Code for "Developing a Scalable Annotation Method for Large Datasets That Enhances Alarms With Actionability Data to Increase Informativeness: Mixed Methods Approach"

This code repository accompanies the following publication:

Klopfenstein S.A.I.[1,2], Flint A.R.[1], Heeren P.[1,4], Prendke M.[1], Chaoui A.[1], Ocker T.[4], Chromik J.[3], Arnrich B.[3], Balzer F.[1, 5], Poncette A.S.[1,4] (2025). **Developing a Scalable Annotation Method for Large Datasets That Enhances Alarms With Actionability Data to Increase Informativeness: Mixed Methods Approach**

[1] Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Institute of Medical Informatics, Charitéplatz 1, 10117 Berlin, Germany

[2] Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Core Facility Digital Medicine and Interoperability, Charitéplatz 1, 10117 Berlin, Germany

[3] Hasso-Plattner-Institute, University of Potsdam, Digital Health - Connected Healthcare, Rudolf-Breitscheid-Straße 187, 14482 Potsdam, Germany

[4] Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Department of Anesthesiology and Intensive Care Medicine, Charitéplatz 1, 10117 Berlin, Germany

[5] Einstein Center Digital Future, Berlin, Germany

https://doi.org/10.2196/65961

Contents

This repository contains the R scripts that were used to process raw *.CSV or *.XML alarm log files from Philips Intellivue MX800 patient monitors. The logs are transformed into a more readable and useful form, the alarms are then classified and linked to their respective patients and finally pseudonymized.

The raw alarm logs have 5 columns, Date, Bedname, Action, Device-Name, Clinical User:

Datum	Bettname	Aktion	Geräte-Name	Klinischer Benutzer
21.05.2020 21:19:24	Bett123	**ABPs 69 <90 21:19:20.	PIIC iX: gwm123	

Using RegEx information is extracted from these logs, like type and severity of alarm, if it is a start, stop or pause, the start/end time of the alarm, as well as the actual measured value and the threshold. Since alarms do not have identifiers in the logs, the pause and alarm durations can only be determined with a best effort guess, by sorting the alarm events in chronological order. The result would be a table row like this (example filled with random values):

Column Name	Data Type	Description	Example Value
Datum	DATETIME	log entry timestamp	2019-10-12 07:49:05
Bettname	String	bed name	WAB01
Aktion	String	logged alarm string	**SpO2r 85 <90 beendet.
Devicename	String	name of device	PIIC iX: GWWAWR
station	String	name of icu unit	PAN-PACU
weekday_name	String	linguistic name of weekday	Monday
weekday_int	Integer	number of day of the week starting Mondays	1
total_beds	Integer	sum of unique beds in icu unit in log	22
shift	String	working shift label	early
alarm_level	Integer	alarm severity level in range 1-3	2
alarm_color	String	alarm color label (blue, yellow, red)	yellow
alarm_start	Integer	is alarm start boolean	0
alarm_end	Integer	is alarm end boolean	1
alarm_start_time	DATETIME	time of alarm start	2019-10-12 07:49:04
alarm_type	String	type of alarm	AF
value_measured	Integer	measured value	4
alarm_threshold	Integer	alarm threshold	8
threshold_excess	Integer	margin by which alarm threshold was exceeded	4
alarm_group	String	category of alarm type	EKG
limit_exceeded	String	binary indicator of upper/lower limit exceeded	LOWER
pause	String	BINARY indicator for pause start/stop	START
alarm_indication	String	alarm type grouping label for annotation	hf_low
pause_duration	Integer	duration of pause in seconds (best effort guess)	12

Column Name	Data Type	Description	Example Value
time_lag	Integer	lag between alarm time and log timestamp in seconds	5
alarm_duration	Integer	duration of alarm in seconds (best effort guess)	19
AUFN	DATETIME	time of hospital admission	2019-10-10 13:15:00
ENTL	DATETIME	time of hospital discharge	2019-10-14 08:00:59
subject_id	Integer	patient identifier	895532155478511
hadm_id	Integer	Hospital stay identifier	232654224158824
alarmlog_id	String	alarm log entry identifier -- MD5 of Datum + subject_id + Aktion + Devicename, not strictly unique	3edcdfa6e2a520b619dbcf3d74d548e4

While the log transformation and mapping is rather generic, the patient linking and pseudonymization code is highly specific for our local infrastructure. It would not work in other settings, but it might still give an idea of how to tackle the task.

Usage

Place the *.XML or *.CSV alarm log files in a ./input/raw/[stationname] subfolder, where stationname will be the short name of the ward where the logs come from. Then adjust paths where necessary and run the main.R script (or run scripts separately in the ./source folder).

Acknowledgement

The scripts for the alarm log transformation and alarm classification are enhanced versions of R scripts by Poncette et al (https://zenodo.org/records/4560041#.YJcyX7VKg2w) which were developed for the publication **Patient Monitoring Alarms in an Intensive Care Unit: Observational Study With Do-It-Yourself Instructions** (https://doi.org/10.2196/26494).

BMBF Grant

The INALO project, funded by the German Federal Ministry of Education and Research (BMBF) under the "KMU Innovativ" program, operates under grant number 16SV8559.