Computers in Biology and Medicine 188 (2025) 109798

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Multitask learning approach for PPG applications: Case studies on signal quality assessment and physiological parameters estimation

Mohammad Feli [a,*], Kianoosh Kazemi [a], Iman Azimi [b], Pasi Liljeberg [a], Amir M. Rahmani [b,c]

[a] Department of Computing, University of Turku, Turku, Finland
[b] Department of Computer Science, University of California, Irvine, USA
[c] School of Nursing, University of California, Irvine, USA

ARTICLE INFO

Keywords:
Photoplethysmography
Multitask learning
Wearable devices
Health monitoring
Deep learning

ABSTRACT

Wearable technology has expanded the applications of photoplethysmography (PPG) in remote health monitoring, enabling real-time measurement of various physiological parameters, such as heart rate (HR), heart rate variability (HRV), and respiration rate (RR). While existing studies mainly focus on individual parameters derived from PPG, they often overlook the shared characteristics among these physiological parameters. Multitask learning (MTL) offers a promising solution by training a single model to perform multiple related tasks, leveraging their interdependencies. However, the potential of MTL has not been thoroughly investigated in the context of PPG analysis. In this paper, we develop MTL approaches that exploit shared underlying characteristics across PPG-related tasks to improve the performance of PPG-based applications. We propose customized multitask deep learning models for two applications: (1) PPG quality assessment for HR and HRV features collected in free-living conditions and (2) simultaneous HR and RR estimation from PPG. Our models are evaluated on a PPG dataset collected from 46 subjects wearing smartwatches during their daily activities. Results demonstrate that the proposed MTL methods significantly outperform baseline single-task models, achieving higher accuracy in quality assessment and reduced error rates in HR and RR estimation.

1. Introduction

Recent advancements in wearable technologies have significantly widened the applications of photoplethysmography (PPG) in remote health monitoring systems. PPG is a non-invasive optical technique for measuring blood volume changes in the peripheral circulation [1]. This method operates by emitting light onto the skin and recording the amount of light absorbed or reflected by the blood vessels. The intensity of the reflected or transmitted light varies with blood volume changes with each heartbeat, enabling PPG sensors to capture the pulsatile variations that form the PPG signal.

PPG is widely used in health monitoring applications, particularly in commercial wearable devices such as smartwatches, smart rings, and fitness trackers, to monitor cardiovascular health [2]. This technology provides real-time insights into physiological parameters, including heart rate (HR) and heart rate variability (HRV), during everyday activities [3,4]. Additionally, it has been employed to measure respiration rate (RR) [5], blood oxygen saturation (SpO2) [6], and blood pressure [7]. PPG is also utilized in sleep monitoring [8], stress detection [9], mental health monitoring [10], and maternal health assessment [11]. These diverse applications make PPG a versatile tool, enabling the simultaneous extraction of multiple health parameters from a single signal.

Current studies primarily focus on a single physiological parameter when analyzing PPG signals. Such single-task approaches overlook the potential synergies and shared characteristics among these physiological parameters. For instance, various HRV parameters are highly correlated as they all reflect autonomic nervous system activity [12]. Additionally, research has demonstrated an inverse correlation between HR and HRV [13], while HR and RR exhibit a positive correlation, indicating the interplay between the cardiovascular and respiratory systems [14]. Single-task approaches cannot benefit from these interdependencies that could enhance the accuracy and efficiency of health monitoring systems.

To address these limitations, multitask learning (MTL) presents a promising solution. MTL is a machine learning approach where a single model is trained to perform multiple related tasks simultaneously, generating corresponding multiple outputs [15]. Compared to single-task learning (STL), MTL enhances generalization by learning shared features across related tasks, resulting in more robust and generalized

* Corresponding author.
E-mail address: mohammad.feli@utu.fi (M. Feli).

https://doi.org/10.1016/j.compbiomed.2025.109798
Received 2 October 2024; Received in revised form 30 January 2025; Accepted 3 February 2025
Available online 12 February 2025

models. This approach has been successfully applied in various health monitoring applications [16–23].

MTL's applications in health monitoring have primarily focused on electrocardiogram (ECG) signals. These applications include arrhythmia classification [16–19], blood pressure estimation [20], and combinations of ECG with other signals like acceleration (ACC) and PPG, for measuring respiration rate [24] and blood pressure [21]. Despite these advancements, the potential of employing MTL for PPG-specific applications remains largely underexplored. Only a limited number of studies [25–27] have utilized MTL for PPG, primarily focusing on blood pressure estimation. However, they ignore the broader range of PPG applications, including HR, HRV, and RR analysis. Additionally, to the best of our knowledge, most existing research [16–21] has developed MTL methods on data from controlled clinical settings, often overlooking the challenges of noise encountered in real-world, free-living conditions. MTL offers the potential to address these challenges by leveraging the inherent relationships between physiological parameters derived from PPG. We believe further exploration of this approach could significantly enhance the accuracy and reliability of PPG-based health monitoring systems.

In this paper, we develop MTL approaches for multivariate PPG-based health monitoring. These approaches are designed to leverage shared underlying characteristics across different PPG-related tasks. To achieve this, we propose customized multitask deep learning models that harness both the spatial and temporal dependencies within PPG signals. Our MTL models are implemented for two case studies: one for HR/HRV-guided PPG signal quality assessment (SQA) and another for HR and RR estimation from PPG. For the SQA case study, we develop a Convolutional Neural Networks-based MTL model to classify PPG signals into "Reliable" and "Unreliable" classes according to their reliability for the HR and HRV parameters. In the second case study, we design another MTL model to simultaneously estimate HR and RR from PPG signals. We evaluate the performance of these MTL models using a dataset comprising PPG, ECG, and ACC signals recorded from 46 subjects via smartwatches in free-living conditions. ECG signals serve as the reference for quality assessment and HR estimation, while the ACC signal is considered the reference for RR estimation. We benchmark our MTL models against state-of-the-art STL methods. In summary, the major contributions of the paper are as follows:

1. Development of MTL approaches for PPG-based health monitoring to utilize the shared characteristics among various PPG tasks.
2. Design of two MTL models for two PPG applications: PPG SQA for HR and HRV parameters, and simultaneous HR and RR estimation from PPG signals.
3. Evaluating our MTL models using a dataset including PPG signals collected using wearable devices in free-living conditions.
4. Benchmarking the performance of our MTL approaches against baseline STL methods to demonstrate their accuracy.

2. Related works and motivation

Various studies have applied MTL technique to remote health monitoring, with a primary focus on ECG signals [16–20]. For instance, Hughes et al. [16] developed an MTL method for heartbeat classification using 12-lead ECG recordings from a clinical dataset. Similarly, Shahin et al. [17] introduced an adversarial MTL approach for ECG heartbeat classification. Geng et al. [18] proposed an ECG classification approach based on the MTL technique and contextual transformer attention mechanism. In another study [19], an MTL model was introduced for simultaneous arrhythmia reconstruction and classification using ECG signals. Additionally, in a study by Fan et al. [20], an MTL approach was developed for continuous blood pressure estimation from clinical ECG data.

Beyond ECG-specific approaches, a few research studies proposed MTL methods on combinations of ECG with other signals like ACC and PPG. For example, Rathore et al. [24] introduced an MTL method for respiration rate estimation by integrating ECG and ACC signals. In another study by Jeong et al. [21], an MTL method was developed for continuous blood pressure estimation using features extracted from ECG and PPG signals.

In addition, MTL techniques have been employed for heart rate estimation using remote PPG (rPPG), where physiological parameters are captured from the skin using camera-based technologies [28]. Huang et al. [23] proposed an MTL model to estimate infants' HR and rPPG signals from clinical videos. Similarly, Yin et al. [22] developed another MTL model to estimate HR and rPPG signals from facial video data while focusing on reducing noise and interference. Additionally, MTL has been explored in other domains, such as sleep monitoring from ECG and PPG [29] and stress assessment using ECG signals [30].

While most MTL applications in health monitoring have focused on ECG signals—particularly for tasks like arrhythmia classification [16–19]—or on combinations of ECG with other signals [21,24], there has been limited exploration of MTL specifically for PPG-based applications. A few studies [25–27] have demonstrated the effectiveness of MTL for predicting blood pressure using PPG as a single input. However, these limited efforts merely open the door to further possibilities. PPG signals are widely used for various applications, such as measuring HR, HRV, and RR, which presents an opportunity for MTL to harness the inherent correlations between these physiological parameters.

To emphasize the interconnections between the parameters extracted from PPG, we provide a motivational example through correlation analysis of HR and HRV features within a PPG dataset (details provided in Section 4.4.1). The Pearson and Spearman correlation matrices for HR and HRV parameters in this dataset (n=84,032, p-values<0.0001) are shown in Fig. 1. This analysis reveals strong inverse relationships between HR and several HRV metrics, such as average of NN intervals (AVNN) (Pearson: −0.94, Spearman: −0.99), standard deviation of NN intervals (SDNN) (Pearson: −0.57, Spearman: −0.55), and root mean square of successive NN intervals differences (RMSSD) (Pearson: −0.66, Spearman: −0.72). Additionally, significant positive correlations are observed among HRV parameters, including AVNN, SDNN, and RMSSD. By incorporating these correlations, an MTL approach can harness shared features, potentially improving performance across multiple PPG-related tasks.

Moreover, most existing MTL methods in health monitoring [16–21] have been developed using clinical data, which often fails to account for the challenges posed by noise and artifacts in real-world, free-living conditions. PPG signals, in particular, are vulnerable to motion artifacts and environmental noise when collected via wearable devices during everyday activities. These challenges necessitate tailored MTL methods to ensure accurate and reliable PPG-derived measurements under such conditions. We believe there is a significant opportunity to advance the development of MTL techniques to enhance the robustness and effectiveness of PPG-based health monitoring methods.

3. Background

3.1. Photoplethysmogram (PPG)

PPG is a non-invasive optical technique for measuring blood volume changes in the microvascular bed of tissues [1]. It involves illuminating the skin with a light-emitting diode (LED) and capturing the amount of light absorbed or reflected by the blood vessels. In the absorption method, a photodetector measures the intensity of light passing through the tissue, while in the reflection method, the photodetector captures the intensity of the light reflected from the tissue. The pulsatile changes in light intensity caused by variations in blood volume with each heartbeat generate the PPG signal.

