PhysioTag: An Open-Source Platform for Collaborative Annotation of Physiological Waveforms

Lucas McCullum[1], Hasan Saeed[1,2], Benjamin Moody[1], Diane Perry[1,3], Eric Gottlieb[3,4], Tom Pollard[1], Xavier Borrat Frigola[1,5], Qiao Li[6], Gari Clifford[7], Roger G Mark[1], Li-wei H Lehman[1]

[1] Laboratory for Computational Physiology, MIT, Cambridge, MA, USA
[2] University of Michigan, Ann Arbor, MI, USA
[3] Beth Israel Medical Center, Boston, MA, USA
[4] Renal Section, Brigham and Women's Hospital, Boston, MA, USA
[5] Anesthesiology and Critical Care Department, Hospital Clinic de Barcelona, Barcelona, Spain
[6] Department of Biomedical Informatics, Emory University, Atlanta, GA, USA
[7] Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Abstract

To develop robust algorithms for automated diagnosis of medical conditions such as cardiac arrhythmias, researchers require large collections of data with human expert annotations. Currently, there is a lack of accessible, open-source platforms for human experts to collaboratively develop these annotated datasets through a web interface. In this work, we developed a flexible, generalizable, web-based framework to enable multiple users to create and share annotations on multi-channel physiological waveforms. Using the developed annotation platform, we carried out a pilot study to assess the validity of ventricular tachycardia (VT) alarms from multiple commercial monitors. Thus far, four clinical experts have used this annotation tool to annotate a total of 5,658 VT alarm events, among which approximately 44% (N=2,468) have been labeled by two independent annotators.

1. Introduction

Electrocardiographic (EKG) monitoring is becoming increasingly commonplace, with modern ambulatory devices allowing long-term, continuous capture from a broad population. To develop robust algorithms for automated diagnosis and characterisation of medical conditions such as ventricular tachycardia (VT), researchers require high-quality annotations. These annotations typically need to be provided by human experts. Currently there is a lack of freely available, maintained, open-source software to enable collaborative, human annotation of physiological waveforms such as EKG. Prior annotation tools, such as MetaAnn [1] and WAVE [2], focused on a single-user mode annotation in a local desk-top setting without support for remote Web-based access. Web-based EKG annotators such as WaveformECG [3] and LabelECG [4] were developed to enable remote visualization and annotation of ECG waveforms.

We developed an open-source annotation platform, PhysioTag, that enables experts to collaboratively annotate physiological waveform records using a standard web browser. The software, freely available on PhysioNet (`https://physionet.org/content/physiotag/`) [5], is simple to install, following best practice in Python packaging, and offers a range of features, including: user management and task customization; a programmatic interface for data import and export; and a leaderboard for annotation progress tracking. Using the platform, we carried out a pilot study to assess the validity of ventricular tachycardia (VT) alarms from several commercial hospital monitors.

Our eventual goal is to integrate the Web-based annotation platform with the PhysioNet system to encourage and enable PhysioNet research communities to collaboratively develop and share annotated datasets. In the remainder of this paper, we first describe different components of our platform, and the rationales behind our architectural design. We then describe the results from the VT pilot study, followed by our conclusions and discussions for future work.

2. Methods

The design of our platform is driven by the following objectives and considerations:

- Scalable collaborative annotation: allow multiple concurrent annotators, potentially from diverse geographic

locations, to annotate datasets in parallel.
- Fast response time in waveform visualization with customizable display settings.
- Flexible adjudication user interface to enable human adjudicators to review and resolve conflicting annotator decisions.
- Functional annotation management to track project progress and manage user accounts and annotations.
- Lightweight and easy to deploy annotation servers.
- Open data format using standard WFDB Python library to enable remote access to datasets on PhysioNet.

In the remainder of this section, we detail the design and implementations of our system to achieve the above objectives.

2.1. Web-based access: server/client architecture

Since the purpose of this project was to create as many annotations as possible, our goal was to minimize the amount of time and effort, on the part of individual annotators, to make the software work. To avoid the need to install a custom application on the annotator's machine, we opted to build the system as a web application that can be used from any modern web browser. The interface is designed to present the viewer with the most relevant information, allow them to make their decision quickly, and save their responses to the server automatically.

