtered with a median filter	.csv .csv .csv	Attention (0-100), meditation (0-100), blink strength, 5 EEG waves (dB), local date and time
Smartwatch	filter_ACC filter_Gyro filter_Heart	Huawei Watch 2 signals filtered using a Butterworth low-pass filter or a moving average filter	.txt .txt .txt	Acceleration on X, Y, Z axes (m/s^2), angular velocity on X, Y, Z axes ($^{\circ}/s$), heart rate (bpm), UNIX timestamp
Fitbit	Heart	Heart rate signal from the FitBit Sense filtered with a Butterworth low-pass filter	.csv	Heart rate (bpm), timestamp

Table 4. Descriptions of the processed edBB files included in the edBB Processed Data folder.

Folder Name	Files	Description	Format	File Contents
Logge	file_events_LS file_marks file_position user_data file_pretest	Activities and events logs	.csv .csv .csv .csv .csv	Key name for the activity/event, start timestamp, end timestamp, marks of assignments, mobile phone position in table, medical issues, group, pretest mark, posttest mark, self-reported stress before and after the session, self-reported distraction before and after the session, self-reported difficulty of session, self-reported performance
edX	file	edX logs	.csv	Action name, context, session, time, event type

Table 5. Description of files collected by the edX platform and LOGGE tool, included in the edX and LOGGE folders.

Folder Name	Files	Description	Format	File Contents
MouseCapture	Mouse_Event Mouse_Move Mouse_Dragged Mouse_Wheel Mouse_Click	Mouse click events and movement tracking	.csv .csv .csv .csv .csv	Press/Release events, mouse button pressed, position (pixels), drag-and-drop detection, scroll movement, click count, Unix timestamp, frame from the webcams
Keyboard	Keyboard_Capture Keyboard_Key_Typed	Keyboard data capture and typed Unicode characters	.csv .csv	Press/Release events, key code, native key code, associated native Unicode character, Unix timestamp, frame from the webcams
MindWave	file_ATT file_MED file_BLINK file_BANDPOWER	EEG data	.csv .csv .csv .csv	Attention (0-100), meditation (0-100), blink strength, 5 EEG waves (dB), local date and time
Smartwatch	ACC Gyro Heart	Huawei Watch 2 data	.txt .txt .txt	Acceleration on X, Y, Z axes (m/s^2), angular velocity on X, Y, Z axes ($^\circ/\text{s}$), heart rate (bpm), UNIX timestamp
Fitbit	Heart EDA	FitBit Sense data	.csv .csv	Skin conductance levels, heart rate (bpm), timestamp
Eye Tracker	Learner ID	All information obtained from the eye tracker	.tsv	Recording timestamp (μs), recording date (local date of the recording), recording start time (local start time), recording resolution (pixels), gaze point (pixels), pupil diameter (mm), eye movement type, validity
VideoMonitor	monitor Videomonitor	Screen recording	.csv .mp4	Frame and timestamps recorded in both UNIX and local time
VideoWebcam	webcam WebcamCapture	Side webcam recording	.csv .mp4	Frame number and timestamps recorded in both UNIX and local time
VideoWebcam2	webcam WebcamCapture	Overhead webcam recording	.csv .mp4	Frame number and timestamps recorded in both UNIX and local time
RealSense	Time Color Depth Left_Infrared Right_Infrared	Front-facing RGB, NIR, and depth cameras	.csv .mp4 .mp4 .mp4	Frame number and timestamps recorded in both UNIX and local time for each camera
SoundCapture	Record	Session audio recording	.wav	—
PCInformation	PCInformationCapture	PC information	.csv	Computer Name, IP address, MAC address, operating system and version, system architecture, main HDD, free HDD, screen resolution, keyboard language
StudentData	StudentData	Session and learner information	.csv	Learner ID, session number, session start/end time (local time), session completion status indicator

Table 6. Description of files collected by the edBB platform and contained in the edBB folder.

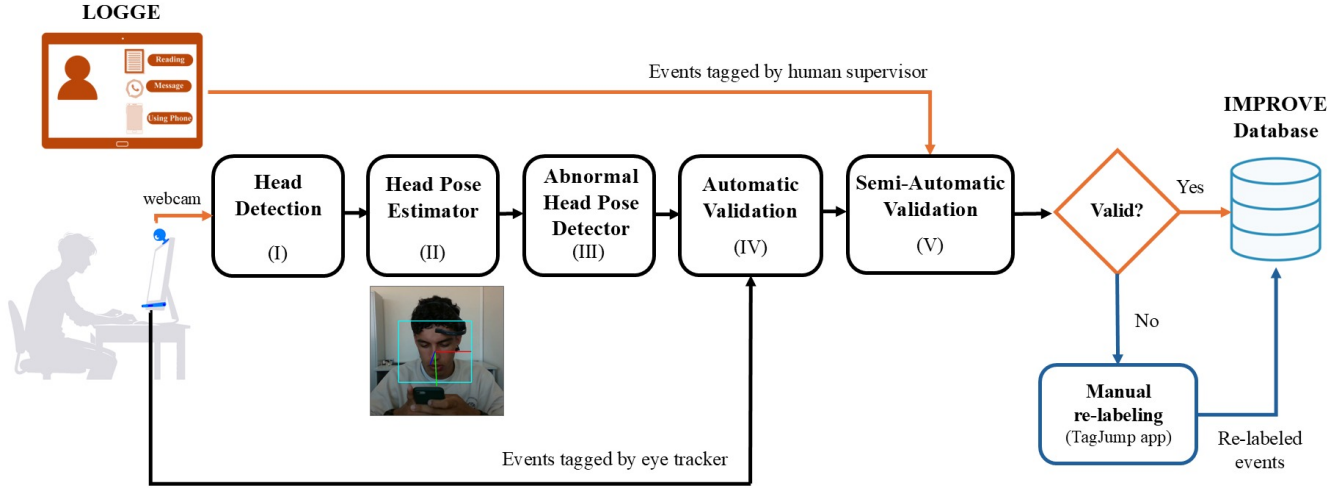


Figure 5. Diagram of the proposed semi-supervised method for curating mobile phone usage event data. The participant provided consent for the publication of this image.

Technical Validation

For the technical validation of the dataset, we conducted five analyses, organised into three subsections. The first subsection, Mobile Phone Event Labelling Accuracy, evaluates the precision of mobile phone event annotations and the changes in physiological and biometric signals during these events. It includes: (i) Semi-supervised Event Data Curation, and (ii) Statistical Validation of Signal Changes During Mobile Phone Usage Events. The second subsection, Signal Quality Assessment, assesses the overall quality and consistency of the signals recorded by the sensors through: (iii) Statistical Distributions of Physiological and Biometric Signals, and (iv) Verification of Signals and Video Recordings with M2LADS. Finally, the third subsection, Inter-Device Comparisons, analyses the consistency of heart rate estimations across different smartwatch devices: (v) Heart Rate Data Interoperability Between Devices.

Mobile Phone Event Labelling Accuracy

i) Semi-supervised Event Data Curation.

