ARTICLE

Keep Me in the Loop: Real-Time Feedback with Multimodal Data

Daniele Di Mitri[1] ⬡ · **Jan Schneider[1]** · **Hendrik Drachsler[1,2]**

Abstract

This paper describes the CPR Tutor, a real-time multimodal feedback system for cardiopulmonary resuscitation (CPR) training. The CPR Tutor detects training mistakes using recurrent neural networks. The CPR Tutor automatically recognises and assesses the quality of the chest compressions according to five CPR performance indicators. It detects training mistakes in real-time by analysing a multimodal data stream consisting of kinematic and electromyographic data. Based on this assessment, the CPR Tutor provides audio feedback to correct the most critical mistakes and improve the CPR performance. The mistake detection models of the CPR Tutor were trained using a dataset from 10 experts. Hence, we tested the validity of the CPR Tutor and the impact of its feedback functionality in a user study involving additional 10 participants. The CPR Tutor pushes forward the current state of the art of real-time multimodal tutors by providing: (1) an architecture design, (2) a methodological approach for delivering real-time feedback using multimodal data and (3) a field study on real-time feedback for CPR training. This paper details the results of a field study by quantitatively measuring the impact of the CPR Tutor feedback on the performance indicators and qualitatively analysing the participants' questionnaire answers.

✉ Daniele Di Mitri
dimitri@dipf.de

Jan Schneider
schneider.jan@dipf.de

Hendrik Drachsler
drachsler@dipf.de

[1] DIPF - Leibniz Institute for Research and Information in Education, Rostocker Straße 6 60323 Frankfurt am Main, Germany

[2] Open University of The Netherlands, Valkenburgerweg 177, 6419 AT Heerlen, Netherlands

Introduction

In learning science, there is an increasing interest in collecting and integrating data from multiple modalities and devices to analyse learning behaviour (Blikstein & Worsley, 2016; Ochoa & Worsley, 2016; Cukurova et al., 2020). This phenomenon is witnessed by the rise of multimodal data experiments especially in the contexts of project-based learning (e.g. Spikol et al. 2018), lab-based experimentation for skill acquisition (e.g. Giannakos et al. 2019, Worsley & Blikstein 2018), young children learning (Crescenzi-Lanna, 2020), game-based learning (Emerson et al., 2020) and simulations for mastering psychomotor skills (Santos, 2019; Martinez-Maldonado et al., 2018). Most of the existing studies using multimodal data for learning stand at the level of "data geology", i.e. the tendency as in geology science to measure phenomena at a superficial level and consequently try to infer deeper level dynamics: in this case, using the multimodal data as evidence of deeper level cognitive processes or other learning outcomes.

In some cases, related studies trained machine learning models with the collected data for classifying or predicting outcomes such as emotions (D'Mello & Kory, 2015) or learning performance (e.g. Olsen et al. 2020). The existing research that uses multimodal and multi-sensor systems for training different types of psychomotor skills features neither personalised nor adaptive feedback (Santos, 2016).

In this study, we aimed at filling this gap by *exploring how we can use multimodal data to support psychomotor skill development by providing real-time feedback*. We implemented the CPR Tutor using a *design-based research* approach. This study is based on the insights of (Di Mitri et al., 2019b) where we introduced a detection system for CPR mistakes related to chest compressions (CC). Such a system monitors standard CPR mistakes typically tracked by commercial CPR training tools like the Laerdal ResusciAnne manikin[1]. The mistakes related to (1) incorrect speed of the chest compressions (class *classRate*), (2) incorrect depth of the compression (class *classDepth*) and (3) incomplete release of the hands from the manikin at the end of the compression (class *classRelease*). Additionally, in the preliminary study (Di Mitri et al., 2019b), we showed that it is possible to extend the standard mistake detection with the CPR tutor. We were able to detect the correct locking of the arms (class *armsLocked*) and the proper use of the body weight when performing the CCs (class *bodyWeight*). We obtained the mistake-detection models by training multiple long-short term memory networks (LSTMs), using the multimodal data as input and the presence or absence of the CPR mistakes as output (Di Mitri et al., 2019b).

In the conference publication (Di Mitri et al., 2020) we outlined the real-time mistake detection and feedback intervention system of the CPR tutor. The contribution of this paper is to extend on the previously presented work by providing the reader with additional background information on the CPR Tutor, including (1) additional references to related works, (2) further analysis on the relevance and impact of the feedback provided, (3) the results of the participants' questionnaire of the feedback

---

[1] https://laerdal.com/products/simulation-training/resuscitation-training/resusci-anne-qcpr/

study, (4) further discussions on limitations, and (5) future optimisations that we can consider for the further development of the CPR Tutor. In line with the previous contributions, the purpose of this study is to illustrate the case of the CPR Tutor, a prototypical multimodal system that can *close the feedback loop* by providing automatic support to the learner practising a physical task in the medical domain.

Background

Multimodal Data for Learning

With the term "multimodal data", we refer to the data sources derived from multimodal and multi-sensor interfaces that go beyond the typical mouse and keyboard interactions (Oviatt et al., 2018). These data sources can be collected using wearable sensors, depth cameras or Internet of Things devices. The promise of multimodal data for learning lies in providing additional empirical evidence of the learning process by expanding the visible area (Di Mitri et al., 2018). The multimodal data can enrich the learner model and task description and therefore provide better adaptation and personalisation. For instance, learning technologies can employ modern sensor devices to collect evidence of the learner's motoric movements, the physiological responses such as heart rate or electrodermal activation, or contextual information related to the learning environment or activity the learner is performing and connected interaction with physical and digital objects (Di Mitri et al., 2018). The exploration of these novel data sources inspired the Multimodal Learning Analytics (MMLA) research (Ochoa & Worsley, 2016). MMLA focuses on the collection and analysis of various traces obtained from different aspects of learning processes for better understanding and improving those processes (Chan et al., 2020).

The MMLA common hypothesis is that combining and analysing data from multiple modalities can provide valuable insights to the educational actors, informing them about the learning dynamics and supporting them to design more valuable feedback (Blikstein & Worsley, 2016). Researchers have found out that using an MMLA approach can better predict learning performance during desktop-based game playing (Giannakos et al., 2019). The MMLA approach is also thought to be useful for assessing complex and more ill-defined learning constructs such as creativity or empathy or social awareness (Cukurova et al., 2019) to investigate their efficacy. Recent MMLA prototypes have been developed for modelling classroom interactions (Ahuja et al., 2019; Prieto et al., 2018) or for estimating success in group collaboration (Spikol et al., 2018; Olsen et al., 2020).

Nevertheless, the contribution of multimodal data to learning and MMLA is still a research topic under exploration. Learning is a complex multidimensional process: in the taxonomy proposed by Bloom et al., learning spans across cognitive, affective and psychomotor domains (Bloom, 1956). In this study, we are interested in the support that multimodal data can provide to the psychomotor domain of learning, i.e., acquiring mental models that lead toward mastering practical skills. This idea has been recently reformulated as "embodied learning", the notion that learning and skill acquisition are grounded in the body and the environment in which it is

operating (Juntunen, 2020). Multimodal data were also employed to design MMLA for psychomotor tasks and physical learning activities that require complex body coordination (Martinez-Maldonado et al., 2018). Santos reviewed existing studies using sensor-based applications in diverse psychomotor disciplines for training specific movements in different sports and martial arts (2019). Limbu et al. reviewed current studies that modelled the experts to train apprentices using recorded expert performance (2018b).

Multimodal Intelligent Tutors

We are interested in the application of multimodal data for providing automatic and real-time feedback for psychomotor tasks. This aim is pursued by the Intelligent Tutoring Systems (ITS) and Artificial Intelligence in Education (AIED) research. Historically, ITSs have been designed for highly structured learning activities. The task sequence is clearly defined, and the assessment criteria and the range of learning mistakes that ITS can detect. Related ITS research looked primarily at the cognitive domain, in terms of estimation and assessment of learner's prior knowledge (Koedinger & Corbett, 2006) as well as affective domains of learning, such as the detection of learners' emotional states (e.g. D'Mello et al. 2008, Arroyo et al. 2009). ITSs have been used in combination with MMLA for emotions and idea improvement analysis (Zhu et al., 2019), as well as detection of mind-wandering (Hutt et al., 2019). A recent literature review presents a comprehensive analysis of ITSs (Alqahtani & Ramzan, 2019).

ITSs and, more generally, educational technologies focused primarily on desktop-based systems where the learners interact with a computer-based learning platform or system. However, especially in practical skills learning, the learning interactions are not confined within the boundaries of the learning platform. Still, they are distributed across a variety of locations, contexts or devices across physical and digital spaces (Martinez-Maldonado et al., 2019). There exist some examples *multimodal* ITSs in the psychomotor domain of learning which goes beyond "mouse and keyboard". In medical robotics and surgical simulations, various training robots and ITSs have been developed to allow aspiring surgeons to train surgical skills in safe environments (Taylor et al., 2016). The surgical training scenario is distinguished from other medical skills for its high relevance and a high degree of structure. Similar systems have been designed for general skills assessment in surgery (Levin et al., 2019). For example, they produce adaptive feedback orthopaedic surgery using haptic interfaces (Luengo & Mufti-Alchawafa, 2013) or within virtual reality during temporal-bone surgery (Davaris et al., 2019).

