MIMIC-III-Ext-CA: A MIMIC-III Derived Dataset of Cardiac Arrests in Photoplethysmographs

Roelof G Hup[1], Xi Long[1], Rik Vullings[1]

[1] Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract

Photoplethysmography (PPG) is increasingly considered for detecting out-of-hospital cardiac arrest (OHCA), but publicly available datasets remain scarce. We present a method to identify cardiac arrest episodes captured by PPG in the MIMIC-III database, combining automated screening with manual annotation of waveform and clinical data. Using this approach, we compiled 36 annotated cardiac arrest episodes from 31 patients. The dataset will be publicly available on PhysioNet and provides a resource for developing and validating wearable OHCA detection technologies. Our method is also adaptable to other clinical events.

1. Introduction

Photoplethysmography (PPG) is an optical method to determine the relative changes in blood volume in the skin by measuring light absorption. In the context of wearable devices, it has gained interest as a method to measure cardiovascular parameters continuously and non-invasively, including heart rate, oxygen saturation, pulse rate variability, cuffless blood pressure, and blood glucose levels [1]. Special interest has been shown in the detection of obstructive sleep apnea [2], signs of infection [3], heart failure [4], and atrial fibrillation [5, 6].

One emerging application is the detection of out-of-hospital cardiac arrest (OHCA). As 29.7% to 63.4% of the OHCA episodes go unwitnessed [7], the use of automated monitoring with PPG could improve survival rates and patient outcomes. Several research teams, such as BECA [8], DETECT [9], HEART-SAFE [10], and Google Research [11], are actively working on this technology.

A major challenge is the availability of PPG measurements during cardiac arrests. Although the above-mentioned projects mostly simulate or approximate (out-of-hospital) cardiac arrest, real data remains vital for algorithm development and validation. To our knowledge, until now, PhysioNet has published only a single dataset containing PPG and annotated life-threatening arrhythmias, as

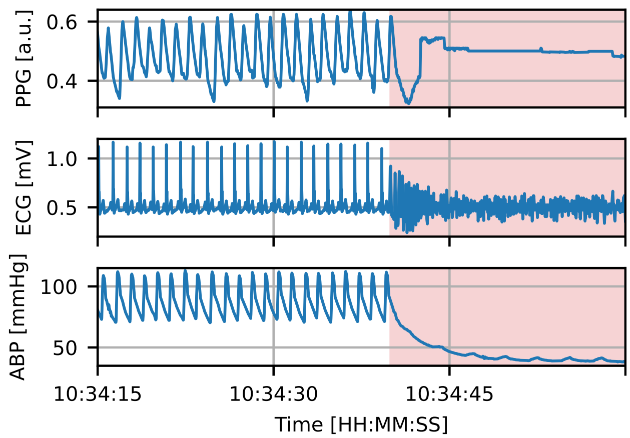

Figure 1. An example of an episode of ventricular fibrillation (marked by a red background) in the MIMIC-III database. Photoplethysmography (PPG), electrocardiogram (ECG) and arterial blood pressure (ABP) are shown.

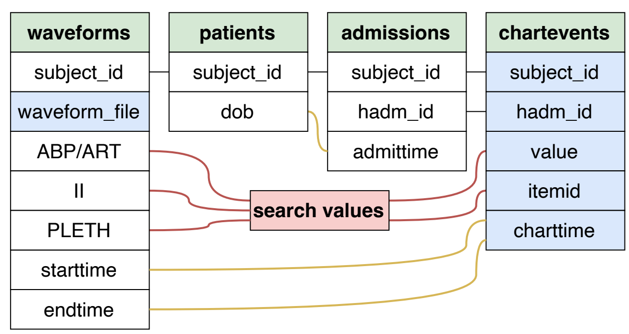

Figure 2. Relational database tables with the columns of interest (marked blue), inter-table relationships (black lines), comparison operations (yellow lines) and value searches (red lines).

part of the PhysioNet/Computing in Cardiology Challenge 2015 [12]. Other arrhythmia datasets, such as the MIT-BIH Arrhythmia Database, do contain occurrences of cardiac arrest, but not in PPG signals [13].

One PhysioNet dataset sparks interest about this topic: the MIMIC-III critical care database, which contains data from 38,597 adult ICU patients [14]. MIMIC-III contains

 ISSN: 2325-887X DOI: 10.22489/CinC.2025.135

not only bedside monitor waveforms, including PPG, but also anonymized electronic medical record (EMR) data. Given the vast number of patients and their health conditions, episodes of cardiac arrest are expected to occur in this dataset. However, the size of the dataset makes it impractical to manually screen for cardiac arrest episodes.