A semi-supervised method was employed for the data curation of the mobile phone usage events. To this end, we used the eye tracker data and head pose estimation through video (see Fig. 5) for the automatic detection of mobile phone usage events. These automatically annotated events were validated against the events labeled by the human supervisor, who manually labeled the start and end of each event.

The semi-supervised data curation method processes video data from the RGB webcam to detect head pose changes and moments when the learner was not looking at the computer, which with high probability, could correspond to mobile phone usage (see Fig. 5). To process the videos and detect head pose changes, we used two state-of-the-art modules based on CNNs, which are described in more detail in the Data Processing section.

- In the first module (I), the video frames were processed to detect the learner's face by selecting the most centered and widest bounding box for each frame, using MediaPipe's BlazeFace facial detector⁴⁹.
- In the second module (II), the WHENet head pose estimator⁵² was applied to obtain the Euler angles (pitch, roll, and yaw).
- In the third module (III), the method presented by Becerra et al.⁵⁷ was followed to detect abnormal head pose events. The process involves calculating the Euler angles (pitch, yaw, and roll) for each video frame. A sliding window (W_l) methodology was then applied to compute the average angles within each temporal window. An event was detected when the local average of an angle in a window exceeded a predefined threshold, which was determined by a significant deviation from the global average. This method ensures that only meaningful head pose changes are flagged as events, minimizing false positives from minor movements. The method consists of the following steps:

1. The mean (μ) and standard deviation (σ) for each Euler angle were calculated for all sessions.

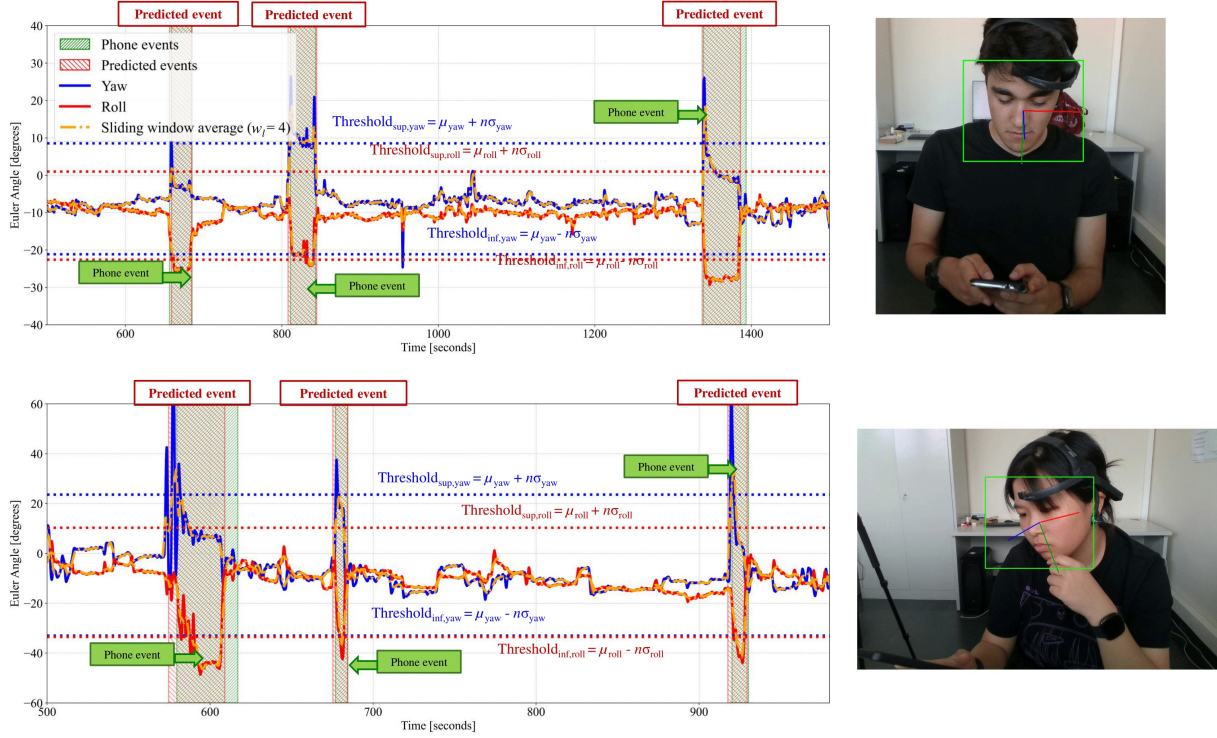


Figure 6. Example of the abnormal head pose detector for two students. The graphs show the Euler angles obtained during the session, along with the average for each angle in a window $W_l = 4s$. The set thresholds for each angle are also shown; values exceeding these thresholds are marked as events. The detector exhibits a high overlap with the manually labeled events. The images show the students' pose while using a mobile phone. Yaw is represented by the solid blue line, and roll by the solid red line. The temporal window average is indicated by the dashed yellow line. Predicted event labels are framed in red, while manually labeled events are framed in green. Thresholds are represented by dotted lines with the corresponding color for each angle. The participants provided consent for the publication of these images.

2. For each temporal window W_l , the average μ of the Euler angles was calculated, providing a representative value for pitch, yaw, and roll within that specific time frame.
3. A temporal window W_l was labeled as an event when the (μ_{W_l}) exceeded a predefined threshold, which was set based on a significant deviation from the global mean (μ) of the Euler angles calculated over all sessions. The following equation shows the threshold: $|\mu_{W_l}^i - \mu^i| > n\sigma^i$, where i represents the angle (yaw, pitch, or roll) and n is the threshold factor that defines the sensitivity of the event detection, determining how many standard deviations away from the mean a particular value must be to be considered an event.
4. All overlapping events were merged.

Figure 6 shows the detection of abnormal head pose events applied in module (III), for two learners from the IMPROVE dataset.

- In the fourth module (IV), information from the eye tracker was used to discard abnormal head pose false detections in cases where the learner continued to look at the computer screen.
- In the fifth module (V), the information from the LOGGE tool was used to determine when mobile phone usage events were labeled. This information helped select only the predicted events from our abnormal head pose detector that occurred within the time window W_s indicated by the supervisors, including a buffer $\Delta t = 10s$ before and after the event. The assumption is that manual labeling was generally accurate but involved slight delays when marking the start and end of each event. The conditions for maintaining the supervisor's labeling were as follows: the abnormal head pose

