MDDT Summary of Evidence and Basis of Qualification for

The University of California San Francisco (UCSF) Lethal Arrhythmia Database (LAD)

BACKGROUND

MDDT NAME: The University of California San Francisco (UCSF) Lethal Arrhythmia Database (LAD)

SUBMISSION NUMBER: U220727

DATE OF SUBMISSION: October 3rd, 2023

CONTACT: Fabio Badilini, Director Center for Physiologic Research
Center for Physiologic Research,
Division of Cardiology, University of California San Francisco,
505 Parnassus Avenue, Room M1182A, Box 0124
San Francisco, CA 94143-0124
Email: fabio.badilini@ucsf.edu

TOOL DESCRIPTION AND PRINCIPLE OF OPERATION

The University of California San Francisco (UCSF) Lethal Arrhythmia Database (LAD) Medical Device Development Tool (MDDT) version 1.X.X is designed to test and report the performance of computerized methods to detect lethal cardiac arrhythmias in patient monitoring systems. USCF LAD is composed of a set of digital signals (ECG waveforms, SpO2, invasive arterial blood pressure and transthoracic impedance recordings), acquired in consecutive patients admitted to an Intensive Care Unit (ICU) which includes a large set of annotated lethal cardiac arrhythmias, specifically: asystole (AS), ventricular tachycardia (VT) and ventricular fibrillation (VF).

Performance of the algorithm under evaluation is based on comparison between the test algorithm annotation files and the reference annotation files.

The total number of annotations in the reference database (named “LAD” – Lethal Arrhythmia database) is: 6568 VTs in 643 patients; 32 VFs in 22 patients; and 446 AS in 240 patients.

Summary Results of the LAD v1.0.0

Database Attributes	VT	VF	AS
Total # of patients	5,302*	5,319	5,319
Total # of records	30,547*	31,004	31,004
# of events	6,568	32	446
# of subjects with >= 1 event	643	22	240
# of records with >= 1 event	1,156	25	248
% subjects with >= 1 event	12.1%	0.41%	4.5%
% records with >= 1 event	3.78%	0.08%	0.79%

Table 1: ** 17 patients with ambiguous VT excluded, corresponding to 457 records*

The process to test a submitted algorithm is as follows: the submitting organization will upload the test algorithm to Center for Physiological Research UCSF-CPR, who will run the algorithm on the database and generate a report summarizing the performance of the algorithm.

Therefore, manufacturers utilizing this tool will remain blind to the database signals and ground truth annotations.

The output of the MDDT is a report that includes descriptive and graphical statistics to summarize performance. The reported performance metrics include sensitivity, positive predictive value (+P), and false positive/negative computations, separately computed for each type of arrhythmia alarm. All statistics are based on arrhythmia events and do not include beat-specific metrics, as the MDDT is intended to address event detection performance, which may not be based on the detection and classification of individual beats.

The statistics are computed separately for ventricular tachycardia (VT), ventricular fibrillation (VF) and asystole (AS). Overall (ALL Events) is also reported pooling VT, VF and AS together, to account for mixed-event mislabeling.

The statistics are reported both in an aggregated (full database, and grouped by ICU, Age, Gender, Day, and Subject) and in a detailed record-by-record basis.

Agreement between the test algorithm output (i.e., detection of relevant events) and the reference LAD database is based on similar methods as those specified in the ANSI/AAMI EC57 standard and is established on the existence of a match between the reference and test events.

The performance metrics are based on sensitivity (Se, i.e. the fraction of reference events which are detected by the test algorithm) and positive predictivity (+P, i.e. the fraction of test events which are true events), and include the numbers of false positives and false negatives. All metrics are separately computed for each of the alarm types and on the pooled set (all-alarms combined). Finally, the performance metrics are reported for each of the subgroups, namely ICU type, Age, Gender, Day, Bed, and by Subject.

