Real-Time Machine Learning-Based Intensive Care Unit Alarm Classification without Prior Knowledge of the Underlying Rhythm

1* Wan-Tai M. Au-Yeung, PhD
1* Rahul K. Sevakula, PhD
2 Ashish K. Sahani, PhD
1 Mohamad Kassab, MD
3 Richard Boyer, MD, PhD
4 Eric M. Isselbacher, MD, MSc
1,5 Antonis A. Armoundas, PhD

From:
* Equally contributing authors

1 Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA
2 Center for Biomedical Engineering, Indian Institute of Technology Ropar, Punjab, India
3 Anesthesia Department, Massachusetts General Hospital, Boston, MA
4 Healthcare Transformation Lab, Massachusetts General Hospital, Boston, MA
5 Institute for Medical Engineering and Science, Massachusetts Institute of Technology Cambridge, MA

Correspondence:
Antonis A Armoundas, PhD
Cardiovascular Research Center
Massachusetts General Hospital
149 13th Street
Charlestown, MA 02129
Email: aarmoundas@partners.org

© The Author(s) 2021. Published by Oxford University Press on behalf of the European Society of Cardiology. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Abstract

Background - This work attempts to develop a standalone heart rhythm alerting system for the intensive care unit (ICU), where life-threatening arrhythmias have to be identified/alerted more precisely and more instantaneously (i.e., with lower latency) than existing bedside monitors.

Methods and Results - We use the dataset from the PhysioNet 2015 Challenge, which contains records that led to true and false arrhythmic alarms in the intensive care unit. These records have been re-annotated as one of 8 classes, namely (1) asystole, (2) extreme bradycardia, (3) extreme tachycardia, (4) ventricular fibrillation, (5) ventricular tachycardia, (6) normal sinus rhythm, (7) sinus tachycardia, and (8) noise/artifacts. Arrhythmia-specific features and features that measure the signal quality were extracted from all the records. To improve ventricular fibrillation detection, an improved, over an existing, single-lead R-wave detection was developed that takes into account the R-waves detected in all ECG leads. To avoid false R-wave detection due to pacing spikes, ECG signals were filtered with a low pass filter prior to R-wave detection, while the raw signals were used for feature extraction. Random Forest was used as the classifier, and 10 times 5-fold cross-validation, resulted in a macro-average sensitivity of 81.54%.

Conclusions - In conclusion, comparing with the bedside monitors used in the PhysioNet 2015 competition, we find that our method achieves higher positive predictive values for asystole, extreme bradycardia, ventricular tachycardia and ventricular fibrillation; furthermore, our method is able to alert the presence of arrhythmia instantaneously i.e., up to 4 seconds earlier.

Key words: Bedside monitors; multi-class classification; artificial intelligence; machine learning; feature engineering; signal processing
Introduction

Electrocardiographic (ECG) signals remain the most important means of capturing cardiac activity in real-time for the monitoring of intensive care unit (ICU) patients. While existing bedside monitors are designed to raise alarms whenever the ECG recordings go out of the normal range, a persistent drawback is that most of the alarms raised by the bedside monitors are false alarms, creating an unnecessarily noisy environment and contributing to alarm fatigue.

The “PhysioNet/Computing in Cardiology Challenge 2015: Reducing False Arrhythmia Alarms in the ICU” was specifically designed to foster the development of new methods to filter out false alarms for asystole, bradycardia, tachycardia, ventricular tachycardia (VT), and ventricular fibrillation (VF). The PhysioNet data included physiological signals of 2-lead ECG, arterial blood pressure (BP), and photoplethysmography (PPG). From this challenge, several methods were created that based upon the prior knowledge of the alarms’ annotation from the bedside monitor; i.e., they were all designed to be used as a second stage classifier/filter, after the bedside monitor has raised an alarm. Since these methods were only built to filter out the false alarms of five specific heart rhythms, the design/training of these methods were limited to small data sets corresponding to the signal records that the bedside monitors considered to be abnormal. This raises the question on whether these methods could, by themselves, truly identify an abnormal heart rhythm from all other heart rhythm classes.

Multi-class heart rhythm classification of ECG signals has been a persistently challenging problem, and machine learning techniques have been essential in providing improved outcomes. One of the first attempts for arrhythmia analysis was by Guvenir et al, who used supervised machine learning to classify 12-lead ECG signals to 16 classes, and achieved a 10-fold cross validation accuracy of 68%. The authors published the extracted feature values from their data to the University of California Irvine repository, and later other studies have tested their methods on these data. Recently, the “China Physiological Signal Challenge (CPSC) 2018” provided a large, open repository of 6877 raw 12-lead ECG recordings, with the intent of building a classifier that can accurately sort the ECG recordings into 9 classes (normal heart rhythm or one of the 8 abnormal heart rhythms). The winner of the challenge used convolutional neural networks to achieve a median F-1 score of 0.84. Many popular studies on multi-class
classification of ECG rhythms/beats have utilized the data recorded from ambulatory devices or single-lead ECG devices. Among these, Hannun et al\textsuperscript{15} used a 34 layered convolutional neural network for detecting 10 heart rhythm classes from one sec of ECG signal from an ambulatory device. Due to the one second limitation, the model missed out some important heart rhythms that require more time for detection, e.g., sinus bradycardia, asystole. The “AF Classification from a Short Single Lead ECG Recording - The PhysioNet Computing in Cardiology Challenge 2017”\textsuperscript{19} encouraged the development of methods which could identify from a single short ECG lead recording (30-60 seconds), whether it shows normal sinus rhythm (NSR), atrial fibrillation (AF), an alternative rhythm, or is too noisy to be classified. The two best scorers\textsuperscript{20,21} from this challenge used stacked classifiers, wherein first a deep learning based classifier was used to obtain the classification result and the confidence score with which the classification was made; if the confidence score was below a defined threshold, the second classifier using handcrafted features was used to determine the class. It should be noted that most arrhythmia classification methods have analyzed only the morphology of ECG signals to determine the heart rhythm. However, critical life-threatening arrhythmias—such as VT and VF—can be detected more accurately by including the BP signal\textsuperscript{22}.

