Enhancing CPR Quality through Smartwatch: A Neural Network Approach

Gaurav Rao, David W. Savage, Gabrielle Erickson, Nathan Kyryluk, Pawan Lingras, Vijay Mago

Submitted to: JMIR mHealth and uHealth
on: February 17, 2024

Table of Contents

Enhancing CPR Quality through Smartwatch: A Neural Network Approach

Gaurav Rao[1] BCA, MCA, MTEI; David W. Savage[2]; Gabrielle Erickson[2]; Nathan Kyryluk[2]; Pawan Lingras[1]; Vijay Mago[3]

[1]saint mary's university Halifax CA
[2]NOSM Univeristy Thunder Bay CA
[3]York University Toronto CA

Corresponding Author:
Gaurav Rao BCA, MCA, MTEI
saint mary's university
923 Robie St.
Halifax
CA

Abstract

Background: In the event of cardiac arrest, providing immediate, high-quality cardiopulmonary resuscitation (CPR) and applying a defibrillator are crucial for patient care. High-quality CPR is defined by chest compressions at a rate of 100–120 per minute and a compression depth of 50–60 mm. However, during an emergency, monitoring the count and depth of compressions poses a significant challenge for individuals administering CPR.

Objective: This study introduces a neural network model designed to predict and assess the quality of CPR utilizing accelerometer data from a participant's smartwatch.

Methods: This research involved collecting real-world chest compression data from 83 participants performing CPR on a mannequin, with accelerometer data captured via smartwatches worn by the participants. This data was employed to train the model against a gold-standard dataset from the mannequin. The accelerometer-derived compression data were aligned with those from the mannequin dataset. Subsequently, the data were segmented into five-second intervals to facilitate training the neural network models.

Results: Throughout the study, 1,226 neural network models were developed, incorporating variations in hyperparameters and the dataset. The optimal model demonstrated the capability to accurately predict the number of compressions and the average compression depth within a five-second interval, achieving an accuracy of ±3.8 mm and an average deviation in compression count of 0.8.

Conclusions: The study validates the efficacy of a neural network model in accurately predicting CPR metrics, outperforming other models discussed in the literature and involving a considerably large participant base. Clinical Trial: The ethics application for this research received approval from TBRHSC (REB 2022519), allowing the collection and use of participant data for research purposes. Furthermore, all participants gave written consent for their data to be collected and used in this study.

(JMIR Preprints 17/02/2024:57469)
DOI: https://doi.org/10.2196/preprints.57469

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Enhancing CPR Quality through Smartwatch: A Neural Network Approach

Abstract

Background: In the event of cardiac arrest, providing immediate, high-quality cardiopulmonary resuscitation (CPR) and applying a defibrillator are crucial for patient care. High-quality CPR is defined by chest compressions at a rate of 100–120 per minute and a compression depth of 50–60 mm. However, during an emergency, monitoring the count and depth of compressions poses a significant challenge for individuals administering CPR.
Objective: This study introduces a neural network model designed to predict and assess the quality of CPR utilizing accelerometer data from a participant's smartwatch.
Methods: This research involved collecting real-world chest compression data from 83 participants performing CPR on a mannequin, with accelerometer data captured via smartwatches worn by the participants. This data was employed to train the model against a gold-standard dataset from the mannequin. The accelerometer-derived compression data were aligned with those from the mannequin dataset. Subsequently, the data were segmented into five-second intervals to facilitate training the neural network models.
Results: Throughout the study, 1,226 neural network models were developed, incorporating variations in hyperparameters and the dataset. The optimal model demonstrated the capability to accurately predict the number of compressions and the average compression depth within a five-second interval, achieving an accuracy of ±3.8 mm and an average deviation in compression count of 0.8.
Conclusions: The study validates the efficacy of a neural network model in accurately predicting CPR metrics, outperforming other models discussed in the literature and involving a considerably large participant base.

Trial Registration: The ethics application for this research received approval from TBRHSC (REB 2022519), allowing the collection and use of participant data for research purposes. Furthermore, all participants gave written consent for their data to be collected and used in this study.

Keywords: CPR performance, CPR feedback, neural network, Smart health, Smartwatch, Sudden Cardiac Arrest (SCA)

Introduction

Sudden Cardiac Arrest (SCA) is identified as a principal cause of mortality in North America, particularly among young athletes [1–4]. It can affect individuals irrespective of their lifestyle or health status, leading to either irregular or non-existent heart rhythms [5,6]. As a consequence, the flow of blood to major organs is halted, depriving them of essential oxygen, resulting in tissue damage and potentially culminating in organ failure. The probability of survival decreases by about 10% with each passing minute without intervention; therefore, immediate and effective treatment is crucial to enhance survival rates by minimizing damage to tissues and organs [7,8].

Efforts to deliver prompt care to SCA victims continue, with Emergency Medical Services (EMS) prioritizing SCAs to expedite emergency responses [9,10]. Public data indicate that the target for EMS response time to an SCA event is within 8–10 minutes [11,12]. Survival rates significantly decline if response times greatly exceed 10 minutes [5,13,14]. The initial treatment involves performing Cardiopulmonary Resuscitation (CPR) and utilizing an Automated External Defibrillator (AED) to assist heart pumping and maintain blood flow to the brain and other vital organs. In certain instances, an AED can also restore the heart's electrical activity [6,15].

Organizations such as the American Heart Association (AHA) and the Red Cross play a pivotal role in providing CPR and AED training to the public, empowering bystanders to administer early care to SCA victims until EMS arrives. Studies indicate that bystanderadministered CPR significantly improves survival rates compared to scenarios where CPR is not administered [16–18]. The quality of CPR is critical, as high-quality CPR is associated with increased chances of survival. The AHA defines high-quality CPR as having a compression depth of 5-6 cm and a rate of 100–120 compressions per minute, standards that are also supported by the European Resuscitation Council [19,20].

