Hemodynamic-Impact-Based Prioritization of Ventricular Tachycardia Alarms

Kalpit Desai, Michael Lexa, Brett Matthews, Sahika Genc

***Abstract*— Ventricular tachycardia (V-tach) is a very serious condition that occurs when the ventricles are driven at high rates. The abnormal excitation pathways make ventricular contraction less synchronous resulting in less effective filling and emptying of the left ventricles. However, almost half of the V-tach alarms declared through processing of patterns observed in electrocardiography are not clinically actionable. The focus of this study is to provide guidance on determining whether a technically-correct V-tach alarm is clinically-actionable by determining its "hemodynamic impact". A supervisory learning approach based on conditional inference trees to determine the hemodynamic impact of a V-tach alarm based on extracted features is described. According to preliminary results on a subset of Multiparameter intelligent monitoring in intensive care II (MIMIC-II) database,true positive rate of more than 90% can be achieved.**

I. Introduction

Ventricular tachycardia (V-tach) are among the most frequent critical arrhythmias. In fact, in a recent intensive care unit study [1], 35% of all critical ECG arrhythmia alarms were V-tach alarms. V-tach is a very serious condition that occurs when the ventricles prematurely contract at high rates. The abnormal excitation pathways make ventricular contraction less synchronous with artrial contraction and can lead to ineffectual filling and emptying of the left ventricles [2, 3]. V-tach can be a precursor to ventricular fibrillation which may be fatal unless quickly corrected by cardiac conversion [4].

Despite being categorized as critical, not all V-tach alarms are *actionable*. In fact, in a recent large study of intensive care alarms only 15% of the total number of alarms (885 out of 5934) were deemed to be clinically-relevant or actionable [5]. *Non-actionable* alarms can be divided into two groups: 1) *technically-false* and 2) *clinically non-actionable*. Technically-false alarms typically occur when there is noise (e.g., interference from other devices) or artifacts (e.g., motion) present in the signal. Clinically non-actionable alarms occur when the alarm is correctly detected and isolated by the monitoring device, but the alarm does not necessiate clinicial intervention (e.g., does not necessiate the administration of medication).

Previous efforts to reduce technically-false alarms have focused on improving signal acquisition to reduce noise, incorporating accelerometers to reduce motion artifacts, and calculating signal quality to quantify confidence in the alarm [1]. Previous efforts to reduce clinically non-actionable alarms have employed adaptive thresholding and emphasized multi-parameter analysis. The focus of this study is to provide guidance on whether a technically-correct (true positive) V-tach alarm is clinically-actionable or not by automatically rating the alarm in terms of its "hemodynamic impact". A V-tach alarm declared by the monitoring device may be correct based on morphology, but may be clinically non-actionable because it may not significantly degrade the heart's ability to perfuse the body's organs. For example, an isolated short-duration V-tach causing a small temporary drop in systolic blood pressure may only have a minimal hemodynamic impact. The idea is that the additional hemodynamic information may help further categorize and prioritize V-tach alarms thereby improving alarming and aiding hospital workflow.

Figure 1 shows an example of a high and a low impact event. The top plots are electrocardiographic (ECG) waveforms. The middle plots show the corresponding arterial blood pressure (ABP) waveforms along with summary numerics (systolic, diastolic, and mean pressures). Note that for the event on the left the arrhythmia not only causes a sharp drop in pressure but causes the ABP waveform to completely lose its pulsatile nature during a portion of the event. This behavior signifies a high hemodynamic impact and is in contrast to the event on the right hand side where the pulsatile nature of the ABP waveform is maintained. The V-tach on the right still effects the systolic, diastolic, and mean pressures but to a lesser extent than the event on the left. Comparatively then the event on the right has a low hemodynamic impact.

The paper is organized as follows. First, we describe our technical approach. Then, we provide preliminary results on a subset of records from the Multi-Parameter Intelligent Monitoring for Intensive Care II (MIMIC-II) database [6]. Finally, we provide future research directions.