Outside of the medical field, Multimodal ITSs have been developed for training 21st-century skills. For example, the Presentation Trainer (Schneider et al., 2015) provides real-time and retrospective feedback about posture and prosody while people are delivering a presentation. The Calligraphy Tutor is a tablet-based application whose purpose is to train people to write in a foreign alphabet using a capacitive pen (Limbu et al., 2018a). In addition to 21st-century skills, multimodal ITSs were developed for training sport disciplines. The KUMITRON system (Echeverria & Santos, 2021) uses physiological signals and computer vision to monitor

the performance in Karate combat. The Table Tennis Tutor (Mat Sanusi et al., 2021) captures motion data and tracks the skeletal points of the body to observe the execution of fundamental table tennis strokes. Finally, the system proposed by (Santos & Corbí, 2019) introduces an ITS to learn physics concepts with Aikido movements.

Real-time Feedback Architectures

To maximise muscles memory and shorten training time, embodied learning tasks require repeated practice and *real-time* feedback. Contrarily to delayed (retrospective) feedback, the real-time feedback is provided *just-in-time* for steering the learner towards the ideal learning performance by preventing and modifying incorrect movements and reinforcing correct executions. The feedback approaches may consist of nudges, prompts or other forms of automatic intervention. In related literature, automated feedback provision systems have been investigated from various perspectives. In the autonomous agents and engineering literature, a feedback system often refers to the interaction of two (or more) dynamical systems connected and influencing each other. Feed-forward systems can use sets of sensors and actuators to sense and manipulate external environmental conditions. In these systems, the human is seen as one of the agents in the feedback loop (Åström & Murray, 2021). Education science and psychology focus instead on the human notion of feedback, as the information is given to a person to modify her behaviour. Research in multimodal interaction seeks to combine both the engineering and the human learning side of feedback. It aims to design reliable processing systems that can understand multiple communication means in real-time (Dumas et al., 2009). The proposed architectures combine human-agent verbal interaction with other multimodal behavioural cues, such as pointing fingers to near objects to achieve naturalistic interactions with situated robots or virtual avatars (e.g. Krishnaswamy & Pustejovsky 2019). The CPR Tutor, in its first iteration, implements user interaction via movement detection and assessment rather than with verbal exchange. Similarly to other real-time multimodal feedback architectures, the main challenges are the fusion of different types of data and the real-time processing and temporal constraints imposed on information processing (Dumas et al., 2009). The closest example of a real-time feedback architecture based on the Kinect v2 sensor is the Presentation Trainer (Schneider et al., 2015) (PT). The PT uses a rule-based approach for detecting specific movements. The feedback is calculated instantaneously and shown to the user visually on the screen. As the authors of the PT pointed out, it is more beneficial and less confusing for the user to fire one feedback message at a time with sufficient lags in between to prevent cognitive overloading. This feedback approach is naturalistic, as it mimics the way how a human coach would give feedback. The feedback design of the CPR Tutor got inspired by the PT.

Cardiopulmonary Resuscitation (CPR)

This study focuses on one of the most frequently applied and well studied medical simulations: Cardiopulmonary Resuscitation. CPR is a lifesaving technique used in

many emergencies, including a heart attack, near drowning or in the case of stopped heartbeat or breathing. CPR is mandatory for healthcare professionals and several other professions, especially those developing exposed to the general public. CPR training is an individual learning task with a highly standardised procedure consisting of a series of predefined steps and criteria to measure the quality of the performance. We refer to the European CPR Guidelines (Perkins et al., 2015).

There are various commercial tools for supporting CPR training, which can track and assess CPR execution. One popular choice is the Laerdal ResusciAnne QCPR manikin[2]. ResusciAnne manikins can be extended with feedback devices: (1) the SkillReporter app, which provides a real-time dashboard that counts the repetitions and the duration of the CCs and visualises in real-time ventilation scores, compression rate, depth and release; (2) the SimPad PLUS, used in this study, include the functionalities of the SkillReport and collects data for performance scoring and evaluation; (3) the TeamReporter app which is designed for CPR team training. Another commercial CPR manikin is the Innosonian Brayden Pro[3] implements light-based visual feedback on the manikin. It also checks the repetitions and duration of the CCs, ventilation rate, compression rate, depth, release and correct hand positioning. As an alternative to human-shaped manikins, CPR training solutions also include devices such as the True CPR Coaching Device[4]. The TrueCPR offers real-time feedback during CCs and recommendations for improvement after the CPR session. It measures compression depth and provides metronome sounds for the correct compression rate. The study from Dick-Smith et al. (Dick-Smith et al., 2020) compared the three commercial devices just described looking at which modality combination (among visual, audio and embodied) support best undergraduate nursing students to acquire CPR skills. The combination of visual and audio of the TrueCPR appeared to be the most effective.

In commercial CPR training tools, we noticed that some CPR performance indicators are neglected. We assumed this choice is due to the need to downscaling the cost of the CPR training devices. Examples of these indicators are the use of the body weight or the locking of the arms while doing the CCs. So far, these mistakes need to be corrected by human instructors creating a feedback gap for the learner and higher responsibility for the course instructors.

In our study, it is essential to point out that we focused on one specific aspect of the CPR training procedure in developing the CPR Tutor: performing correct chest compressions. We did not cover other relevant elements of CPR such as rescue ventilation, preparation before the CPR or use of the Automated External Defibrillator. The reason for this choice was that CPR Tutor, in the stage of development presented, was not aimed to become a comprehensive CPR training tool, but rather proof the concept of a new CPR technology that enables real-time feedback provision using multimodal data and machine learning methods.

---

[2] https://laerdal.com/us/products/simulation-training/resuscitation-training/resusci-anne-qcpr/

[3] https://www.innosonian.us/brayden-pro/

[4] https://www.strykeremergencycare.com/products/devices/truecpr/

System Architecture of the CPR Tutor

The System Architecture of the CPR Tutor implements the five-step approach introduced by the *Multimodal Pipeline* (Di Mitri et al., 2019c) a framework for the collection, storing, annotation, processing and exploitation of data from multiple modalities. We optimised the System Architecture to the selected sensors and for the specific task of CPR training. The five steps, proposed by the *Multimodal Pipeline* are numbered in the graphical representation of the System Architecture in Fig. 1. The architecture also features three layers: (A) the Presentation Layer interfacing with the user (either the learner or the expert); (B) the Application Layer, implementing the logic of the CPR Tutor; (C) the Data Layer, consisting of the data used by the CPR Tutor. In the CPR Tutor, we can distinguish two main phases, which have two corresponding data-flows: (1) the *offline training* of the machine learning models and (2) the *real-time exploitation* in which the real-time feedback system is activated. Following the framework for psychomotor learning systems proposed by Santos (2016), the offline training phase includes the *movement modelling*, i.e. the training of machine learning models that detect training mistakes. The offline phase also includes the *feedback design*, i.e. the rules and the priority by which the feedback is fired. The real-time exploitation phase includes the *sensing movements* instead, i.e. detecting when a CC occurs, and of the *delivering of the feedback*, i.e. the right timing and form of the feedback prompts.
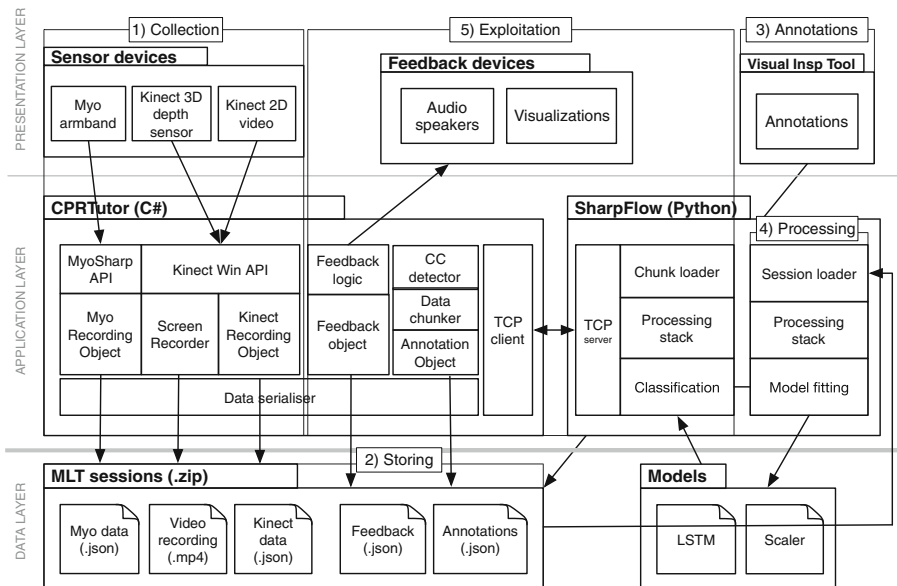

Fig. 1 The System Architecture of the CPR Tutor

Data Collection

The first step corresponds to the collection of the dataset. The main system component responsible for the data collection is the CPR Tutor, a C# application running on a Windows 10 computer. The CPR Tutor collects data from two main devices: (1) the Microsoft Kinect v2 depth camera[5] and (2) the Myo electromyographic (EMG) arm-band[6]. In the graphic user interface, the user of the CPR Tutor can 'start' and 'stop' the recording of the session. The CPR Tutor collects the data of the user in front of the camera wearing the Myo. The collected data consist of:

– the 3D kinematic data (x,y,z) of the body joints (excluding ankles and hips)
– the 2D video recording from the Kinect RGB camera
– 8 EMG sensors values, 3D gyroscope and accelerometer of the Myo.