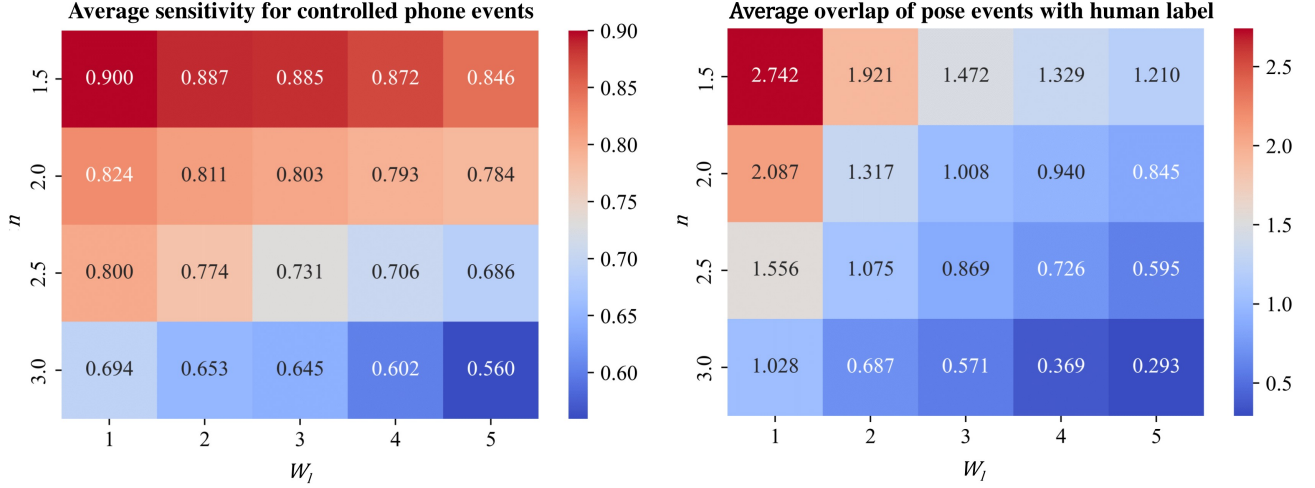


Figure 7. Average event detection sensitivity obtained by the abnormal head pose detection approach to detect mobile phone usage across all users (left). Average number of detected events by the same approach across all users within the window $W_s \pm \Delta t$ (right). The x and y axes in both heatmaps represent the tested parameters W_l and n , respectively, which were varied to analyze their effect on detection performance.

detector must report the same number of events within $W_s \pm \Delta t$ as those recorded by LOGGE, and the overlap between the predicted and manually labeled events must have a delay of less than 4 seconds. If these conditions are not met, manual re-labeling is required using the TagJump application we developed. TagJump allows displaying the video of the session at the start and end times of events reported by LOGGE and those predicted by the abnormal head pose detector. It also enables quick navigation through nearby frames to facilitate re-labeling. The protocol of this application first shows the LOGGE-labeled event, and if the event is not found within the first 4 seconds, the event predicted by the detector is displayed. This approach simplifies the supervisor's labeling process.

To fine-tune the parameters W_l and n , a preliminary study was conducted using data labeled by the LOGGE tool (see the previous subsection Data, Platforms, and Sensors), specifically focusing on the controlled events, which were presumed to have been labeled with greater precision by the supervisors. The goal was to optimize the parameters to maximize the detection of mobile phone usage events while minimizing false positives, ensuring that the predicted windows closely matched those labeled by the supervisors. Fig. 7 presents two heatmaps for controlled events. The first heatmap illustrates the event detection accuracy, while the second shows the average number of detected events within the windows $W_s \pm \Delta t$. Both are dependent on the selected parameters and the temporal window $W_s \pm \Delta t$. Based on the data analysis, we selected $W_l = 4s$ and $n = 2$ as the optimal parameters. However, this choice is not straightforward, and other combinations of parameters could also be valid depending on the specific use case.

Following this protocol, 80% of the LOGGE events were re-labeled. Only 20% had a delay of less than 4 seconds. The LOGGE labels and those detected by the abnormal pose detector complemented each other. In 24.80% of the events, when one of the labels had a delay of more than 4 seconds, the other label had a delay of less than 4 seconds. This complementarity between the two labeling systems facilitated faster re-labeling with the TagJump application. 27.2% of the re-labeled events showed a difference of less than 3 seconds, 15.2% had a difference between 3 and 4 seconds, while 57.6% showed a delay of more than 4 seconds compared to the LOGGE reports.

The re-labeled events were checked to determine whether they improved the event detection accuracy of the abnormal head pose detector. For controlled events, accuracy improved by approximately 3%, and for uncontrolled events, by around 5%, demonstrating that the detector becomes even more accurate with proper re-labeling.

ii) Statistical Validation of Signal Changes Due to Mobile Phone Usage:

The absolute mean and standard deviation of the biometric and physiological signals were calculated both during mobile phone usage events and when the mobile phone was not in use. A statistical analysis was then conducted to determine whether significant changes occurred in the signals during mobile phone usage. The results in Table 7 indicate that, according to the Student's t-test, almost all signals have p-values below 0.05, suggesting a low likelihood that the observed differences between phone and no-phone conditions are due to chance. Overall, these findings suggest that significant impacts on these variables are caused by mobile phone usage.

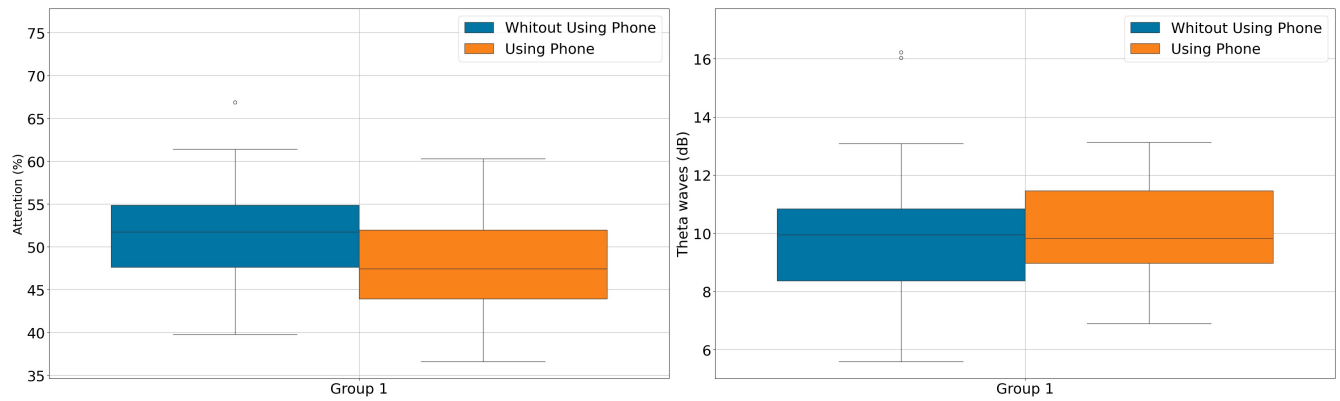


Figure 8. Boxplots comparing mobile phone usage and non-usage conditions for attention levels (left) and theta wave activity (right).

Notably, the accelerometer and gyroscope signals exhibit highly significant differences between the phone and no-phone conditions, as expected due to the increased movement associated with mobile phone use, which contrasts with typical movements during an online session. Heart rate also shows a significant difference, indicating that it is affected by mobile phone use. Most EEG signals, including attention, also show significant differences. However, the meditation, beta, and theta signals do not exhibit statistically significant differences in their mean values. Theta waves are related to relaxation and wakeful states, so it is unsurprising that the Meditation signal shows no significant differences, suggesting that relaxation or meditation is not notably affected by mobile phone usage. However, further research is required to confirm this.