In this work, we present a method for identifying episodes of PPG-captured cardiac arrest in the MIMIC-III database, for example, as shown in Figure 1. We show that the use of the relational database of clinical data and automated bedside monitor heart rhythm classifications point to cardiac arrest episodes. Finally, we describe the derived dataset, which is published on PhysioNet.

2. Methods

2.1. Data acquisition

We acquired the EMR data from the *MIMIC-III Clinical Database* (version 1.4) [15] and the waveform data, matched to the EMR data, from the *MIMIC-III Waveform Database Matched Subset* (version 1.0) [16]. The EMR data were inserted directly into a SQLite relational database. For each waveform record, we extracted the subject ID, start and end time of the recording, and the available waveform signal names from the metadata, and inserted this into the WAVEFORMS table in the SQLite database. A diagram of the relevant tables and column headers is presented in Figure 2.

2.2. Identifying candidate cardiac arrests

Potential cardiac arrests were identified by searching for bedside monitor-annotated cardiac arrest events in the CHARTEVENTS database table. Inspection of the D_ITEMS and CHARTEVENTS tables led us to the required values of CHARTEVENTS.*itemid* and CHARTEVENTS.*value* that indicate ventricular tachycardia/fibrillation or asystole in both hospital systems used. For the CareVue hospital system, we require *itemid* $=$ 212 and *value* $\in$ {*Asystole*, *Vent. Tachy*, *Ventricular Fib*}. For the MetaVision system, we require *itemid* $=$ 220048 and *value* $\in$ {*Asystole*, *VT (Ventricular Tachycardia)*, *VF (Ventricular Fibrillation)*}.

The waveform files were matched to the EMR data using the *subject_id* field. To exclude bedside monitor rhythm classifications that were not accompanied by waveform signals, we set the condition that the classification event should be charted between the start and end time of the matched waveform records, e.g., WAVEFORMS.*starttime* $\leq$ CHARTEVENTS.*charttime* $\leq$ WAVEFORMS.*endtime*.

Another condition is the presence of a PPG signal (in MIMIC-III: *PLETH, PLETH_L, PLETH_R, PLETHl, PLETHr*). Furthermore, to allow for verification of cardiac arrests in the PPG signal, we added the requirement that either the electrocardiogram (ECG, lead II/II+) or the continuous arterial blood pressure (ABP/ART) signal had to be present as a reference.

Lastly, we excluded any patient who was underage (<18 years) at the time of hospital admission, e.g., ADMISSIONS.*admittime* - PATIENTS.*dob* $\geq$ 18 years.

2.3. Manual annotation of cardiac arrests

After identifying candidate cardiac arrest events, we manually reviewed these events by inspecting the PPG, ECG, and ABP waveforms around the event time stamp, and if present, the discharge notes in the NOTEEVENTS table corresponding to the hospital admission (NOTEEVENTS.*hadm_id* = CHARTEVENTS.*hadm_id*). To ensure that only clinically meaningful cardiac arrest episodes were retained, events were excluded whenever one of the following conditions was met:

- The PPG was unavailable/unreliable at the event timestamp, for example, due to motion artifacts or sensor disconnect.
- Both the ECG and ABP were unavailable/unreliable at the event timestamp, for example, due to clipping.
- Ventricular tachycardia is present, but is not associated with loss of pulse (non-life-threatening VT).
- There is no VF/VT/asystole present between one hour before the event timestamp and one hour after (false detection).
- The event points to the same cardiac arrest episode as another event (duplicate).

All events that were not excluded are considered part of the final dataset. For each of these events, the start and end of the cardiac arrest episode have also been determined. If a waveform record ended before the end of cardiac arrest, the end timestamp was set to the end of the record.

2.4. Population characteristics

After cardiac event selection, population characteristics were derived from other tables in the dataset. The patient's sex was taken from the PATIENTS.*gender* column, and the age was calculated based on the time difference between the date of birth and the date of admission (ADMISSIONS.*admittime* - PATIENTS.*dob*). The initial diagnosis at admission (*admissions.diagnoses*) was categorized based on medical specialty.