II. Technical approach

The flow diagram of our technical approach is shown in Figure 2. First, we extact features from segments of ECG and ABP waveforms that were recorded during true positive V-tach events. Then, using these features, we apply a supervised learning approach to train a classifier that categorizes the events' hemodynamic impact. The approach yields a transparent classification scheme that labels (true positive) V-tach events as having low, medium, or high hemodynamic impact. The approach also includes a mechanism to weigh the features individually so that the classifier's performance can be further optimized.

In particular, we propose an adaptive classification approach based on *conditional inference (CI) trees*. A CI tree [7] is a relatively new machine learning method that produces classifiers structured as binary trees. The approach,

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

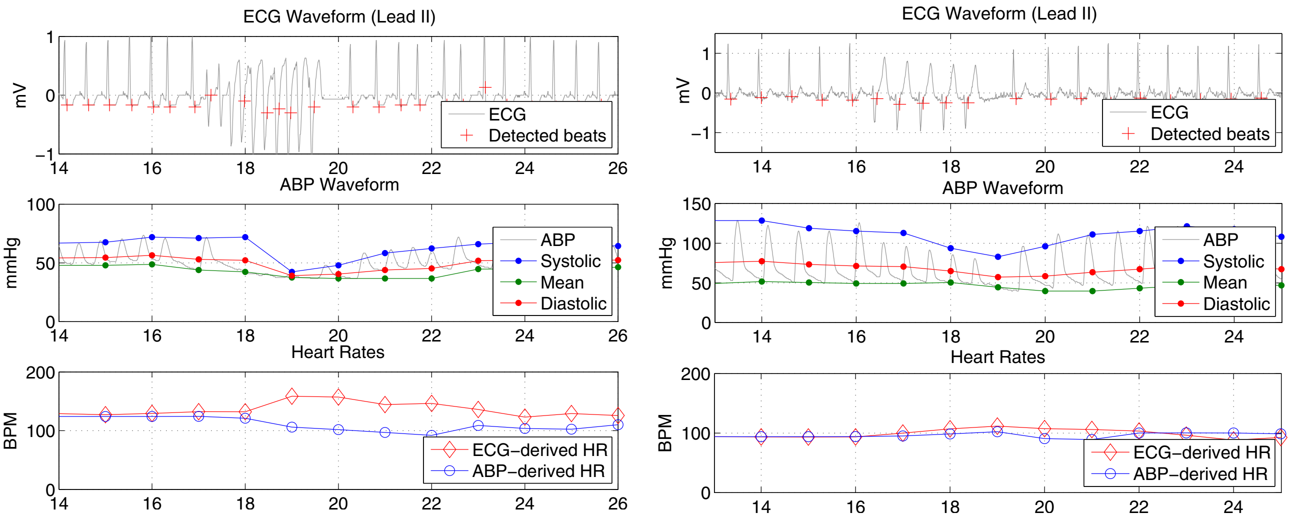

Fig. 1. Examples of V-tach events having high (left) and low (right) hemodynamic impact.

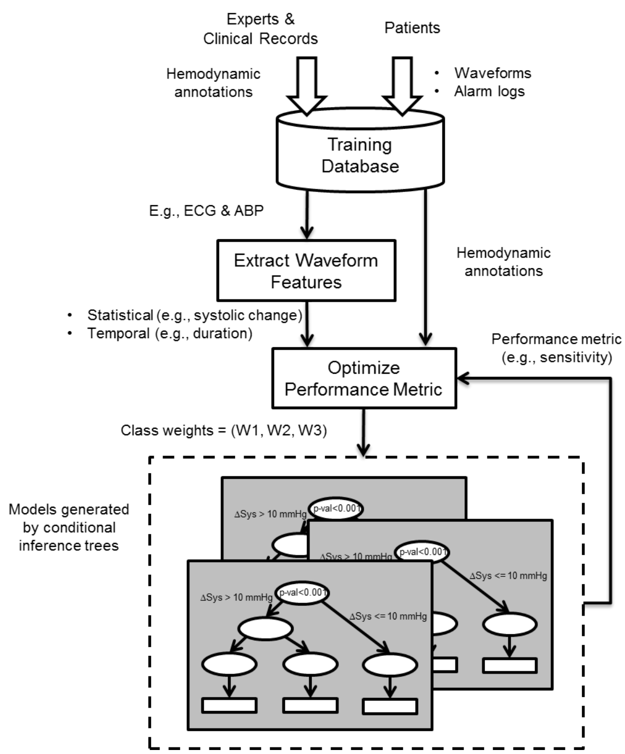

Fig. 2. The overview of training the conditional inference tree.