Beta waves are associated with focus, normal waking consciousness, and alertness^{37,38}. For instance, beta waves occur during problem-solving, decision-making, and anxious thinking. Nevertheless, beta did not show statistically significant differences in this study.

Figure 8 presents boxplots for attention and theta waves across two experimental conditions: "Without Using Phone" and "Using Phone." Attention levels tend to be lower when the mobile phone is used, while the range of values is broader when the phone is not in use. In contrast, theta waves show a slight increase during mobile phone usage.

In summary, most signals are affected by mobile phone usage, validating the correct labeling of these events and supporting the potential to study the effects of mobile phone use through the IMPROVE dataset.

Sensors	Variable (Unit)	Mean (Std)			Statistical Tests	
		No Phone	Phone	All Sessions	t	p-value
Smartwatch	Heart Rate (bpm)	82.68 (10.3)	84.03 (10.29)	82.85 (10.25)	-2.56	0.0153
	Acc _x (m/s ²)	0.74 (0.48)	2.52 (0.99)	0.98 (0.48)	-10.82	0.0000
	Acc _y (m/s ²)	4.58 (1.55)	8.05 (1.03)	5.05 (1.39)	-12.15	0.0000
	Acc _z (m/s ²)	8.17 (1.00)	3.31 (1.25)	7.51 (0.89)	16.80	0.0000
	Gyro _x (°/s)	0.05 (0.02)	0.22 (0.09)	0.07 (0.03)	-11.25	0.0000
	Gyro _y (°/s)	0.03 (0.01)	0.09 (0.04)	0.04 (0.01)	-10.50	0.0000
	Gyro _z (°/s)	0.04 (0.01)	0.09 (0.04)	0.04 (0.02)	-9.70	0.0000
EEG Band	Attention (0-100)	51.28 (5.5)	48.47 (8.05)	50.91 (5.68)	3.03	0.0048
	Meditation (0-100)	56.33 (4.4)	56.80 (6.51)	56.39 (4.46)	-0.59	0.5626
	Alpha (dB)	6.31 (1.72)	6.85 (1.78)	6.38 (1.69)	-3.57	0.0011
	Beta (dB)	3.30 (0.95)	3.45 (0.98)	3.32 (0.94)	-2.00	0.0541
	Delta (dB)	1.69 (0.43)	1.80 (0.49)	1.70 (0.43)	-2.35	0.0250
	Gamma (dB)	1.86 (0.80)	2.03 (0.72)	1.89 (0.78)	-2.36	0.0243
	Theta (dB)	9.99 (2.36)	10.18 (2.13)	10.02 (2.31)	-0.88	0.3839

Table 7. Absolute mean and standard deviation for physiological, biometric, and movement signals across phone and no phone conditions, with t-values and p-values highlighting statistical significance, for group 1 in the IMPROVE dataset.

Sensors	Variable (Unit)	Mean (Std)			
		Group 1	Group 2	Group 3	Overall
Smartwatch	Heart Rate (bpm)	82.85 (10.25)	79.82 (13.70)	78.26 (14.30)	80.31 (12.75)
	Acc _x (m/s ²)	0.98 (0.48)	0.81 (0.58)	0.75 (0.50)	0.85 (0.52)
	Acc _y (m/s ²)	5.05 (1.39)	4.19 (1.54)	4.82 (1.39)	4.69 (1.44)
	Acc _z (m/s ²)	7.51 (0.89)	8.46 (0.94)	8.13 (0.95)	8.03 (0.92)
	Gyro _x (°/s)	0.07 (0.03)	0.04 (0.03)	0.04 (0.02)	0.05 (0.03)
	Gyro _y (°/s)	0.04 (0.01)	0.02 (0.01)	0.02 (0.01)	0.03 (0.01)
	Gyro _z (°/s)	0.04 (0.02)	0.03 (0.02)	0.03 (0.01)	0.03 (0.02)
EEG Band	Attention (0-100)	50.91 (5.68)	54.98 (4.40)	51.21 (6.54)	52.37 (5.54)
	Meditation (0-100)	56.39 (4.46)	54.25 (5.48)	54.58 (5.79)	55.07 (5.24)
	Alpha (dB)	6.38 (0.62)	5.47 (0.61)	6.23 (0.61)	6.03 (0.61)
	Beta (dB)	3.32 (0.62)	3.28 (0.58)	3.67 (0.58)	3.42 (0.59)
	Delta (dB)	1.70 (0.97)	1.59 (1.00)	1.59 (0.95)	1.63 (0.97)
	Gamma (dB)	1.89 (0.90)	1.97 (0.82)	2.06 (0.75)	1.97 (0.82)
	Theta (dB)	10.02 (0.69)	9.21 (0.71)	9.84 (0.71)	9.69 (0.70)

Table 8. Absolute mean and standard deviation for physiological and biometric signals across the different groups in the IMPROVE dataset.

Signal Quality Assessment

iii) Statistical Distributions of Physiological and Biometric Signals:

The mean and standard deviation of each biometric and physiological signal were calculated to confirm the absence of general errors and to verify that the values fell within expected ranges for a similar environment (see Table 8). These results were also compared with the mEBAL2 dataset²², which was obtained in a similar e-learning environment and utilized the same EEG band. The values obtained for the physiological and biometric signals were found to fall within the expected ranges for learners in an e-learning context. The average heart rate across the three groups ranged from 78.26 to 82.85 bpm, consistent with values reported in previous studies on students during virtual learning sessions^{28, 14}, and with the normal human resting heart rate, typically between 60-100 bpm⁵⁸.

The inertial sensors (accelerometer and gyroscope) in the Huawei Watch 2 smartwatch, worn on the hand used to operate the mouse, provided the following observations. The average acceleration in the X and Y axes was low, which is expected during a learning session where most of the time is characterized by either no movement or minor mouse movements. In contrast, the higher values on the Z axis were primarily attributed to the influence of gravitational force. Similar results were observed in the gyroscope, with mean values close to zero across all axes, reflecting the same reasons.

For attention and meditation, the means were 52.37 and 55.07, respectively, aligning with previous studies in e-learning environments, such as the mEBAL2 dataset^{8, 22}, which also report values around 50% for attention and 53.88% for meditation. These findings are in line with moderate-length sessions, where attention generally starts high but gradually decreases over time, stabilizing around the mean without extended periods of low attention¹¹. EEG signals (alpha, beta, delta, gamma, theta): The average values of the EEG frequency bands in IMPROVE are consistent and comparable to those obtained in the mEBAL2 dataset, where alpha (8.09 dB), beta (5.22 dB), delta (2.00 dB), gamma (2.59 dB), and theta (9.93 dB) are reported. In both datasets, the order of the signals according to their relative power remains constant, and the value ranges are very similar. The slight differences observed can be attributed to the varying learning activities performed in each dataset, as well as the intrinsic differences between users.

iv) Verification of Signals and Video Recordings with M2LADS

M2LADS²³ was used to validate the signals and video recordings captured during the session. Through the system's visualization and the synchronization it provides, all data were checked, and no corrupt files were reported by the system. Table 9 presents the results of the validation process carried out using M2LADS.