In this manuscript, we propose a standalone, real-time processing platform that can detect and categorize life-threatening heart rhythms by analyzing the ECG, BP, and PPG signals of ICU patients, into one of the multiple classes, such as VT, VF, extreme tachycardia, extreme bradycardia and asystole. We use signal processing techniques and machine learning to provide more precise and instantaneous alerts of life-threatening arrhythmias. Additionally, sinus tachycardia, NSR, and noise/artifacts have been included as one of the possible output classes.

\textbf{Methods}

\textbf{Dataset}

We used data from the “PhysioNet/Computing in Cardiology Challenge 2015: Reducing False Arrhythmia Alarms in the ICU”. The challenge data was sourced from four hospitals in the USA and Europe, and employed bedside monitors from three manufacturers; furthermore, it was ensured that no manufacturer or hospital contributed to more than half of the records. The data was made public and is available at https://physionet.org/content/challenge-2015/1.0.0/.
database includes records of 300 second long physiological signals from ICU patients, taken just prior to the time when an alarm was raised by the bedside monitor. The physiological signals included were: 2-lead ECG, BP, and the PPG. We re-annotated each of these events as belonging to one of the following categories: (1) asystole, (2) extreme bradycardia, (3) extreme tachycardia, (4) VF, (5) VT, (6) NSR, (7) sinus tachycardia, and (8) noise/artifact. The definitions/criteria for asystole, extreme bradycardia, extreme tachycardia, VT and VF were taken from the PhysioNet 2015 challenge, and are listed in Table 1. More information regarding the dataset is available in the Online Supplement.

**R-wave detection**

We used the R-wave peak detection method described by Martinez et al. The method identifies peaks based on the morphological characteristics of the ECG signal, and works well in most ECG records. To further improve R-wave peak detection that could impact the accurate detection of complex rhythms e.g. VF, we have developed a multi-lead moving window method, which is described in detail in the Online Supplement.

To avoid the scenario, that pacing spikes and other high-frequency noise are identified as R-wave peaks, we designed a finite impulse response low-pass filter in MATLAB. The filter has 51 taps and a cut-off frequency of 15 Hz. The raw ECG signal is, accordingly, passed through the low-pass filter, and the filtered signal is used by the multi-lead moving window based R-wave peak detector. Once the R-wave peak locations are identified, features are extracted from the raw ECG signal.

**Feature Extraction**

Feature extraction was performed on ECG, BP, and PPG signals. We extracted a set of signal quality indexes (SQIs) to indicate if the signals are appropriate for further processing or are noisy. Furthermore, a set of arrhythmia-specific features, which characterize each arrhythmia class, was also extracted (Table 3).

Many of these features have been used in our recent study on the reduction of false arrhythmia alarms in the ICU when one has prior knowledge of the type of the alarm; more
details about the features can be found in the online supplement of that paper. A major objective of our work is to perform real-time identification of patients’ heart rhythms. An implied sub-objective is that detection of these heart rhythms has to happen as soon as the arrhythmia pattern criteria is met, so that a timely alarm is raised. While observing the data/annotations from the PhysioNet 2015 challenge data, we realized that the alarms raised by the bedside monitors can sometimes be delayed several seconds after the abnormal heart rhythm criteria have been met. To achieve “instant” detection, we defined window length requirements for each heart rhythm class. For example, 4 sec of vital sign signals is sufficient to determine whether or not the patient is having VT/VF. Similarly, 8 sec is sufficient to determine if a patient is experiencing asystole. Finally, we found that 15 sec is an appropriate window-length to determine if the patient is suffering from extreme bradycardia or extreme tachycardia. Accordingly, the corresponding arrhythmia-specific features of extreme tachycardia, extreme bradycardia, asystole, VT and VF were computed from signals of 15 sec, 15 sec, 8 sec, 4 sec, and 4 sec window-lengths, respectively. The idea is pictorially presented in Figure 1, and the motivation for having different window lengths is illustrated in the Online Supplement.

Whenever one or more physiological signals were absent or noisy, while computing features the corresponding feature values were assigned NaNs (missing values). The SQIs were also designed in a way that the SQI values for noisy signals would either have low values or be rendered NaNs.