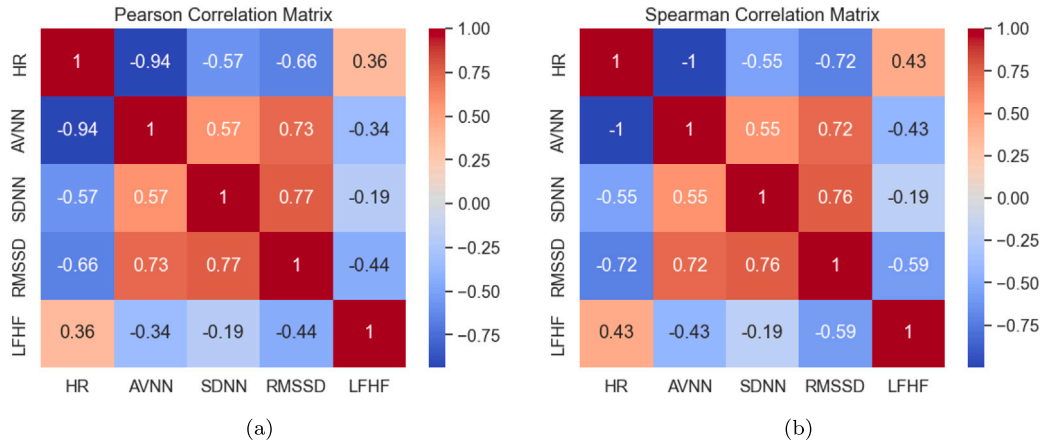


Fig. 1. Pearson (a) and Spearman (b) Correlation matrices of HR and HRV parameters ($n = 84,032$, p-values < 0.0001).

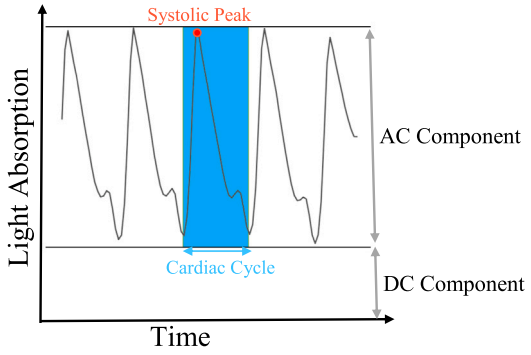


Fig. 2. A typical PPG signal with its components.

Fig. 2 shows a typical PPG signal, indicating its two main components: the alternating current (AC) and the direct current (DC) components. The AC component reflects cardiovascular fluctuations synchronized with each heartbeat and is essential for deriving parameters such as HR and HRV. These parameters are estimated from the time intervals between successive peaks, similar to ECG analysis [31,32]. The DC component represents the slowly varying baseline level of blood volume and is also utilized to assess respiratory functions. The combination of AC and DC components makes PPG a convenient tool for assessing various physiological parameters for health monitoring applications.

3.2. PPG Signal Quality Assessment (SQA)

Although PPG is a popular method for remote health monitoring, it is highly susceptible to motion artifacts and noise [33]. Factors such as subject movement, environmental interferences, and poor sensor contact can compromise signal quality, particularly in real-world, free-living conditions. When PPG signals are corrupted, vital signs cannot be accurately extracted, potentially leading to incorrect decisions. Therefore, assessing PPG signal quality post-collection is crucial for distinguishing between reliable and unreliable signals, preventing erroneous decision-making and potentially life-threatening consequences.

Numerous studies have proposed methods for PPG SQA [34–39]. Some approaches involve developing decision rules based on physiological parameters and signal characteristics to differentiate between reliable and unreliable signals [34,35]. Additionally, conventional machine learning techniques utilizing handcrafted feature extraction have been proposed for PPG SQA [36,37]. Recent research has also developed deep learning models, demonstrating high accuracy in assessing signal quality [40,41]. These advancements are crucial for ensuring the reliability of the PPG method for health monitoring systems.

4. Methodology

In this section, we develop MTL approaches for PPG-based health monitoring tasks. First, we outline the foundational concept of the MTL technique and its application in our study. We then present the deep learning architectures used for multitask analysis to harness both the spatial and temporal dependencies within PPG signals. Subsequently, we design MTL models for two case studies: (1) HR/HRV-guided PPG SQA and (2) HR and RR estimation from PPG. Each case study highlights the application of MTL in addressing specific challenges in PPG-based health monitoring. Finally, we provide implementation details for the proposed MTL models to ensure clarity and reproducibility.

4.1. Multitask learning (MTL)

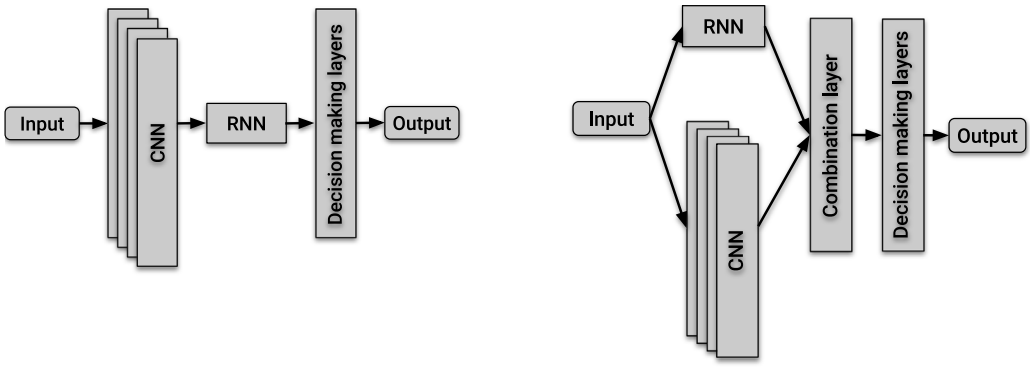
In our study, we develop MTL approaches to leverage potential shared underlying characteristics among PPG-related tasks. By capturing the inherent correlation between these tasks, MTL can improve the performance of PPG-based methods. While MTL is not exclusively developed with deep learning techniques, it is commonly associated with deep learning due to the ability of deep neural networks to learn and generalize shared representations across multiple tasks. In this work, we design two deep learning models to apply MTL for PPG PPG-related tasks.

MTL aims to train a single model to perform multiple interrelated tasks simultaneously, utilizing shared representations between tasks [15]. Instead of training separate models for each task, MTL allows the model to learn and extract advantageous features for all tasks involved. This shared learning process can lead to better generalization and improved performance across tasks, particularly in scenarios where interdependencies and shared characteristics exist.

During the training process, a joint objective function that encompasses all concurrently learned tasks is optimized. Throughout this process, the model receives input data for all tasks and adjusts its parameters to minimize the aggregate loss across tasks. By adopting this joint optimization strategy, the model is able to extract shared patterns across tasks while also capturing task-specific parameters.

4.2. Deep learning architectures for multitask analysis of PPG

We design hybrid Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) models to apply MTL for PPG-related tasks. CNN layers efficiently exploit the spatial aspects of PPG signals, while RNN layers process the temporal dependencies and dynamics inherent in the cardiovascular activity captured by PPG. This combination enables a more robust and accurate analysis for different PPG tasks.



(a) Serial Combination (b) Parallel Combination

Fig. 3. Overview of hybrid CNN-RNN architectures for multitask PPG analysis.

Various architectures can be employed to develop deep learning models for PPG analysis [42–44]. Common architectures include CNNs, RNNs, transformer models, and hybrid models. Hybrid architectures combine the strengths of multiple model types. They are effective for timeseries applications, as they capture different patterns in the data, enhancing generalization and robustness.

One of the most popular architectures for hybrid models is the combination of CNN and RNN layers. CNN layers extract spatial features, while RNN layers derive temporal features from the signal. Integrating these two types of layers ensures comprehensive feature extraction. In a hybrid model, CNN and RNN layers can be combined in two main ways: serially and in parallel. In a serial combination, the model connects the layers sequentially, where the output of one layer type feeds directly into the other type. Typically, CNN layers are positioned at the beginning, followed by the RNN layers. In a parallel model, CNN and RNN layers process the same input simultaneously but independently. The outputs from these layers are then merged and commonly fed to subsequent decision-making layers of the model. Fig. 3 provides an overview of these serial and parallel CNN-RNN architectures, which serve as the foundation for the MTL approaches developed in this study. The detailed design of these architectures is discussed in the following section.

4.3. Methodological case studies

In this section, we present the proposed MTL methods for two case studies in PPG-based health monitoring. First, we develop a multitask deep learning model employing a serial CNN-RNN hybrid architecture to assess PPG signal quality for HR and HRV parameters. Subsequently, we develop another MTL approach for estimating HR and RR from PPG using a parallel CNN-RNN hybrid model configuration.

4.3.1. Case study 1: MTL approach for HR/HRV-guided PPG sqa

We propose an MTL method to assess the quality of PPG signals for HR and HRV parameters. Our objective is to design an accurate MTL model capable of classifying PPG signals as either "Reliable" or "Unreliable" for HR and various HRV metrics simultaneously. While a PPG signal might demonstrate reliability for one parameter (e.g., HR), it could be unreliable for another (e.g., HRV metrics) [45]. Therefore, an effective SQA approach must consider the signal's reliability across various parameters. To address this challenge, we propose an MTL technique that concurrently evaluates PPG signal quality for both HR and several HRV parameters. This approach leverages shared underlying characteristics among HR and HRV parameters, allowing the model to exploit correlations and enhance classification accuracy. In the following, we provide a detailed description of the development of this methodological case study.

Data Preparation: To develop the MTL SQA model, we construct a dataset where PPG signals serve as inputs, and the corresponding quality assessments for each HR and HRV parameter are utilized as labels. Fig. 4 illustrates the pipeline of this data preparation process. Below, we describe the various components of this pipeline.

- **Preprocessing:** This initial stage involves the segmentation and filtering of both the PPG and ECG signals. We employ a sliding window approach to divide the signals into 5-minute segments with a stride of 10 s. This window duration is chosen based on established literature, indicating its suitability for extracting HRV parameters [12]. Subsequently, the signals are filtered to remove unwanted frequencies, particularly baseline wander, using a Butterworth high-pass filter with a 0.5 Hz cutoff.
- **Peak Detection:** We identify systolic peaks in the PPG signals and R-peaks in the ECG signals. These peaks and the time intervals between them are crucial for extracting HR and HRV parameters. We utilize the Elgendi et al. method [46] for peak detection. To ensure accuracy, we adopt a rule-based filter, wherein any R-peak deviating by more than 20% from the average RR intervals is considered a false peak. We then discard ECG segments containing more than 30% false peaks. It should be noted that the removal of an ECG segment due to false peaks results in the elimination of its corresponding PPG segment.
- **HR/HRV Extraction:** In this stage, HR and HRV parameters are computed from the peaks and inter-beat intervals (IBI) derived in the previous step. We compute HRV parameters, including root mean square of successive NN intervals differences (RMSSD), average of NN intervals (AVNN), standard deviation of NN intervals (SDNN), percentage of successive NN intervals that differ by more than 50 ms (pNN50), and the ratio of low-frequency to low-frequency power (LFHF) [12].
- **Annotation:** PPG signals are annotated into the "Reliable" and "Unreliable" categories for each HR and HRV parameter. We consider ECG signals as the gold reference and calculate the pairwise absolute error between the HR and HRV parameters extracted from PPG and the reference ECG. Error thresholds for each parameter are defined based on the range of the feature [47]. Consequently, a PPG signal is labeled as "Unreliable" for a parameter if its error with reference ECG exceeds the predefined error threshold of that parameter.