Data Storing

The CPR Tutor adopts the data storing logic of the *Multimodal Learning Hub* (Schneider et al., 2018), a core component of the *Multimodal Pipeline*. As the sensor applications collect data at different frequencies, at the 'start' of the session, each sensor application is assigned to a *Recording Object* a data structure arbitrary number of *Frame Updates*. In the case of the CPR Tutor, two main streams are coming from the Myo and the Kinect. The *Frame Updates* contain the relative timestamp starting from the moment the user presses the 'start' until the 'stop' of the session. Each *Frame Update* within the same *Recording Object* shares the same set of sensor attributes, in the case of the CPR Tutor, 8 attributes for Myo and 32 for Kinect, corresponding to the raw features that can be gathered from the public API of the devices. The video stream recording from the Kinect uses a particular type of *Recording Object*, specific for video data. At the end of the session, when the user presses 'stop', the data gathered in memory in the *Recording Objects* and the *Annotation Object* is automatically serialised into the custom format introduced by the LearningHub: the *MLT-JSON* session (Meaningful Learning Task). For the CPR Tutor, the custom data format consists of a zip folder containing: the Kinect and Myo sensor file and the 2D video in MP4 format. Serialising the sessions is necessary for creating the dataset for the offline training of the machine learning models. The advantage of using the LearningHub is that the video and the sensor data are automatically synchronised.

Data Annotation

The annotation can be carried out by one researcher retrospectively using the *Visual Inspection Tool* (VIT) (Di Mitri et al., 2019a). In the VIT, the researcher can load the *MLT Session* files one by one to triangulate the video recording with the sensor data. The researcher can select and plot individual data attributes and inspect

---

[5]https://developer.microsoft.com/en-us/windows/kinect/

[6]https://developerblog.myo.com/

visually how they relate to a video recording. The VIT is also a tool for collecting annotations by the researcher. The annotations are given as properties of every time interval, which corresponds to a single chest compression in the case of CPR Tutor. Normally, to add an annotation, the researcher needs first to create a time interval by segmenting the plot of the data attributes. Then, the researcher can select one or multiple time intervals and assign an attribute name and a corresponding value, e.g. *ArmsLocked* = 1, meaning that the arms were correctly locked in the selected CCs. The process of segmenting all recordings in time intervals is time-consuming for the researcher. In the case of CPR Tutor, this problem was mitigated with a semi-automatic approach. From the SimPad of the ResusciAnne manikin, we extracted the performance results of each recorded session. With a Python script, we processed the data from the ResusciAnne manikin in a JSON annotation file, which we added to each recorded session using the VIT. This procedure allowed us to have the performance metrics of the ResusciAnne manikin as "ground truth" for training the classifiers for the expert dataset. As previously mentioned, the Simpad tracks the chest compression performance monitoring three indicators, the correct compression rate *classRate*, correct release *classRelease* and depth of the compression *classDepth*. However, the researcher can extend these indicators by adding custom annotations in attribute-value pairs by using the VIT. For this study, we use the target custom classes *armsLocked* and *bodyWeight* corresponding to two performance indicators, currently not tracked by the ResusciAnne manikins. The expert dataset collected in the first phase of the study and the feedback intervention dataset collected in the second phase preserves the same annotation scheme, five annotations per CC corresponding to the five target classes. However, these two datasets differ in the way the CCs were segmented. In the expert dataset, the segmentation was derived from the SimPad of the ResusciAnne manikin and the compression rate, release and depth used for ground truth as the model training. In the feedback intervention dataset, the segmentation of the CCs was executed automatically by the CPR Tutor monitoring the vertical movements of the shoulder joints as explained in section 2. We took this design choice to prove that the CPR Tutor can work independently from the ResusciAnne manikin and virtually on any CPR manikin. Although the CPR Tutor CC segmentation was calibrated to resemble one of the ResusciAnne manikins as much as possible, there is no exact correspondence between the two sets of annotations. For this reason, it was not possible to compare the classification results of the neural networks to the baseline of ResusciAnne manikin for the feedback intervention dataset.

Data Processing

For data processing, we developed a Python script named SharpFlow[7]. This component is used for offline training and validation of the mistake detection classifiers and the real-time classification of the single CCs. In the training phase, the entire

---

[7]Code available on GitHub (https://github.com/dimstudio/SharpFlow)

dataset (MLT-JSON sessions with their annotations) is loaded into memory and transformed into two Pandas data frames, one containing the sensor data and the other containing the annotations. The sensor data came from devices with different sampling frequencies, and the sensor data frame had many missing values. To mitigate this problem, we resampled the data frame into a fixed number corresponding to the median length of each sample. We obtained, therefore, a 3D tensor of shape (#samples × #attributes × #intervals). The dataset was divided into 85% for training and 15% for testing using random shuffling. A part of the training set (15%) was used as the validation set. We also applied feature scaling using min-max normalisation with a range of -1 and 1. The scaling was fitted on the training set and applied to the validation and test sets. The model used for classification was a Long-Short Term Memory network (Hochreiter & Schmidhuber, 1997) (LSTM), a particular type of recurrent neural network. The choice of LSTMs was justified because these networks are proven to be better suited to capture temporal dependencies that are peculiar for the kinematic and myographic data collected by the CPR Tutor (Vohra et al., 2015). Alternative supervised machine learning models such as Random Forests or Support Vector Machines treat the learning samples as independent and identically distributed data samples, disregarding the observations' time-dependency. Besides neural networks, another class of time-sensitive models suited for supervised classification is the Hidden Markov Models (HMMs). These models are not specifically designed for machine learning but need proportionally fewer data samples to converge; therefore are better suited for supervised learning from limited data samples. However, in the CPR Tutor use case, the HMMs turned out to be not as flexible as the neural networks in dealing with the high-dimensional multivariate dataset.

Implementation was performed using PyTorch. The architecture of the model (reported in Fig. 2) chosen was a sequence of two stacked LSTM layers followed by two dense layers:

- a first LSTM with input shape 17x52 (#intervals × #attributes) and 128 hidden units
- a second LSTM with 64 hidden units
- a fully-connected layer with 32 units with a sigmoid activation function
- a fully connected layer with 5 hidden units (number of target classes)
- a sigmoid activation.

Our classes have a binary type, so we use a binary cross-entropy loss for optimisation and train for 30 epochs using an Adam optimiser with a learning rate of 0.01.

Real-time Exploitation

Real-time data exploitation is the run-time behaviour of the System Architecture of the CPR Tutor. This phase is a continuous communication loop between the CPR Tutor, the SharpFlow application, and the feedback prompting. We can summarise it in three steps (1) detection, (2) classification and (3) feedback.
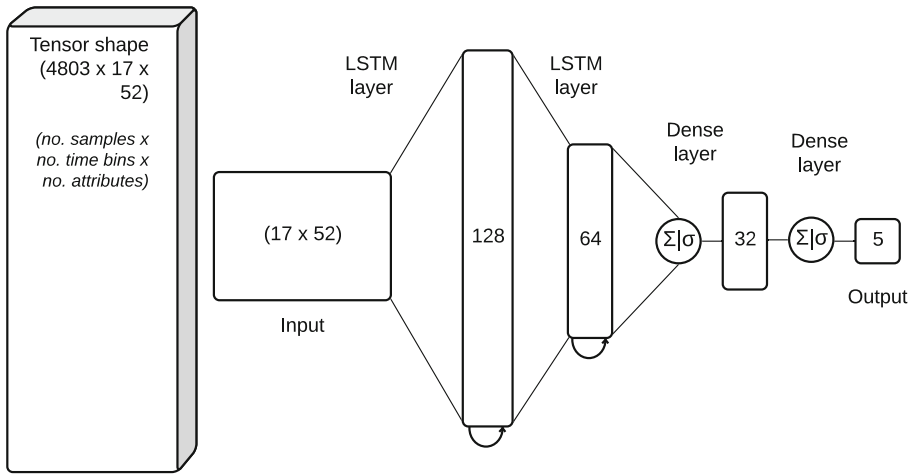
Fig. 2 The chosen LSTM architecture

Chest Compression Detection

For being able to assess a particular action and possibly detect if some mistake occurs, the CPR Tutor has to be sure that the learner has performed a CC and not something different. The approach chosen for action detection is rule-based. While recording, the CC detector continuously checks the presence of CCs by monitoring the vertical movements of the shoulder joints from the Kinect data. These rules were calibrated manually so that the CC detector finds the beginning and the end of the CCs. The CPR Tutor pushes the entire data chunk to SharpFlow via a TCP client at the end of each CC.