Figure 1. Laerdal Skill Trainer application displaying data from the manikin sensor. Source: https://laerdal.com/ca/products/simulation-training/manage-assess-debrief/skillreporter-app/

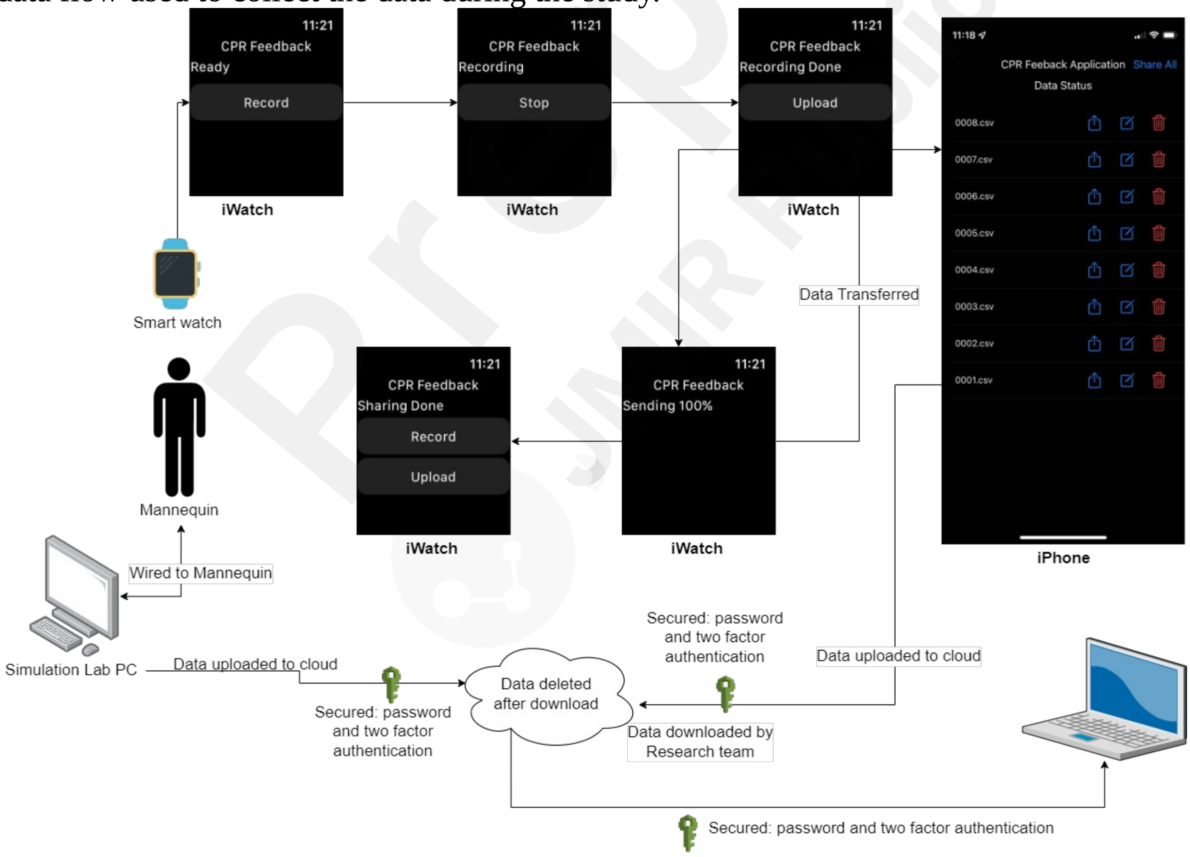
Figure 2. Laerdal CPRMeter2 device used to measure compression performance. Source: https://laerdal.com/

Training the general public in CPR administration has emerged as a key objective for numerous organizations aiming to improve survival rates for SCA patients [21,22]. If a significant portion of the populace receives training in CPR, SCA victims stand a higher chance of receiving necessary immediate care until Emergency Medical Services (EMS) arrive, thereby enhancing their survival prospects. Public training programs instruct participants on identifying SCA victims and initiating high-quality CPR. In these sessions, participants engage in CPR practice on mannequins while trainers offer real-time feedback on their performance, ensuring proficiency [23–25]. With advancements in technology, CPR performance can now be quantified, allowing feedback to be grounded in these measurements. Technologies utilized in training encompass sensor-equipped mannequins and CPR feedback devices.

Mannequins equipped with sensors accurately assess compression depth and frequency, providing trainers with metrics to deliver precise feedback to trainees. Figure 1 presents a sample output from the Laerdal Skill Trainer application. Furthermore, CPR feedback devices, placed on the mannequin's chest and compressed by the trainee, include sensors such as accelerometers and pressure sensors to evaluate CPR performance. Based on the performance, these devices provide auditory or visual feedback to promote high-quality CPR.

For SCA incidents, emergency call operators utilize technologies like Tele CPR and video CPR to assist individuals in achieving high-quality CPR [26–28]. Upon identifying an SCA situation, EMS is dispatched, and the caller is guided over the phone on performing CPR. Nevertheless, the operator cannot gauge the quality of compressions administered, providing only verbal instructions to help the caller maintain an appropriate pace, which leaves compression depth uncertain. Video CPR, as an advanced approach, transitions communication from audio to video, allowing the operator to observe the CPR performance and offer real-time feedback. This technology requires advancements in emergency response systems to support video calling capabilities and necessitates video capability on the caller's part. A limitation is its dependency on two bystanders: one to execute CPR and another to capture the action on film.

Figure 3. The flow diagram shows the screenshots of iPhone and Apple Watch applications, and the data flow used to collect the data during the study.



Researchers have delved into next-generation CPR feedback technologies, including virtualreality (VR) and augmented-reality (AR) based devices [29,30]. These devices employ integrated cameras to capture and analyze compression depth and rate, displaying real-time statistics on the device screen to aid users in enhancing their CPR performance [31, 32]. However, the widespread adoption of such advanced technologies encounters obstacles: they are either underused or not always accessible to those who do use them.

Therefore, the challenge lies in identifying a device suitable for real-world emergencies that can precisely measure CPR performance and provide appropriate feedback without the necessity for specialized hardware. The literature identifies smartphones and smartwatches as two potential solutions [33–35]. Both types of devices are equipped with sensors capable of evaluating CPR performance. Song et al. introduced a mobile application that leverages accelerometer data from smartphones to assess compression quality and provide feedback via screen displays and audio cues [36]. Nonetheless, this approach faces limitations: it necessitates attaching the smartphone to the user’s arm, lacks details on data cleaning and noise removal, and does not consider variations in device orientation [37]. Similar challenges are evident in other algorithms designed for mobile CPR applications [38,39], with many focusing solely on training scenarios rather than real emergencies.

Gruenerbl et al. proposed a smartwatch application capable of measuring CPR parameters and offering visual feedback [40]. This application analyzes accelerometer data to evaluate compression quality, identifying each positive peak on the y-axis as a compression and calculating the differences in y-axis peaks to determine compression depth. However, the study does not provide detailed algorithmic and data cleaning methodologies for replication and comparison.
Lu et al. also introduced a smartwatch application alongside an algorithm for evaluating compression metrics [41]. They tested using a Resusci Anne QCPR training manikin (Laerdal) and an Android ASUS ZenWatch 2 (model WI501Q). The developed polynomial model predicts compression depth and rate from smartwatch accelerometer data. Although data was collected, its limited variability—compression counts between 80–140 and depths of 4–7cm—fails to cover the wider range expected in real-world scenarios, nor does it elaborate on data cleaning or handling irregularities.