  Let us denote $PPG_{HRV} = [ppg_{hrv_1}, ppg_{hrv_2}, \ldots, ppg_{hrv_n}]$, where ppg_{hrv_i} represents the i_{th} HRV parameter extracted from the PPG signal. The pairwise absolute error between the PPG and ECG signals for an HRV parameter can be calculated as:

$$\text{Error}_{hrv_i} = |\text{ppg}_{hrv_i} - \text{ecg}_{hrv_i}| \quad (1)$$

  where ecg_{hrv_i} represents the ith HRV parameters extracted from the reference ECG. If the absolute error for an HRV parame-

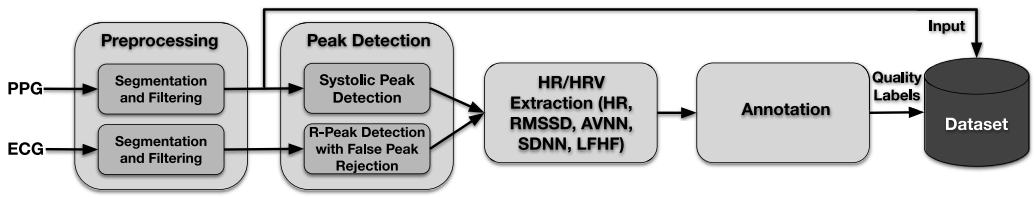


Fig. 4. Data preparation pipeline for case study 1: PPG SQA.

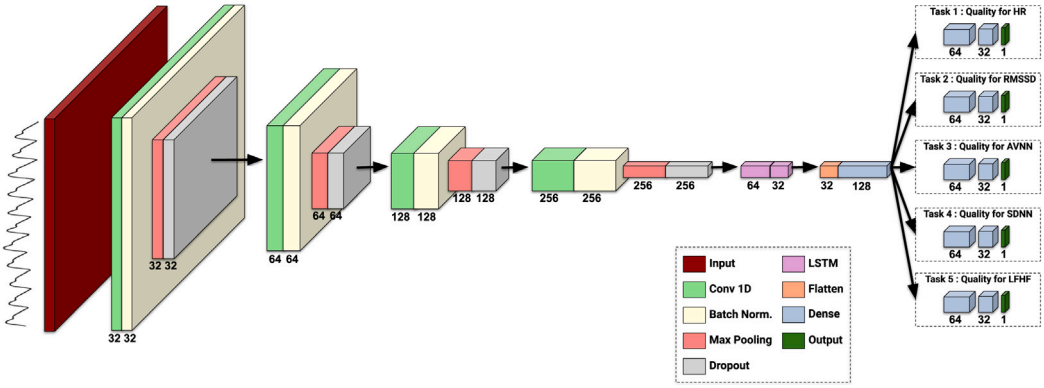


Fig. 5. Proposed multitask deep learning model for HR/HRV-guided PPG SQA, utilizing a serial hybrid architecture.

ter exceeds the error threshold τ_{hrv_i}, the PPG signal is labeled as "Unreliable" for that parameter; otherwise, it is labeled as "Reliable":

$$\text{PPG Signal Quality}_{hrv_i} = \begin{cases} \text{"Unreliable"} & \text{if Error}_{hrv_i} > \tau_{hrv_i} \\ \text{"Reliable"} & \text{otherwise} \end{cases} \quad (2)$$

MTL SQA Model: We develop a multitask deep learning model to assess PPG signal quality for HR and various HRV parameters simultaneously. Let $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ be a set of PPG segments, where each X_i is a time series. Our aim is to develop a model $f : \mathbf{X} \rightarrow \mathbf{Y}$ that evaluates the quality of these PPG segments for HR and HRV parameters. The outputs $\mathbf{Y}$ consist of multiple tasks: the signal quality for HR, denoted as y_{hr} and HRV, denoted as y^j_{hrv}, where $j = 1, \ldots, m$ represents different HRV parameters.

For each PPG segment X_i, the model predicts the probability that the segment is "Reliable" or "Unreliable" for HR and each HRV parameter, denoted as $P(Y_{HR} = 1 \mid X_i)$ and $P(Y^j_{HRV} = 1 \mid X_i)$ for each HRV parameter j.

The designed deep learning model is a hybrid CNN-RNN architecture, serially combining CNN layers with Long Short-Term Memory (LSTM) layers. This multitask model leverages shared underlying characteristics among HR and HRV parameters, exploiting correlations between them to assess PPG signal quality. The model's schematic architecture is demonstrated in Fig. 5.

The model consists of several Conv1D layers with different numbers of filters (32, 64, 128, 256) and a filter size of 3, followed by BatchNormalization, MaxPooling1D, and Dropout layers to prevent overfitting. These convolutional layers extract high-level features from the input PPG segments while the pooling layers downsample the spatial dimensions. Dropout layers also regularize the model and prevent it from being influenced by noise in the training data. Following the convolutional layers, two LSTM layers with 64 and 32 units are added to capture temporal dependencies. Subsequently, the output of the LSTM layers is flattened and passed through a fully connected dense layer with the ReLU activation function to prepare the extracted features for the final task-specific output layers.

The model includes multiple output branches, each corresponding to a specific task (HR and HRV parameters). For each task, we add dense layers with ReLU activation functions to further process the shared features. Finally, output layers comprise dense layers with a single neuron and sigmoid activation functions to predict the probability of the input PPG segment belonging to either the "Reliable" or "Unreliable" class.

4.3.2. Case study 2: MTL approach for HR and RR estimation from PPG

We propose an MTL approach for simultaneously estimating HR and RR from PPG signals. The aim of this methodological case study is to design an MTL model that leverages the inherent relationships between HR and RR to improve estimation. Both HR and RR offer valuable insights into the cardiovascular and respiratory systems. These two vital signs are related and synchronously fluctuate under normal physiological conditions [14]. Accordingly, our proposed MTL technique aims to leverage these associations for simultaneous HR and RR estimation. Below, we outline the development of this methodological case study.

Data Preparation: We construct a dataset to develop our MTL HR/RR estimation model. This dataset utilizes PPG signals as inputs, with the corresponding HR and RR values serving as the ground-truth labels. The pipeline of this data preparation is demonstrated in Fig. 6. In the following, we present different components of this pipeline.

- **Preprocessing:** In this stage, we first apply segmentation to the PPG, ECG, and triaxial ACC signals by employing a sliding 30-second window with a 2-second stride. Then, we utilize a Butterworth high-pass filter with a cutoff frequency of 0.5 Hz to remove baseline wander noise from the ECG signals. Additionally, the triaxial ACC signals are filtered by a total variation filter introduced by Rudin et al. [48], which has been demonstrated as effective for respiratory rate extraction [49].
- **PPG Stream:** PPG segments are processed to generate inputs for the dataset. We first employ a well-established PPG SQA method proposed in our previous work [38] to determine the reliability of the signal. Reliable segments undergo further processing using a Butterworth high-pass filter with a 0.1 Hz cutoff frequency, preserving relevant information for respiratory analysis. Unreliable segments, along with their corresponding ECG and ACC data, are excluded from further analysis.
- **ECG Stream:** We extract ground truth HR labels from ECG segments. Employing the Elgendi et al. method [46], we identify

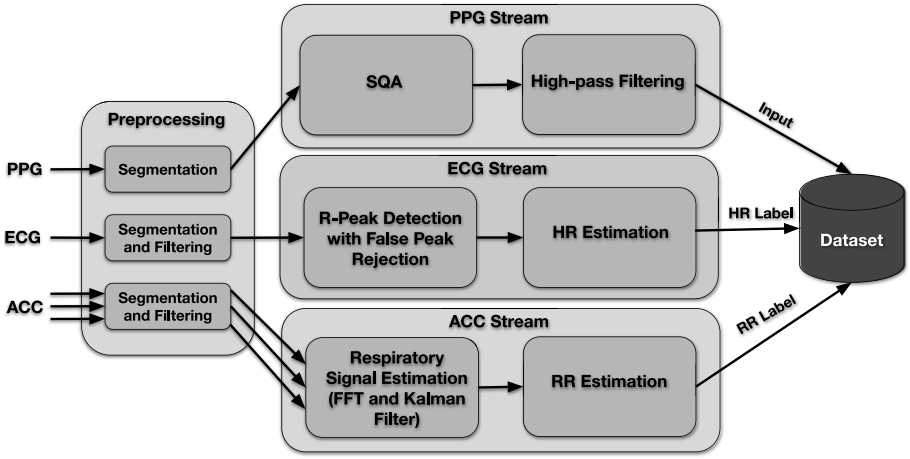


Fig. 6. Data preparation pipeline for case study 2: HR/RR estimation.

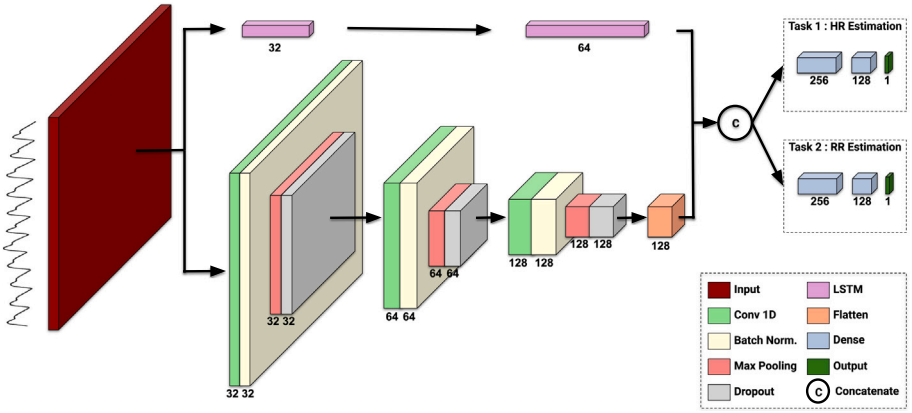


Fig. 7. Proposed multitask deep learning model for HR and RR estimation from PPG, utilizing a parallel hybrid architecture.

R-peaks, followed by a false peak detector where ECG segments containing more than 30% false peaks are removed from further analysis. The HR labels are then computed from the valid R-peaks and assigned as ground-truth HR values in the dataset.

- **ACC Stream:** We derive ground truth RR values from ACC signals using the method proposed by Sun et al. [49]. This method applies a Fast Fourier Transform (FFT) to identify the RR frequency and employs a multi-axis fusion approach by Kalman filter.