**Machine Learning-Random Forest Algorithm**

We tried multiple supervised machine learning algorithms for classification, namely artificial neural networks (ANN), support vector machine (SVM) and random forest (RF) classifiers. We have found that the RF classifier gave the best classification performance.
In a scenario in which one or more vital sign signal could be noisy, resulting in many missing feature values, the ability of a classifier to handle missing data is of high importance. The RF classifier has an in-built ability to directly handle missing data, while other classification algorithms, namely SVM and ANN need data imputation techniques to fill in the missing values, before they can be used for training and testing over the data. RF is an ensemble learning method that can be used for either regression or classification. More details regarding the machine learning random forest algorithm, are included in the Online Supplement.

Results

10-time 5-fold Cross-Validation

We performed k-fold cross-validation (with stratified random sampling) on the available data to estimate our method’s performance on the unseen data. To account for the stochastic differences during k-fold cross-validation, we assessed the 5-fold cross-validation performance of our method ten times, and then took the mean performance.

Performance Metrics

The sum of the confusion matrices derived from each 5-fold cross-validation, were summed up, and are presented in Table 4. Performance metrics, namely overall accuracy, macro-average sensitivity, macro-average positive predictive value, sensitivity of each class, positive predictive value (PPV) of each class, and the most common misclassification (when misclassified, which class instead is most chosen as the result), were derived from the confusion matrix, and have been noted in Table 5. Overall accuracy is defined as the ratio of all the true positives to the total number of all records. Sensitivity for each class is defined as the ratio of the number of true positives to the total number of records of that class. PPV for each class is defined as the ratio of the number of true positives to the total number of true positives and false positives of that class. Macro-average sensitivity is simply the sum of sensitivities of all classes divided by the total number of classes; similarly, macro-average PPV is the sum of PPVs of all classes divided by the number of classes.
**Algorithm Performance Evaluation**

Our algorithm consistently performed well in identifying arrhythmias. The confusion matrix of the 8-class classification results after 10-times 5-fold cross-validation is presented in Table 4. We have been able to achieve an overall accuracy of 74.7%. The macro-average sensitivity equaled 81.5% and the macro-average positive predictive value equaled 66.1%. For asystole, extreme bradycardia, and extreme tachycardia, the method achieved a sensitivity of 91.4% or above. The method achieved a sensitivity of 83.8% for VF and a sensitivity of 84.3% for VT. An important note here is that all the unidentified VF cases were classified as VT, which also is a serious life-threatening arrhythmia. In other words, an alarm was raised for 100% of the VF cases.

Figure 2 shows the heat map of our results, to pictorially depict the confusion matrix. It can be observed that the classifier did well in classifying the five types of life-threatening arrhythmias, with sensitivities exceeding 83.8%. Many of the errors made by the classifier were understandable. The most common misclassification of extreme tachycardia was sinus tachycardia, of sinus tachycardia was extreme tachycardia, and of extreme bradycardia was NSR; all these classes are similar morphologically and differ only by heart rate. We found that most of these misclassifications were actually borderline cases, in which the heart rate hovered around cutoff between classes. If required, the classification result can be enhanced by performing a simple secondary check on the heart rate. Another understandable error is that the most common misclassification result for asystole is extreme bradycardia. The error is understandable because asystole causes the mean heart rate to drop significantly. Furthermore, since extreme bradycardia is also a life-threatening arrhythmia and an alarm is raised for it, the error is not very concerning clinically.

The proposed method did not perform well in classifying normal sinus rhythm (NSR) and noise/artifacts. In our attempt to achieve high sensitivity in detecting arrhythmia, many NSR cases were misclassified as arrhythmia; incidentally, the majority of the false positives in asystole, extreme bradycardia, and sinus tachycardia classes belonged to the NSR class. Our experiments were performed on acausal records of different arrhythmic events and NSR; studies suggest that
such experiments may present over-optimistic results on false alarm suppression, and when implemented in real-time, they may exhibit higher false alarm rates and poorer NSR classification performance. To assess this algorithm for such scenario, we identified long, continuous NSR segments (noise free signals, at a heart rate 40-100 bpm) from the Physionet 2015 competition, that were not used for training/validation (amounting to ~28.5 K sec of data), and evaluated the five classifiers trained over the five-fold training-sets, on these new NSR segments. The classifiers identified the heart rhythm in a causal manner, once every sec, and alarms raised in close succession (i.e. within the next 15 seconds) were treated as the same alarm (no new alerts were raised). Our algorithm achieved a mean classification accuracy of 84.29% in identifying NSR, which is higher than that reported in Table 4. With the above settings, we estimate experiencing ~374 false alarms per patient, per day.

The class with the poorest performance was noise/artifacts, with a sensitivity of only 60.8%. Not surprisingly, we found that correctly identifying noise/artifacts is challenging, as noise appears in a wide variety of forms, such as flat line, high frequency noise, or fluctuating readings. We therefore found it difficult to design features that can consistently identify all forms of noise. One other error of particular concern was that 7.74% of VT events were misclassified as sinus tachycardia, which does not raise an alarm and thus could be of clinical consequence.

In Table 2, we can see that there is significant imbalance in the number of cases for each arrhythmia. The classification results however indicate that the supervised machine learning (ML) model is fairly not affected by the class imbalance.