Utilizing smartwatch accelerometer data presents challenges, including noise from various sources—sensor white noise, gravitational movements, hand shaking, and watch movement due to fit. Filtering such noise from sensor data remains a significant challenge [42]. Employing noisy datasets can result in substantial variations in output results over time. To tackle noise filtration, we propose a neural network model trained on noisy data to predict compression performance, as detailed in Methods section.

However, assessing compression depth and counting compressions during an emergency proves to be impractical for responders. Various solutions have been proposed to measure and provide real-time feedback for high-quality CPR, including Tele CPR, feedback devices, and Video CPR, though these are constrained by geographical and additional factors [27,43,44]. An innovative approach involves utilizing smartwatches to gauge CPR performance and offer immediate feedback, yet research on smartwatch applications is limited and often lacks the necessary data and algorithmic details for replication and improvement [35,45].

Methods

This section presents the technology and equipment used during the data collection, cleaning, and processing phases. A total of 83 participants from Thunder Bay Regional Health Sciences Centre (TBRHSC) took part in the data collection by performing CPR in a controlled simulation setting, using an Apple Watch, Apple iPhone, and a Laerdal mannequin for data capture [46,47]. The ethics application for this research received approval from TBRHSC (REB 2022519), allowing the collection and use of participant data for research purposes. Furthermore, all participants gave written consent for their data to be collected and used in this study.

Table 1. Demographic details of participants in the data collection process.

Description	No. of participants
No. of mannequin-smartwatch datasets	124
No. of participants	83
Participants performing 2 rounds of CPR	41
Sex	
Female	56
Male	27
Age Classes	
<30	42
30-39	27
40-49	8
>50	6
Previous CPR attempts	
0	12
1-9	23
10-19	10
20-29	14
30+	24

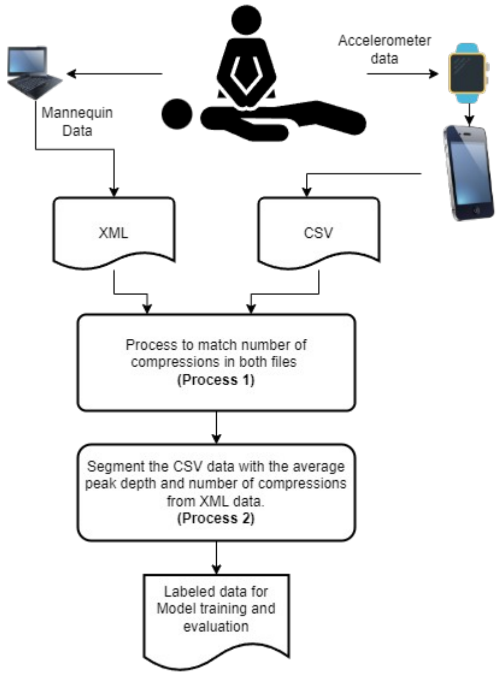
Equipment/Hardware

Participants collected accelerometer data with an Apple Watch Series 7 while performing CPR, which was then stored on an Apple iPhone 14. These devices, chosen for their significant presence in the wearable segment, were equipped with specially developed apps for data collection and analysis [48,49]. This study differs from previous ones by focusing on the potential of the Apple ecosystem for healthcare applications, rather than on Android devices. The apps were designed for compatibility with Apple iPhone 10 and newer, as well as Apple Watch Series 3 and later models.

CPR performance practice occurred on the Laerdal SimMan 3G Mannequin, which is outfitted with sensors that record and provide feedback on key compression quality metrics such as depth, rate, and recoil. This aids in improving CPR quality [50]. Laerdal mannequins were chosen, as in other studies, for their precise data capture capabilities [36,37,40].

Data analysis was performed using a laptop with an i7 Intel processor, 32 GB RAM, and a 1TB hard drive, using Microsoft Excel and Python for data cleaning, processing, model training, and evaluation. Final analyses were done on a cloud server (Amazon Web Services) equipped with an NVIDIA A10G GPU (24 GB GPU Memory), 32 GB RAM, and 8 vCPUs [51].

Figure 4. The flow diagram illustrates the process for preparing the data for model training.



Data collection

Data for this investigation were derived from two principal sources: an application on the Apple Watch and data exported from a Laerdal mannequin. These datasets were recorded in distinct files, with a uniform unique identifier applied to each, ensuring their association was maintained throughout the processes of data cleaning and preprocessing. Figure 4 illustrates the procedural flowchart for preparing the raw data for subsequent model training.

Data exported from the mannequin, encompassing a comprehensive list of compressions performed during the sessions, was acknowledged as the benchmark for this research due to the sensors' precision in measuring critical parameters like compression depth, rate, and recoil. Concurrently, the Apple Watch app provided time-stamped accelerometer readings across the X, Y, and Z axes, formatted as comma-separated values. Both datasets included detailed compression information, facilitating the analysis of two-minute CPR sessions. For analyses targeting shorter durations, segmentation of data from the accelerometer and the mannequin was required.

While the provision of real-time feedback to users was not within the immediate scope of this project, the collected data were thoroughly examined to assess their potential future use for such feedback. A five-second window for compression analysis was identified as optimal for delivering preliminary feedback, with high-quality CPR—characterized by an estimated ten compressions within this timeframe—serving as the basis for feedback. Following the reception of feedback, it was anticipated that users would need a brief period to digest the information and adjust their technique accordingly before the subsequent round of feedback.

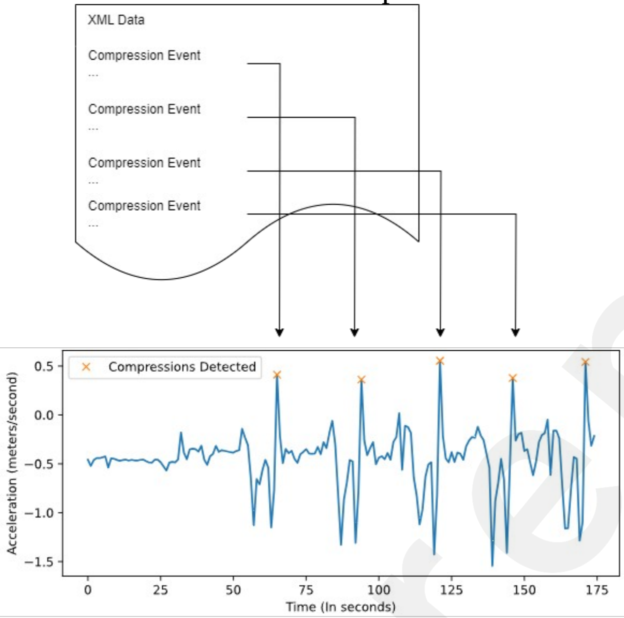
Data Cleaning and Preparation for Model Training

This study utilized data from two sources: an Apple Watch app and exported data from a mannequin. Each source's data were recorded in separate files, but a unique identifier linked them for consistency throughout the data cleaning and preprocessing stages. Figure 4 illustrates the process of preparing the raw data for model training.