In this MDDT, a match is defined by any event in which the onsets of the reference and test events are less than five seconds apart.

The report generated by this tool does not necessarily constitute sufficient validation data to support the acceptable performance of the algorithm in a device regulatory submission. Device manufacturers should determine whether additional data is necessary for device validation, given the specifics of the device technology and proposed device indications for use. Additional data representative of specific device use and technology may be necessary (e.g., in case of different use cases such as ambulatory patients, novel technologies like machine learning/artificial intelligence, devices that may directly deliver or drive care, devices for forecasting future events, etc.). It is sole responsibility of the submitting organization to determine and document the adequacy of the LAD database for any specific use.

The tool does not supersede requirements in other standards and/or guidance documents.

CONTEXT OF USE

The UCSF LAD is designed to test and report the performance of existing and new algorithms for the detection of three types of lethal cardiac arrhythmia alarms (ventricular tachycardia, ventricular fibrillation and asystole) during hospital-based adult patient monitoring inside 3 different types of ICU (Cardiac, Medical/Surgical and Neurological).

The primary target audience of the tool are patient monitoring system device or algorithm developers, who will be able to evaluate the performance of their algorithm for identifying the three cardiac arrhythmias against an independent dataset. The results generated by the tool could be used as non-clinical test data supporting algorithm performance in a regulatory submission.

SUMMARY OF EVIDENCE TO SUPPORT QUALIFICATION

UCSF-LAD is based on real-world data collected over 19 months from 5,319 patients, 77 beds in 5 ICUs of three types (cardiac, medical/surgical and neurological) at UCSF Medical Center.

Data collected included both ECG signals and non-ECG waveforms, such as the plethysmogram (SpO2), thoracic impedance, and invasive arterial blood pressure. To prevent potential algorithm over-training, database waveforms and annotations are not accessible to the submitting organization.

Performance of the tested algorithm is based on comparison between the test annotation files (algorithm outputs, e.g., times when the algorithm has determined that the arrhythmia event have begun) and the reference annotation files (LAD). The method for generating the reference annotation files and ensuring the accuracy of the annotated arrhythmias is summarized below.

Performance data (e.g., sensitivity, positive-predictive-value, true/false positive/negative, etc.) is consistent with existing consensus standards (e.g., ANSI/AAMI EC57:2012) and reported for the various events, both in aggregate (full database, and grouped by ICU, Age, Gender, Day and Subject) and in a detailed record-by-record basis.

The LAD has been annotated by the Center of Physiologic Research (CPR), a Center established within the Department of Physiologic Nursing at the UCSF School of Nursing using both a set of automated lethal arrhythmia algorithms and a multi-level human reader annotation protocol. Indeed, the potential events were initially identified by existing computer algorithms designed to detect lethal arrhythmia and then reviewed by the human teams.

Human review of the arrhythmia alarms relied on two expert teams:

1. The Nursing Review Team (NRT): composed of five experienced faculty-based nurse scientists.
2. The CPR-based Expert Review Team (ERT): composed of two engineers with years of experience in signal processing methods used in hospital-based monitoring and by an expert cardiologist.

The human annotations review workflow employed was slightly different for the three types of alarm annotations:

VT Annotations

The dataset was annotated by the NRT and by ERT using the following steps/protocol:

1. Each alarm was randomly assigned to three NRT annotators.
2. VT for which the three assigned annotators concurred were finalized.
3. VTs for which agreement had not been achieved were re-assigned to a 4th NRT annotator:
    a. If the disagreement was resolved the VT was finalized.
    b. If the disagreement was not resolved the alarm was assigned the ERT team.