Sensor / Platform	Data Issues	Group 1 (IDs)	Group 2 (IDs)	Group 3 (IDs)	Count
Camera	Corrupt depth video	202303281	-	-	1
	Corrupt left NIR video	-	202305233	-	1
edX Platform	Missing edX logs	-	202303241 202303242	202305185	3
Fitbit Sense	Missing all heart rate	-	202303272 202303274	202303273	3
	Incorrect timestamp	-	202305102 202304283	-	2
Huawei Watch 2	All data missing	-	-	202304143	1
	1-minute signal loss	202303283 202305192	-	202303231	3
EEG Band	1-minute signal loss	202303281 202304272 202306161 202304184 202305311 202306075	202304202 202305103 202306141 202304115 202304181 202305092 202304183 202304282 202304284 202306055 202307271	202303291 202304212 202304125 202304131 202305055 202306074 202304143 202304283 202305042 202305051 202305231	28

Table 9. Overview of missing or corrupted data detected through the M2LADS validation, including signal losses, timestamp inconsistencies, and corrupted recordings. The information is grouped by sensor or platform and categorized according to the learners' assigned group based on mobile phone usage.

Inter-Device Comparisons

v) Heart Rate Data Interoperability between Devices:

The heart rate data from both smartwatches (Huawei Watch 2 and Fitbit Sense) were compared to ensure consistent measurements, as similar values would indicate reliable data recording. The heart rate sampling frequency differs between the two devices, with data being captured by the Huawei Watch 2 at 1 Hz, while the Fitbit Sense operates at a lower frequency of 0.2 Hz. A difference of 3.70 bpm was found between the two signals (see Fig. 9), which falls within the reasonable range for this type of device, typically around ± 3 -5 bpm⁵⁹. Additionally, it should be noted that the smartwatches were worn on different wrists, with the dominant hand experiencing more movement, which may have affected heart rate estimation. This, along with the variation in sampling frequencies, may have contributed to the observed differences.

Usage Notes

Suggestions: It is strongly recommended to use the pre-filtered EEG band files or apply appropriate filtering techniques using programs like Matlab or Python. In Python, the **SciPy**⁶⁰ library (<https://scipy.org>) provides algorithms for signal processing, including various filtering algorithms. The EEG band is often affected by artifacts caused by movement and eyeblinks, so it is recommended to apply a moving average or median filter to improve the quality of the signals⁸.

The **Pandas** library (<https://pandas.pydata.org/>) proved very useful for manipulating and analyzing large datasets, while **NumPy**⁶¹ (<https://numpy.org>) was used for processing various types of signals and videos. For training deep learning models or neural networks, **TensorFlow**⁶² (<https://www.tensorflow.org/>) and **PyTorch**⁶³ (<https://pytorch.org/>) are two of the most powerful and flexible platforms for this purpose, and were essential for running the models used in video processing.

For video processing, the **Dlib**⁶⁴ library (www.dlib.net), which includes pre-trained models for facial detection, facial

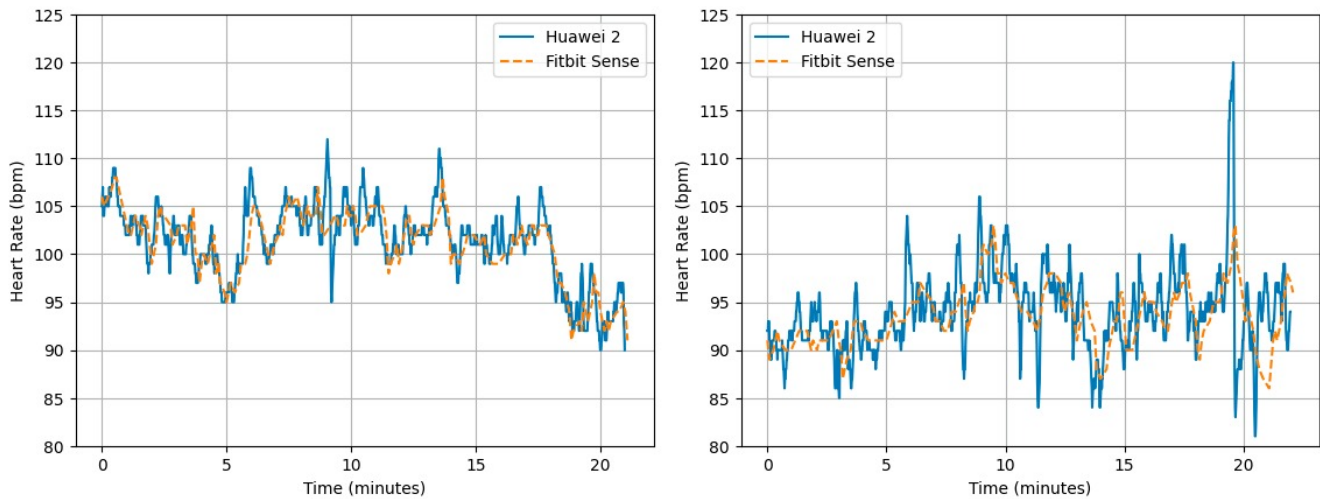


Figure 9. Comparison of heart rate measurements between the Huawei 2 and Fitbit Sense for two students. The solid blue line represents the Huawei 2, while the orange dashed line corresponds to the Fitbit Sense.

landmark estimation, and more, was used. This library integrates well with **OpenCV** (<https://opencv.org/>), which facilitates video reading, filtering, distortion correction, and also provides additional pre-trained models. This combination was used to process the videos, alongside **MediaPipe**⁶⁵ (<https://github.com/google-ai-edge/mediapipe?tab=readme-ov-file>), a library that specializes in fast models for face detection, facial landmarks, image segmentation, and more. MediaPipe was employed to extract the facial bounding boxes and landmarks available in IMPROVE dataset. The use of **RetinaFace**⁶⁶ (<https://github.com/serengil/retinaface>) is also recommended for facial detection with landmark localization.

WHENet⁵² ([www.github.com/Ascend-Research/HeadPoseEstimation-WHENet](https://github.com/Ascend-Research/HeadPoseEstimation-WHENet)) is a fast and efficient method for head pose estimation, which was used to obtain the Euler angle labels in the IMPROVE dataset. Another effective alternative for head pose estimation is **RealHePoNet**⁶⁷ (https://github.com/rafabs97/headpose_final).

For processing eye-tracker data, the official software **Tobii Pro Lab** (<https://www.tobii.com/products/software/behavior-research-software/tobii-pro-lab>) is recommended, which was used in IMPROVE. Another interesting solution is **PsychoPy**⁶⁸ (<https://www.psychopy.org/>), a free tool offering support for the creation and execution of experiments involving eye tracking.

Data Access: The IMPROVE dataset is available for research use only, including both academic and legitimate commercial research and development, provided that the data are not redistributed in any form (e.g., original files, encrypted files, or extracted features). Due to the inclusion of biometric signals and indirect identifiers (e.g., gender, age), access to the dataset is restricted and subject to the signing of a Data Usage Agreement (DUA). Researchers can request access through two available methods. First, the DUA is available on the project's GitHub (<https://github.com/BiDALab/IMPROVE>); a signed and scanned copy must be emailed to atvs@uam.es, following the instructions provided in the repository. Alternatively, access requests can be submitted via the Science Data Bank⁵⁶ (<https://doi.org/10.57760/sciencedb.14565>), where detailed application procedures are provided.