MTL HR/RR Estimation Model: We design a multitask deep learning model to estimate both HR and RR concurrently from PPG signals. Let $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ represent a set of PPG segments, where each X_i is a time series. The objective is to develop a model $f : \mathbf{X} \rightarrow \mathbf{Y}$, where $\mathbf{Y} = \{Y_{HR}, Y_{RR}\}$ represents the estimation outputs for HR and RR. Each output Y_{HR_i} and Y_{RR_i} is a continuous value corresponding to the HR and RR estimated from segment X_i, respectively.

For a given PPG segment X_i, the model estimates HR, denoted as $f_{HR}(X_i)$ and simultaneously RR, denoted as $f_{RR}(X_i)$ where f_{HR} and f_{RR} are the HR and RR estimation functions modeled by the network.

The designed model combines CNN and LSTM networks to capture spatial and temporal dependencies in the data. This model exploits the correlation between HR and RR and improves estimation accuracy by capturing shared underlying features between these two parameters. The model's architecture, as illustrated in Fig. 7, features two parallel streams—one consists of CNN layers and the other of LSTM networks.

The CNN stream extracts spatial features from the input PPG through multiple Conv1D layers with different numbers of filters (32, 64, 128) and a filter size of 3, followed by BatchNormalization, MaxPooling1D, and Dropout layers. Meanwhile, the LSTM stream captures temporal dependencies using two LSTM layers with 32 and 64 units, enabling the model to learn sequential patterns. The output of both streams is concatenated to integrate the extracted spatial and temporal features.

This MTL model includes two output branches: one for HR estimation and the other for RR estimation. These branches comprise further processing incorporating two dense layers to capture task-specific features. The output dense layer in each branch includes a single neuron with a linear activation function to predict HR and RR values.

4.4. Implementation

In this section, we detail the implementation of the proposed MTL methods. We describe the dataset used in this study, including the study design, participant information, data collection approach, and ethical considerations. Additionally, we outline the training and testing settings and provide specifics on the model implementation.

4.4.1. Dataset: Study design, participants, data collection, and ethics

We implement our MTL approaches using a dataset from a remote health monitoring study [50]. In this study, participants wore a Samsung Gear Sport watch [51] on their non-dominant hand and a Shimmer3 device [52] on their chest, capturing their vital signs continuously. The data were collected over a 24-hour monitoring period, covering individuals' daily routines and activities.

The study was conducted in southern Finland between July and August 2019. Data were gathered objectively from 46 participants (23 women and 23 men), including students from the University of Turku,

Table 1
Participants' background information.

Characteristic	Type	Values
Age, mean (SD)	Men	33 (6)
	Women	31.5 (6.6)
BMI, mean (SD)	Men	24.4 (5.6)
	Women	25.5 (2.9)
Physical activity, *n* (%)	Almost daily	12 (27)
	Once a week	9 (20)
	>Once a week	21 (47)
Employment status, *n* (%)	Working	32 (71)
	Unemployed	1 (2)
	Student	8 (18)
	Other	1 (2)

along with additional individuals recruited through a snowball sampling technique. Recruitment inclusion criteria included good health with no cardiovascular diseases, an age range of 18 to 55, the ability to use wearable devices at work, and no physical activity limitations. The study's objective was explained to the candidates in face-to-face meetings. Table 1 summarizes the participants' background information, covering 42 individuals, as data for 4 participants is unavailable.

The Samsung Gear Sport watch [51] was used to record PPG signals from the wrist. This watch is lightweight and waterproof, with dimensions of 44.6 ×42.9× 11.6 mm and a total weight of 67 g including the strap, making it suitable for long-term data collection during daily activities. It runs on the open-source Tizen operating system, with a three-day battery life. The watch is equipped with both optical and inertial measurement unit (IMU) sensors, with the optical sensors recording PPG signals at a sampling frequency of 20 Hz.

The Shimmer3 device [52] was utilized to capture ECG and ACC data from the chest. This portable device offers sufficient internal storage and battery life for continuous data collection over a 24-hour period. It was programmed to record 12-channel ECG signals from four electrodes placed on the torso and triaxial ACC data at a sampling frequency of 512 Hz. This high sampling rate ensures accurate and detailed monitoring of both cardiac and accelerometer data.

The participants were instructed to wear the watch on their non-dominant hand and the Shimmer3 device on their chest for a full day. PPG signals were recorded for 16 min every 30 min, while ECG and ACC signals were recorded continuously for 24 h. Data collection was performed as participants engaged in their daily activities.

In our study, we only use Lead II ECG for analysis, serving as the reference for ground-truth HR and HRV calculation. Additionally, chest ACC data is used as the reference to obtain ground-truth RR values. Previous studies have demonstrated that chest-worn ACC is robust and resilient to noise in RR estimation [53,54]. It should be noted that for sensor calibration purposes, the first and last minutes of each recording are discarded in our analysis.

Ethics: This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and the Finnish Medical Research Act (#488/1999). The study protocol was approved by the Ethics Committee for Human Sciences at the University of Turku (Statement #44/2019). Participants were informed about the study both orally and in writing before giving their consent. Participation was voluntary, and all participants had the right to withdraw from the study at any time without giving a reason.

4.4.2. Training and testing settings

We performed inter-patient experiments to evaluate the proposed MTL methods, in which different participants were chosen for training and testing. This approach ensures robust model generalization across different individuals. We employ 24 h of data from 22 individuals for training and the same duration from 10 individuals for testing, applied to both SQA and HR/RR estimation models.

Table 2
Number of segments in training and test sets for PPG SQA.

Parameter	Training set		Test set	
	Reliable	Unreliable	Reliable	Unreliable
HR	24 839	32 749	12 336	14 108
RMSSD	9698	47 890	4741	21 703
AVNN	25 953	31 635	12 934	13 510
SDNN	11 772	45 816	5433	21 011
LFHF	9698	47 890	4741	21 703

For the SQA model, the training set includes 57,588 five-minute segments, while the testing set contains 26,444 five-minute segments of each PPG and ECG signal. Details of the distribution of "Reliable" and "Unreliable" labels across various HR and HRV parameters are presented in Table 2.

For the HR/RR estimation model, the training data comprises 52,339 thirty-second segments, and the testing data includes 27,328 thirty-second segments of each PPG, ECG, and ACC signal. Histograms illustrating the HR and RR values in our training and test sets are shown in Fig. 8 and Fig. 9, respectively.

The models were trained on a Linux machine equipped with an AMD Ryzen Threadripper 2920X 12-Core processor, 126 GB of RAM, and an NVIDIA TITAN RTX GPU with 24 GB of memory. We implemented the models using Keras Sequential API from TensorFlow on Python.

For the SQA model, we utilized the Adam optimizer with a learning rate of 0.001 and a binary cross-entropy loss function for each task output. Similarly, the HR/RR estimation model was compiled with the Adam optimizer at the same learning rate, but using a mean squared error loss function for the HR and RR estimation outputs. Both models were trained for up to 100 epochs with a batch size of 5.

To monitor model performance and prevent overfitting, we employed early stopping, configured with a patience of 10 epochs to halt training if the validation loss showed no improvement. This approach helped to identify the optimal stopping point during training and efficient use of computational resources.

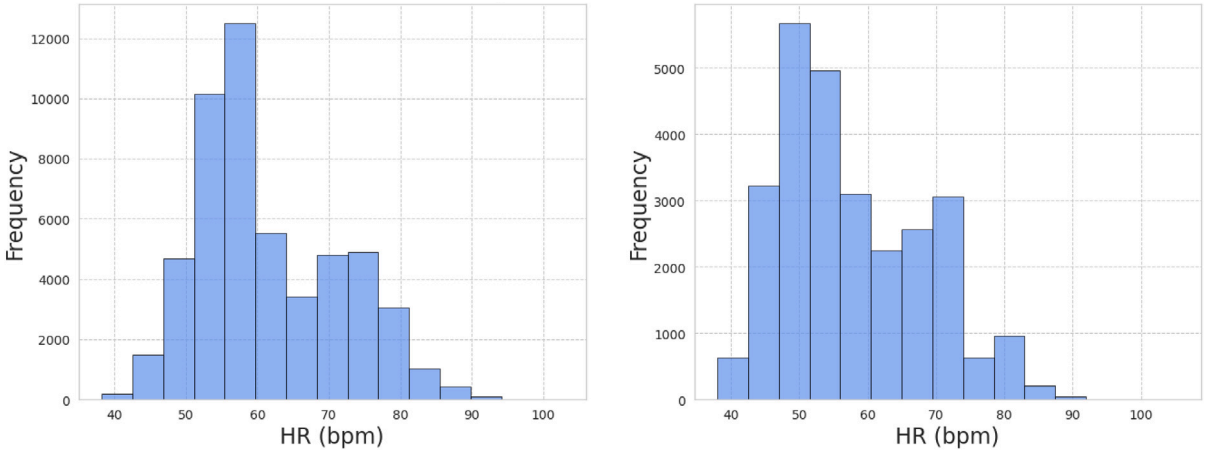
5. Baseline methods

To benchmark our MTL approaches, we compare them against state-of-the-art STL SQA and HR/RR estimation methods. We provide a brief description of these methods in the following.

5.1. Baseline methods for case study 1: SQA

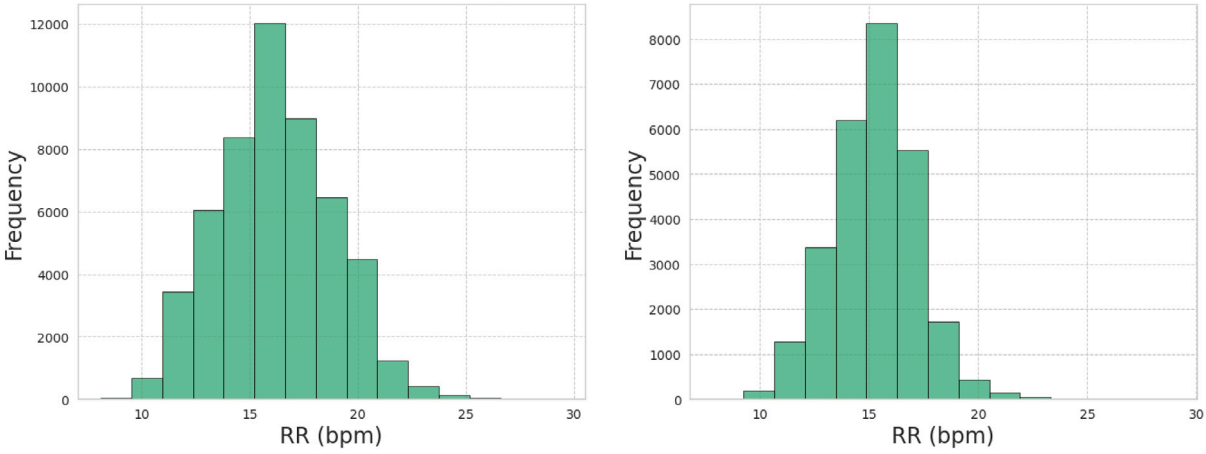
Our MTL SQA case study is compared with two STL SQA approaches:

1. **Shin's Method** [41]: This method employs a multi-layer CNN architecture, including several 1D convolutional layers. It begins with a layer that uses 32 filters of size 10, followed by subsequent layers with 64, 128, and 256 filters, each with decreasing kernel sizes of 6, 4, and 2, respectively. Additionally, batch normalization is applied after each convolutional layer. The network then transitions to a flattened layer connected to fully connected layers and a dropout layer to prevent overfitting. Finally, a single neuron with a sigmoid activation function is employed to classify the PPG segments into "Reliable" and "Unreliable" classes. We implement this model separately for each HR and HRV feature (RMSSD, AVNN, SDNN, and LFHF) to assess PPG signal quality.
2. **Naeini et al.'s Method** [45]: This model is specifically proposed for assessing PPG signal quality across different HR and HRV features. It consists of two convolutional layers equipped with 32 and 64 filters and kernel sizes of 3. Each convolutional layer is followed by a batch normalization layer. The output from the second layer undergoes max-pooling and is then flattened to be fed into fully connected layers. This method is similarly implemented separately for each HR and HRV feature.