3. Results

Given the more than 5 million bedside monitor heart rhythm classifications present in the CHARTEVENTS table, our methodology resulted in the identification of 36

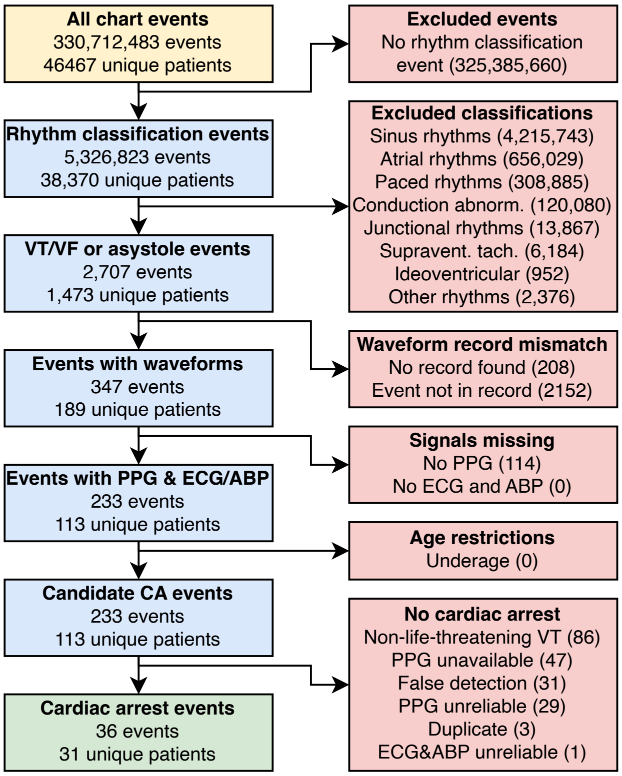

Figure 3. Flowchart of the study population. The left column shows the (left-over) population, while the right column shows the excluded population.

cardiac arrest episodes in 31 unique patients, highlighting the rarity of (PPG-captured) in-hospital cardiac arrest. An example of a cardiac arrest episode is shown in Figure 1. A flowchart of the study population per exclusion step is shown in Figure 3.

Looking at the characteristics of the study population in Table 1, most of the cardiac arrests occurred in men (74%). The average age of the population is 67 ($\pm$ 15) years. A majority of the patients were initially admitted with cardiovascular-related diagnoses.

The dataset derived in this work will be made publicly available on PhysioNet under the PhysioNet Credentialed Health Data License 1.5.0.

4. Discussion

In this work, we have shown that the use of heart rhythm classifications from bedside monitors can be used to find a reasonable number of PPG-captured cardiac arrest episodes in the MIMIC-III database. Automatic filtering of events based on EMR data drastically reduces the workload of manual annotations. The methodology could also be generalized to other types of events, such as other cardiac arrhythmias.

However, one always has to consider the source of the

Table 1. Population description

Characteristic	*n* (%)
Sex	
Male	23 (74%)
Female	8 (26%)
Age [years]	
18 - 29	2 (6%)
30 - 49	0 (0%)
50 - 59	6 (19%)
60 - 69	8 (26%)
70 - 79	7 (23%)
$\geq$ 80	8 (26%)
Initial diagnosis	
Cardiac	16 (52%)
Neurologic	4 (13%)
Respiratory	4 (13%)
Oncologic	2 (6%)
Gastrointestinal	2 (6%)
Other	3 (10%)

EMR data used. In this work, the cardiac monitoring functionality of the bedside monitors was used as a guideline for cardiac arrest timestamps, but these monitors produced false alarms (in our case: 31) and probably missed or did not register cardiac events. For example, approximately 1600 patients do not have entries in CHARTEVENTS, making the presented methodology unsuitable for finding cardiac arrests in the records of those patients.

Furthermore, only a third of the waveform records in the MIMIC-III Waveform Database have been matched to the Clinical Database, which results in the inability to match many cardiac arrest events in the Clinical Database to a specific waveform record. Figure 3 illustrates this limitation, as a drastic decrease in events occurs after matching waveform records (2,707 to 347 events). The presented dataset should therefore not be considered as if it contains all cardiac arrest episodes in the MIMIC-III database, but just a subset of episodes found in the context of bedside monitor algorithm sensitivity/specificity.

To broaden the search for cardiac arrest episodes in the MIMIC-III database, one should look beyond the use of only EMR data. Therefore, future work could include the application of ECG/ABP-based detection algorithms, combined with manual annotation, to find more PPG-captured cardiac arrests.