The annotation platform was implemented in Python using Django, enabling ease of deployment, as well as ease of developing new features and customizations for specific annotation projects. The user interface uses *django-plotly-dash* to provide a featureful, efficient, cross-browser display of the waveforms.

On the back end, the server uses the WFDB Python Package [6] to read the input waveform files. This allows the system to be used to annotate any waveform records stored in WFDB format (which includes most waveform data published on PhysioNet). To provide an interface to open-access web-based data interfaces such as PhysioNet, an Application Programming Interface (API) was developed using GraphQL to extract annotations from the platform's database for external applications.

2.2. Waveform visualization and annotation workflow

In order to build a "gold standard" corpus of event annotations, each event needs to be reviewed by two annotators independently. However, the experience of past annotation projects has shown that we rarely know in advance how many total records will be available for annotation, how many expert reviewers will be able to participate in the project, or how quickly each annotator will be able to work. So, rather than deciding in advance which annotators will review each event, events are assigned to annotators dynamically. When a new annotator joins the project, they are assigned 100 randomly-selected events that have not yet been annotated. After they have finished reviewing those 100 events, they can assign themselves a new batch of 100, and so on. This "self-assignment" strategy is intended to ensure a diverse annotation dataset between the annotators, while encouraging productivity through multiple small tasks instead of one large task.

After registering and logging in, an annotator can use the annotation interface to label waveform records, enter comments, bookmark samples to return to them later, and continue to the next waveform (Figure 1). For our task, users were given the options of "True" for when they believe the alarm was correct, "False" for when they believe the alarm was incorrect, "Uncertain" for when they are unsure which annotation to assign, "Reject" for when the alarm is unreadable due to noise, artifacts, or other hindrance, and "Save for Later" for when the user would like to return to annotate this event at another time.

To provide self-evaluation and training, we provided a small set of expert-annotated sample waveforms as a "practice set". New users are encouraged to review and annotate these sample records in order to practice using the platform and to refresh their knowledge. At the end of the practice test, the user can compare their answers with those of the experts.

2.3. Adjudication

Significant inter-annotator variability can exist among manual labels by clinical experts [7]. In order to resolve conflicts between two annotator decisions, an adjudication framework was added to recruit extra opinions on particularly difficult alarm events. Project administrators can designate trusted annotators to serve as adjudicators based on the annotator's level of expertise. Disagreements between annotators can then be resolved either through an active one-on-one discussion between the annotators involved, or by an adjudicator voting to breaking the tie.

2.4. Annotation management

Throughout the annotation process, users will have the ability to view their current complete, incomplete, and save-for-later annotations and make changes (i.e., adjust their comments, change their decision). At any point, the user can adjust the settings for the annotator interface (i.e., signal thickness, time before / after the alarm event, colors, downsampling, etc.) to optimize their efficiency in completing the annotations.

A public leaderboard was created to show the ranking of each annotator over the past day, week, month, and