Chest Compression Classification

SharpFlow runs a TCP server implemented in Python, which is continuously listening for incoming data chunks by the CPR Tutor. If a new chunk is received, SharpFlow checks if it has a correct data format and is not truncated. It then resamples the data chunks and feeds them into the min-max scaler loaded from memory to ensure that the new instance is normalised correctly. Once ready, the transformed data chunk is fed into the layered LSTMs also saved in memory. The results for each of the five target classes are serialised into a dictionary and sent back to the CPR Tutor, where they are saved as annotations of the CC. SharpFlow takes on average 70 milliseconds to classify one CC.

Real-Time Feedback

Every time the CPR Tutor receives a classified CC, it computes a performance and an Error Rate (ER) for each target class. The performance is calculated with a moving average with a window of 10 seconds, meaning it considers only the CCs

performed in the previous 10s. The Error Rate is calculated as the inverse sum of the performance:

$$ER_j = 1 - \sum_{i=0}^{n} \frac{P_{i,j}}{n} \tag{1}$$

where j is one of the five target classes, n is the number of CCs in one time window of 10s.

Not all the mistakes in CPR are, however, equally important. Although we did not find any objective list of CPR mistakes sorted by their relevance, we assumed that the most important indicators are the ones considered by the standard CPR tools such as the Laerdal ResusciAnne, i.e. the rate, the speed and the release of the chest compression. Some mistakes are correlated or can even cause others: e.g. not locking the arms or not using the bodyweight correctly can result in shallow CCs. Furthermore, in the first dataset collected, the ratio of mistakes and correct executions was not balanced among the five target classes, resulting in varying Error Rates (we report the frequencies in the following section). For example, the mistake in the compression depth was the most frequent, and for this reason, it could activate the corrective feedback the most often, shadowing the feedback on the other target classes. To mitigate the class imbalance, we decided to handcraft feedback rules using five activation thresholds to balance the number of feedback prompts among the five target classes. The thresholds were manually tuned up after multiple attempts. Their values are correlated to the ratio of mistakes: e.g. the ArmsLocked class featured a 38.9% mistake ratio and a threshold value of 5; ClassDepth had a 64.6% mistake ratio and a threshold value of 60. If the Error Rate is equal or greater than these thresholds, the feedback is fired; otherwise, the next rule is checked. The rules are parsed sequentially in the following order:

1. $ER_{armsLocked} >= 5$,
2. $ER_{bodyWeight} >= 15$,
3. $ER_{classRate} >= 40$,
4. $ER_{classRelease} >= 50$,
5. $ER_{classDepth} >= 60$.

Although every CC is assessed immediately after 0.5s, we set the feedback frequency to 10s to avoid overloading the user with too much feedback. In this way, we set the input to fire one message at a time and not more than once every ten seconds. This decision was grounded on previous evidence showing that users prefer to receive a straightforward form of feedback as opposed to multiple feedback messages at a given time (Schneider et al., 2015). The modality chosen for the input was sound. We considered the auditory sense the least occupied channel while doing CPR and resembling the most the feedback given by the human during CPR Training.

We created the following audio messages for the five target classes:

– *classRelease*: "release the compression"
– *classDepth*: "improve compression depth"
– *armsLocked*: "lock your arms"
– *bodyWeight*: "use your body weight"
– *classRate*: *metronome sound at 110 bpm*.

Method

In light of the research gap on providing real-time feedback from multimodal systems, we formulated the following research hypothesis, which guided our scientific investigation.

H1: The proposed architecture allows the provision of real-time feedback for CPR training.

H2: The real-time feedback of the CPR Tutor has a positive impact on the considered CPR performance indicators.

Study Design

To test H1, we developed the CPR tutor with a real-time feedback component based on our design-based research cycle insights. We planned a quantitative intervention study in collaboration with the Aachen Interdisciplinary Training Centre for Medical Education, part of RWTH University in Aachen, Germany[8]. The study took place in two phases: (1) Expert data collection involving a group of 10 expert participants, in which the dataset was collected; (2) a Feedback intervention study involving a new group of 10 participants. A snapshot of the study setup for both phases is shown in Fig. 3. All participants in the study were asked to sign an informed consent letter detailing all the experiment details and the collected data treatment following the new European General Data Protection Regulation (2016/679 EU GDPR).

Phase 1 - Expert Data Collection

The expert group counted 10 participants (M: 4, F: 6), having an average of 5.3 previous CPR courses per person. We asked the experts to perform four sessions of 1-minute duration. In two of these sessions, they had to perform correct CPR, while in the reminder two sessions, they had to perform incorrect executions, not locking their arms and not using their body weight. In fact, from the previous study (Di Mitri et al., 2019b) we noticed it was difficult to obtain the full span of mistakes the learners can perform. Asking the experts to mimic the mistakes was, thus, the most sensible option for obtaining a dataset with a balanced class distribution. We, therefore, collected around 400 CCs per participant. We set the duration to one minute to prevent that physical fatigue influenced the novice's performance. Once we completed the data collection, we inspected each session individually using the *Visual Inspection Tool*. We annotated the CC detected by the CPR Tutor by triangulating with the performance metrics from the ResusciAnne manikin. The *bodyWeight*
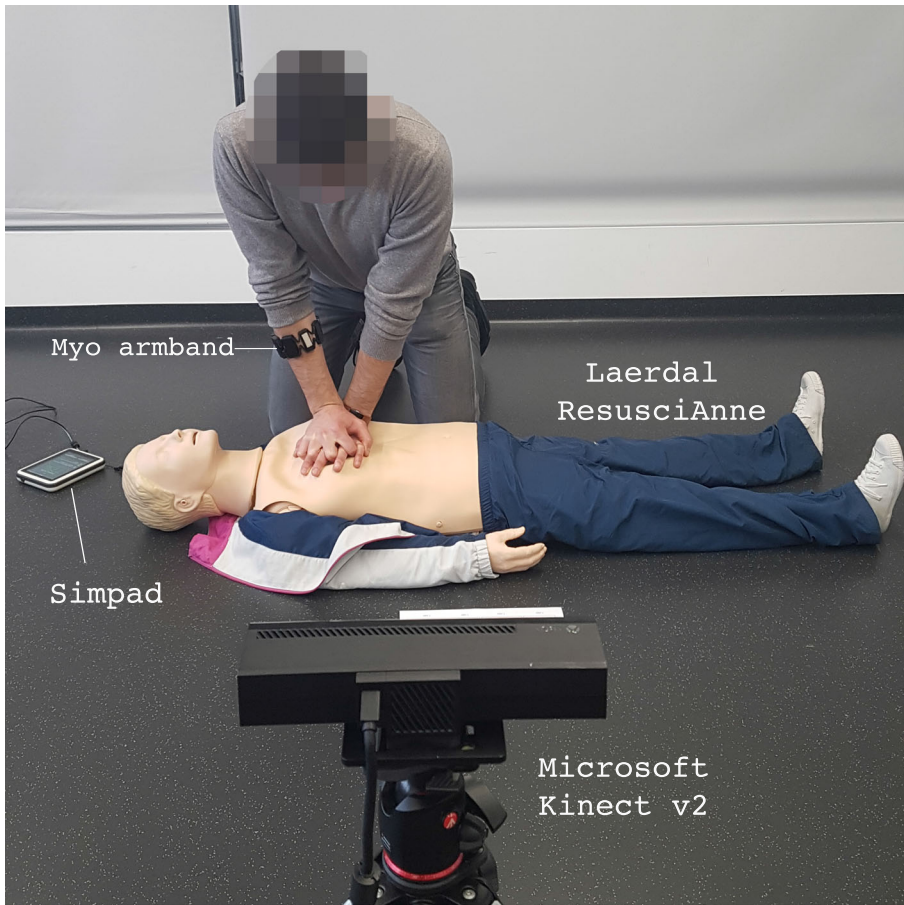

Fig. 3 Study design of the CPR Tutor

and *armsLocked* were instead annotated manually by one component of the research team.

Phase 2 - Feedback Intervention

The feedback intervention phase counted 10 participants (M: 5, F: 5) having an average of 2.3 previous CPR courses per person. Those were not absolute novices but were recruited among the students who needed to renew their CPR certificates. The last CPR training for these participants was, therefore, older than one year. Each participant in the feedback intervention group performed two sessions of 1 minute, one with feedback enabled and one without feedback.

Table 1 Five target classes distribution and performance of corresponding LSTM models trained on the expert dataset

class	Class distribution	Accuracy	Precision	Recall	F1-score
classRelease	0: (1475, 42.9%), 1: (1959, 57.1%)	0.90	0.89	0.95	0.93
classDepth	0: (2221, 64.6%), 1: (1213, 35.4%)	0.95	0.96	0.95	0.95
classRate	0: (1457, 42.5%), 1: (1977, 57.5%)	0.90	0.82	0.82	0.82
armsLocked	0: (1337, 38.9%), 1: (2097, 61.1%)	0.98	0.98	1	0.99
bodyWeight	0: (1206, 35.1%), 1: (2228, 64.9%)	0.97	0.97	0.99	0.98

Results

Expert Group Dataset

The collected dataset from the expert group consisted of 4803 CCs. Each CC was annotated with five classes. With the methodology described in Section 2, we obtained a tensor of shape (4803, 17, 52). As the distribution of the classes was too unbalanced, we downsampled the dataset to 3434 samples (-28.5%). We manually downsampled the dataset by reducing the majority classes (additional details concerning the downsampling strategy are provided in Section 2. In Table 1, we report the new distribution for each target class. In addition, we report the results of the LSTM training reporting for each target class the accuracy, precision, recall and F1-score.