The mannequin's data, exported as a comprehensive list of compressions during each session, served as the gold standard due to the sensors' precise measurements of compression depth, rate, and recoil. Meanwhile, the Apple Watch provided timestamped accelerometer data across the X, Y, and Z axes, formatted as comma-separated values. Both datasets included detailed compression information, enabling analysis of two-minute CPR sessions.

While real-time user feedback was beyond the scope of this project, the data were examined to assess future applicability for this purpose. A five-second analysis window was selected for potential feedback, where high-quality CPR—defined by approximately ten compressions within this period—was used as the feedback standard. After receiving feedback, users would need time to comprehend and adjust their technique before the next feedback prompt.

Figure 5. Process 1: The diagram illustrates a representative output showcasing the matching of accelerometer data with compression information from the mannequin files.



Algorithm 1: Pseudocode to generate matching between compressions in Mannequin and accelerometer data files.

```
Input  : AccelerometerFile (C); MannequinFile (X);
         AccelerationDifferenceRange (AD),
         MinimumAccelerationPeakRange (MAD),
         WindowSizeRange (WS)
Output: Dictionary (CompressionMatching)
         mapping compression peaks in accelerometer
         data to corresponding records in mannequin
         data
PeaksInMannequin =
  findCompressionsCountInMannequin(X) ;
for ad ∈ AD do
    for mad ∈ MAD do
        for ws ∈ WS do
            peaksFound = findPeaks(C, ad, mad, ws) ;
            CompressionMatching =
              getMatchingData(X, C, peaksFound) ;
            if peaksFound == PeaksInMannequin
              then
                return CompressionMatching ;
            else
                updateRangeVariables(AD, MAD, WS)
            end
        end
    end
end
```

Data matching (Process 1)

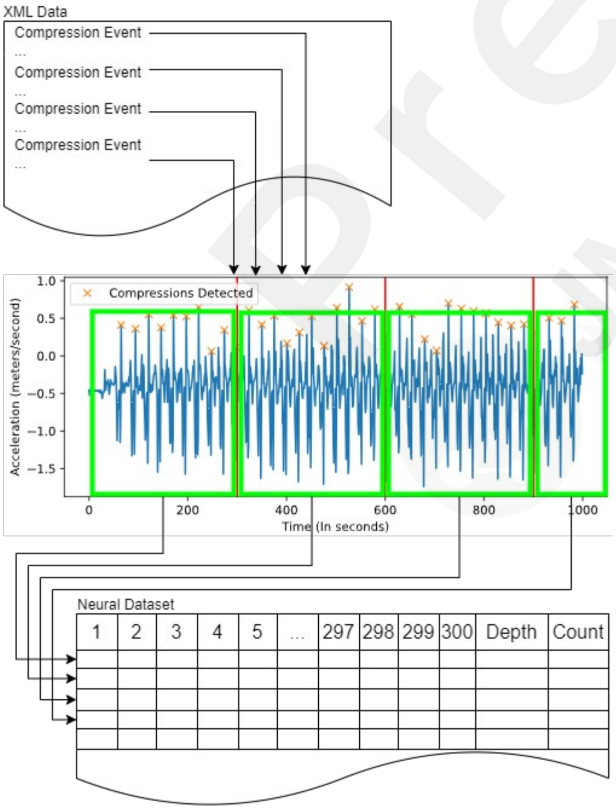
To ensure accuracy in training, it was crucial to synchronize the compression data logged by the mannequin with the accelerometer readings from the smartwatch captured in the datasets. To achieve this synchronization, an algorithm was conceived to identify the optimal blend of three specific parameters: 1) acceleration difference; 2) window size; and 3) minimum acceleration peak, thereby precisely matching the number of compressions recorded by the mannequin. These parameters were initially assigned a range based on a detailed manual examination of the acceleration patterns observed in the accelerometer data. Employing the "find peaks" function from the Python Library Scipy, the algorithm endeavored to pinpoint the peaks, determining those that corresponded to compression events within the scope of this research [52]. The procedure involved iteratively adjusting these parameters, methodically refining their ranges until the algorithm successfully identified the exact count of peaks matching those captured by the mannequin. The pseudocode presented in Algorithm 2.3 details the methodology employed to achieve alignment between the mannequin data and accelerometer-derived compression readings, with Figure 5 displaying the results of this data matching effort.

Within the context of Algorithm 1, three distinct variables were meticulously optimized for the dataset, reflecting the inherent variability of the data. The window size (ws) parameter was crucial for defining the interval expected to contain a singular compression depth peak. While theoretically, up to two compressions might occur within one second, variations were anticipated, with novices potentially performing CPR at a slower rate of one compression per second and experts capable of achieving three compressions in the same timeframe. The minimum acceleration peak (mad) parameter was indicative of the acceleration threshold necessary to signal the initiation of a compression, its variability likely mirroring the user's compression speed and the achieved depth. Conversely, the acceleration difference (ad) parameter captured the peak acceleration during a compression, with its variability potentially reflecting the user's application of speed or force. Given the potential for significant variation, especially among less experienced CPR performers, initial ranges were established for each variable. Subsequent explorations of all viable combinations within these ranges were conducted to identify the parameter values that most accurately correlated with the compressions documented in the smartwatch data.

Data Chunking (Process 2)

For each session involving a pair of mannequin and smartwatch datasets, a total duration of two minutes was recorded. The data were segmented into smaller portions to facilitate prompt feedback, potentially enabling the provision of real-time feedback. This segmentation process involved three critical steps: 1) aligning the datasets from the mannequin and the smartwatch to ensure their data corresponded accurately for analysis; 2) determining an appropriate duration for the data chunks; and 3) executing the chunking algorithm. The procedure for aligning the data is detailed in the subsection titled ”Data matching (Process 1).” A five-second interval was chosen as the optimal chunking duration, with the rationale for this selection explained in subsection 2.3. Subsequently, the aligned datasets underwent segmentation into chunks for detailed analysis. Figure 6 illustrates the segmentation process applied to a representative dataset. Following the alignment of accelerometer data with the mannequin’s compression peaks, the accelerometer readings were divided into 300-point segments (indicated by a green box). Each segmented block (denoted by a green box) was then documented as a single record within the neural dataset, formatted in CSV. These records encompassed acceleration data for a five-second interval (equivalent to 300 data points at a 60Hz sampling rate), alongside the average compression depth and the total number of compressions recorded within that timeframe. For instances where the dataset was less than five seconds in length, the acceleration data was supplemented with zeros to standardize it to a five-second duration. The datasets prepared through this method were subsequently utilized to train the neural network model.