While extracting the arrhythmia specific features from different sized windows of signals, though rare, a case may arise wherein two arrhythmia criteria are being satisfied simultaneously. For example, a case of asystole (window-length 8 sec) may simultaneously satisfy the criterion of extreme bradycardia (window-length of 15 sec). This is not a major limitation, since asystole and extreme bradycardia, are both life-threatening arrhythmias and an alarm is raised for both heart rhythms.

Algorithm Performance Comparison with Bed-Side Monitors
While it is hard to compare our classifier’s sensitivities for life-threatening arrhythmias with ICU monitors (these data are not available), sensitivities of at least 83.8% are nonetheless very encouraging. We now compare the PPVs of our proposed system to those of the bedside monitors used in the PhysioNet 2015 challenge\(^2\), MIMIC II study\(^26\), and the alarm fatigue study from UCSF\(^27\) (Table 5). It is important to note that the PPVs mentioned here for the bedside monitors are based purely on the empirical observations of the number of true alarms to the total number of alarms. Furthermore, we are assuming that the alarm records included in each of these studies accurately reflected the monitors’ actual abilities to discriminate true from false alarms. All PPVs were computed by taking the ratio of the number of true alarms to the total number of alarms raised for the relevant arrhythmia. In comparing the various PPVs, the proposed algorithm achieves higher PPVs for most life-threatening arrhythmias. Across the five life-threatening arrhythmias, the proposed algorithm achieved a macro average PPV of 62.46%; in comparison, the monitors in the PhysioNet 2015 challenge achieved a PPV of 39.63% and the Philips monitors in the MIMIC II study achieved a PPV of 57.26%. With regard to the UCSF study, we compared the algorithm performance across only asystole, VF and VT; the macro-average PPV of our algorithm was 51.53% and that of the GE bedside monitors in UCSF study was 18.06%. Following AAMI (Association for the Advancement of Medical Instrumentation) recommendations, Philips Healthcare demonstrated their “ST/AR” (ST segment and arrhythmia) algorithm (used in Philips bedside monitors) on the American Heart Association (AHA) and MIT-BIH (Massachusetts Institute of Technology-Beth Israel Hospital) databases, and claimed that their algorithm achieved 100% sensitivity in generating red alarms during episodes of VF. GE Healthcare similarly claims that their bedside monitors demonstrated >95% sensitivity and >95% PPV in detecting VF episodes in the AHA and MIT-BIH databases. In the present study our algorithm has also been able to generate a red alert for 100% of VF cases (83.8% as VF and 16.2% as VT. It is essential to note that an algorithm’s performance on such test databases do not necessarily reflect their performance in the real-world clinical settings. For example, GE Healthcare monitoring system achieved a PPV of >95% in identifying VF in the AHA and MIT-BIH databases, whereas it achieved a PPV of only 67.72% in identifying VF in the UCSF study.
In the previous subsection, we estimated that our method could raise ~374 false alarms per patient, per day. Placing these results in context, Cho et al.'s study\textsuperscript{28} observed ~697 false alarms per patient per day, Graham et al.'s study\textsuperscript{29} observed 942 alarms per patient per day, and The John Hopkins Hospital observed ~350 alarms per patient per day\textsuperscript{30}. It should be noted that the conditions considered as alarm worthy differed from one study to another; furthermore, most studies have only mentioned the total count of alarms, but not the count of false alarms. Thus, while it is difficult to make an exact comparison between these studies, our method’s high PPV in life-threatening arrhythmias, provides superior false alarm reduction.

**Algorithm Efficiency Performance**

On a regular desktop with an Intel i5 processor, our proposed system took only few milliseconds to identify a given heart rhythm; the system is therefore realistic for real-world implementation and the heart rhythm classification may be evaluated/updated every sec. Moreover, our classification system captures arrhythmia-specific, vital-sign features that manifest themselves in differently sized windows. A sample recording from the PhysioNet 2015 Challenge that resulted in a VF alarm is presented in Figure 3; notably, our system was able to detect the presence of VF in this very sample more quickly—in fact 4 seconds faster—than the bedside monitors used in the challenge.

**Discussion**

Although monitors sound alarms when ICU patients experience abnormal heart rhythms, unfortunately the majority of such alarms are found to be false.\textsuperscript{27} Excessive number of false alarms can create a noisy environment, and their effects on patients and families pertain to sleep disturbance, delirium, increased blood pressure and heart rate, negative effects on the immune system, slower healing and recovery process, increased length of stay, and impact on patient satisfaction\textsuperscript{27,31}. With respect to the effects on staff, noise increases occupational stress (irritation, fatigue, and tension headaches), reduces staff work performance and work satisfaction, delays recognition and response to medical device alarm signals, which may
negatively affect patient safety, and impairs oral communication and increases errors, which has a direct impact on patient safety\(^{32,33}\). Too many false alarms can also lead to desensitization among caregivers, which could potentially cause them to mistakenly ignore true alarms as well. A growing number of research studies, some of which employ ML \(^{3,4,34,35}\), have aimed to reduce the number of such false alarms, yet clearly better solutions are still needed. In this report, we present an arrhythmia alerting system that identified life-threatening arrhythmias more precisely and more efficiently than existing bedside monitors.