Real-time Feedback Intervention

In the feedback group, we collected a dataset of 20 sessions from 10 participants with 2223 CCs detected by the CPR Tutor and classified automatically. For each participant, we collected two sessions, one with feedback function enabled and one without (in alternating order). The feedback was present in 10 out of 20 sessions and it was fired a total of 16 times: with the following frequency: *classRelase* 2 times, *classDepth* 5 times, *classRate* 5 times, *armsLocked* 1 time, *bodyWeight* 1 time. In Table 2, we report the class distributions and frequencies of the User Group.

Table 2 Class distribution, Average Error Rates and their derivatives for the five target classes in the sessions with feedback

Target	Class distribution	$\overline{E}$	$\overline{\delta E}$	$|F|$	$\overline{\delta_{-10s}}$	$\overline{\delta_{+10s}}$
classRelease	0: (475, 21.4%), 1: (1746, 78.6%)	0.21	0.17	2	-0.11	-0.13
classDepth	0: (704, 31.7%), 1: (1517, 68.3%)	0.38	0.30	5	0.17	-0.04
classRate	0: (475, 21.4%), 1: (1746, 78.6%)	0.26	-0.09	5	0.24	-0.34
armsLocked	0: (3, 0.1%), 1: (2218, 99.9%)	0.01	-0.01	1	0.05	-0.08
bodyWeight	0: (69, 3.1%), 1: (2152, 96.9%)	0.03	0.03	3	0.17	-0.14

We calculated the Average Error Rate for all the 20 sessions of the feedback group find out that the *classDepth* was the class with highest Error Rate (37.9%), followed by *classRate* (25.7%), by *classRelease* (20.8%). Error rate in *bodyWeight* and *armsLocked* were far less frequently with respectively 3% and 0.1%.

In our first analysis, we focused on the subset of sessions where feedback was fired. To have a feeling of the overall tendency of the Error Rate for the entire group, we computed the first derivative of the Error Rate for each target class for each session; later, we grouped by target class, computing the Average Error Rate for all the sessions. We found out that the Error Rate in *classDepth* tended to increase (29.9%) in average, as well as *classRelease* (17.4%) while *classRate* tended to decrease (-8.9%). This data is reported in Table 2.

As further analysis, we generated plots of the Error Rates for each target class for each session. In Fig. 4, we provide an example plot of a session having five feedback interventions (vertical dashed lines) matching the same colours of the target classes. Analysing the plots, we noticed that the feedback was fired nearly every time the Error Rate for the targeted mistake was subject to a drop. We analysed, therefore, the effect of CPR Tutor feedback by focusing on the short-term changes in Error Rate for the mistakes targeted by the CPR Tutor. We calculated the first derivative specifically for the ten seconds before ($\overline{\delta_{-10s}}$) and ten seconds after ($\overline{\delta_{+10s}}$) the feedback was fired. We found out that, on average, the Error Rates of all five target classes are decreasing soon after the feedback is prompted. As shown graphically in Fig. 5 and in Table 2, the greatest decrease is for *classRate*, followed by *bodyWeight*,
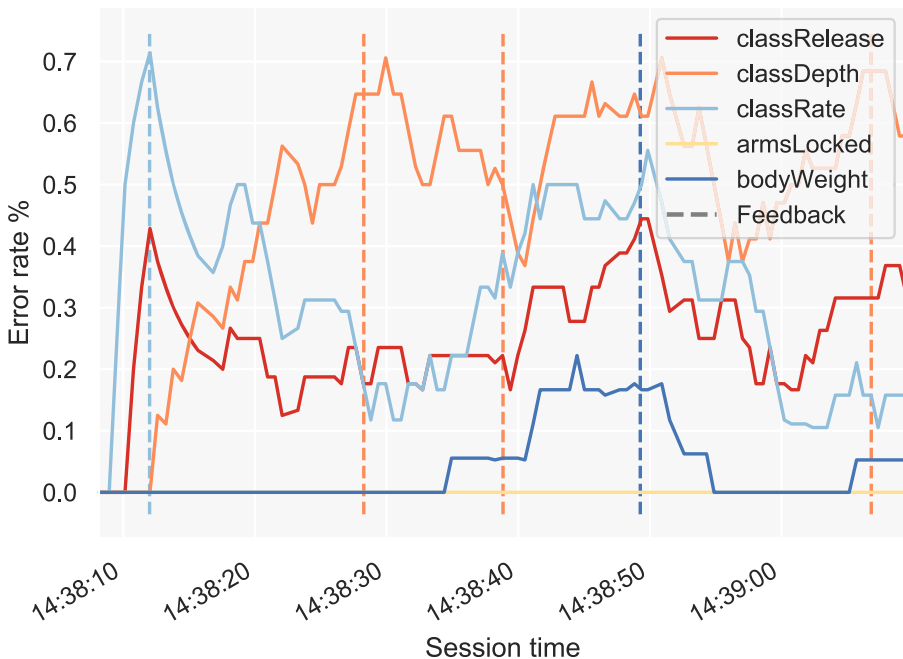

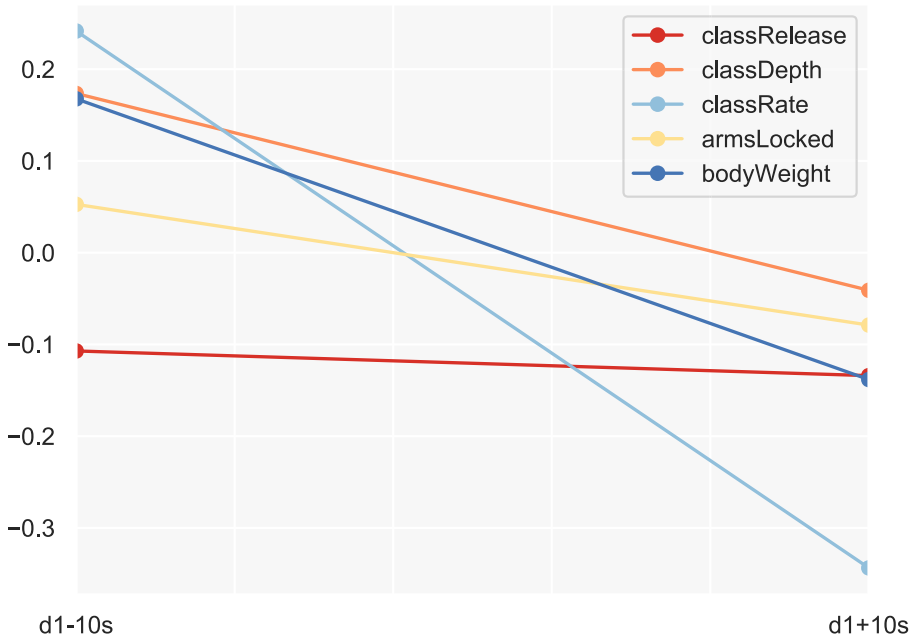**Fig. 4** Plot of the error rates for one session

Fig. 5 Error rate derivative for each target class 10 seconds before and 10 seconds after the feedback was prompted. Results are the average of all the sessions

classDepth, *armsLocked* and *classRelease*. The audio feedback using the metronome sound correct was the most effective feedback intervention.

We compared the sessions where feedback was fired (Feedback group) with the ones where the feedback was purposely deactivated (Control group). In Fig. 6, we plot both the mean Error Rate for each target class for both feedback and control groups. Considering only the mean value flattens all the fluctuations of the Error Rates computed by the CPR Tutor. Nevertheless, we can observe that the distribution of observed Error Rates is almost similar in the sessions with feedback and those without feedback.

To have a general indication of the effectiveness of the CPR Tutor's feedback, we counted the frequency of mistakes regardless of their type for each CC. The results of this analysis are shown in Table 3. We then performed a two-sided Kolmogorov-Smirnov (KS) test to compare the two groups. In this test, the null hypothesis is that the two distributions are identical. We obtained a KS statistic of 0.034 and a p-value of 0.54, meaning we could not reject the null hypothesis. In other words, there was no significant difference in the mistake frequency between the control and the feedback group.