Figure 6. Process 2: Illustration of combining accelerometer and mannequin data to create the dataset for training the neural network model



Model to Predict CPR Performance

The aim of this study was to evaluate CPR performance by analyzing accelerometer data collected from smartwatches, which inherently included various forms of sensor noise, such as vibrations due to hand movements and accelerations not associated with chest compressions. The direct elimination of this noise was deemed impractical; therefore, the decision was made to employ a supervised learning approach, specifically a neural network model, to tackle this intricate challenge.

This neural network model processed 300 accelerometer data points, corresponding to a duration of five seconds at a 60Hz sampling rate, with the objective of predicting both the quantity of compressions and their average depth during this interval. The development of the model utilized the TensorFlow library within the Python programming environment [53]. To fine-tune the model for optimal performance with the dataset at hand, a comprehensive grid search was conducted, focusing on adjusting the model's hyperparameters, which included the number of hidden layers, the sizes of these layers, the number of epochs, batch sizes, and the rates of dropout. This meticulous process was aimed at discovering the most effective combination of hyperparameters that would enable the model to accurately interpret the accelerometer data, despite the presence of sensor noise, and make reliable predictions regarding CPR quality based on the smartwatch data collected.

Results

A series of experiments were systematically conducted, exploring various hyperparameters and datasets, until no notable improvements in outcomes were observed. This investigative process entailed 12 different experiments to pinpoint the optimal set of hyperparameters. Within each trial, data on the variations in hyperparameters were meticulously documented, noting the discrepancies between anticipated and actual figures for compression depths and counts. These discrepancies enabled the calculation of statistical measures such as medians, averages, minimums, and maximums. Given that the variance between expected and observed results could manifest in negative figures, potentially skewing average calculations, absolute values were employed for a more accurate determination of the absolute average and median. These analytical steps informed adjustments in hyperparameters for ensuing experiments aimed at enhancing performance.

Throughout this extensive testing phase, a total of 1,226 iterations were performed, each exploring different hyperparameter settings. Details regarding these settings and the premier outcomes from each experiment are tabulated in Table 2. The initial trial set the hyperparameters as follows: epochs ranged from 10 to 1,000, batch sizes varied from 3 to 11, and layer sizes were set between 10 and 1,200 across three hidden layers, with a 10% reduction in connection dropout implemented after every two hidden layers. This structure established the input layer to accommodate 300 points, while the output layer was configured to include two points. The strategy for structuring hidden layers dictated that the central layer would achieve the maximum size, with sizes incrementally increasing from the input towards this midpoint before decreasing towards the output layer (300-700-1000-100-2). An illustrative configuration for a three-layer model might progress from 300 to 1,000 at the peak, then taper down to 100 at the output layer.

The first experiment involved 64 different model iterations, showcasing the impact of varying layer sizes, with the largest hidden layer reaching 4,000 and the epoch count set at 10. This particular iteration yielded the most accurate prediction for compression depth, with an average absolute deviation of 4.8mm and for compression count, an average absolute deviation of 1.2 counts. Notably, this experiment underscored the critical influence of layer size configurations on model outcomes, as identical batch sizes and epoch durations led to significantly diverse results based on the arrangement of the final hidden layer.

In the second experiment, layer sizes were intentionally decreased to evaluate their impact on performance, while introducing variability into other hyperparameters such as batch size, epochs, and the number of layers, resulting in 36 distinct combinations of these variables. Among these iterations, the most successful outcomes were observed when the batch size and epoch count were set to 500, with the model utilizing four hidden layers. This configuration yielded the most accurate prediction of compression depth, registering an average absolute deviation of 4.0mm, and a similar accuracy in compression count with an average absolute deviation of 0.9 counts.

Table 2. Table summarizing hyperparameters for each experiment and the best results achieved.

Experi-ment	**Epochs**	**Batch size**	**Layer sizes**	**No. of layers**	**Connection Drop Out**	**Best Result (Compression depth, count)**
1	10	3, 5, 7, 9, 11	10, 35, 60, 85, 100, 400, 700, 1000, 1200	3	10% after every 2 layers	± 4.8mm, 1.2 counts
2	100, 500	6, 12, 18	10, 35, 60, 85, 100, 400, 700, 1000, 1200	3, 4	10% after every 2 layers	± 4.0mm, 0.9 counts
3	500, 750	18, 24	5, 40, 100, 500, 900, 1300	4, 6	10% after every 2 layers	± 4.8mm, 1.2 counts
4	500, 750	18, 24	5, 40, 100, 500, 900, 1300	4, 6	10% after every 2 layers	± 3.9mm, 0.9 counts
5	500, 750	18, 24	5, 10, 60, 100, 400, 700, 1000	4	10% after every 2 layers	± 3.9mm, 0.8 counts
6	500, 750	6, 24	5, 10, 50, 100, 500, 1000, 4000	4, 5	10% after every 2 layers	± 6.2mm, 1.2 counts
7	1000	3	5, 10, 50, 100, 500, 1000, 4000	5	10% after every 2 layers	± 4.0mm, 0.9 counts
8	1000	12	5, 10, 50, 100, 500, 1000, 4000	5	No dropouts	± 3.9mm, 0.9 counts
9	1000	3, 12	5, 10, 50, 100, 500, 1000, 4000	3	No dropouts	± 4.1mm, 0.9 counts
10	100, 500, 1000	3, 9	5, 10, 25, 50, 75, 100, 250, 500, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3250, 3500, 3750, 4000	3	10% after every 2 layers	± 4.1mm, 0.9 counts
11	100, 1000	3128	5, 10, 25, 50, 75, 100, 250, 500, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3250, 3500, 3750, 4000	5	10% after every 2 layers	± 3.9mm, 0.8 counts
12	**1000, 2000**	**256, 1024**	**5, 10, 25, 50, 75, 100, 250, 500, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3250, 3500, 3750, 4000**	**5**	**10% after every 2 layers**	**± 3.8mm, 0.8 counts**

For the third experiment, adjustments were made to increase epochs to [500, 700], batch sizes to [18, 24], and the number of hidden layers to [4, 6], while also simplifying layer options to ascertain the impact of size reduction on performance. A strategic modification was applied to the layer combination approach, prioritizing the largest hidden layer initially and progressively diminishing layer sizes towards the output. This methodology facilitated 30 unique iterations exploring various configurations of hyperparameters and layer structures. The premier iteration within this set achieved an average absolute loss of 3.9mm for compression depth and 0.9 counts for compression count, with the top three iterations all featuring four hidden layers, batch sizes of 24, and 750 epochs.