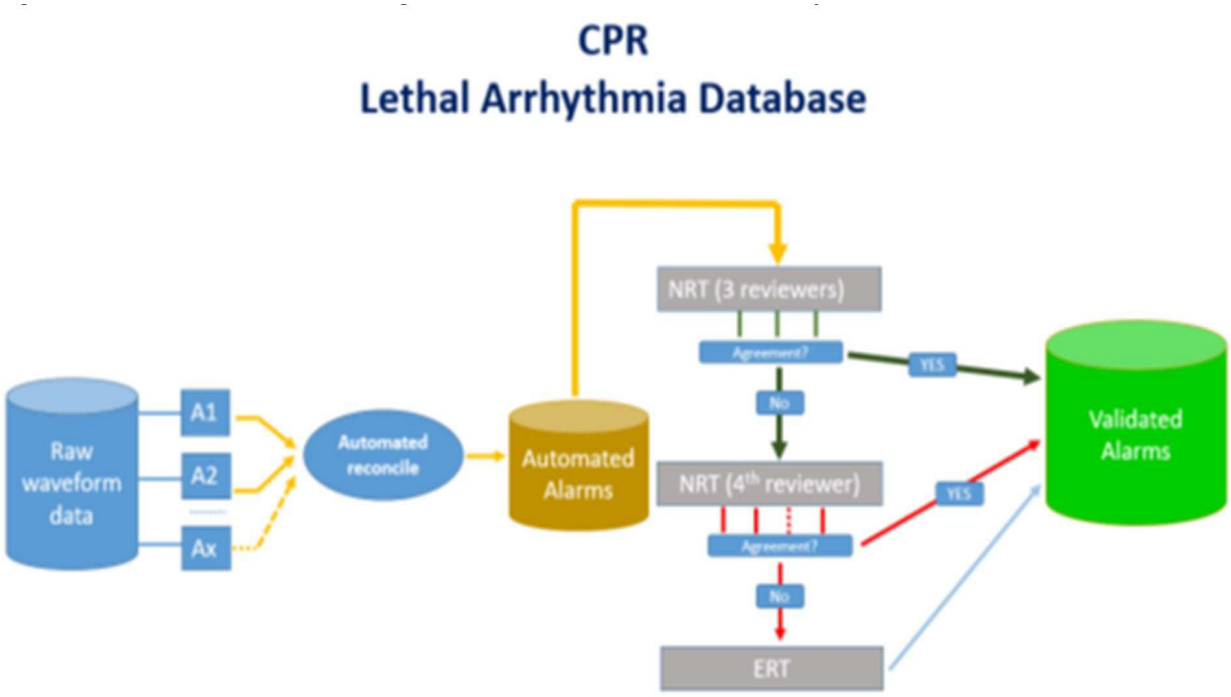

Figure 1. Overview of Validating Annotations in the Lethal Arrhythmia Database

VF and AS Annotations

The potential VF and AS alarms were directly assessed by the ERT team in conjunct review sessions.

Missing events

Considering the size of the LAD dataset, a complete manual analysis of all records to identify true events would not be practically feasible. Identification of true event is based on automated analysis of the dataset followed by manual review, and it is likely that not all the events were identified in this process. To evaluate the amount of positive events

missed, a random sampling of the dataset was manually analyzed for the solely VT events. The results from this analysis allow the estimate that missing VT events are less than 400 across the entire database including 6568 positive VT events. A similar analysis was not conducted for other event types as their lower prevalence would have required the analysis of a significantly larger random sample; however considering that other events are identified with similar methods, it is reasonable to expect that a similar proportion of events may be missed for other types (VF, AS).

DISCUSSION OF THE STRENGTH OF EVIDENCE SUPPORT QUALIFICATION

The UCSF-LAD is a large annotated dataset to be used to compute lethal arrhythmia performance metrics for new and existing algorithms used in bedside monitoring. The paradigm is similar to the databases used in currently available consensus standards such as ANSI/AAMI EC57:2012, but proposing a benchmark of data with a series of important advantages such as:

1. The size: 19 months of continuous data.
2. The population is from the very same context of the intended usage of the targeted algorithms.
3. No selection bias: all patients included.
4. The data has been collected from FDA-cleared bedside monitoring equipment, i.e. precisely the context intended by the MDDT.
5. The annotation effort followed a very strict protocol, specifically designed for the purpose of the MDDT.
6. The MDDT tool is supported by a large University Hospital, thus assuring a long-term and independent source of support to its potential usage.
7. UCSF-LAD prevents over-training and performance overestimation, as the database waveforms and annotations are not accessible to the submitting organization.