Questionnaire

At the end of the study, we asked the participants to fill out a questionnaire. Given the small number of participants, we decided to ask open-ended questions related to
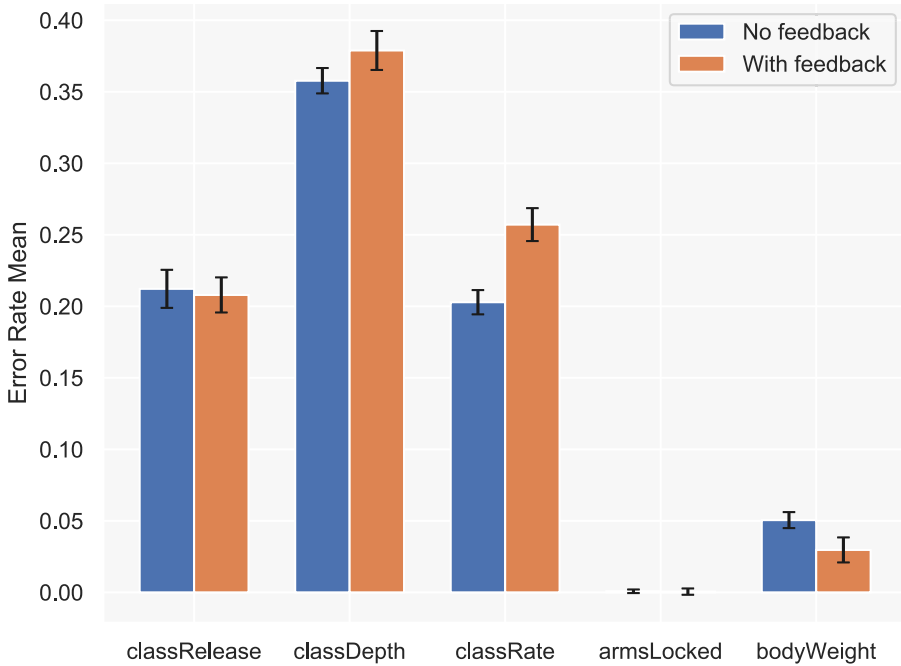
Fig. 6 Error rate mean comparison between sessions with feedback (in orange) and sessions without feedback (in blue) with corresponding standard errors

the CPR Tutor to collect their feedback and facilitate further iterations. We asked to mention up to three aspects of the CPR Tutor they liked, did not like and what they wish to have in the future version of the CPR Tutor. In the following list, we summarise the answers.

What aspects of the CPR Tutor did they like? (no. of times it was mentioned).

– Train CPR without the trainer (4)
– Direct feedback (3)

Table 3 Frequency of mistakes relative to total number of chest compressions compared among the Control group (feedback not activated) and the Feedback group

No. of mistakes	Control group	Feedback group
Zero mistakes	583/1204 (48.4%)	457/1019 (44.8%)
One mistake	398/1204 (33.1%)	356/1019 (34.9%)
Two mistakes	160/1204 (13.3%)	158/1019 (15.5%)
Three mistakes	58/1204 (4.8%)	42/1019 (4.1%)
Four mistakes	5/1204 (0.4%)	4/1019 (0.3%)
Five mistakes	0 (0%)	0 (0%)

- – The feedback is prompted only in the presence of a mistake (1)
- – The feedback is automatic (1)
- – The feedback is self-explanatory (1)
- – The feedback consists of vibrations and audio (1)
- – Feedback on the release of the compression (1)
- – The metronome sound (1)

What aspects of the CPR Tutor did they not like?

- – Feedback was not clear (how to correct the mistake?) (2)
- – Compression depth (1)
- – No positive feedback (1)
- – No general feedback at the end (1)
- – No feedback if pausing the CPR for some seconds (1)

What aspects of the CPR Tutor do they wish to have?

- – Continuous metronome feedback (2)
- – Continuous positive feedback (e.g. green smiley) (2)
- – More explanation on the feedback (2) (what mistake are you making?)
- – Visual feedback for how to release compression (1)
- – Feedback at the end of the session (1)

Discussion

In H1, we hypothesised that the proposed architecture for real-time feedback is suitable for CPR training. With the System Architecture outlined in Section 2, we implemented a functional system that we can use both for the offline model training of the CPR mistakes as well as for the real-time multimodal data exploitation. The proposed architecture exhibited reactive performances by classifying one CC in about 70 milliseconds. The System Architecture proposed is the first complete implementation of the *Multimodal Pipeline* (Di Mitri et al., 2019c), and it shows that it is possible to close the feedback loop with real-time multimodal feedback.

In H2, we hypothesised that the CPR Tutor with its real-time feedback function could positively impact the performance indicators. With a first feedback study involving 10 participants, we noticed a short-term positive influence of the real-time feedback on the detected performance, witnessed by a decrease of Error Rate in the 10 seconds after the feedback was fired (Table 2). This effect is confirmed in three out of five target classes. The remaining two classes show opposite behaviour. In these two cases, Error Rate's increase is less compared to the former target classes. The positive influence of the feedback seems to be, however, only short-term. When we compare the sessions where feedback was fired with the ones without feedback (Fig. 6), we see a similar Error Rate distribution in both groups. This tendency is confirmed by the analysis of mistake frequency between the feedback and control groups reported in Table 3. The two distributions did not exhibit significant differences, suggesting that the CPR Tutor cannot reduce the overall mistake frequency during CPR training.

It is essential to point out that when comparing Error Rates with their derivatives and the mistake frequencies, we do not discriminate whether the feedback was fired for a particular session and target class. We can interpret these results as the overall medium-term effectiveness of the feedback strategy of the CPR Tutor.

A separate discussion point needs to be made for the classes armsLocked and bodyWeight. We suppose that the low detected Error Rates are linked to their extreme class distribution. In turn, this distribution can be because the second group participants were not beginners and, therefore did, not perform common mistakes such as not locking the arms or not using their body weight correctly.

Finally, we think that the observations drawn from the results need to be carefully interpreted and cannot be generalised due to the small number of participants tested for the study.

Questionnaire

The main positive aspects identified in the CPR Tutor by the participants were the ability to practice individually without a trainer and the possibility of receiving automatic direct feedback.

The participants also wished to improve the feedback explanations, particularly concerning what to do when the CPR Tutor recognised a mistake and feedback was fired. The choice to model the mistakes in a binary fashion (presence or absence of the mistake) is not optimal for some performance indicators. For instance, an incorrect CC depth could be corrected both with deeper or softer CCs.

Among the features that some participants wished to have, there was a lack of positive and validating feedback when the CPR Tutor detects no or low amount of mistakes. Moreover, after completing the study, the participants were not given the possibility to overview their performance. Retrospective feedback at the end of the session was mentioned as a desirable feature.

Moreover, the metronome feedback seemed quite an appreciated feature in line with the finding that the metronome was the most effective intervention. This argument is also justified because getting the right CC rate is a critical point for CPR training.

Further Considerations

Since this study aimed to generate real-time feedback using supervised machine learning techniques, we needed several mistakes to train the neural network classifiers. Learning from the expert group was, however, not trivial. Experts tend to make few mistakes. To deal with this problem, we asked the experts to mimic common CPR mistakes. Despite this, the collected data still exhibited unbalanced distribution. The bias in the training data seemed to produce an *amplifying effect* in the classification of the new instances: the majority class in the training set tended to prevail even more in the test set. The solution chosen to mitigate this problem was to downsample the dataset. The downsampled dataset, however, was not easy to reach. As the target classes were five, downsampling one class also affected the other ones. A fair balance between the classes was challenging to achieve. On the other hand,

oversampling was not an option with the time-series dataset: generating synthetic time-series is not straightforward and can undermine the prior class distribution. Future iterations of the model training will consider other options to reduce the dataset imbalance, including propensity score matching or coarsened exact matching.

A limitation of this study is related to splitting the training, test and validation just by random shuffling the data samples. This approach has probably led the neural networks to overfit the expert dataset. One expected consequence could be the poor generalisation of the real-time feedback dataset leading to inaccurate classifications. In future experiments for the CPR Tutor, we recommend training the neural networks using a *leave-one-out* approach, i.e. iteratively keeping data of one expert as test set and while using the data of the remaining experts as for the test set.

There seems to be an implicit hierarchy of importance among the selected performance indicators in the CPR domain. Some mistakes, such as the incorrect locking of the arms, are less frequent but more critical than others. For this reason, the more critical mistakes need to be corrected first. Some other mistakes, such as the incorrect compression depth, are more frequent but not critical since they do not seriously compromise the whole CPR procedure. An important lesson learned from the development of the CPR Tutor was that despite the intention of designing a Multimodal Feedback System using machine learning, the feedback strategy required a rule-based expert system. In the end, the feedback prioritisation that we chose was more dictated by the class distribution of the mistakes rather than the objective importance of CPR training mistakes.

When it comes to the feedback frequency, the decision was to set it to ten seconds intervals. We chose this value as more frequent feedback would distract or confuse the participant. To further prevent confusion, we explained to the participants beforehand which feedback they would receive to know what to expect and what each audio feedback message meant. The automatic generated audio feedback seemed to have a *short-term* positive influence on the CPR performance on the considered target classes. Nevertheless, when comparing the sessions where feedback was activated with those without feedback, we cannot conclude that the CPR Tutor had a consistent *medium-term* positive influence. When looking at session-level results through the mean values of the Error Rates, we can conclude that only the Error Rate in the classRate seemed to be decreasing, which can be linked to the overall effectiveness of the CPR Tutor. This positive influence was not present in the other target classes.

It is important to consider that from a learning sciences standpoint, the performance does not coincide with learning (Soderstrom & Bjork, 2015). Performance refers to the temporary fluctuations in behaviour or knowledge that are observed and measured during training or instruction; learning is the relatively permanent change in behaviour or knowledge that support long-term retention and transfer (Soderstrom & Bjork, 2015). For this reason, we can measure the *long-term* positive influence of the CPR Tutor on learning only with repeated longitudinal studies that last multiple days. This investigation fell out of the scope of the study here presented.