The fourth experiment maintained the hyperparameter settings of the third experiment but introduced a significant modification to the dataset used for training. Instead of solely relying on the raw accelerometer data, this trial incorporated a smoothed version of each data record, effectively doubling the dataset with one subset remaining raw and the other subjected to a smoothing process using a window size of five. Despite this alteration in the dataset, the impact on the experiment's outcomes was minimal, resulting in findings that closely paralleled those of the third experiment.

In the fifth experiment, attention was directed towards evaluating smaller hidden layer sizes. The approach for configuring layers reverted to positioning the central hidden layer as the largest, with preceding layers incrementally increasing towards this peak and subsequent layers diminishing towards the output. This adjustment led to 30 viable hyperparameter combinations. The most accurate iteration within this trial achieved an average absolute deviation in compression depth of 3.9mm and in compression count of 0.8 counts. Among the top-performing configurations, four hidden layers, 750 epochs, and batch sizes set at 24, and 18 were common. When comparing the leading outcomes from the fourth and fifth experiments, a slight improvement was noted in the latter.

The sixth experiment introduced alterations in hyperparameters, specifying batch sizes of 3 and 12, hidden layer counts of 4 and 5 and adding a layer size of 4,000. Unfortunately, this led to a decline in performance, with the optimal prediction exhibiting a substantial average absolute loss in compression depth of 6.2mm.

In the subsequent seventh experiment, epochs were elevated to 1,000, batch size was minimized to 3, and the model was structured with five layers. These modifications markedly enhanced the results, yielding the finest prediction of an average absolute compression depth of 4.0mm and an average absolute compression count of 0.9.

The eighth experiment saw further adjustments from the seventh, notably increasing the batch size to 1,000 and eliminating connection dropouts. While the best outcomes of the seventh and eighth experiments were closely matched, a comparative analysis revealed that all iterations from the eighth experiment consistently outperformed the third-highest result of the seventh, indicating a refined optimization of hyperparameters in the latter experiments.

For experiment nine, batch sizes were designated as [3,12], and the model was configured with three hidden layers. The outcomes from this trial were notably inferior to those observed in the eighth experiment. The most accurate iteration within this series estimated the compression depth with an average absolute deviation of 4.1mm and the compression count with an average absolute deviation of 0.9 counts.

The tenth experiment explored a broader range of hyperparameters, setting epochs at [100,500,1000], batch sizes at [3,9], and maintaining the number of layers at three. A 10% dropout rate was applied after every two layers, and layer sizes varied from 5 to 4,000, applying a strategy where layer sizes increased towards the midpoint before decreasing towards the output. This extensive exploration resulted in 519 potential hyperparameter combinations. The iteration yielding the most accurate results within this framework predicted the compression depth with an average absolute loss of 4.1mm and the compression count with an average absolute loss of 1.0 counts, indicating that the findings from this extensive trial did not surpass the performance metrics established in earlier experiments.

In the eleventh experiment, adjustments were made to three hyperparameters: epochs were set to [100,1000], batch size to [3,128], and layer size was fixed at [5]. The strategy for layer configuration aimed to establish the initial hidden layer as larger than the input layer, with each subsequent layer diminishing in size relative to its predecessor. This approach generated 107 potential model iterations. The optimal iteration within this series estimated the compression depth with an average absolute deviation of 3.9mm and the compression count with an average absolute deviation of 0.8 counts. Compared to the results from the tenth experiment, the eleventh experiment showcased superior performance, yielding ten models that outperformed the top model from experiment ten.

In the twelfth experiment, modifications were applied to two hyperparameters: epochs were adjusted to [1000,2000], and batch sizes were set to [256,1024]. This trial yielded slight enhancements in predicting both compression depth and count over its predecessor. The premier iteration from this experiment achieved a compression depth prediction with an average absolute loss of 3.8mm and a compression count prediction with an average absolute loss of 0.8 counts.

Summarizing the extensive testing, the most effective model across all 1,226 iterations was characterized by the following hyperparameters: hidden layers configured as [2000,1000,250,100,50], a batch size of 256, 2,000 epochs, and a dropout rate of 10% after every two layers. Predictions from this model demonstrated an average absolute loss in compression depth of 3.8mm and in compression count of 0.8 counts, with a median absolute loss for compression depth at 2.59mm and for compression count at 0.03 counts. Throughout the course of the experiments, 27 models were identified as capable of predicting an average absolute compression depth below 4.0mm, highlighting the meticulous optimization process undertaken to refine CPR assessment accuracy based on smartwatch data.

Discussion

High-quality Cardiopulmonary Resuscitation markedly enhances the likelihood of survival for patients experiencing sudden cardiac arrest. Achieving high-quality CPR involves administering chest compressions at a frequency of 100–120 per minute and ensuring a compression depth of 50–60mm [19,20]. In emergency situations, expecting laypersons to accurately count compressions and gauge their depth is impractical. Numerous studies have highlighted significant enhancements in CPR quality when performers receive real-time feedback through intelligent technologies [37,54,55]. Addressing this challenge, the current study introduces a neural network model capable of utilizing accelerometer data from a smartwatch to guide users in delivering high-quality CPR. To facilitate this, applications for both Apple Watch and iPhone were developed to gather accelerometer data during CPR. This research compiled data from 83 participants employing Apple Watches to perform CPR on a mannequin, resulting in a total of 27,844 chest compressions.

Principal Results

Considering mannequin data as the benchmark, the neural network was trained with the accelerometer data to emulate the mannequin's feedback. After undergoing 1,226 iterations of model refinement, the optimal model emerged, capable of estimating compression depth with an average discrepancy of 3.8mm and compression count with an average discrepancy of 0.8 counts across five-second intervals. This investigation underscores the potential of a neural network model to surpass existing models in predicting CPR quality, underpinned by a more extensive and realistic dataset than those previously reported in the literature [37,40,41,55–57].

Smartwatches, widely adopted for personal health and activity monitoring, hold the potential for hosting neural network models capable of executing computationally intensive tasks directly on the device. Such advancements would enable the proposed model to operate on smartwatches, offering immediate feedback to individuals performing CPR and significantly boosting the survival prospects for sudden cardiac arrest victims.

Future directions for this research include deploying the model directly onto smartwatches to assess the impact of real-time feedback on the quality of CPR performed by users. Plans are underway to incorporate a broader range of smartwatches, including models from Samsung Galaxy and Google, and to collaborate with Apple developers to address the limitations posed by the device's operating system. Additionally, enhancing the dataset to include scenarios where CPR is administered while standing or on softer surfaces, like mattresses, is anticipated.

Limitations

The study presents several limitations worth noting. Firstly, the developed model focuses on measuring compression depth and count but does not account for compression recoil, a critical element of effective CPR that facilitates proper chest re-expansion and enhanced air intake. Secondly, the Apple Watch Series 7 was the exclusive device used; given the variability in accelerometer sensors across different smartwatch models, modifications to the model might be necessary to accommodate data from Android devices or other smartwatches. The observed data frequency fluctuated between 48–60Hz, deviating from the expected 60Hz, indicating potential discrepancies in data collection. Additionally, an inherent constraint within the Apple Watch's operating system, which automatically shuts down any active application after 30 seconds of inactivity, poses a challenge for continuous CPR monitoring.