(a) HR Histogram for Training Set (b) HR Histogram for Test Set

Fig. 8. HR values histograms for HR estimation.



(a) RR Histogram for Training Set (b) RR Histogram for Training Set

Fig. 9. RR values histograms for RR estimation.

5.2. Baseline methods for case study 2: HR estimation

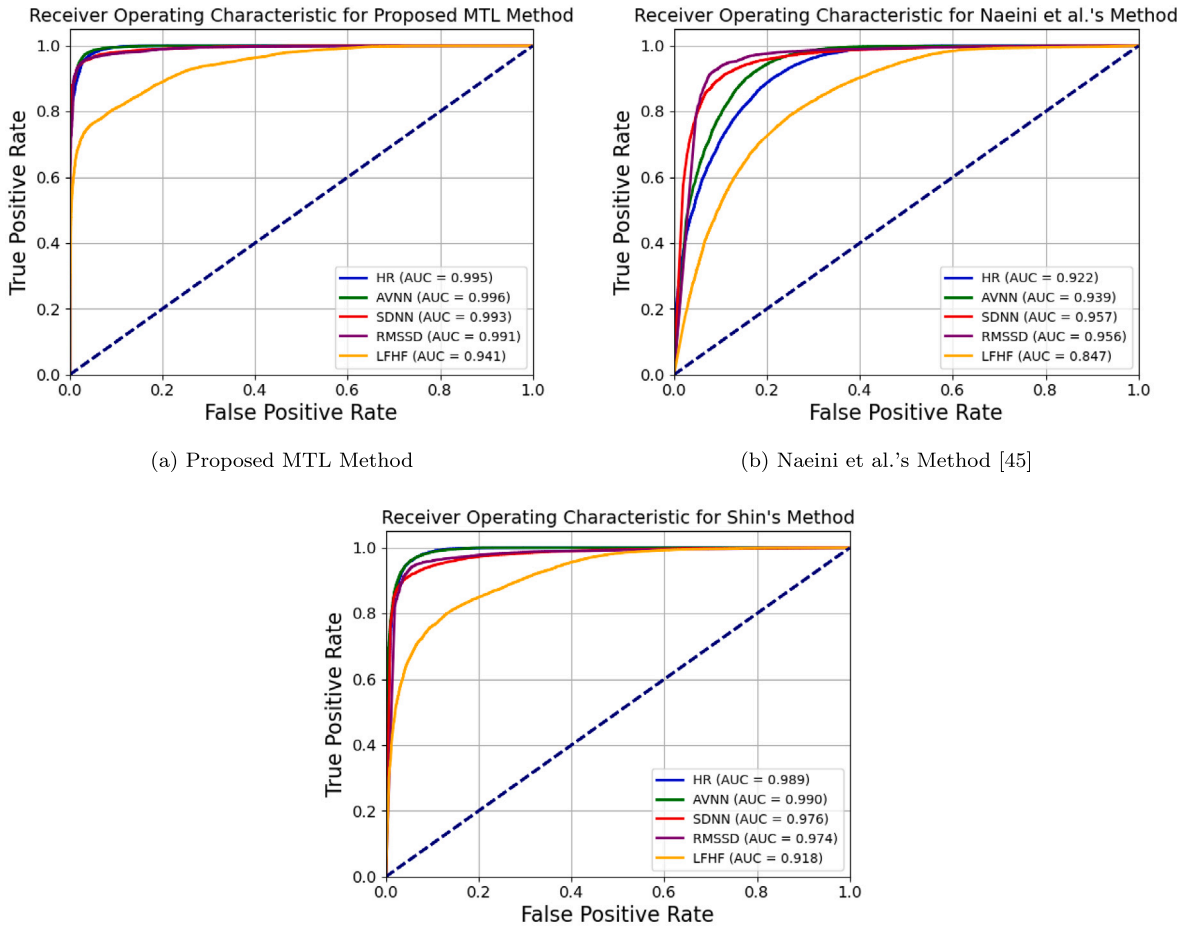
We compare the performance of our MTL HR/RR estimation method in estimating HR with two STL HR estimation approaches:

1. **PP-Net** [55]: This method begins with segmentation by sliding an 8-second window with a 2-second stride over the signals. HR estimation is then performed using a hybrid model that combines CNN and LSTM layers. The model starts with two CNN layers, each containing 20 filters with a kernel size of 9, followed by max-pooling and dropout layers. Subsequent layers include two LSTM layers with 64 and 128 units, ending with a fully connected layer with a single neuron and linear activation to predict the output.
2. **PPGnet** [56]: Similar to PP-Net, this method employs a sliding window segmentation of 8 s with a 2-second stride. PPG signals are filtered using a band-pass filter with cutoff frequencies of 0.5–5 Hz. HR is predicted from PPG signals using a deep learning model structured into two primary feature extractors: a CNN section that includes an inception block with parallel convolutions (16 filters with kernel sizes of 5, 20, 40, 60, and 80) followed by sequential blocks for hierarchical feature extraction, and an LSTM section comprising two LSTM layers. The outputs from these parallel sections are concatenated and processed through additional LSTM layers before generating the final estimation via a dense layer.

5.3. Baseline methods for case study 2: RR estimation

The performance of our MTL HR/RR estimation method in estimating RR is compared with two STL RR estimation approaches:

1. **Bian et al.'s Method** [57]: This method utilizes a CNN-based model to estimate RRs using raw PPG signals. Their proposed network employs ResNet blocks as the core architecture, followed by fully connected layers for RR prediction. The ResNet architecture follows the design specifications found in the literature [57], incorporating 1D convolution layers activated by ReLU. The model processes PPG signals as input, formatted as a vector with dimensions (600,1), representing a 30-second segment sampled at 20 Hz. During training, we experimented with three different learning rates, as detailed in [57], with the best results obtained at a learning rate of 10^{-5}.
2. **Aqajari et al.'s Method** [58]: In this work, a cycle-generative adversarial network (cycle-GAN) is proposed to reconstruct respiratory signals from PPG signals that leveraged the ability of the cycle-GAN to reconstruct signal methods. In the first step, PPG and respiratory signals were converted into images



(a) Proposed MTL Method

(b) Naeini et al.'s Method [45]

(c) Shin's Method [41]

Fig. 10. ROC curves of the proposed MTL and baseline STL PPG SQA methods.

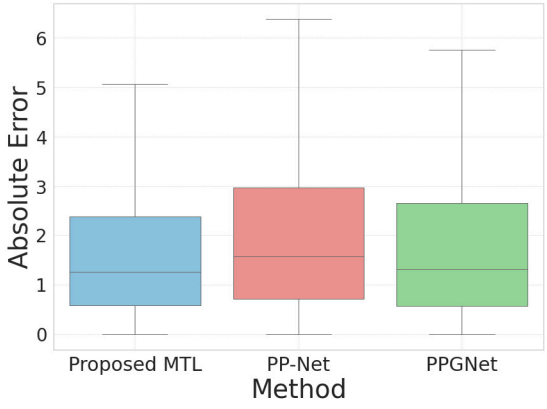


Fig. 11. Box plot of HR absolute errors for proposed MTL and baseline STL methods.

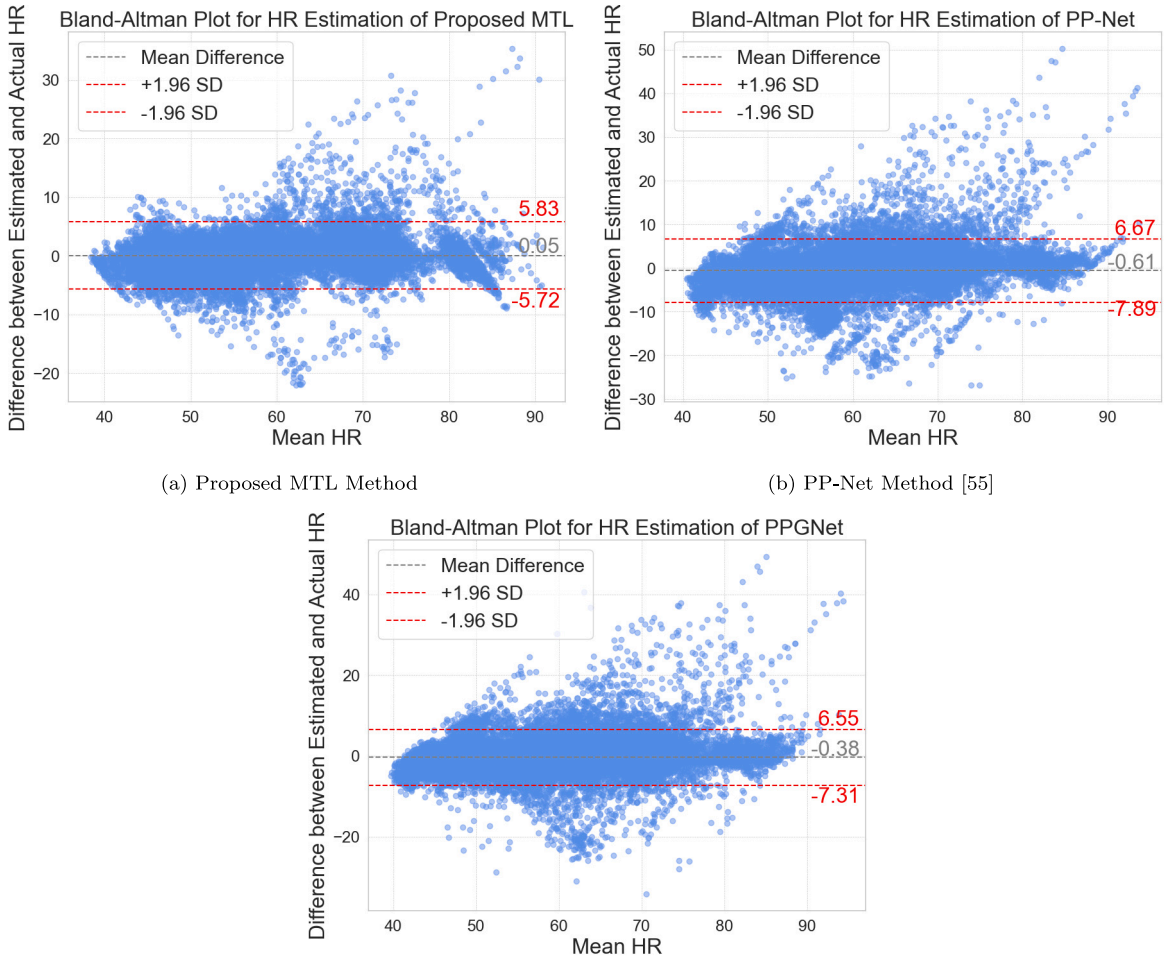
using an image translator. These images were then employed to train the model. The optimization function included adversarial losses, cycle consistency losses, and RR losses to ensure the reconstructed signal maintained respiratory characteristics. The approach incorporated three types of convolutions: stride-2 convolutions, residual blocks, and fractionally-scaled convolutions for the generative networks, as well as 70×70 PathGANs for the discriminator networks. Finally, the reconstructed respiration signals were fed to BreathMetrics [59], an open-source library for respiration signal analysis, to estimate the RR.