In sum, the results need to be interpreted, taking into account the limited number of participants. The amount of data both in the expert and participant group collected from 10 experts was limited. While we acknowledge the findings can not be generalised, we believe they provide some indication that the real-time feedback of the

CPR tutor has a short-term positive influence on the CPR performance in the target classes.

Future Work

Future empirical testings of the CPR Tutor should explore the longer-term influence of the feedback on the target performance indicators. To achieve that, we would need to (1) collect data from more participants; (2) increase the number of sessions per participant; (3) select participants with less CPR experience so their performance is not optimal and feedback is fired more frequently.

To improve the user experience, the next iteration of the CPR Tutor should implement some features suggested by the questionnaire answers such as (1) better explanations on how to correct the mistakes taking into account multiple classes of mistakes (e.g. 'too deep' or 'too shallow' CCs), (2) positive feedback in case no or a negligible amount of mistakes are detected, (3) visual summary of the performance at the end of the session; (4) warnings in case the CPR is suspended.

The mistake detection models of the CPR Tutor can also improve in future iterations. While LSTMs are generally suitable for detecting mistakes such as the correct body positions and the locking of the arms, they are not optimal for discovering regularities such as the correct *classRate*. An alternative method has been used by Lins et al. (2019) who fitted a sinusoids function using differential evolution, leveraging the rhythmic nature of the CPR movement. Future model architecture improvements can consider alternative neural network architectures, such as Convolutional Neural Networks or Multivariate Hidden Markov Models tuned for supervised classification.

Another challenge that we can address in further iterations of the CPR Tutor is better detecting the CCs. From rule-based, this could become machine-driven. The multimodal sensor streams have to be fed in a CC classifier using sliding windows to achieve that. This would require changing the SharpFlow's "online exploitation" from a TCP-based *batch approach*, to a UDP or MQTT *streaming approach*.

Also, the sensor hardware could be reconsidered. For instance, it would be interesting to see whether it is possible to retrieve kinematic data for regular RGB webcams using body pose estimation libraries such as OpenPose or PoseNet. That would allow to use webcam instead of a Kinect depth camera and make the hardware requirements more accessible to the final user.

To perform the correct evaluation of the CCs, we need to check further the data required's fidelity. Based on this fidelity, we can predict which technology can be best suited for continuous streaming. Also, based on the fidelity, we can say if solutions like OpenPose or PoseNet are sufficient to make predictions or need more powerful software libraries.

Conclusions

We presented the design and the development of real-time feedback architecture for CPR Tutor. Building upon existing components, we developed an open-source data

processing tool (SharpFlow) that implements a neural network architecture and a TCP server for real-time CCs classification. The architecture was employed in a first study aimed at expert data collection and offline training. The second study for real-time feedback intervention allowed us to prove our first hypothesis. Regarding H2, we collected findings that, while they cannot be generalised, indicate that the feedback of the CPR tutor had a short-term positive influence on the CPR performance in the target classes. To sum up, the architecture used for the CPR Tutor allowed for the provision of real-time multimodal feedback, and the generated feedback seemed to have a short-term positive influence on the considered CPR performance indicators.

References

Ahuja, K., Agarwal, Y., Kim, D., Xhakaj, F., Varga, V., Xie, A., Zhang, S., Townsend, J.E., Harrison, C., & Ogan, A (2019). EduSense: Practical Classroom Sensing at Scale. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3(3):1–26 https://doi.org/10.1145/3351229. http://dl.acm.org/citation.cfm?doid=3361560.3351229.

Alqahtani, F., & Ramzan, N. (2019). Comparison and efficacy of synergistic intelligent tutoring systems with human physiological response. Sensors (Switzerland), 19(3), https://doi.org/10.3390/s19030460.

Arroyo, I., Cooper, D.G., Burleson, W., Woolf, B.P., Muldner, K., & Christopherson, R (2009). Emotion sensors go to school. Frontiers in Artificial Intelligence and Applications 200(1):17–24, https://doi.org/10.3233/978-1-60750-028-5-17.

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238. https://doi.org/10.18608/jla.2016.32.11.

Bloom, B.S. (1956). *Taxonomy of educational objectives handbook 1: cognitive domain*, 2nd edn. New York; New York; London: Addison-Wesley Longman Ltd.

Chan, M.CE., Ochoa, X., & Clarke, D (2020). Multimodal Learning Analytics in a Laboratory Classroom. In M. Virvou, E. Alepis, G.A. Tsihrintzis, & L.C. Jain (Eds.) *Machine Learning Paradigms: Advances in Learning Analytics, Intelligent Systems Reference Library, Springer International Publishing, Cham, pp 131–156* https://doi.org/10.1007/978-3-030-13743-4_8.

Crescenzi-Lanna, L. (2020). Multimodal Learning Analytics research with young children: A systematic review. British Journal of Educational Technology 51(5):1485–1504, https://doi.org/10.1111/bjet.12959. https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12959.

Cukurova, M., Kent, C., & Luckin, R (2019). Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring. British Journal of Educational Technology pp bjet.12829–bjet.12829 https://doi.org/10.1111/bjet.12829. https://onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12829.

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R (2020). The promise and challenges of multimodal learning analytics. British Journal of Educational Technology 51(5):1441–1449, https://doi.org/10.1111/bjet.13015. https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.13015.

Davaris, M., Wijewickrema, S., Zhou, Y., Piromchai, P., Bailey, J., Kennedy, G., & O'Leary, S (2019). The importance of automated real-Time performance feedback in virtual reality temporal bone surgery training. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.) *Artificial Intelligence in Education, Springer International Publishing, Cham, Lecture Notes in Computer Science, pp 96–109.* https://doi.org/10.1007/978-3-030-23204-7_9.

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H (2018). From signals to knowledge: a conceptual model for multimodal learning analytics. Journal of Computer Assisted Learning 34(4):338–349 . http://doi.wiley.com/10.1111/jcal.12288, publisher: Wiley/Blackwell (10.1111). https://doi.org/10.1111/jcal.12288.

Di Mitri, D., Schneider, J., Klemke, R., Specht, M., & Drachsler, H (2019a). Read Between the Lines: An Annotation Tool for Multimodal Data for Learning. In: Proceedings of the 9th International Conference on Learning Analytics & Knowledge - LAK19, ACM, New York, NY, USA, pp 51–60 https://doi.org/10.1145/3303772.3303776, http://dl.acm.org/citation.cfm?doid=3303772.3303776.

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H (2019b). Detecting mistakes in CPR training with multimodal data and neural networks. Sensors (Switzerland) 19(14):1–20, https://doi.org/10.3390/s19143099.

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H (2019c). Multimodal Pipeline: A generic approach for handling multimodal data for supporting learning. In *AIMA4EDU Workshop in IJCAI 2019 AI-based Multimodal Analytics for Understanding Human Learning in Real-world Educational Contexts, pp 2–4.*

Di Mitri, D., Schneider, J., Trebing, K., Sopka, S., Specht, M., & Drachsler, H (2020). Real-Time Multimodal Feedback with the CPR Tutor. In I. Bittencourt, M. Cukurova, & K. Muldner (Eds.) *Artificial Intelligence in Education (AIED'2020), Springer, Cham, Cham, Switzerland, pp 141–152* https://doi.org/10.1007/978-3-030-52237-7_12.

Dick-Smith, F., Elliott, D., Martinez-Maldonado, R., & Power, T (2020). Comparing Real-Time Feedback Modalities to Support Optimal Cardiopulmonary Resuscitation for Undergraduate Nursing Students: A Quasi-Experimental Cross-Over Simulation Study. Clinical Simulation in Nursing 44:59–67 https://doi.org/10.1016/j.ecns.2020.01.007, https://www.sciencedirect.com/science/article/pii/S1876139920300104.

D'Mello, S., & Kory, J. (2015). A Review and Meta-Analysis of Multimodal Affect Detection Systems. ACM Computing Surveys 47(3):43:1–43:36 https://doi.org/10.1145/2682899.

D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., El Kaliouby, R., Picard, R.W., & Graesser, A (2008). Autotutor detects and responds to learners affective and cognitive states. *IEEE Transactions on Education*, *48*(4), 612–618. https://doi.org/10.1109/TE.2005.856149.

Dumas, B., Lalanne, D., & Oviatt, S (2009). Multimodal Interfaces: A Survey of Principles, Models and Frameworks. In D. Lalanne, & J. Kohlas (Eds.) *Human Machine Interaction, vol 5440, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 3–26* https://doi.org/10.1007/978-3-642-00437-7_1. http://link.springer.com/10.1007/978-3-642-00437-7_1, series Title: Lecture Notes in Computer Science.

Echeverria, J., & Santos, O. (2021). KUMITRON: Artificial Intelligence System to Monitor Karate Fights that Synchronize Aerial Images with Physiological and Inertial Signals. In *26th International Conference on Intelligent User Interfaces, Association for Computing Machinery, New York, NY, USA, pp 37–39,* https://doi.org/10.1145/3397482.3450730.