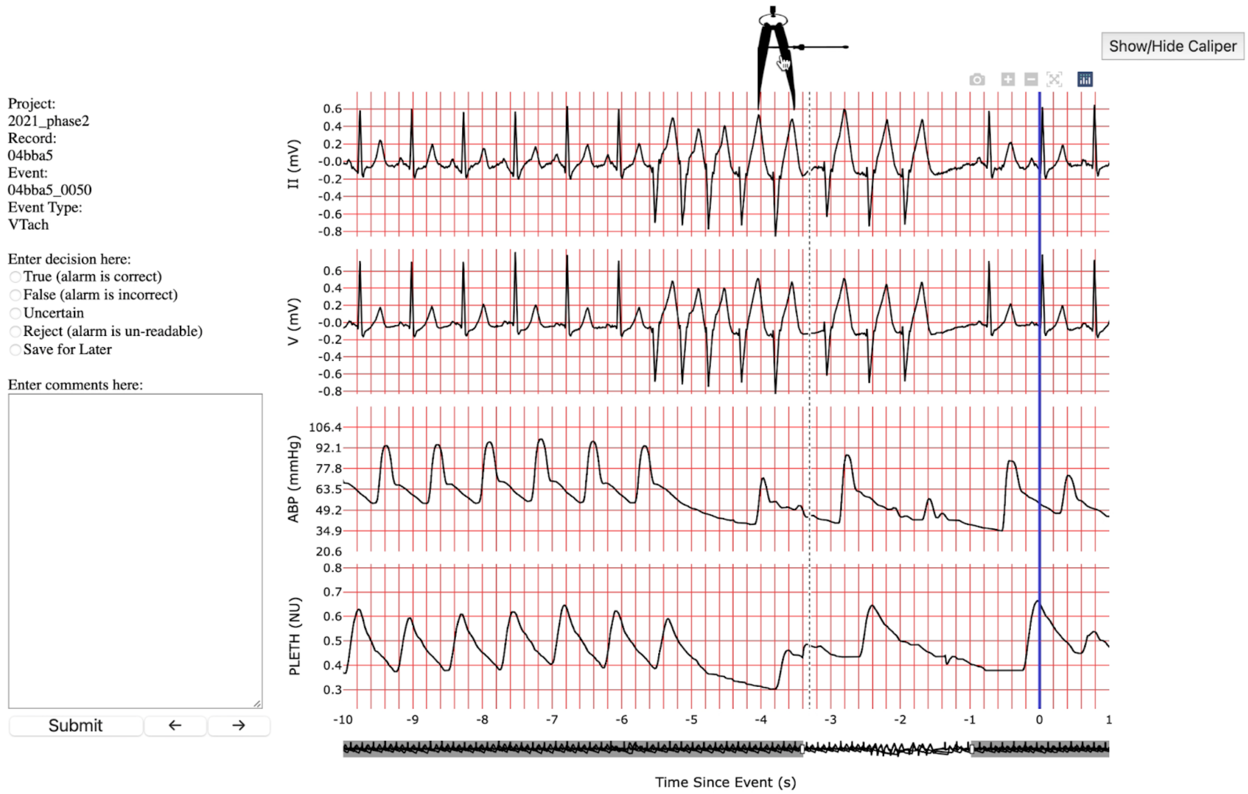

Figure 1. The annotator's annotation interface with the current annotation project 2021_phase2 patient record (04bba5), patient event (04bba5_0050), arrhythmia label (VTach for ventricular tachycardia), annotation decisions, accompanying comments, and a button to submit and arrows move to the previous or next annotation in the left column. In the right column are the EKG waveforms and supporting physiological signals with an optional scalable and draggable caliper (black object where mouse is hovering) for precise interval measurement. At the top of the figure, small buttons can be seen which take a screenshot of the current waveforms, zoom in, zoom out, and reset the view while, at the bottom, a compressed version of the entire signal range can be seen and dragged to move the signal window through time surrounding the VT event. The user also has the option to manually adjust the y-axis scaling for each waveform by clicking on the y-axis labels. The vertical blue line extending across each signal represents the time of the alarm, while the vertical dashed black line represents the current cursor location for comparing signals at the same time point.

all time in the number of annotations completed to motivate healthy competition between the annotators to stimulate productivity. Further, pie charts are displayed on the leaderboard to track the ratio of alarm annotation decisions and completion statistics for all the possible alarm events in the dataset. Admin users of the platform have access to an admin console which has functionality to invite new users, view current user waveform annotator settings, and assign current users as an adjudicator or admin. Also shown on the admin console are all the complete and incomplete annotations and adjudications along with their associated user, decision, comments, and timestamp.

3. Pilot study: VT alarm annotation

One of the most difficult arrhythmias for algorithms to detect reliably is ventricular tachycardia[8, 9]. Despite considerable effort from algorithm developers, false arrhythmia alarms in ICUs are still a major problem [10]. Thus, we conducted a pilot project to develop an annotated database with human-confirmed VT episodes. Following the PhysioNet Challenge 2015, we define a VT episode as five or more consecutive ventricular beats with heart rate higher than 100 beats-per-minute (bpm) [8]. Using a customized viewer on our platform, participants were shown waveform segments and asked to classify the monitor-identified arrhythmias as "true", "false", "uncertain" or "reject" if a decision cannot be made due to low data quality. Each alarm event will be presented to two annotators. In the event of disagreement, the alarm will be annotated by an adjudicator. Four expert cardiac clinicians were recruited to participate in the pilot study to annotate 3,486 VT alarm events.