Finally, in the context of OHCA detection, a fundamental limitation of the current work is found in the differences between data acquisition settings. MIMIC-III waveform data is captured from ICU (in-hospital cardiac arrest, IHCA) patients with finger clip PPG, while OHCA detection technology is applied to detect OHCA and could capture PPG at more convenient sites, such as the wrist. This limits direct generalization to out-of-hospital settings.

5. Conclusion

In this work, we presented a method for identifying cardiac arrest episodes, captured by PPG, in the MIMIC-III database. The method can be easily adapted to identify other types of events. Applying this approach, we compiled a dataset consisting of 36 cardiac arrest episode annotations in 31 unique patients, which could be helpful in the development and validation of wearable OHCA detection technologies. The dataset will be available on PhysioNet.

Acknowledgments

This work is financed by the PPP Allowance made available by Top Sector Life Sciences & Health to the Dutch Heart Foundation to stimulate public-private partnerships, grant number 01-003-2021-B005, and by Philips Electronics Nederland B.V.

References

[1] Kim KB, Baek HJ. Photoplethysmography in wearable devices: A comprehensive review of technological advances, current challenges, and future directions, 7 2023.

[2] Kim MW, Park SH, Choi MS. Diagnostic performance of photoplethysmography-based smartwatch for obstructive sleep apnea. Journal of Rhinology 11 2022;29(3):155–162. ISSN 23844361.

[3] Conroy B, Silva I, Mehraei G, Damiano R, Gross B, Salvati E, et al. Real-time infection prediction with wearable physiological monitoring and AI to aid military workforce readiness during COVID-19. Scientific Reports 12 2022; 12(1). ISSN 20452322.

[4] Shah AJ, Isakadze N, Levantsevych O, Vest A, Clifford G, Nemati S. Detecting heart failure using wearables: A pilot study. Physiological Measurement 4 2020;41(4). ISSN 13616579.

[5] Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. New England Journal of Medicine 2019;381(20):1909–1917. ISSN 0028-4793.

[6] Lubitz SA, Faranesh AZ, Selvaggi C, Atlas SJ, McManus DD, Singer DE, et al. Detection of atrial fibrillation in a large population using wearable devices: The Fitbit Heart study. Circulation 11 2022;146(19):1415–1424. ISSN 15244539.

[7] Nishiyama C, Kiguchi T, Okubo M, Alihodžić H, Al-Araji R, Baldi E, et al. Three-year trends in out-of-hospital cardiac arrest across the world: Second report from the International Liaison Committee on Resuscitation (ILCOR). Resuscitation 2023;186(December 2022):109757. ISSN 03009572.

[8] Hup RG, Linssen EC, Eversdijk M, Verbruggen B, Bak MA, Habibovic M, et al. Rationale and design of the BECA project: Smartwatch-based activation of the chain of survival for out-of-hospital cardiac arrest. Resuscitation Plus 3 2024;17. ISSN 26665204.

[9] Edgar R, Scholte NT, Ebrahimkheil K, Brouwer MA, Beukema RJ, Mafi-Rad M, et al. Automated cardiac arrest detection using a photoplethysmography wristband: Algorithm development and validation in patients with induced circulatory arrest in the DETECT-1 study. The Lancet Digital Health 3 2024;6(3):e201–e210. ISSN 25897500.

[10] Schober P, van den Beuken WM, Nideröst B, Kooy TA, Thijssen S, Bulte CS, et al. Smartwatch based automatic detection of out-of-hospital cardiac arrest: Study rationale and protocol of the HEART-SAFE project. Resuscitation Plus 12 2022;12:100324. ISSN 26665204.

[11] Shah K, Wang A, Chen Y, Munjal J, Chhabra S, Stange A, et al. Automated loss of pulse detection on a consumer smartwatch. Nature 6 2025;642(8066):174–181. ISSN 0028-0836.

[12] Clifford GD, Silva I, Moody B, Li Q, Kella D, Shahin A, et al. The PhysioNet/Computing in Cardiology Challenge 2015: Reducing false arrhythmia alarms in the ICU. Computing in Cardiology 2015;42:273–276. ISSN 2325-8861.

[13] Moody G, Mark R. The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine 2001;20(3):45–50. ISSN 07395175.

[14] Johnson AE, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific Data 5 2016;3. ISSN 20524463.

[15] Johnson A, Pollard T, Mark R. MIMIC-III clinical database (version 1.4), 2016.

[16] Moody B, Moody G, Villarroel M, Clifford GD, Silva I. MIMIC-III waveform database matched subset (version 1.0), 2020.

Address for correspondence:

Roelof G. Hup
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands
r.g.hup@tue.nl