Several conclusions can be drawn from this study: \emph{first}, obtaining arrhythmia-specific features from different length windows made the detection algorithm more efficient; \emph{second}, our method can identify these arrhythmias with high sensitivity (a minimum of 83.8%); \emph{third}, the RF classifier is not much affected by the class imbalance of the records; \emph{fourth}, the algorithm does not have high computational requirements, therefore facilitating real-time implementation, and; \emph{fifth}, the algorithm is at least as robust as those of several commercial bedside monitors.

Many prior heart rhythm classification studies \(^{11,12,14,17,18}\) have relied on analyzing the ECG morphology only. This approach is limiting because arrhythmias, such as VT and VF, can be better detected with the inclusion of the BP signal. Even though all VT events may not be accompanied by hemodynamic instability, a fall in blood pressure is expected w.r.t. to the baseline blood pressure. Our heart rhythm classification algorithm utilizes the ECG, BP and PPG signals. Many heart rhythm classification studies that use ECG signals \(^{9,13}\) have not included VF as one of the arrhythmia classes detected. Our study not only included VF, but also alerted all cases of VF as life-threatening and did so 4 seconds faster than the bedside monitors used in PhysioNet 2015 challenge. Finally, some heart rhythm classification algorithms \(^{15}\) employ a short window of ECG segments to identify the arrhythmia., and are therefore unable to identify rhythms that need a longer duration of ECG analysis, such as extreme bradycardia, whereas our algorithm is capable to detect such rhythms. By comparing the performance of our algorithm with those of the bedside monitors, one observes that the proposed algorithm achieves PPV for most life-threatening arrhythmias that are at least as high as the commercial monitors (Table 5).

\textit{Conclusions}
We have built a standalone algorithm that can successfully utilize live-streaming data in the ICU to successfully identify multiple life-threatening arrhythmias in real-time, and may be of significant clinical benefit. False alarms could be further reduced, by wrapping the proposed method with false alarm detection mechanisms\textsuperscript{22,35}; however, it should be noted that false alarm detection mechanisms are generally built over small data sets\textsuperscript{2}, and methods built for one type of monitor may not work well for another. In our study, we designed features that measured the signal quality of ECG, BP and PPG, and designed arrhythmia specific features to identify the presence of a particular characteristic and/or morphology. The engineered features provided a reasonably good classification performance. Moving forward, we would like to explore augmenting the proposed system with features derived from deep learning architectures, in order to improve classification performance and clinical utility.

**Study limitations**

In the present study, the sample size is small, especially for VF. Also, the ratios of the number of records among all these classes are not representative of real-life situations, e.g., even though a large number of NSRs have been included during training and testing, in reality the ratio of NSR occurrences to the number of other heart rhythm occurrences, is expected to be even more skewed, which may affect performance metrics. We believe a larger scale study on real-world ICU patient data would be very informative and enable us to refine our algorithms further.

**Ethical Approval and Consent to participate**

We used data from the “PhysioNet/Computing in Cardiology Challenge 2015: Reducing False Arrhythmia Alarms in the ICU”. The data used is open source and is available at https://physionet.org/content/challenge-2015/1.0.0/.

**Consent for publication**
Availability of supporting data
The data will be available to any investigator upon request.

Competing interests
The authors have no competing interests, financial and non-financial interests of any kind.

Funding Support
The work was supported by a Grant-in-Aid (#15GRNT23070001) from the American Heart Association (AHA), the Institute of Precision Medicine (17UNPG33840017) from the AHA, the RICBAC Foundation, NIH grants 1 R01 HL135335-01, 1 R21 HL137870-01 and 1 R21EB026164-01, and a Founders Affiliate Post-doctoral Fellowship (#19POST34450149) from the AHA.

Author contributions
Wan-Tai M. Au-Yeung, PhD: Participated in the development of the algorithms, data analysis, and writing the manuscript
Rahul K. Sevakula, PhD: Participated in the development of the algorithms and writing the manuscript
Ashish K. Sahani, PhD: Participated in the development of the algorithms, data analysis, and writing the manuscript
Mohamad Kassab, MD: Participated in the data analysis and writing the manuscript
Richard Boyer, MD, PhD: Participated in the data analysis and writing the manuscript
Eric M. Isselbacher, MD, MSc: Participated in the conception of the study and writing the manuscript

Antonis A. Armoundas, PhD: Participated in the conception of the study, the development of the algorithms, data analysis, and writing the manuscript

Acknowledgements

None
Figure Captions

**Figure 1.** Presentation of the different window-length strategy used in extracting arrhythmia specific features characterizing different heart rhythms.

**Figure 2.** Heat map to pictorially show our method’s mean classification performance across 10 times 5-fold cross-validation.