Code Availability

Example code is available on GitHub (<https://github.com/BiDALab/IMPROVE>) for filtering physiological and biometric signals, as well as for detecting pose events. The repository includes:

- A Python script for detecting abnormal head poses using previously extracted Euler angles. These angles are obtained through facial detection with MediaPipe and head pose estimation with WHENet. For full functionality, users should install the corresponding dependencies from the official GitHub repositories of these tools.
- MATLAB scripts for filtering EEG, inertial, and heart rate signals. These scripts replicate the processing applied during dataset preparation, including median filtering, moving average smoothing, and Butterworth low-pass filtering.

The provided code is intended to support the research community in processing the IMPROVE dataset. These example

scripts were used to process the raw data included in the dataset release and can serve as a reference for reproducing or extending the preprocessing pipeline.

Acknowledgements

This work has been supported by projects: HumanCAIC (TED2021-131787B-I00 MICINN), BIO-PROCTORING (GNOSS Program, Agreement Ministerio de Defensa-UAM-FUAM dated 29-03-2022), SNOLA (RED2022-134284-T), e-Madrid-CM (S2018/TCS4307). Research partially funded by the Autonomous Community of Madrid. Roberto Daza is supported by a FPI fellowship from MINECO/FEDER. A. Morales is supported by the Madrid Government (Comunidad de Madrid-Spain) under the Multiannual Agreement with Universidad Autónoma de Madrid in the line of Excellence for the University Teaching Staff in the context of the V PRICIT (Regional Programme of Research and Technological Innovation).

Author Contributions Statement

R.D.: Conceptualization, Data Collection, Data Curation, Data Analysis, Investigation, Software, Technical Validation, Writing—Original Draft, Writing—Review & Editing. **A.B.:** Data Collection, Data Curation, Software, Data Analysis, Investigation, Writing—Review & Editing. **R.C.:** Funding Acquisition, Investigation, Supervision, Writing—Review & Editing. **J.F.:** Funding Acquisition, Supervision, Writing—Review & Editing. **A.M.:** Project Administration, Funding Acquisition, Supervision, Writing—Review & Editing.

Competing interests

The authors declare no competing interests.

References

1. Daza, R. *et al.* edBB-Demo: Biometrics and Behavior Analysis for Online Educational Platforms. In *Proc. AAAI Conf. on Artificial Intelligence (Demonstration)*, [10.1609/aaai.v37i13.27066](https://doi.org/10.1609/aaai.v37i13.27066) (2023).
2. Global Market Insights. E-learning Market Size. <https://www.gminsights.com/industry-analysis/elearning-market-size> (2023).
3. Ma, L. & Lee, C. S. Investigating the Adoption of MOOCs: A Technology–User–Environment Perspective. *J. Comput. Assist. Learn.* **35**, 89–98, [10.1111/jcal.12314](https://doi.org/10.1111/jcal.12314) (2019).
4. Tindell, D. R. & Bohlander, R. W. The Use and Abuse of Cell Phones and Text Messaging in the Classroom: A Survey of College Students. *Coll. teaching* **60**, 1–9, [10.1080/87567555.2011.604802](https://doi.org/10.1080/87567555.2011.604802) (2012).
5. Huey, M. & Giguere, D. The Impact of Smartphone Use on Course Comprehension and Psychological Well-being in the College Classroom. *Innov. High. Educ.* **48**, 527–537, [10.1007/s10755-022-09638-1](https://doi.org/10.1007/s10755-022-09638-1) (2023).
6. Andrews, S., Ellis, D. A., Shaw, H. & Piwek, L. Beyond Self-report: Tools to Compare Estimated and Real-world Smartphone Use. *PloS One* **10**, e0139004, [10.1371/journal.pone.0139004](https://doi.org/10.1371/journal.pone.0139004) (2015).
7. Smith, A. Americans and Text Messaging. <https://www.pewresearch.org/internet/2011/09/19/americans-and-text-messaging/> (2011).
8. Daza, R. *et al.* DeepFace-Attention: Multimodal Face Biometrics for Attention Estimation With Application to e-Learning. *IEEE Access* **12**, 111343–111359, [10.1109/ACCESS.2024.3437291](https://doi.org/10.1109/ACCESS.2024.3437291) (2024).
9. Altmann, E. M., Trafton, J. G. & Hambrick, D. Z. Momentary Interruptions can Derail the Train of Thought. *J. Exp. Psychol. Gen.* **143**, 215, [10.1037/a0030986](https://doi.org/10.1037/a0030986) (2014).
10. Tanil, C. T. & Yong, M. H. Mobile Phones: The Effect of its Presence on Learning and Memory. *PloS One* **15**, e0219233, [10.1371/journal.pone.0219233](https://doi.org/10.1371/journal.pone.0219233) (2020).
11. Mendoza, J. S., Pody, B. C., Lee, S., Kim, M. & McDonough, I. M. The Effect of Cellphones on Attention and Learning: The Influences of Time, Distraction, and Nomophobia. *Comput. Hum. Behav.* **86**, 52–60, [10.1016/j.chb.2018.04.027](https://doi.org/10.1016/j.chb.2018.04.027) (2018).
12. Froese, A. D. *et al.* Effects of Classroom Cell Phone Use on Expected and Actual Learning. *Coll. Stud. J.* **46**, 323–332 (2012).
13. Cheever, N. A., Rosen, L. D., Carrier, L. M. & Chavez, A. Out of Sight is not Out of Mind: The Impact of Restricting Wireless Mobile Device Use on Anxiety Levels among Low, Moderate and High Users. *Comput. Hum. Behav.* **37**, 290–297, [10.1016/j.chb.2014.05.002](https://doi.org/10.1016/j.chb.2014.05.002) (2014).