During data segmentation, the division based on a fixed number of data points (300) raised the possibility of inaccurately splitting compressions, potentially skewing the compression count and depth by up to two counts.

Conclusions

This study presents a neural network model meticulously designed to gauge the rate and depth of CPR compressions accurately. The model has undergone training on an exceptionally large dataset, ensuring a robust foundation for its predictive capabilities. Its accuracy is remarkably close to the measurements obtained from a CPR mannequin, serving as a benchmark for comparison. Notably, this model demonstrates superior performance when compared with other models discussed within the existing literature, establishing it as a significant advancement in the field of medical emergency training and response analysis. Implementation of this model in real-world scenarios could significantly improve SCA survival rates.

Acknowledgements

We would like to thank Leanne Baird and the Ibn Sina Simulation Lab at the Thunder Bay Regional Health Sciences Centre for providing technical support and access to the simulation mannequin.

Conflicts of Interest

The authors of this paper do not have any conflict of interest. This research was funded by the MITACS grant and received by Gaurav Rao and Avive Solutions Inc. Other research costs are covered by the NSERC Discovery Grant (RGPIN-2017-05377), held by Dr. Vijay Mago.

Abbreviations

AED: Automated External Defibrillator
AHA: American Heart Association
CPR: Cardiopulmonary resuscitation
CSV: Comma-separated value
EMS: Emergency Medical Services
SCA: Sudden Cardiac Arrest
TBRHSC: Thunder Bay Regional Health Sciences Centre

References

1. Hosey RG, Armsey TD. Sudden cardiac death. Clinics in Sports Medicine. 2003;22(1):51–66.
2. Medlineplus.gov . Sudden Cardiac Arrest — Sudden Cardiac Death — MedlinePlus. 2022. https://medlineplus.gov/suddencardiacarrest.html, Accessed September 16, 2022.
3. Heart.org F. 2021 Heart Disease and Stroke Statistics Update Fact Sheet At-aGlance. 2022. https://www.heart.org/-/media/phd-files-2/science-news/ 2/2021-heart-and-stroke-stat-update/2021_heart_disease_and_stroke_ statistics_update_fact_sheet_at_a_glance.pdf, Accessed September 16, 2022.
4. Research CIoH. Research program intended to help Canadians survive sudden cardiacarrest expands to all 10 Canadian provinces - CIHR. 2019. https://cihr-irsc.gc.ca/ e/50090.html, Accessed September 16, 2022.
5. Heart.org C. What is CPR?. 2022. https://cpr.heart.org/en/resources/ what-is-cpr, Accessed September 16, 2022.
6. American Heart Association . Emergency Treatment of Cardiac Arrest. Last Accessed15 January 2019.
7. American Heart Association . Every Second Counts AED fact sheet 2014. 2014. LastAccessed 24 October 2018.
8. Valenzuela TD, Roe DJ, Cretin S, Spaite DW, Larsen MP. Estimating Effectiveness ofCardiac Arrest Interventions. Circulation. 1997;96(10):3308–3313.
9. Weisfeldt ML, Sitlani CM, Ornato JP, et al. Survival after application of automatic external defibrillators before arrival of the emergency medical system: evaluation in the resuscitation outcomes consortium population of 21 million. Journal of the American College of Cardiology. 2010;55(16):1713–1720.
10. Brooks SC, Simmons G, Worthington H, Bobrow BJ, Morrison LJ. The PulsePoint Respond mobile device application to crowdsource basic life support for patients with out-of-hospital cardiac arrest: Challenges for optimal implementation. Resuscitation. 2016;98:20–26.
11. Van de Voorde P, Gautama S, Momont A, Ionescu CM, De Paepe P, Fraeyman N. The drone ambulance [A-UAS]: golden bullet or just a blank?. Resuscitation. 2017;116:46–48.
12. Chien CY, Tsai SL, Tsai LH, et al. Impact of transport time and cardiac arrest centers on the neurological outcome after out-of-hospital cardiac arrest: A retrospective cohort study. Journal of the American Heart Association. 2020;9(11).
13. Ko SY, Ro YS, Shin SD, Song KJ, Hong KJ, Kong SY. Effect of a first responder on survival outcomes after out-of-hospital cardiac arrest occurs during a period of exercise in a public place. PLoS ONE. 2018;13(2).

14. Andelius L, Hansen C, Lippert F, et al. 40 Long ambulance response time is associated with higher incidence of cardiopulmonary resuscitation and defibrillation by dispatched citizen first-responders. BMJ Open. 2019;9(Suppl 2):A15—-A15.
15. Rea T, Blackwood J, Damon S, Phelps R, Eisenberg M. A link between emergency dispatch and public access AEDs: potential implications for early defibrillation. Resuscitation. 2011;82(8):995–998.
16. Sanko S, Kashani S, Lane C, Eckstein M. Implementation of the Los Angeles TieredDispatch System is associated with an increase in telecommunicator-assisted CPR. Resuscitation. 2020;155:74–81.
17. Andelius L, Hansen CM, Lippert F, et al. Will shorter distance between dispatched layrescuer and out-of-hospital cardiac arrest increase cardiopulmonary resuscitation and early defibrillation rates?. Resuscitation. 2018;130:e56–e57.
18. Pijls RW, Nelemans PJ, Rahel BM, Gorgels AP. A text message alert system for trainedvolunteers improves out-of-hospital cardiac arrest survival. Resuscitation. 2016;105:182– 187.
19. Heart.org G. New resuscitation guidelines update CPR chest pushes. 2015. https://www.heart.org/en/news/2018/05/01/new-resuscitation-guidelines-update-cpr-chest-pushes, Accessed September 16, 2022.
20. Perkins GD, Handley AJ, Koster RW, et al. European Resuscitation Council Guidelines for Resuscitation 2015. Resuscitation. 2015;95:81–99.
21. Z`egre-Hemsey JK, Grewe ME, Johnson AM, et al. Delivery of Automated External Defibrillators via Drones in Simulated Cardiac Arrest: Users' Experiences and the HumanDrone Interaction. Resuscitation. 2020;157:83–88.
22. Z`egre-Hemsey JK, Bogle B, Cunningham CJ, Snyder K, Rosamond W. Delivery of Automated External Defibrillators (AED) by Drones: Implications for Emergency Cardiac Care. Current Cardiovascular Risk Reports. 2018;12(11).
23. Tanaka S, Tsukigase K, Hara T, et al. Effect of real-time visual feedback device 'Quality Cardiopulmonary Resuscitation (QCPR) Classroom' with a metronome sound on layperson CPR training in Japan: A cluster randomized control trial. BMJ Open. 2019;9(6).
24. Baldi E, Cornara S, Contri E, et al. Real-time visual feedback during training improves laypersons' CPR quality: A randomized controlled manikin study. Canadian Journal of Emergency Medicine. 2017;19(6):480–487.
25. Wang SA, Su CP, Fan HY, Hou WH, Chen YC. Effects of real-time feedback on cardiopulmonary resuscitation quality on outcomes in adult patients with cardiac arrest: A systematic review and meta-analysis. Resuscitation. 2020;155:82–90.
26. Rao G, Savage DW, Mago V, Lingras P. A Survey on Technologies Used During out of Hospital Cardiac Arrest. HEALTHINF 2023 - 16th International Conference on Health Informatics. 2023;5:477–488.
27. Hardeland C, Sk˚are C, Kramer-Johansen J, et al. Targeted simulation and education to improve cardiac arrest recognition and telephone assisted CPR in an emergency medical communication centre. Resuscitation. 2017;114:21–26.
28. Sanko S, Feng S, Lane C, Eckstein M. Comparison of Emergency Medical DispatchSystems for Performance of Telecommunicator-Assisted Cardiopulmonary Resuscitation among 9-1-1 Callers with Limited English Proficiency. JAMA Network Open. 2021;4(6).
29. Hambly H, Rajabiun R. Rural broadband: Gaps, maps and challenges. Telematics and Informatics. 2021;60:101565.
30. Lee HS, You K, Jeon JP, Kim C, Kim S. The effect of video-instructed versus audioinstructed dispatcher-assisted cardiopulmonary resuscitation on patient outcomes following out of hospital cardiac arrest in Seoul. Scientific Reports. 2021.
31. Meinich-Bache Ø, Engan K, Birkenes TS, Myklebust H. Real-time chest compression