like most tree based methods, recursively partitions the feature space to make each partition as pure (single class) as possible. Moreover, CI trees are easy to interpret because one can transverse the tree and see all the decisions that lead to a particular classification–a quality that is particularly useful in the context of patient monitoring.

CI trees hold two major advantages over the other tree based methods:

- Pruning is not necessary. Other tree based methods first grow a tree to unlimited depth to fit the training data as well as possible, and then prune the tree in a post-hoc manner in order to reduce (ideally eliminate) the overfitting. The CI tree method on the other hand leverages an early-stopping criteria for growing the tree that is based on non-parametric permutation tests.
- The conditional inference framework allows selection of categorical features for making a split in unbiased and statistically principled way. Most other tree based methods have a bias towards selecting the continuous features (or features with many possible values) for making the split. This artificial bias compromises the classification accuracy on unseen data.

While the CI tree approach has advantages, its standard application to the present problem often results in imbalanced performance for the three impact classes, that is, there may be large, unwanted variations in the sensitivity and specificity among the low, medium, and high classes. We address this issue by making the CI approach adaptive in the sense that we tune the class weights to influence performance.

The CI tree approach has an optional weight vector that can be used in the fitting process. Only nonnegative integer valued weights are allowed. The default class weight is one. Each node of the CI tree is represented by a vector of case weights having nonzero elements when the corresponding observations are elements of the node and are zero otherwise. The splitting of a node is based on the sum of these weights. Thus, by changing these class weights, the splitting can be manipulated to improve the sensitivity and specificity of a specific class of hemodynamic impact.

For implementation of the CI tree approach, we use the *ctree* module in the R package *party* [8]. A CI tree offers several hyperparameters that can be tuned. We used 10-fold cross-training. In our preliminary experiments, we set the maximum p-value for the split to 0.01, and set the maximum depth along any branch to 5. In the future, we plan to optimize the values to improve the specificity. Let $W0$, $W1$ and $W2$ be the weights of the low-impact class, medium-impact class and high-impact class respectively. In the training phase, we take the previously determined values of the hyperparameters (e.g. relative class weights), and run the training algorithm as outlined in Algorithm 1. Given the inputs and the constraints, the core algorithm also describes the process used for the training step of cross-validation, that we use for tuning weights as well as for evaluating the algorithm performance (next section).

Algorithm 1 Training a Conditional Inference Tree

```
Input Relative Class Weights: W0 = 1; W1 = W1*; W2 = W2*
Input Dataset: D_train : (X_i, y_i) for i = 1,..,N_train ; y_i ∈ low, med, high
Input Split Decision Constraints: p-value ≤ 0.01, Depth ≤ 5, SumParentWeights ≥ 20, SumChildWeights ≥ 7
repeat
  STEP 1: Compute association of features with the response
  for all j ∈ (1, 2, ..., m) do
    Test the null hypothesis H_0 of independence between Y and X_j
    Using non-parametric permultation tests, obtain p-value for H_0 for X_j, i.e. P_j
  end for
  STEP 2a. If H_0 cannot be rejected with the statistical significance level specified by p-value, STOP
  STEP 2b. Determine the feature X_{j*} with strongest association to y. X_{j*} = argmin_j P_j.
  STEP 3. Search for the value of X_{j*} that provides best binary split of the data at the current node.
  STEP 4. Recursively repeat the above steps for both child nodes.
until H_0 cannot be rejected for any feature.
```

III. PRELIMINARY RESULTS

The data set used in the current study is a subset of the data set described in [1] where a collection of 447 waveform records were isolated from the MIMIC II Waveform Database (version 2). Each record within this super set contains ECG and ABP waveform recordings during the time at which critical ECG arrhythmia alarms were issued. Among the total of 5386 alarms, 1015 were identified by a team of experts as being true positive (TP) V-tach events (see [1] for details about how these annotations were performed). V-tach and extreme tachycardia were, respectively, the first (35.3%) and second (34.8%) most common critical arrhythmias within this data set.