To ensure a fair comparison across all HR and RR estimation methods, including both proposed and baseline approaches, we utilize the same SQA technique [38] used in our proposed HR/RR estimation model. This strategy ensures that each method is evaluated under the same SQA conditions for data preparation. Moreover, we utilize the same methodologies to extract ground-truth HR and RR labels from reference ECG and ACC signals for all methods. It ensures that any observed differences in performance are due to the models' characteristics and not variations in data preprocessing.

6. Evaluation metrics

We assess our MTL and baseline STL methods using different evaluation metrics for classification and prediction tasks. To evaluate the performance of the MTL SQA models, we employ the Receiver Operating Characteristic (ROC) curves, Area Under the Curve (AUC), Accuracy, and F1-score. These metrics effectively quantify the model's capability to correctly classify PPG segments as "Reliable" or "Unreliable". For all evaluations, the "Reliable" class is treated as the positive class, while "Unreliable" is considered negative.

ROC curves represent the trade-off between the true positive rate (TPR) and the false positive rate (FPR) at different threshold values. The AUC measures the model's overall performance, where higher values show better discrimination between "Reliable" and "Unreliable" segments.



(a) Proposed MTL Method

(b) PP-Net Method [55]

(c) PPGNet Method [56]

Fig. 12. Bland-Altman plots of the proposed MTL and baseline STL HR estimation methods.

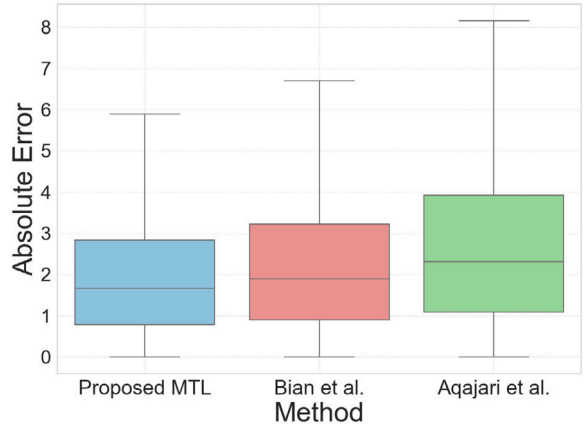


Fig. 13. Box plot of RR absolute errors for proposed MTL and baseline STL methods.

Accuracy measures the proportion of correctly classified instances among the total number of instances. It is defined mathematically as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP denotes true positives, TN true negatives, FP false positives, and FN false negatives.

F1-score provides a balance between precision ($\frac{TP}{TP+FP}$) and recall ($\frac{TP}{TP+FN}$), which is especially important in cases of uneven class distribution, as observed in our data (refer to Table 2). It is calculated as the harmonic mean of precision and recall:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

For the HR and RR estimation models, we utilize Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Median Absolute Difference (MAD), and the Bland-Altman method to assess the accuracy of HR and RR estimations:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

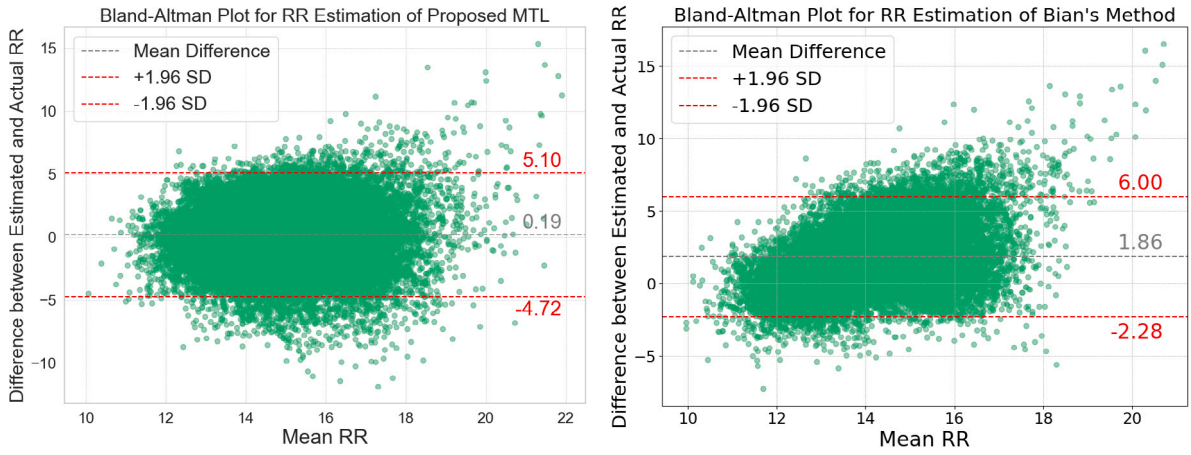
where n is the number of samples, y_i represents the ground-truth values, and $\hat{y}_i$ is the predicted value of HR or RR.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{7}$$

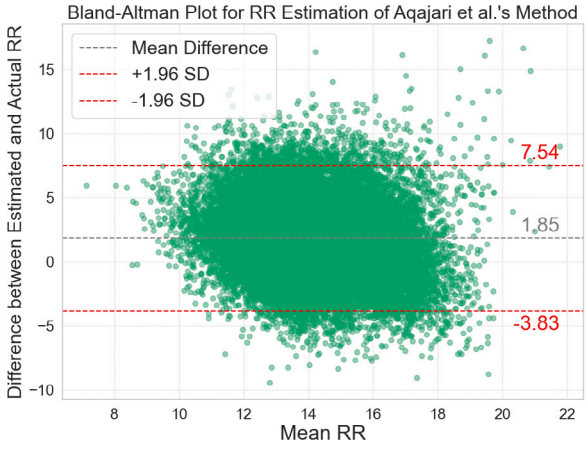
$$\text{MAD} = \text{median}(|y_i - \hat{y}_i|) \tag{8}$$

The Bland-Altman method further analyzes the agreement between estimated and actual HR/RR values by plotting the differences against the averages of these measurements. It provides insights into any possible biases in the model's predictions.



(a) Proposed MTL Method

(b) Bian et al.'s Method [57]



(c) Aqajari et al.'s Method [58]

Fig. 14. Bland-Altman plots of the proposed MTL and baseline STL RR estimation methods.

7. Test set results

In this section, we present the experimental results obtained from the proposed MTL and baseline STL methods. These results include findings from two methodological case studies: (1) PPG SQA for HR and HRV parameters and (2) HR and RR estimation from PPG.

7.1. Results for case study 1: PPG SQA

We detail the findings from our first case study, which focuses on classifying PPG signals according to their reliability for HR and HRV parameters. Fig. 10 demonstrates the ROC curves and the AUC values for our MTL SQA approach and baseline STL SQA methods. These curves are plotted for HR, AVNN, SDNN, RMSSD, and LFHF. Ideally, a perfect ROC curve rises sharply to the top-left corner of the plot, indicating a TPR of 1 and an FPR of 0. The results reveal that the ROC curves for our MTL approach bend more towards the ideal corner for most parameters compared to the baseline methods. However, the deviation towards the ideal corner is slightly less for LFHF. Moreover, the AUC values for our MTL model outperform those of the STL models across all parameters, demonstrating the superior predictive performance of our MTL approach.

In addition, the comparison of the SQA methods in terms of Accuracy, F1-score, and AUC is summarized in Table 3. As indicated, our MTL model demonstrates superior performance across all evaluated metrics compared to two baseline STL models. Specifically, the Accuracy, F1-score, and AUC values for HR, RMSSD, AVNN, and SDNN in our MTL approach are significantly higher. While the performance metrics for LFHF are lower than other parameters in our model, they still exceed those of the baseline methods. It should be emphasized that among the STL models, Shin's method [41] achieved comparable performance, surpassing the SQA method introduced by Naeini et al. [45].

7.2. Results for case study 2: HR estimation

We present the results of HR estimation from our MTL model and baseline STL models. The comparative evaluation in Table 4 highlights the performance metrics of the methods. As indicated, our MTL model achieves the lowest error rates, with an MAE of 1.89, RMSE of 2.95, MAPE of 0.03, and MAD of 1.26. These results demonstrate that our MTL model significantly outperforms the STL models in all error evaluation metrics for HR estimation, showing its superior accuracy.

Fig. 11 illustrates the box plots showing the distribution of absolute errors in HR estimation of our MTL and baseline methods. As shown, the proposed MTL method exhibits more accurate HR predictions by a lower median absolute error. Additionally, the significantly narrower interquartile range (IQR) and whiskers of our MTL model demonstrate a more consistent and reliable performance with less variation in errors. In contrast, PP-Net and PPGNet show higher absolute errors with wider IQRs and whiskers.

Table 3
Performance comparison of the proposed MTL and baseline STL methods for PPG SQA.

Feature	Naeini et al. [45]			Shin [41]			Proposed MTL		
	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
HR	81.08	81.07	92.20	95.34	95.34	98.88	**96.18**	**96.18**	**99.50**
RMSSD	93.76	93.84	95.59	94.97	95.04	97.42	**96.21**	**96.27**	**99.09**
AVNN	87.41	87.35	93.86	95.26	95.25	99.03	**96.87**	**96.87**	**99.62**
SDNN	94.97	90.73	95.69	93.67	93.76	97.62	**96.00**	**96.07**	**99.26**
LFHF	76.55	76.58	84.71	83.15	83.17	91.76	**85.16**	**85.18**	**94.10**

Table 4
Comparison of evaluation metrics for the proposed MTL and baseline STL methods for HR estimation.