**Figure 3.** The figure shows the plot of recording f5631 between 285 to 301 sec. The figure for this recording then compares the latency of bedside monitors and that of proposed system in alerting for ventricular fibrillation. The bedside monitor identifies the presence of ventricular fibrillation at 300th sec, as visualized by the red dashed line. On the other hand, the proposed system can identify the presence of ventricular fibrillation at 296th sec itself (4 sec earlier), as visualized by the green line.
References

1. Cvach, M. Monitor alarm fatigue: an integrative review. Biomedical instrumentation & technology 46, 268-277 (2012).
2. Clifford, G. D. et al. The PhysioNet/Computing in Cardiology Challenge 2015: Reducing False Arrhythmia Alarms in the ICU. Comput Cardiol (2010) 2015, 273-276, doi:10.1109/CIC.2015.7408639 (2015).
3. Plesinger, F., Klimes, P., Halamek, J. & Jurak, P. Taming of the monitors: reducing false alarms in intensive care units. Physiological measurement 37, 1313-1325, doi:10.1088/0967-3334/37/8/1313 (2016).
4. Kalidas, V. & Tamil, L. S. Cardiac arrhythmia classification using multi-modal signal analysis. Physiol Meas 37, 1253-1272, doi:10.1088/0967-3334/37/8/1253 (2016).
5. Eerikäinen, L. M., Vanschoren, J., Rooijakkers, M. J., Vullings, R. & Aarts, R. M. Reduction of false arrhythmia alarms using signal selection and machine learning. Physiol Meas 37, 1204-1216, doi:10.1088/0967-3334/37/8/1204 (2016).
6. Eerikäinen, L. M., Vanschoren, J., Rooijakkers, M. J., Vullings, R. & Aarts, R. M. in 2015 Computing in Cardiology Conference (CinC). 293-296 (IEEE).
7. Sevakula, R. K. et al. State-of-the-Art Machine Learning Techniques Aiming to Improve Patient Outcomes Pertaining to the Cardiovascular System. Journal of the American Heart Association 9, e013924 (2020).
8. Bazoukis, G. et al. Machine learning versus conventional clinical methods in guiding management of heart failure patients—a systematic review. Heart failure reviews, 1-12 (2020).
9. Guvenir, H. A., Acar, B., Demiroz, G. & Cekin, A. in Computers in Cardiology 1997. 433-436 (IEEE).
10. Dua, D. & Graff, C. in University of California, Irvine, School of Information and Computer Sciences (2017).
11. Jadhav, S. M., Nalbalwar, S. L. & Ghatol, A. A. Modular neural network based arrhythmia classification system using ECG signal data. International Journal of Information Technology and Knowledge Management 4, 205-209 (2011).
12. Jadhav, S. M., Nalbalwar, S. & Ghatol, A. in 2010 International Conference on Electronics and Information Engineering. V1-228-V221-231 (IEEE).
13. Liu, F. et al. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. Journal of Medical Imaging and Health Informatics 8, 1368-1373 (2018).
14. Chen, T.-M., Huang, C.-H., Shih, E. S., Hu, Y.-F. & Hwang, M.-J. Detection and Classification of Cardiac Arrhythmias by a Challenge-Best Deep Learning Neural Network Model. Iscience 23, 100886 (2020).
15. Hannun, A. Y. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine 25, 65-69, doi:10.1038/s41591-018-0268-3 (2019).
16. Kampouraki, A., Manis, G. & Nikou, C. Heartbeat time series classification with support vector machines. IEEE Transactions on Information Technology in Biomedicine 13, 512-518 (2008).
17. Rahman, Q. A. et al. Utilizing ECG-based heartbeat classification for hypertrophic cardiomyopathy identification. IEEE transactions on nanobioscience 14, 505-512 (2015).
Luz, E. J. d. S., Schwartz, W. R., Cámar-Chávez, G. & Menotti, D. ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer methods and programs in biomedicine* **127**, 144-164 (2016).

Clifford, G. D. *et al.* in *2017 Computing in Cardiology (CinC)*. 1-4 (IEEE).

Teijeiro, T., García, C. A., Castro, D. & Félix, P. Arrhythmia classification from the abductive interpretation of short single-lead ecg records. *Comput. Cardiol* **44**, 1-4 (2017).

Plesinger, F., Nejedly, P., Viscor, I., Halamek, J. & Jurak, P. Parallel use of a convolutional neural network and bagged tree ensemble for the classification of Holter ECG. *Physiological measurement* **39**, 094002 (2018).

Au-Yeung, W.-T. M., Sahani, A. K., Isselbacher, E. M. & Armoundas, A. A. Reduction of false alarms in the intensive care unit using an optimized machine learning based approach. *NPJ Digital Medicine* **2**, 1-4 (2019).

Martinez, J. P., Almeida, R., Olmos, S., Rocha, A. P. & Laguna, P. A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Trans Biomed Eng* **51**, 570-581, doi:10.1109/TBME.2003.821031 (2004).

Vistisen, S. T., Johnson, A. E. & Scheeren, T. W. (Springer, 2019).

Chen, C., Liaw, A. & Breiman, L. Using random forest to learn imbalanced data. *University of California, Berkeley* **110**, 24 (2004).

Aboukhalil, A., Nielsen, L., Saeed, M., Mark, R. G. & Clifford, G. D. Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. *Journal of biomedical informatics* **41**, 442-451 (2008).

Drew, B. J. *et al.* Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. *PloS one* **9**, e110274 (2014).

Cho, O. M., Kim, H., Lee, Y. W. & Cho, I. Clinical alarms in intensive care units: Perceived obstacles of alarm management and alarm fatigue in nurses. *Healthcare informatics research* **22**, 46 (2016).

Graham, K. C. & Cvach, M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. *American Journal of Critical Care* **19**, 28-34 (2010).