quality measurements by smartphone camera. Journal of Healthcare Engineering. 2018;2018.
32. Ali S, Athar M, Ahmed S. A randomised controlled comparison of video versus instructor-based compression only life support training. Indian Journal of Anaesthesia. 2019;63(3):188–193.
33. Jeon SA, Chang H, Yoon SY, et al. Effectiveness of smartwatch guidance for highquality infant cardiopulmonary resuscitation: A simulation study. Medicina (Lithuania). 2021;57(3):1–10.
34. Sevil H, Bastan V, Gu¨ltu¨rk E, El Majzoub I, Go¨ksu E. Effect of smartphone applicationson cardiopulmonary resuscitation quality metrics in a mannequin study: A randomized trial. Turkish Journal of Emergency Medicine. 2021;21(2):56–61.
35. An M, Kim Y, Cho WK. Effect of smart devices on the quality of CPR training: A systematic review. Resuscitation. 2019;144:145–156.
36. Song Y, Chee Y, Oh J, Ahn C, Lim TH. Smartwatches as chest compression feedback devices: A feasibility study. Resuscitation. 2016;103:20–23.
37. Semeraro F, Taggi F, Tammaro G, Imbriaco G, Marchetti L, Cerchiari EL. iCPR: A new application of high-quality cardiopulmonary resuscitation training. Resuscitation. 2011;82(4):436–441.
38. Chan T, Wan K, Chan J, Lam H, Wong Y, Kan P. New Era of CPR: Application of I-Technology in Resuscitation. Hong Kong Journal of Emergency Medicine. 2012;19(5):305–311.
39. Park CS, Kang IG, Heo SJ, et al. A randomised, cross over study using a mannequin model to evaluate the effects on CPR quality of real-time audio-visual feedback provided by a smartphone application. Hong Kong Journal of Emergency Medicine. 2014;21(3):153–160.
40. Gruenerbl A, Pirkl G, Monger E, Gobbi M, Lukowicz P. Smart-Watch life saver: SmartWatch interactive-feedback system for improving bystander CPR. ISWC 2015 - Proceedings of the 2015 ACM International Symposium on Wearable Computers. 2015:19–26.
41. Lu TC, Chen Y, Ho TW, et al. A novel depth estimation algorithm of chest compression for feedback of high-quality cardiopulmonary resuscitation based on a smartwatch. Journal of Biomedical Informatics. 2018;87:60–65.
42. Asada HH, Jiang HH, Gibbs P. Active noise cancellation using MEMS accelerometersfor motion-tolerant wearable bio-sensors. Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings. 2004;26 III:2157–2160.
43. Riou M, Ball S, O'Halloran KL, et al. Hijacking the dispatch protocol: When callers preempt their reason-for-the-call in emergency calls about cardiac arrest. Discourse Studies. 2018;20(5):666–687.
44. Case R, Cartledge S, Siedenburg J, et al. Identifying barriers to the provision of bystander cardiopulmonary resuscitation (CPR) in high-risk regions: A qualitative review of emergency calls. Resuscitation. 2018;129:43–47.
45. King CE, Sarrafzadeh M. A Survey of Smartwatches in Remote Health Monitoring. Journal of Healthcare Informatics Research. 2018;2(1-2):1–24.
46. Apple . App Store - Apple (CA). 2023. https://www.apple.com/
47. Laerdal.com . Laerdal Medical. 2023. https://laerdal.com/
48. Statista.com . Statista Mobile OS share. 2023. https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/.
49. Statista.com . Statista Watch OS share. 2023. https://www.statista.com/statistics/910862/worldwide-smartwatch-shipment-market-share/
50. Laerdal.com . Laerdal Medical AnneQCPR. 2023. https://laerdal.com/ca/products/simulation-training/resuscitation-training/little-anne-qcpr/

51. Amazon Web Services. G5 instance types. 2023. https://aws.amazon.com/ec2/instance-types/g5/
52. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020;17:261–272.
53. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Software available from tensorflow.org.
54. Truszewski Z, Szarpak L, Kurowski A, et al. Randomized trial of the chest compressions effectiveness comparing 3 feedback CPR devices and standard basic life support by nurses. American Journal of Emergency Medicine. 2016;34(3):381–385.
55. Song Y, Oh J, Chee Y. A New Chest Compression Depth Feedback Algorithm for High-Quality CPR Based on Smartphone. Telemedicine and e-Health. 2015;21(1):36–41.
56. Ahn C, Lee J, Oh J, et al. Effectiveness of feedback with a smartwatch for high-quality chest compressions during adult cardiac arrest: A randomized controlled simulation study. PLOS ONE. 2017;12(4):e0169046.
57. Lu TC, Chang YT, Ho TW, et al. Using a smartwatch with real-time feedback improves the delivery of high-quality cardiopulmonary resuscitation by healthcare professionals. Resuscitation. 2019;140:16–22.