Emerson, A., Cloude, E.B., Azevedo, R., & Lester, J (2020). Multimodal learning analytics for game-based learning. British Journal of Educational Technology 51(5):1505–1526, https://doi.org/10.1111/bjet.12992. https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12992.

Giannakos, M.N., Sharma, K., Pappas, I.O., Kostakos, V., & Velloso, E (2019). Multimodal data as a means to understand the learning experience. International Journal of Information Management 48(February):108–119 https://doi.org/10.1016/j.ijinfomgt.2019.02.003, publisher: Elsevier.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation 9(8):1735–1780 https://doi.org/10.1162/neco.1997.9.8.1735, http://www7.informatik.tu-muenchen.de/hochreitidsia.ch/juergen.

Hutt, S., Krasich, K., Mills, C., Bosch, N., White, S., Brockmole, J.R., & D'Mello, SK (2019). Automated gaze-based mind wandering detection during computerized learning in classrooms. User Modeling and User-Adapted Interaction 29(4):821–867 https://doi.org/10.1007/s11257-019-09228-5.

Juntunen, M.L. (2020). Embodied Learning Through and for Collaborative Multimodal Composing: A Case in a Finnish Lower Secondary Music Classroom. International Journal of Education & the Arts 21(Number 29) https://doi.org/10.26209/ijea21n29. http://www.ijea.org/v21n29/index.html.

Koedinger, K., & Corbett, A. (2006). Cognitive Tutors: Technology Bringing Learning Science to the Classroom.

Krishnaswamy, N., & Pustejovsky, J. (2019). Multimodal Continuation-style Architectures for Human-Robot Interaction. arXiv:190908161 [cs] http://arxiv.org/abs/1909.08161, arXiv: 1909.08161 version: 1.

Levin, M., McKechnie, T., Khalid, S., Grantcharov, T.P., & Goldenberg, M (2019). Automated Methods of Technical Skill Assessment in Surgery: A Systematic Review. Journal of Surgical Education 76(6):1629–1639 https://doi.org/10.1016/j.jsurg.2019.06.011. http://www.sciencedirect.com/science/article/pii/S1931720419301643.

Limbu, B., Schneider, J., Klemke, R., & Specht, M (2018a). Augmentation of practice with expert performance data: Presenting a calligraphy use case. In *3rd International Conference on Smart Learning Ecosystem and Regional Development - The interplay of data, technology, place and people, pp 1–13*.

Limbu, B.H., Jarodzka, H., Klemke, R., & Specht, M (2018b). Using sensors and augmented reality to train apprentices using recorded expert performance: A systematic literature review. Educational Research Review 25(June 2017):1–22 https://doi.org/10.1016/j.edurev.2018.07.001, publisher: Elsevier.

Lins, C., Eckhoff, D., Klausen, A., Hellmers, S., Hein, A., & Fudickar, S (2019). Cardiopulmonary resuscitation quality parameters from motion capture data using Differential Evolution fitting of sinusoids. *Applied Soft Computing Journal*, *79*, 300–309. https://doi.org/10.1016/j.asoc.2019.03.023.

Luengo, V., & Mufti-Alchawafa, D. (2013). Target the controls during the problem solving activity, a process to produce adapted epistemic feedbacks in ill-defined domains. In: CEUR Workshop Proceedings, p 8. http://ceur-ws.org/Vol-1009/0803.pdf.

Martinez-Maldonado, R., Echeverria, V., Santos, O.C., Santos, A.DPD., & Yacef, K (2018). Physical learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge, ACM, New York, NY, USA, pp 375–379* https://doi.org/10.1145/3170358.3170379. http://dl.acm.org/citation.cfm?doid=3170358.3170379, issue: May.

Martinez-Maldonado, R., Hernández-Leo, D., & Pardo, A (2019). Preface to the special issue on learning analytics and personalised support across spaces. User Modeling and User-Adapted Interaction 29(4):751–758 https://doi.org/10.1007/s11257-019-09243-6.

Mat Sanusi, K.A., Mitri, D.D., Limbu, B., & Klemke, R (2021). Table Tennis Tutor: Forehand Strokes Classification Based on Multimodal Data and Neural Networks. Sensors 21(9):3121 https://doi.org/10.3390/s21093121. https://www.mdpi.com/1424-8220/21/9/3121, number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Ochoa, X., & Worsley, M. (2016). Augmenting Learning Analytics with Multimodal Sensory Data. Journal of Learning Analytics 3(2):213–219 https://doi.org/10.18608/jla.2016.32.10, http://learning-analytics.info/journals/index.php/JLA/article/view/5081.

Olsen, J.K., Sharma, K., Rummel, N., & Aleven, V (2020). Temporal analysis of multimodal data to predict collaborative learning outcomes. British Journal of Educational Technology 51(5):1527–1547 https://doi.org/10.1111/bjet.12982. https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12982.

Oviatt, S., Schuller, B., Cohen, P.R., Sonntag, D., Potamianos, G., & Krüger, A (2018). The Handbook of multimodal-Multisensor interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 2. [s.n.] . https://dl.acm.org/citation.cfm?id=3015783, publication Title: The Handbook of Multimodal-Multisensor Interfaces: Foundations User Modeling, and Common Modality Combinations - 2. https://doi.org/10.1145/3107990.

Perkins, G.D., Handley, A.J., Koster, R.W., Castrén, M., Smyth, M.A., Olasveengen, T., Monsieurs, K.G., Raffay, V., Gräsner, J., Wenzel, V., Ristagno, G., Soar, J., Bossaert, L.L., Caballero, A., Cassan, P., Granja, C., Sandroni, C., Zideman, D.A., Nolan, J.P., . . . Greif, R (2015). European Resuscitation

Council Guidelines for Resuscitation 2015: Section 2. Adult basic life support and automated external defibrillation. Resuscitation 95:81–99 https://doi.org/10.1016/j.resuscitation.2015.07.015, publisher: Elsevier.

Prieto, L., Sharma, K., Kidzinski, L., Rodríguez-Triana, M., & Dillenbourg, P (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. Journal of Computer Assisted Learning https://doi.org/10.1111/jcal.12232.

Santos, O.C. (2016). Training the body: the potential of AIED to support personalized motor skills learning. *International Journal of Artificial Intelligence in Education*, *26*(2), 730–755. https://doi.org/10.1007/s40593-016-0103-2.

Santos, O.C. (2019). Artificial Intelligence in Psychomotor Learning: Modeling Human Motion from Inertial Sensor Data. International Journal on Artificial Intelligence Tools 28(04):1940006–1940006 https://doi.org/10.1142/s0218213019400062.

Santos, O.C., & Corbí, A. (2019). Can Aikido Help With the Comprehension of Physics? A First Step Towards the Design of Intelligent Psychomotor Systems for STEAM Kinesthetic Learning Scenarios. IEEE Access 7:176458–176469 https://doi.org/10.1109/ACCESS.2019.2957947, conference Name: IEEE Access.

Schneider, J., Börner, D., van Rosmalen, P., & Specht, M (2015). Presentation Trainer, your Public Speaking Multimodal Coach.

Schneider, J., Di Mitri, D., Limbu, B., & Drachsler, H (2018). Multimodal Learning Hub: A Tool for Capturing Customizable Multimodal Learning Experiences. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer, Cham, Switzerland, vol 11082 LNCS, pp 45–58* https://doi.org/10.1007/978-3-319-98572-5_4.

Soderstrom, N.C., & Bjork, R.A. (2015). Learning Versus Performance: An Integrative Review. Perspectives on Psychological Science 10(2):176–199 https://doi.org/10.1177/1745691615569000, http://journals.sagepub.com/doi/10.1177/1745691615569000.

Spikol, D., Ruffaldi, E., Dabisias, G., & Cukurova, M (2018). Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning*, *34*(4), 366–377. https://doi.org/10.1111/jcal.12263.

Åström, K.J., & Murray, R.M. (2021). *Feedback systems: an introduction for scientists and engineers second edition*, 2nd edn. Princeton: Princeton University Press.

Taylor, R.H., Menciassi, A., Fichtinger, G., Fiorini, P., & Dario, P (2016). Medical Robotics and Computer-Integrated Surgery. In: Springer Handbook of Robotics, Springer International Publishing, Cham, pp 1657–1684 https://doi.org/10.1007/978-3-319-32552-1_63. http://link.springer.com/10.1007/978-3-319-32552-1_63.

Vohra, R., Goel, K., & Sahoo, JK (2015). Modeling temporal dependencies in data using a DBN-LSTM. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp 1–4* *https://doi.org/10.1109/DSAA.2015.7344820*.

Worsley, M., & Blikstein, P. (2018). A Multimodal Analysis of Making. International Journal of Artificial Intelligence in Education 28(3):385–419 https://doi.org/10.1007/s40593-017-0160-1, publisher: International Journal of Artificial Intelligence in Education.

Zhu, G., Xing, W., Costa, S., Scardamalia, M., & Pei, B (2019). Exploring emotional and cognitive dynamics of Knowledge Building in grades 1 and 2. User Modeling and User-Adapted Interaction 29(4):789–820 https://doi.org/10.1007/s11257-019-09241-8.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.