Sendelbach, S. & Funk, M. Alarm fatigue: a patient safety concern. *AACN advanced critical care* **24**, 378-386 (2013).

Lawless, S. T. Crying wolf: false alarms in a pediatric intensive care unit. *Crit Care Med* **22**, 981-985 (1994).

Wilken, M. *et al.* in *GMDS*. 107-111.

Donchin, Y. & Seagull, F. J. The hostile environment of the intensive care unit. *Current opinion in critical care* **8**, 316-320 (2002).

Eerikäinen, L. M., Vanschoren, J., Rooijakkers, M. J., Vullings, R. & Aarts, R. M. Reduction of false arrhythmia alarms using signal selection and machine learning. *Physiological measurement* **37**, 1204 (2016).

Fernandes, C. O., Miles, S., De Lucena, C. J. P. & Cowan, D. Artificial Intelligence Technologies for Coping with Alarm Fatigue in Hospital Environments Because of Sensory Overload: Algorithm Development and Validation. *Journal of medical Internet research* **21**, e15406 (2019).
Physiological Signals

- ECG
- ABP
- PPG

Extracting arrhythmia specific features to identify the heart rhythm at time, $t$

Different window lengths used for extracting arrhythmia specific features:

| Arrhythmia Type                  | Window Length |
|----------------------------------|--------------|
| Ventricular Tachycardia, VF      | 4 seconds    |
| Asystole, Signal Quality Indices | 8 seconds    |
| Extreme Bradycardia, Tachycardia| 15 seconds   |

Algorithm Performance

- Asystole
- Extreme Brady
- Extreme Tachy
- VF
- VT
- NSR
- Sinus Tachy
- Noise

Graphical Abstract

Downloaded from https://academic.oup.com/ehjdh/advance-article/doi/10.1093/ehjdh/ztab058/6312556 by guest on 06 July 2021
Different window lengths used for extracting arrhythmia specific features

| Condition                                | Length  |
|------------------------------------------|---------|
| Ventricular Tachycardia, Ventricular Fibrillation | 4 seconds |
| Asystole, Signal Quality Indices         | 8 seconds |
| Extreme Bradycardia, Extreme Tachycardia | 15 seconds |
Table 1 Definitions of the 8 classes.

| Class               | Definition                                                                 |
|---------------------|---------------------------------------------------------------------------|
| Asystole            | A gap of at least four seconds between two successive R-waves             |
| Extreme Bradycardia | A heart rate lower than 40 beats per minute for 4 consecutive beats       |
| Extreme Tachycardia | A heart rate higher than 140 beats per minute for 17 consecutive beats    |
| Ventricular Fibrillation | An oscillatory ECG waveform of at least 4 seconds            |
| Ventricular Tachycardia | 5 consecutive VT beats with heart rate of at least 100 beats per minute |
| Normal Sinus Rhythm | A heart rate between 60 and 100 beats per minute for 15 seconds          |
| Sinus Tachycardia   | A heart rate between 100 and 140 beats per minute for 17 consecutive beats |
| Noise/Artifacts     | All physiological signals are filled with noise or artifacts.            |
Table 2 Number of events in each of the 8 classes

| Class                  | Number of Events |
|------------------------|------------------|
| Asystole               | 19               |
| Extreme Bradycardia    | 50               |
| Extreme Tachycardia    | 118              |
| Ventricular Fibrillation| 8               |
| Ventricular Tachycardia| 93               |
| Normal Sinus Rhythm    | 503              |
| Sinus Tachycardia      | 222              |
| Noise or artifacts     | 71               |
Table 3  List of extracted features that were fed to the 8-class Random Forest Classifier. Column “Signal” indicates the physiological signal from which the features are extracted, and column “Feature” provides the name of each feature. Column “Category” provides the context behind the features’ utility. The features have been broadly categorized into six categories: (i) Signal quality indices which indicate the quality of the signal (clean or noisy) and are computed over 8 sec of signal, (ii) asystole features which characterize asystole and are computed over 8 sec of signal, (iii) extreme bradycardia features which characterize extreme bradycardia and are computed over 15 sec of signal, (iv) extreme tachycardia features which characterize extreme tachycardia and are computed over 15 sec of signal, (v) ventricular fibrillation (VF) features which characterize VF and are computed over 4 sec of signal, and (vi) ventricular tachycardia (VT) features which characterize VT and are computed over 4 sec of signal. The set of ECG features were computed from each lead, separately. Thus, a total of 74 features (26 features from ECG lead 1, 26 feature from ECG lead 2, 10 features from BP and 12 features from PPG) were used here, for heart rhythm classification.