14. Hernandez-Ortega, J., Daza, R., Morales, A., Fierrez, J. & Ortega-Garcia, J. edBB: Biometrics and Behavior for Assessing Remote Education. In *Proc. AAAI Workshop on Artificial Intelligence for Education* (2020).
15. Baró-Solé, X. *et al.* Integration of an Adaptive Trust-based e-assessment System into Virtual Learning Environments—The TeSLA Project Experience. *Internet Technol. Lett.* **1**, e56, [10.1002/itl2.56](https://doi.org/10.1002/itl2.56) (2018).
16. Bhattacharjee, S., Ivanova, M., Rozeva, A., Durcheva, M. & Marcel, S. Enhancing Trust in eAssessment—the Tesla System Solution. In *Proc. on Technology Enhanced Assessment* (2018).
17. Cobos, R. Self-Regulated Learning and Active Feedback of MOOC Learners Supported by the Intervention Strategy of a Learning Analytics System. *Electronics* **12**, 3368, [10.3390/electronics12153368](https://doi.org/10.3390/electronics12153368) (2023).
18. Topali, P., Cobos, R., Agirre-Uribarren, U., Martínez-Monés, A. & Villagrà-Sobrino, S. ‘instructor in action’: Co-design and evaluation of human-centred la-informed feedback in moocs. *J. Comput. Assist. Learn.* [10.1111/jcal.13057](https://doi.org/10.1111/jcal.13057) (2024).
19. Daza, R., Shengkai, L., Morales, A., Fierrez, J. & Nagao, K. SMARTe-VR: Student Monitoring and Adaptive Response Technology for e-Learning in Virtual Reality. In *Proc. AAAI Workshop on Artificial Intelligence for Education* (2025).
20. Nagao, K. Virtual Reality Campuses as New Educational Metaverses. *IEICE Transactions on Inf. Syst.* **106**, 93–100, [10.1587/transinf.2022ETI0001](https://doi.org/10.1587/transinf.2022ETI0001) (2023).
21. Zhou, Y., Suzuki, K. & Kumano, S. State-Aware Deep Item Response Theory using Student Facial Features. *Front. Artif. Intell.* **6**, 1324279, [10.3389/frai.2023.1324279](https://doi.org/10.3389/frai.2023.1324279) (2024).
22. Daza, R., Morales, A., Fierrez, J., Tolosana, R. & Vera-Rodriguez, R. mEBAL2 Database and Benchmark: Image-based Multispectral Eyeblink Detection. *Pattern Recognit. Lett.* **182**, 83–89, [10.1016/j.patrec.2024.04.011](https://doi.org/10.1016/j.patrec.2024.04.011) (2024).
23. Becerra, Á. *et al.* M2LADS: A System for Generating MultiModal Learning Analytics Dashboards in Open Education. In *Proc. Annual Computers, Software, and Applications Conference (COMPSAC) in the Workshop on Open Education Resources*, [10.1109/COMPSAC57700.2023.00241](https://doi.org/10.1109/COMPSAC57700.2023.00241) (2023).
24. Daza, R. *et al.* ALEBk: Feasibility Study of Attention Level Estimation via Blink Detection applied to e-learning. In *Proc. AAAI Workshop on Artificial Intelligence for Education* (2022).
25. Rafiqi, S. *et al.* PupilWare: Towards Pervasive Cognitive Load Measurement using Commodity Devices. In *Proc. of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, 1–8, [10.1145/2769493.2769506](https://doi.org/10.1145/2769493.2769506) (2015).
26. Intel Corporation. Intel RealSense D400 Series Datasheet. <https://www.intelrealsense.com/wp-content/uploads/2020/06/Intel-RealSense-D400-Series-Datasheet-June-2020.pdf> (2020). Accessed: 2025-06-10.
27. Hosseini, S. *et al.* A Multimodal Sensor Dataset for Continuous Stress Detection of Nurses in a Hospital. *Sci. Data* **9**, 255, [10.1038/s41597-022-01361-y](https://doi.org/10.1038/s41597-022-01361-y) (2022).
28. Hernandez-Ortega, J., Daza, R., Morales, A., Fierrez, J. & Tolosana, R. Heart Rate Estimation from Face Videos for Student Assessment: Experiments on edBB. In *Proc. of Annual Computers, Software, and Applications Conference*, 172–177, [10.1109/COMPSAC48688.2020.00031](https://doi.org/10.1109/COMPSAC48688.2020.00031) (2020).
29. Coşkun, B. *et al.* A Physiological Signal Database of Children with Different Special Needs for Stress Recognition. *Sci. Data* **10**, 382, [10.1038/s41597-023-02272-2](https://doi.org/10.1038/s41597-023-02272-2) (2023).
30. Romero-Tapiador, S. *et al.* AI4FoodDB: a Database for Personalized e-Health Nutrition and Lifestyle through Wearable Devices and Artificial Intelligence. *Database* **2023**, baad049, [10.1093/database/baad049](https://doi.org/10.1093/database/baad049) (2023).
31. Acien, A., Morales, A., Fierrez, J. & Vera-Rodriguez, R. BeCAPTCHA-Mouse: Synthetic Mouse Trajectories and Improved Bot Detection. *Pattern Recognit.* **127**, 108643, [10.1016/j.patcog.2022.108643](https://doi.org/10.1016/j.patcog.2022.108643) (2022).
32. Kirschstein, T. & Köhling, R. What is the Source of the EEG? *Clin. EEG Neurosci.* **40**, 146–149, [10.1177/155005940904000305](https://doi.org/10.1177/155005940904000305) (2009).
33. Hall, J. E. & Hall, M. E. (eds.) *Guyton and Hall Textbook of Medical Physiology e-Book* (Elsevier, 2020).
34. Chen, C.-M. & Wang, J.-Y. Effects of Online Synchronous Instruction with an Attention Monitoring and Alarm Mechanism on Sustained Attention and Learning Performance. *Interact. Learn. Environ* **26**, 427–443, [10.1080/10494820.2017.1341938](https://doi.org/10.1080/10494820.2017.1341938) (2018).
35. Li, Y. *et al.* A Real-Time EEG-based BCI System for Attention Recognition in Ubiquitous Environment. In *Proc. Intl. Workshop on Ubiquitous Affective Awareness and Intelligent Interaction*, 33–40, [10.1145/2030092.2030099](https://doi.org/10.1145/2030092.2030099) (2011).
36. Daza, R. *et al.* MATT: Multimodal Attention Level Estimation for e-learning Platforms. In *Proc. AAAI Workshop on Artificial Intelligence for Education* (2023).