ASSESSMENT OF ADVANTAGES/DISADVANTAGES OF QUALIFICATION

The technology for bedside monitoring systems has experienced continuous innovation, with new and more sophisticated hardware components being implemented. Despite these advancements, the development of new alarm detection algorithms (or simply the improvement of existing ones) has been stagnated. Specifically, and more critically, the software underlying the detection of critical and lethal cardiac arrhythmias is often designed on old technology or based on outdated signal contexts, where only one or two electrocardiographic (ECG) data channels represent the only input used by the algorithms.

Part of the problem is that current requirements for the validation of novel cardiac alarms algorithms are based on the AAMI/ANSI-EC57:2012 and IEC 60601-2-47:2012 Standards. These standards rely on three existing databases of ECG signals (the AHA-ECRI, the CUBD, and the MIT/BIH Databases) which were developed in the mid 1970s and do not include digitally acquired ECGs or any of the other signals currently acquired in modern patient monitoring systems. These data are limited in duration, number, and quality of recordings. These datasets (signals and annotations) have been publicly available for decades, which also raise concerns of potential over-training whenever these datasets may have been used as part of algorithm development process.

Regulatory agency and private stakeholders have expressed the need for newer and better tools for assessing the performance of arrhythmia detection algorithms in the different contexts they need to be implemented.

A critical aggravating factor is the large false positive incidence rate of critical arrhythmia detection by currently implemented algorithms that can easily elicit “alarm fatigue” on bedside/telemetry nursing personnel. This phenomenon has been described and exquisitely quantified, with false arrhythmia alarm ratios for lethal, or “critical” cardiac events, reaching levels as high as 90%.

The core of this tool is the UCSF LAD database, whose data was collected in the real-world clinical context of an ICU that it targets to serve. It is a unique repository in terms of size, acquired waveforms, diversity of stored events and thorough annotations by human experts. It is evolutionary in nature, and because of its design, the UCSF LAD database is able to expand its scope and applicability to meet the developmental, evaluative, and regulatory needs required for both academic and commercial developers of new sophisticated ECG-based bedside monitoring technologies. Lastly, the organization supporting the MDDT is housed in one of the leading academic medical centers in the world, and is financially independent, thus, is capable of supporting long-term advances in the field of arrhythmia algorithm development.

One disadvantage is related to that fact that in its current implementation, the UCSF LAD MDDT is only suitable for three lethal cardiac arrhythmia types. This limitation will eventually be resolved in future versions and/or in other more specific MDDTs. Under this qualification minor changes or additions to the database may be made under version 1.X.X to improve consistency between the automated analysis and the manual review of missed true events.

CONCLUSIONS

This non-clinical assessment model MDDT is qualified within its context of use as well as the advantages and disadvantages indicated above.

This MDDT introduces a novel and unique database that provides a strong benchmark for the validation of algorithms implemented in bedside monitoring.
Among the most relevant, in contrast to other databases, this MDDT prevents over training and performance overestimation, while minimizing patient selection bias, and it is large enough to provide statistically significant metrics.
Furthermore, this is the first ECG database which also includes non-ECG waveforms data will encourage developers to test multiparametric algorithms.

CONTACT INFORMATION FOR ACCESS TO TOOL

Center for Physiologic Research,
Division of Cardiology, University of California San Francisco,
505 Parnassus Avenue, Room M1182A, Box 0124
San Francisco, CA 94143-0124
Email: mddt_cpr@ucsf.edu