| Signal            | Feature                                                                 | Category                                      |
|-------------------|-------------------------------------------------------------------------|-----------------------------------------------|
| Electrocardiogram | Periodicity Measure                                                     | Signal Quality Indices (8 sec)                |
|                   | Sharpness Measure                                                       |                                               |
|                   | Correlation Measure                                                     |                                               |
|                   | Peak Height Stability Measure                                           |                                               |
|                   | Max Period Between Consecutive R waves                                   | Asystole features (8 sec)                     |
|                   | Histogram Analysis Mean                                                 |                                               |
|                   | Histogram Analysis Standard Deviation                                   |                                               |
|                   | Median Neighborhood Swing                                               |                                               |
|                   | Blank Area Swing                                                       |                                               |
|                   | Blank Area Swing to Median Neighborhood Swing ratio                      |                                               |
|                   | Minimum Heart Rate across 4 beats                                        | Extreme bradycardia features (15 sec)         |
|                   | Number of Beats Slower than 46bpm                                       |                                               |
|                   | Maximum Heart Rate across 17 beats                                       | Extreme tachycardia features (15 sec)         |
|                   | Number of Heartbeats within the window of analysis                      |                                               |
| Complexity Measure | VF features (4 sec) |
|--------------------|---------------------|
| Bandwidth          |                     |
| Dominant Frequency |                     |
| Mean Frequency     |                     |
| Median Frequency   |                     |
| Max Power to Total Power Ratio | |
| Number of Peaks with normalized power above 0.2 | |
| Five consecutive VT beats at >100bpm | VT features (4 sec) |
| Sharpness Measure over 5 Beats | |
| Correlation Measure over 5 Beats | |
| Max Heart Rate over 5 Beats | |
| Max Mean Diff LF SUB Peaks | |
| Periodicity Measure | Signal Quality Indices (8 sec) |
| Pulse Pressure Stability | |
| Correlation Measure | |
| Max Period Between Consecutive Onsets of Waveform | Asystole feature (8 sec) |
| Minimum Heart Rate across 4 beats | Extreme bradycardia features (15 sec) |
| Number of Beats Slower than 46bpm | |
| Maximum Heart Rate across 17 beats | Extreme tachycardia features (15 sec) |
| Not Enough Beats for Calculating Maximum Heart Rate | |
| No Peaks | VF/VT features (4 sec) |
| Decreasing Pressure | |
| PPG | |
| Periodicity Measure | Signal Quality Indices (8 sec) |
| Stability Measure | |
| Correlation Measure | |
| Max Period Between Consecutive Onsets of Waveforms | Asystole features (8 sec) |
| Max Amplitude before Onset | |
| Max Amplitude after Onset | |
| Amplitude Decrease | |
| Minimum Heart Rate across 4 beats | Extreme bradycardia features (15 sec) |
|----------------------------------|-------------------------------------|
| Number of Beats Slower than 46bpm |                                     |
| Maximum Heart Rate across 17 beats | Extreme tachycardia features (15 sec) |
| Not Enough Beats for Calculating Maximum Heart Rate | |
| Decreasing PPG | VF/VT features (4 sec) |
Table 4  Confusion matrix of the 8-class classification result after 10-times 5-fold cross-validation

| Ground Truth | Asystole | Extreme Brady | Extreme Tachy | VF  | VT  | NSR | Sinus Tachy | Noise/Artifacts |
|--------------|----------|---------------|---------------|-----|----|-----|-------------|-----------------|
| Asystole     | 178      | 9             | 0             | 0   | 0  | 0   | 0           | 3               |
| Extreme Brady| 10       | 472           | 0             | 0   | 6  | 10  | 1           | 1               |
| Extreme Tachy| 0        | 0             | 1078          | 0   | 22 | 0   | 60          | 20              |
| VF           | 0        | 0             | 0             | 67  | 13 | 0   | 0           | 0               |
| VT           | 0        | 5             | 7             | 2   | 784| 29  | 72          | 31              |
| NSR          | 56       | 198           | 33            | 14  | 580| 3390 | 378         | 381             |
| Sinus Tachy  | 1        | 31            | 193           | 1   | 134| 61  | 1700        | 99              |
| Noise/Artifacts | 42  | 17            | 57            | 32  | 55 | 21  | 54          | 432             |
Table 5 Sensitivity, most common misclassification, positive predictive value (PPV) for each rhythm, and precision observed by bedside monitors

| Rhythm                  | Sensitivity (%) | Most common misclassification (error rate (%)) | PPV (%) | PPV (%) Physionet 2015 Challenge | PPV (%) MIMIC II study | PPV (%) UCSF study |
|-------------------------|-----------------|-----------------------------------------------|---------|----------------------------------|------------------------|-------------------|
| Asystole                | 93.7            | Extreme Bradycardia (4.74)                     | 62.0    | 16.67                            | 9.33                   | 32.83             |
| Extreme Bradycardia     | 94.4            | NSR (2.00)                                    | 64.5    | 50                               | 70.71                  | NA                |
| Extreme Tachycardia     | 91.4            | Sinus Tachycardia (5.08)                      | 78.8    | 94.92                            | 76.93                  | NA                |
| VF                      | 83.8            | VT (16.2)                                     | 57.8    | 10.34                            | 20.33                  | 67.72             |
| VT                      | 84.3            | Sinus Tachycardia (7.74)                      | 49.2    | 26.23                            | 53.42                  | 13.00             |
| NSR                     | 67.4            | VT (11.53)                                    | 96.6    | NA                               | NA                     | NA                |
| Sinus Tachycardia       | 76.6            | Extreme Tachycardia (8.69)                    | 75.1    | NA                               | NA                     | NA                |
| Noise/Artifacts         | 60.8            | Extreme Tachycardia (8.03)                    | 44.7    | NA                               | NA                     | NA                |