Out of the 1015 TP V-tach events, we isolated 430 events from 69 waveform records. We subsequently had two engineers in our group, one with clinical training, annotate each event as having a high, medium, or low hemodynamic impact. We refer to these two sets of annotations as “Version 1” and “Version 2” where the former is biased towards a large number of low impact alarms and the later biased towards high impact alarms. Here the word bias refers to human bias, i.e., one of the annotators labeled a greater number of events as low impact compared to the other annotator who labeled a greater number as high impact events.

The features used by the classifier are

- Median of the pulse rate from ECG during the alarm, i.e., electrical heart rate rate.
- Median of the pulse rate from ABP during the alarm, i.e., mechanical heart rate.
- Median of the difference between pulse rates from ECG and ABP.
- Minimum value of the systolic blood pressure during the alarm.
- Ratio of the minimum value of the systolic blood pressure to the median of the systolic blood pressure.
- Duration of the alarm.
- Standard deviation of the pulse rate from the ECG.
- Standard deviation of the pulse rate from the ABP.
- Standard deviation of the difference between pulse rates from ECG and ABP.

Note that while these features fundamentally derive from the ECG and ABP waveforms, most are taken from 1 Hz summary numerics derived from the waveforms. For example, to compute the median of the ECG pulse rate (first feature), a 1 Hz pulse rate time series would be derived from the ECG waveform and the median of that time series during the V-tach event would become the feature of interest.

To understand why many of the features concern the ECG- and ABP-derived pulse rates, we refer back to Fig. 1. In the figure, the bottom plots show the pulse rates (heart rate) for each event as a function of time. Note that for the high impact event on the left the ECG- and ABP-derived pulse rate diverge. This separation is a consequence of the fact that the pulsitile nature of the ABP waveform is lost during the event (beats are missed in the ABP waveform and the interval between detected beats lengthens causing the rate to decrease). The separation is therefore a feature marking a possibly high impact event. In contrast, the absence of rate diverge is a feature of a lower impact event.

The features that are functions of the pressure numerics are meant to track the relative changes in arterial pressure and naturally provide information about the hemodynamic effects of a V-tach.

The results of our algorithm are shown in Figure 3 for the Version 1 and Version 2 annotations. The classifier achieved a true positive rate of 91% at a false positive rate of 31% for the Version 1 annotations. Here, a true positive event is defined as one where the classifier classifies a V-tach as having a high hemodynamic impact when indeed it is a high impact event. Likewise, a false positive event is defined as one where the classifier classifies a V-tach as having a high impact when in fact it is low impact event. For the Version 2 annotations the classifier is able to achieve a higher true positive rate of 97% but at the cost of a higher false positive rate (45%).

A portion of the CI tree for Version 1 is shown in Fig. 3. The rule that describes the branch leading to the classification than contains most of the “low” hemodynamic impact alarms is highlighted in bold red lines. The interpretation is that the

Annotation Type	Low(Perfusing)	High(Non-perfusing)	TPR and FPR - 10-fold Cross validation	Relative Class Weights
Version 1 (large number of "Low"s)	269	101	91% TPR at 31% FPR	1:20
Version 2 (large number of "High"s)	31	339	97% TPR at 45% FPR	1:16

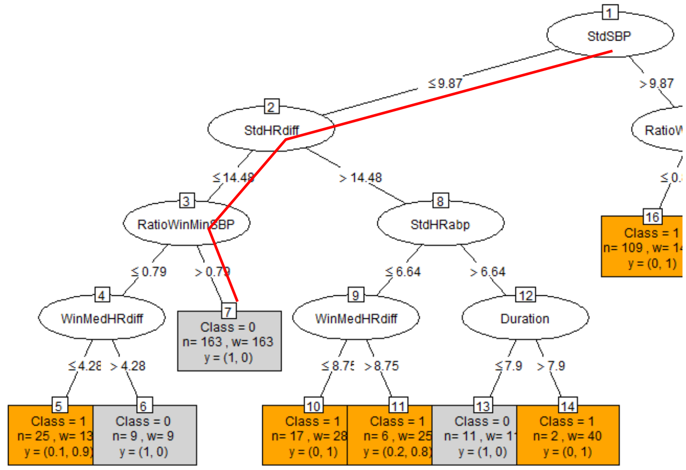

Fig. 3. Performance results of the conditional inference tree for the MIMIC-II data (table) and an illustration of one branch of the classifier.