37. Lin, F.-R. & Kao, C.-M. Mental Effort Detection using EEG Data in E-learning Contexts. *Comput. & Educ.* **122**, 63–79, [10.1016/j.compedu.2018.03.020](https://doi.org/10.1016/j.compedu.2018.03.020) (2018).
38. Chen, C.-M., Wang, J.-Y. & Yu, C.-M. Assessing the Attention Levels of Students by using a Novel Attention Aware System Based On Brainwave Signals. *Br. J. Educ. Technol.* **48**, 348–369, [10.1111/bjet.12359](https://doi.org/10.1111/bjet.12359) (2017).
39. Li, X., Hu, B., Zhu, T., Yan, J. & Zheng, F. Towards Affective Learning with an EEG Feedback Approach. In *Proc. of the First ACM International Workshop on Multimedia Technologies for Distance Learning*, 33–38, [10.1145/1631111.1631118](https://doi.org/10.1145/1631111.1631118) (2009).
40. Daza, R., Morales, A., Fierrez, J. & Tolosana, R. mEBAL: A Multimodal Database for Eye Blink Detection and Attention Level Estimation. In *Proc. Intl. Conf. on Multimodal Interaction*, 32–36, [10.1145/3395035.3425257](https://doi.org/10.1145/3395035.3425257) (2020).
41. Bagley, J. & Manelis, L. Effect of Awareness on an Indicator of Cognitive Load. *Percept. Mot. Ski.* **49**, 591–594, [10.2466/pms.1979.49.2.591](https://doi.org/10.2466/pms.1979.49.2.591) (1979).
42. Sharma, K., Alavi, H. S., Jermann, P. & Dillenbourg, P. A Gaze-based Learning Analytics Model: in-video Visual Feedback to Improve Learner's Attention in MOOCs. In *Proc. Intl. Conference on Learning Analytics & Knowledge*, 417–421, [10.1145/2883851.2883902](https://doi.org/10.1145/2883851.2883902) (2016).
43. Sharma, K., D'Angelo, S., Gergle, D. & Dillenbourg, P. Visual Augmentation of Deictic Gestures in MOOC Videos. In *Proc. Intl. Conference of the Learning Sciences*, 202–209, [10.22318/icls2016.28](https://doi.org/10.22318/icls2016.28) (2016).
44. Andreu-Perez, J., Solnais, C. & Sriskandarajah, K. EALab (Eye Activity Lab): a MATLAB Toolbox for Variable Extraction, Multivariate Analysis and Classification of Eye-Movement Data. *Neuroinformatics* **14**, 51–67, [10.1007/s12021-015-9275-4](https://doi.org/10.1007/s12021-015-9275-4) (2016).
45. Navarro, M. *et al.* VAAD: Visual Attention Analysis Dashboard Applied to e-Learning. In *Proc. Intl. Symposium on Computers in Education (SIIE), IEEE*, 1–6, [10.1109/SIIE63180.2024.10604520](https://doi.org/10.1109/SIIE63180.2024.10604520) (2024).
46. Morales, A. *et al.* KBOC: Keystroke Biometrics Ongoing Competition. In *Proc. Intl. Conf. on Biometrics Theory, Applications and Systems*, 1–6, [10.1109/BTAS.2016.7791180](https://doi.org/10.1109/BTAS.2016.7791180) (2016).
47. Morales, A. *et al.* Keystroke Biometrics Ongoing Competition. *IEEE Access* **4**, 7736–7746, [10.1109/ACCESS.2016.2626718](https://doi.org/10.1109/ACCESS.2016.2626718) (2016).
48. Becerra, A., Daza, R., Cobos, R., Morales, A. & Fierrez, J. M2LADS Demo: A System for Generating Multimodal Learning Analytics Dashboards. In *Proc. AAAI Workshop on Innovation and Responsibility in AI-Supported Education (iRAISE)* (2025).
49. Bazarevsky, Valentin and Kartynnik, Yury and Vakunov, Andrey and Raveendran, Karthik and Grundmann, Matthias. Blazeface: Sub-millisecond Neural Face Detection on Mobile GPUs. In *Proc. CVPR Workshop on Computer Vision for Augmented and Virtual Reality* (2019).
50. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. Mobilenetv2: Inverted Residuals and Linear Bottlenecks. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 4510–4520, [10.1109/CVPR.2018.00474](https://doi.org/10.1109/CVPR.2018.00474) (2018).
51. Duan, K. *et al.* Centernet: Keypoint Triplets for Object Detection. In *Proc. of the IEEE/CVF international conference on computer vision*, 6569–6578, [10.1109/ICCV.2019.00667](https://doi.org/10.1109/ICCV.2019.00667) (2019).
52. Zhou, Yijun and Gregson, James. Whenet: Real-time Fine-Grained Estimation for Wide Range Head Pose. In *Proc. of the British Machine Vision Conference* (2020).
53. Tan, Mingxing and Le, Quoc. Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proc. of the International Conference on Machine Learnin*, 6105–6114 (2019).
54. Zhu, Xiangyu and Lei, Zhen and Liu, Xiaoming and Shi, Hailin and Li, Stan Z. Face Alignment across Large Poses: A 3d Solution. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 146–155, [10.1109/CVPR.2016.23](https://doi.org/10.1109/CVPR.2016.23) (2016).
55. Joo, Hanbyul and Liu, Hao and Tan, Lei and Gui, Lin and Nabbe, Bart and Matthews, Iain and Kanade, Takeo and Nobuhara, Shohei and Sheikh, Yaser. Panoptic Studio: A Massively Multiview System for Social Motion Capture. In *Proc. of the IEEE International Conference on Computer Vision*, 3334–3342, [10.1109/ICCV.2015.381](https://doi.org/10.1109/ICCV.2015.381) (2015).
56. Daza, R., Becerra, A., Cobos, R., Fierrez, J. & Morales, A. IMPROVE dataset. *Science Data Bank* [10.57760/sciencedb.14565](https://doi.org/10.57760/sciencedb.14565) (2024). Accessed: 2025-06-10.

57. Becerra, A. *et al.* Biometrics and Behavior Analysis for Detecting Distractions in e-Learning. In *Proc. Intl. Symposium on Computers in Education (SIIE)*, IEEE, 1–6, [10.1109/SIIE63180.2024.10604582](https://doi.org/10.1109/SIIE63180.2024.10604582) (2024).
58. Palatini, P. & Julius, S. Heart Rate and the Cardiovascular Risk. *J. Hypertens.* **15**, 3–17, [10.1097/00004872-199715010-00001](https://doi.org/10.1097/00004872-199715010-00001) (1997).
59. Hwang, J. *et al.* Assessing Accuracy of Wrist-Worn Wearable Devices in Measurement of Paroxysmal Supraventricular Tachycardia Heart Rate. *Korean Circ. J.* **49**, 437–445, [10.4070/kcj.2018.0323](https://doi.org/10.4070/kcj.2018.0323) (2019).
60. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272, [10.1038/s41592-020-0772-5](https://doi.org/10.1038/s41592-020-0772-5) (2020).
61. Harris, C. R. *et al.* Array Programming with NumPy. *Nature* **585**, 357–362, [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2) (2020).
62. Abadi, M. *et al.* {TensorFlow}: a System for {Large-Scale} Machine Learning. In *Proc. of the Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283 (2016).
63. Paszke, A. *et al.* Pytorch: An Imperative Style, High-Performance Deep Learning Library. In *Proc. of Advances in Neural Information Processing Systems*, vol. 32, 8024–8035 (2019).
64. King, D. E. Dlib-ml: A Machine Learning Toolkit. *The Journal of Machine Learning Research* **10**, 1755–1758 (2009).
65. Lugaresi, C. *et al.* MediaPipe: A Framework for Building Perception Pipelines. *arXiv preprint* [10.48550/arXiv.1906.08172](https://arxiv.org/abs/10.48550/arXiv.1906.08172) (2019).
66. Deng, J. *et al.* RetinaFace: Single-stage Dense Face Localisation in the Wild. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5203–5212 (2020).
67. Berral-Soler, R., Madrid-Cuevas, F. J., Munoz-Salinas, R. & Marín-Jiménez, M. J. RealHePoNet: a Robust Single-Stage ConvNet for Head Pose Estimation in the Wild. *Neural Comput. Appl.* **33**, 7673–7689, [10.1007/s00521-020-05511-4](https://doi.org/10.1007/s00521-020-05511-4) (2021).
68. Peirce, J. W. PsychoPy—psychophysics Software in Python. *J. Neurosci. Methods* **162**, 8–13, [10.1016/j.jneumeth.2006.11.017](https://doi.org/10.1016/j.jneumeth.2006.11.017) (2007).