Method	MAE	RMSE	MAPE	MAD
PP-Net [55]	2.38	3.76	0.04	1.57
PPGnet [56]	2.13	3.55	0.04	1.31
Proposed MTL	**1.89**	**2.95**	**0.03**	**1.26**

Table 5
Comparison of evaluation metrics for the proposed MTL and baseline STL methods for RR estimation.

Method	MAE	RMSE	MAPE	MAD
Bian et al.'s Method [57]	2.24	2.82	0.14	1.90
Aqajari et al.'s Method [58]	2.74	3.44	0.18	2.32
Proposed MTL	**1.98**	**2.51**	**0.13**	**1.66**

Moreover, Fig. 12 shows the Bland-Altman analysis plots for our MTL and baseline STL models. As indicated, our MTL model shows the lowest mean difference of 0.05, indicating a strong agreement with the ground truth values. Additionally, it exhibits a narrower spread of errors around the mean difference. Specifically, the limits of agreement for our MTL model, defined as ±1.96 standard deviations, are narrower at 5.83 and −5.72. In contrast, these limits for PP-Net are wider at +6.67 and −7.89, and for PPGNet, they are +6.55 and −7.31, reflecting a larger dispersion of errors. These results highlight the superior precision and consistency of our MTL model in HR estimation compared to the two STL models, PP-Net and PPGNet.

7.3. Results for case study 2: RR estimation

We provide the RR estimation results from our MTL model and the baseline STL models. The performance comparison on error metrics is summarized in Table 5. As indicated, our MTL approach exhibits the most accurate RR estimations with the lowest errors: MAE of 1.98, RMSE of 2.51, MAPE of 0.13, and MAD of 1.66. Additionally, the RR estimation method developed by Bian et al. [57] shows lower error rates compared to Aqajari et al.'s method [58].

Moreover, Fig. 13 demonstrates the distribution of absolute errors for RR estimation across all methods in box plots. The plots reveal that the proposed MTL model shows the lowest median absolute error. It also has the narrowest IQR and whiskers compared to other methods, highlighting its superior accuracy in predicting RR from PPG signal.

Fig. 14 illustrates the Bland-Altman analysis plots of our MTL and baseline RR estimation models. As shown, the mean difference of our MTL approach is 0.19, which is significantly lower than the 1.86 and 1.85 obtained from Bian et al. [57] and Aqajari et al. [58], respectively. It indicates that the RR predictions from our MTL model are, on average, closer to the actual values with less bias. Furthermore, the agreement limits for our MTL model range from −4.72 to 5.10. While the lower limit is wider than those of STL methods, the upper limit is narrower. The baseline STL methods exhibit a remarkable shift, indicated by their higher mean differences, which shows a significant bias in their RR estimations.

8. Discussion

In this study, we developed MTL approaches to improve the accuracy of PPG-based health monitoring methods. The experimental results demonstrated that MTL models notably outperform STL methods in two case studies: PPG SQA and HR/RR estimation. Significant improvements in performance metrics, such as Accuracy, F1-score, and ROC-AUC for the SQA model, as well as lower error rates for HR and RR estimation, highlight the superiority of the proposed MTL methods.

While our MTL models generally outperformed STL approaches, both MTL SQA and baseline STL models exhibited lower performance for the LFHF parameter. The correlation analysis demonstrated in Fig. 1 also revealed weaker correlations between this parameter and HR and other HRV parameters. This can be attributed to the high sensitivity of the LFHF ratio to noise and artifacts in the PPG signals. Therefore, it may distort underlying physiological relationships that MTL models aim to capture.

One notable advantage of our MTL approach over STL methods is its potential for reducing computational cost. Unlike STL methods, which require separate models for each task, our MTL approaches integrate multiple tasks within a single model. For example, in Naeini et al.'s method [45], five STL models were developed for PPG SQA for HR, AVNN, SDNN, RMSSD, and LFHF. In contrast, our MTL approach handled all these tasks in a single model.

To quantify this efficiency, we measured the execution time of the proposed MTL PPG SQA method and compared it with the baseline STL methods on our test data. The average execution time per five-minute PPG sample for our MTL SQA approach was 5.24 ms, significantly outperforming the Naeini et al.'s [45] and Shin's [41] methods, which required 20.58 and 26.21 ms, respectively. These results demonstrate that multitasking can reduce computational costs while improving efficiency.

However, the performance of our MTL models for real-time analysis on low-powered edge devices remains an important area of exploration. While our MTL approach offers computational advantages compared to STL methods, its real-time performance and energy efficiency on resource-constrained devices require evaluation. In future work, we plan to implement our MTL approaches on low-powered devices to assess their computational time and energy consumption in real-world scenarios.

A significant challenge with PPG signals is their susceptibility to motion artifacts and noise, particularly in free-living conditions. This limitation affected our dataset, as a large proportion of PPG signals were unreliable. As shown in Table 2, both the training and test sets exhibited class imbalance, with the "Unreliable" class being the majority. To ensure that both the proposed and baseline SQA models were not biased, we employed evaluation metrics robust to imbalanced datasets, such as F1-score and ROC-AUC, which provided a robust measure of performance independent of class distributions.

To address the challenge of noise and artifacts in PPG signals for HR and RR estimation, we distinguished reliable signals from unreliable signals in the preprocessing step. For this purpose, we employed a previously validated PPG SQA method with an accuracy of 97%. This ensured that our MTL HR/RR estimation model was trained and evaluated exclusively on reliable PPG signals. Despite that a significant portion of the data was unreliable, we successfully prepared a dataset

comprising 79,667 reliable PPG segments, distributed as shown in Figs. 8 and 9. This amount of reliable data was adequate for effective model training and testing, ensuring the robustness of our findings.

Another challenge was the computational demand for training our multitask deep learning models. The large volume of PPG data with multiple tasks in the models required significant processing power and memory. To address these computational requirements, we leveraged a Linux machine with an AMD Ryzen Threadripper processor, 126 GB of RAM, and an NVIDIA TITAN RTX GPU with 24 GB of memory, ensuring efficient training and testing.

Our study has several limitations. First, our 24-hour data collection period may not capture long-term physiological variations. Future research will extend data collection to several days or weeks to better understand temporal trends. Second, our dataset consisted exclusively of healthy participants, limiting the generalizability of our findings to individuals with cardiovascular conditions, such as atrial fibrillation. Future studies will include data from different population groups, including those with cardiovascular diseases. Lastly, we plan to implement and test our models on wearable and edge devices to evaluate their real-time performance and energy efficiency in real-world settings.

9. Conclusion

Traditional single-task PPG-based health monitoring methods often overlook the potential interdependencies and shared characteristics among PPG-related tasks, which can be leveraged to enhance performance. In this study, we developed MTL approaches to employ these relationships and improve PPG-based health monitoring methods. Specifically, we implemented MTL models for two case studies: HR/HRV-guided PPG Signal Quality Assessment (SQA) and simultaneous HR and RR estimation from PPG signals. The proposed models were evaluated using a dataset including PPG, ECG, and ACC signals collected from 46 individuals wearing smartwatches in real-world conditions. Experimental results demonstrated that MTL models significantly outperformed baseline single-task methods, achieving higher accuracy in PPG SQA and lower error rates in HR and RR estimation. These findings underscore the potential of MTL approaches to advance the field of PPG-based health monitoring.

CRediT authorship contribution statement

Mohammad Feli: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Kianoosh Kazemi:** Writing – review & editing, Formal analysis. **Iman Azimi:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Pasi Liljeberg:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Amir M. Rahmani:** Writing – review & editing, Supervision, Funding acquisition.

Ethics statement

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and the Finnish Medical Research Act (#488/1999). The study protocol was approved by the Ethics Committee for Human Sciences at the University of Turku (Statement #44/2019). Participants were informed about the study both orally and in writing before giving their consent. Participation was voluntary, and all participants had the right to withdraw from the study at any time without giving a reason.

Declaration of competing interest

None Declared.

Acknowledgments

This research was supported by the Academy of Finland through the SLIM Project (grant numbers 316810 and 316811) and the U.S. National Science Foundation through the UNITE Project (grant number SCC CNS-1831918).

Data availability

The data that support the findings of this study are available from the corresponding author (M.F.) upon reasonable request.