We collected multi-channel physiological waveforms, including EKG, arterial blood pressure (ABP), and photoplethysmography (PPG) from multiple manufacturers in clinical ICU settings and converted them to WFDB format. Waveforms were saved in 10 minute segments at 250 Hz with five minutes on each side surrounding the alarm event. Thus far, four clinical experts have used this annotation tool to annotate a total of 5,658 VT alarm events, among which approximately 44% (N=2,468) have been labeled by two independent annotators. Among those, 19% (N=469) received conflicting annotation decisions and thus required adjudication to resolve the conflicts.

4. Discussion and future work

Immediate future work for our annotator and associated data acquisition includes expanding our expert-annotated EKG database and releasing it to PhysioNet for open-access to the community to catalyze the development of novel algorithms and machine learning models for the reduction of monitor-generated EKG arrhythmia false alarms in the ICU.

We also plan to recruit more expert annotators to expedite the annotation and adjudication process as our dataset increases. To do so, we will add expanded functionalities for user account creation and management as well as a robust method to judge annotator quality assurance / control and expertise. Further, work is currently underway to convert this single-purpose tool into a more generalized tool which can be used to interface with and annotate other physiological waveform datasets hosted on PhysioNet such as the electromyography (EMG), electroencephalogram (EEG), and magnetomyography (MMG) or any other new dataset which may be published.

5. Conclusions

We developed a novel, flexible, and generalizable, web-based, interactive, annotation platform and derived an expert-annotated dataset for cardiac arrhythmia false alarm reduction. We foresee that the successful completion of our web-based annotation platform as a generalized annotation tool will foster widespread collaborative effort in the development of large annotated datasets, enabling training of machine learning models, not only for the task of VT false alarm reduction from EKG, but also for a wide range of challenging tasks involving other forms of physiological waveforms. With continuously augmented, freely accessible, datasets derived from a web-based annotator, the coverage of novel algorithm development will reach areas of limited access to essential resources required to collect and annotate complex physiological signals for analysis. The future deployment of this platform onto an open-access web-based data interface like PhysioNet will promote community-based solutions to problems unsuccessfully tackled in "information silo" environments.

Acknowledgments

This research is supported by NIH grant R01EB030362.

References

[1] Moody B. MetaAnn. URL https://github.com/MIT-LCP/metaann.

[2] Moody G. WAVE user's guide, 2014. URL https://physionet.org/physiotools/wug/.

[3] Winslow RL Granite S JC. WaveformECG: A platform for visualizing, annotating, and analyzing ECG data. Comput Sci Eng Sep-Oct 2016;18(5):36–46.

[4] Ding Z, Qiu S, Guo Y, Lin J, Sun L, Fu D, Yang Z, Li C, Yu Y, Meng L, Lv T, Li D, Zhang P. LabelECG: A web-based tool for distributed electrocardiogram annotation. httpsarxivorgabs190806553 August 2019;.

[5] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. circulation 2000;101(23):e215–e220.

[6] Xie C, McCullum L, Johnson A, Pollard T, Gow B, Moody B. Waveform Database Software Package (WFDB) for Python. PhysioNet 2022;URL https://doi.org/10.13026/MMPM-2V55.

[7] Zhu T, Johnson AE, Behar J, Clifford G. Crowd-sourced annotation of ecg signals using contextual information. Annals of Biomedical Engineering Apr 2014;.

[8] Clifford GD, Silva I, Moody B, Li Q, Kella D, Shahin A, Kooistra T, Perry D, Mark RG. The Physionet/Computing in Cardiology Challenge 2015: reducing false arrhythmia alarms in the icu. In 2015 Computing in Cardiology Conference (CinC). IEEE, 2015; 273–276.

[9] Lehman EP, Krishnan RG, Zhao X, Mark RG, Lehman LH. Representation learning approaches to detect false arrhythmia alarms from ecg dynamics. In Machine learning for healthcare conference. PMLR, 2018; 571–586.

[10] Drew BJ, Harris P, Zègre-Hemsey JK, Mammone T, Schindler D, Salas-Boni R, Bai Y, Tinoco A, Ding Q, Hu X. Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. PloS one 2014; 9(10):e110274.

Address for correspondence:

Li-wei Lehman
E25-505, MIT, 77 Mass Ave, Cambridge, MA 02139
lilehman@mit.edu