majority of the TP V-tach alarms have "low" impact if the standard deviation of the systolic pressure is less than or equal to 9.87 AND the standard deviation of the difference in pulse rates is less than or equal to 14.48 AND the ratio of the minimum value of the systolic pressure to the median of the systolic pressure is greater than 0.79. The FPR is still relatively high. In the future, we plan to expand the feature set and also include features from plethysmograph waveform to improve the FPR without compromising from the TPR. We compared the results of the CI tree (not optimized for p-value and depth) to a random forest approach using the Version 1 (large number of "Low"s) annotations. The TPR and FPR for both approaches are shown in Table I. The TPRs are similar but the FPR from the random forest algorithm is twice the FPR of the CI tree approach.

TABLE I

THE COMPARISON OF THE CI TREE AND RANDOM FOREST RESULTS.

Algorithm Type	TPR and FPR - 10-fold Cross validation
CI tree	91% TPR at 31% FPR
Random forest (1000 trees)	93% TPR at 59% FPR

IV. CONCLUSIONS

In this paper, we tailored the conditional inference tree learning method to classify the degree of hemodynamic impact of technically-correct V-tach events. It is hoped that this additional information can help differentiate clinically-actionable, i.e., high impact, T-tach alarms from less critical, non-actionable alarms. According to preliminary results on a subset of the Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) database, a true positive rate of more than 95% can be achieved at the cost of false positive rate of less than 30%. The false positive rate is relatively high. We suspect that this is due to additional features from ECG and ABP waveforms that may be needed to provide differentiation or due to small number of components in the corresponding classes of annotations. In the future, we also plan to include features from the concurrent plethysmograph waveform recordings. We are also in the process of obtaining consensus annotations from clinicians for additional testing of the performance of the algorithm.

REFERENCES

[1] A. Aboukhalil, L. Nielsen, M. Saeed, R. Mark, and G. Clifford, "Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform," *Biomed Inform.*, vol. 41, no. 3, pp. 442–51, 2008.

[2] J. Parker, F. Khaja, and R. Case, "Analysis of left ventricular function by atrial pacing," *Circulation*, vol. 43, pp. 241–252, 1971.

[3] J. Goldstein, i. B. Barzila, T. Rosamond, P. Eisenberg, and A. Jaffe, "Determinants of hemodynamic compromise with severe right ventricular infarction," *Circulation*, vol. 82, no. 2, pp. 359–68, 1990.

[4] D. Mohrman and L. Heller, *Cardiovascular Physiology*, 6th ed. McGraw-Hill, 2006.

[5] S. Siebig, S. Kuhls, M. Imhoff, U. Gather, J. Schlmerich, and C. E. Wrede, "Intensive care unit alarmshow many do we need?" *Critical Care Medicine*, vol. 38, no. 2, pp. 451–456, 2010.

[6] M. Saeed, M. Villarroel, A. Reisner, G. Clifford, L. Lehman, G. Moody, T. Heldt, T. Kyaw, B. Moody, and R. Mark, "Multiparameter intelligent monitoring in intensive care ii (mimic-ii): A public-access icu database," *Critical Care Medicine*, vol. 39, no. 5, pp. 952–960, May 2011.

[7] T. Hothorn, K. Hornik, , and A. Zeileis, "Unbiased recursive partitioning: A conditional inference framework," *Computational and Graphical Statistics*, vol. 15, no. 3, pp. 651–674, 2006.

[8] party: A laboratory for recursive partytioning. [Online]. Available: http://cran.r-project.org/web/packages/party/index.html