References

[1] T. Tamura, Y. Maeda, M. Sekine, M. Yoshida, Wearable photoplethysmographic sensors—past and present, Electronics 3 (2) (2014) 282–302.
[2] J. Allen, Photoplethysmography and its application in clinical physiological measurement, Physiol. Meas. 28 (3) (2007) R1.
[3] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, H. Nazeran, A review on wearable photoplethysmography sensors and their potential future applications in health care, Int. J. Biosens. Bioelectron. 4 (4) (2018) 195.
[4] M. Bolanos, H. Nazeran, E. Haltiwanger, Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals, in: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2006, pp. 4289–4294.
[5] W. Karlen, S. Raman, J.M. Ansermino, G.A. Dumont, Multiparameter respiratory rate estimation from the photoplethysmogram, IEEE Trans. Biomed. Eng. 60 (7) (2013) 1946–1953.
[6] S. Bagha, L. Shaw, A real time analysis of PPG signal for measurement of SpO2 and pulse rate, Int. J. Comput. Appl. 36 (11) (2011) 45–50.
[7] S.S. Mousavi, M. Firouzmand, M. Charmi, M. Hemmati, M. Moghadam, Y. Ghorbani, Blood pressure estimation from appropriate and inappropriate PPG signals using a whole-based method, Biomed. Signal Process. Control. 47 (2019) 196–206.
[8] H. Korkalainen, J. Aakko, B. Duce, S. Kainulainen, A. Leino, S. Nikkonen, I.O. Afara, S. Myllymaa, J. Töyräs, T. Leppänen, Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea, Sleep 43 (11) (2020) zsaa098.
[9] S. Heo, S. Kwon, J. Lee, Stress detection with single PPG sensor by orchestrating multiple denoising and peak-detecting methods, IEEE Access 9 (2021) 47777–47785.
[10] S. Jafarlou, I. Azimi, J. Lai, Y. Wang, S. Labbaf, B. Nguyen, H. Qureshi, C. Marcotullio, J.L. Borelli, N.D. Dutt, et al., Objective monitoring of loneliness levels using smart devices: A multi-device approach for mental health applications, PLoS One 19 (6) (2024) e0298949.
[11] M. Feli, I. Azimi, F. Sarhaddi, Z. Sharifi-Heris, H. Niela-Vilen, P. Liljeberg, A. Axelin, A.M. Rahmani, Preterm birth risk stratification through longitudinal heart rate and HRV monitoring in daily life, Sci. Rep. 14 (1) (2024) 19896.
[12] F. Shaffer, J.P. Ginsberg, An overview of heart rate variability metrics and norms, Front. Public Heal. 5 (2017) 258.
[13] S.Z.H. Kazmi, H. Zhang, W. Aziz, O. Monfredi, S.A. Abbas, S.A. Shah, S.S.H. Kazmi, W.H. Butt, Inverse correlation between heart rate variability and heart rate demonstrated by linear and nonlinear analysis, PLoS One 11 (6) (2016) e0157557.
[14] H.-S. Song, P.M. Lehrer, The effects of specific respiratory rates on heart rate and heart rate variability, Appl. Psychophys. Biof. 28 (2003) 13–23.
[15] Y. Zhang, Q. Yang, A survey on multi-task learning, IEEE Trans. Knowl. Data Eng. 34 (12) (2021) 5586–5609.
[16] J.W. Hughes, T. Sittler, A.D. Joseph, J.E. Olgin, J.E. Gonzalez, G.H. Tison, Using multitask learning to improve 12-lead electrocardiogram classification, 2018, arXiv preprint arXiv:1812.00497.
[17] M. Shahin, E. Oo, B. Ahmed, Adversarial multi-task learning for robust end-to-end ECG-based heartbeat classification, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2020, pp. 341–344.
[18] Q. Geng, H. Liu, T. Gao, R. Liu, C. Chen, Q. Zhu, M. Shu, An ECG classification method based on multi-task learning and cot attention mechanism, in: Healthcare, vol. 11, (7) MDPI, 2023, p. 1000.
[19] S. Tang, Z. Deng, CS-based multi-task learning network for arrhythmia reconstruction and classification using ECG signals, Physiol. Meas. 44 (7) (2023) 075001.
[20] X. Fan, H. Wang, Y. Zhao, Y. Li, K.L. Tsui, An adaptive weight learning-based multitask deep network for continuous blood pressure estimation using electrocardiogram signals, Sensors 21 (5) (2021) 1595.
[21] D.U. Jeong, K.M. Lim, Combined deep CNN–LSTM network-based multitasking learning architecture for noninvasive continuous blood pressure estimation using difference in ECG-PPG features, Sci. Rep. 11 (1) (2021) 13539.

[22] R.-N. Yin, R.-S. Jia, Z. Cui, H.-M. Sun, PulseNet: A multitask learning network for remote heart rate estimation, Knowl.-Based Syst. 239 (2022) 108048.
[23] B. Huang, W. Chen, C.-L. Lin, C.-F. Juang, Y. Xing, Y. Wang, J. Wang, A neonatal dataset and benchmark for non-contact neonatal heart rate monitoring based on spatio-temporal neural networks, Eng. Appl. Artif. Intell. 106 (2021) 104447.
[24] K.S. Rathore, V. Sricharan, S. Preejith, M. Sivaprakasam, MRNet-A deep learning based multitasking model for respiration rate estimation in practical settings, in: 2022 IEEE 10th International Conference on Serious Games and Applications for Health (SeGAH), IEEE, 2022, pp. 1–6.
[25] K. Qin, W. Huang, T. Zhang, Multitask deep label distribution learning for blood pressure prediction, Inf. Fusion 95 (2023) 426–445.
[26] Q. Hu, D. Wang, C. Yang, PPG-based blood pressure estimation can benefit from scalable multi-scale fusion neural networks and multi-task learning, Biomed. Signal Process. Control. 78 (2022) 103891.
[27] H. Xiao, A. Zhao, W. Song, T. Liu, L. Long, Y. Li, H. Li, Advancing cuffless blood pressure estimation: A PPG-based multi-task learning model for enhanced feature extraction and fusion, Biomed. Signal Process. Control 95 (2024) 106378.
[28] D.J. McDuff, J.R. Estepp, A.M. Piasecki, E.B. Blackford, A survey of remote optical photoplethysmographic imaging methods, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2015, pp. 6398–6404.
[29] H.-Y. Chih, T. Ahmed, A.P. Chiu, Y.-T. Liu, H.-F. Kuo, A.C. Yang, D.-H. Lien, Multitask learning for automated sleep staging and wearable technology integration, Adv. Intell. Syst. 6 (1) (2024) 2300270.
[30] H. Kaji, H. Iizuka, M. Sugiyama, ECG-based concentration recognition with multi-task regression, IEEE Trans. Biomed. Eng. 66 (1) (2018) 101–110.
[31] O. Beya, M.M. Hittawe, T. Alashkar, E. Fauvet, O. Laligant, Applying non linear approach for ECG denoising and waves localization, in: 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems, SITIS, IEEE, 2015, pp. 42–47.
[32] O. Beya, M.-M. Hittawe, N. Zegadi, E. Fauvet, O. Laligant, Electrocardiogram signal analysing-delineation and localization of ECG component, in: International Conference on Bio-Inspired Systems and Signal Processing, vol. 5, SCITEPRESS, 2016, pp. 156–161.
[33] C. Orphanidou, Signal Quality Assessment in Physiological Monitoring: State of the Art and Practical Considerations, Springer, 2017.
[34] S. Vadrevu, M.S. Manikandan, Real-time PPG signal quality assessment system for improving battery life and false alarms, IEEE Trans. Circuits Syst. II Express Briefs 66 (11) (2019) 1910–1914.
[35] G.N.K. Reddy, M.S. Manikandan, N.N. Murty, On-device integrated PPG quality assessment and sensor disconnection/saturation detection system for IoT health monitoring, IEEE Trans. Instrum. Meas. 69 (9) (2020) 6351–6361.
[36] T. Pereira, K. Gadhoumi, M. Ma, X. Liu, R. Xiao, R.A. Colorado, K.J. Keenan, K. Meisel, X. Hu, A supervised approach to robust photoplethysmography quality assessment, IEEE J. Biomed. Heal. Inform. 24 (3) (2019) 649–657.
[37] A. Mahmoudzadeh, I. Azimi, A.M. Rahmani, P. Liljeberg, Lightweight photoplethysmography quality assessment for real-time IoT-based health monitoring using unsupervised anomaly detection, Procedia Comput. Sci. 184 (2021) 140–147.
[38] M. Feli, I. Azimi, A. Anzanpour, A.M. Rahmani, P. Liljeberg, An energy-efficient semi-supervised approach for on-device photoplethysmogram signal quality assessment, Smart Heal. 28 (2023) 100390.
[39] M. Feli, K. Kazemi, I. Azimi, Y. Wang, A.M. Rahmani, P. Liljeberg, End-to-end PPG processing pipeline for wearables: From quality assessment and motion artifacts removal to HR/HRV feature extraction, in: 2023 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2023, pp. 1895–1900.
[40] E.K. Naeini, I. Azimi, A.M. Rahmani, P. Liljeberg, N. Dutt, A real-time PPG quality assessment approach for healthcare Internet-of-Things, Procedia Comput. Sci. 151 (2019) 551–558.
[41] H. Shin, Deep convolutional neural network-based signal quality assessment for photoplethysmogram, Comput. Biol. Med. 145 (2022) 105430.
[42] Z. Han, J. Zhao, H. Leung, K.F. Ma, W. Wang, A review of deep learning models for time series prediction, IEEE Sensors J. 21 (6) (2019) 7833–7848.
[43] B. Rim, N.-J. Sung, S. Min, M. Hong, Deep learning in physiological signal data: A survey, Sensors 20 (4) (2020) 969.
[44] M.M. Hittawe, F. Harrou, M.A. Togou, Y. Sun, O. Knio, Time-series weather prediction in the Red sea using ensemble transformers, Appl. Soft Comput. 164 (2024) 111926.
[45] E.K. Naeini, F. Sarhaddi, I. Azimi, P. Liljeberg, N. Dutt, A.M. Rahmani, A deep learning–based PPG quality assessment approach for heart rate and heart rate variability, ACM Trans. Comput. Heal. 4 (4) (2023) 1–22.
[46] M. Elgendi, M. Jonkman, F. De Boer, Frequency bands effects on QRS detection, Biosignals 2003 (2010) 2002.
[47] T. Pereira, P.R. Almeida, J.P. Cunha, A. Aguiar, Heart rate variability metrics for fine-grained stress level assessment, Comput. Methods Programs Biomed. 148 (2017) 71–80.
[48] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D Nonlinear Phenom. 60 (1–4) (1992) 259–268.
[49] X. Sun, L. Qiu, Y. Wu, Y. Tang, G. Cao, Sleepmonitor: Monitoring respiratory rate and body position during sleep using smartwatch, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1 (3) (2017) 1–22.
[50] M.A. Mehrabadi, I. Azimi, F. Sarhaddi, A. Axelin, H. Niela-Vilén, S. Myllyntausta, S. Stenholm, N. Dutt, P. Liljeberg, A.M. Rahmani, et al., Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study, JMIR MHealth UHealth 8 (11) (2020) e20465.
[51] Gear sport 42mm smartwatch, 2025, https://www.samsung.com/us/mobile/wearables/smartwatches/gear-sport-blue-sm-r600nzbaxar/. (Accessed 7 August 2025).
[52] Shimmer3 ECG unit, 2025, https://shimmersensing.com/product/shimmer3-ecg-unit-2/. (Accessed 7 August 2025).
[53] E.P. Doheny, M.M. Lowery, A. Russell, S. Ryan, Estimation of respiration rate and sleeping position using a wearable accelerometer, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2020, pp. 4668–4671.
[54] F. Schipper, R.J. van Sloun, A. Grassi, R. Derkx, S. Overeem, P. Fonseca, Estimation of respiratory rate and effort from a chest-worn accelerometer using constrained and recursive principal component analysis, Physiol. Meas. 42 (4) (2021) 045004.
[55] M. Panwar, A. Gautam, D. Biswas, A. Acharyya, PP-Net: A deep learning framework for PPG-based blood pressure and heart rate estimation, IEEE Sensors J. 20 (17) (2020) 10000–10011.
[56] A. Shyam, V. Ravichandran, S. Preejith, J. Joseph, M. Sivaprakasam, PPGnet: Deep network for device independent heart rate estimation from photoplethysmogram, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2019, pp. 1899–1902.
[57] D. Bian, P. Mehta, N. Selvaraj, Respiratory rate estimation using PPG: A deep learning approach, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2020, pp. 5948–5952.
[58] S.A.H. Aqajari, R. Cao, A.H.A. Zargari, A.M. Rahmani, An end-to-end and accurate PPG-based respiratory rate estimation approach using cycle generative adversarial networks, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2021, pp. 744–747.
[59] T. Noto, et al., Automated analysis of breathing waveforms using BreathMetrics: a respiratory signal processing toolbox, Chem. Senses 43 (8) (2018) 583–597.