False Alarm Suppression in Early Prediction of Cardiac Arrhythmia

Shoumik Roychoudhury, Mohamed F. Ghalwash, Zoran Obradovic*

***Abstract*— High false alarm rates in intensive care units (ICUs) cause desensitization among care providers, thus risking patients' lives. Providing early detection of true and false cardiac arrhythmia alarms can alert hospital personnel and avoid alarm fatigue, so that they can act only on true life-threatening alarms, hence improving efficiency in ICUs. However, suppressing false alarms cannot be an excuse to suppress true alarm detection rates. In this study, we investigate a cost-sensitive approach for false alarm suppression while keeping near perfect true alarm detection rates. Our experiments on two life threatening cardiac arrhythmia datasets from Physionet's MIMIC II repository provide evidence that the proposed method is capable of identifying patterns that can distinguish false and true alarms using on average 60% of the available time series' length. Using temporal uncertainty estimates of time series predictions, we were able to estimate the confidence in our early classification predictions, therefore providing a cost-sensitive prediction model for ECG signal classification. The results from the proposed method are interpretable, providing medical personnel a visual verification of the predicted results. In conducted experiments, moderate false alarm suppression rates were achieved (34.29% for Asystole and 20.32% for Ventricular Tachycardia) while keeping near 100% true alarm detection, outperforming the state-of-the-art methods, which compromise true alarm detection rate for higher false alarm suppression rate, on these challenging applications.**

I. INTRODUCTION

Suppressing high false alarm rates from bedside monitors in intensive care units (ICUs) has been a topic of special interest in the last decade [1]–[5]. Alarm fatigue among care providers inside ICUs due to the high percentages of bedside monitor false alarms, has been identified as one of the top 10 medical hazards [6]. Alarm fatigue results in desensitization among care providers, which ultimately can lead to lower standards of care to patients and also result in fatal consequences [7]. Artifacts, noise and missing values are some primary factors that corrupt the physiological data from bedside monitors, causing high false alarm rates.

Different approaches have been applied to reduce false alarm rates. One direction is to determine the quality of the ECG signal, based on the fact that noisy signals are more prone to trigger false alarms. For example, Behar et al. [3] proposed several novel ways of measuring ECG signal quality. Another direction consists of data fusion methods where extra non-ECG waveform data, such as invasive arterial blood pressure (ABP) and photoplethysmogram (PPG) [2], [8] are incorporated. These non-ECG waveforms are assumed to be highly correlated to ECG, and consequently could be used to identify the alarm types. Recently, several methods were developed to suppress false ventricular tachycardia alarms without the need for additional non-ECG waveforms, which resulted in reduction of true alarm detection [3], [4]. Their approach is based on features extracted from the ECG signal 20 seconds prior to a triggered alarm. All aforementioned methods extract statistical features from the ECG signals and feed them to a classifier, which often results in a black-box approach. However, in medical applications, it is important not only to provide accurate prediction but also to provide interpretable results, such that medical experts get insights about the prediction.

Shoumik Roychoudhury is a PhD student in Computer and Information Science, Temple University, PA, USA. shoumik.rc@temple.edu

Mohamed F. Ghalwash is a Research Assistant Professor in Computer and Information Science, Temple University and Assistant Professor at Faculty of Science, Ain Shams University, Cairo, Egypt. mohamed.ghalwash@temple.edu

Zoran Obradovic (*Contact author) is Professor at Computer and Information Science, Temple University and Director of Center for Data Analytics and Biomedical Informatics. zoran.obradovic@temple.edu

TABLE I

WEIGHTED ACCURACY COMPARISON BETWEEN THE PROPOSED APPROACH TO THE STATE-OF-THE-ART ON ASYSTOLE CARDIAC ALARMS (ASYS) AND VENTRICULAR TACHYCARDIA ALARMS (VTACH)

Dataset	Behar et al. [3]	Our approach
ASYS	45.96±14.33	**65.05±7.55**
VTACH	32.80±8.32	**48.56±7.41**

In this study we characterize a *cost-sensitive* classification model for *early* and *interpretable* prediction of life threatening arrhythmia alarms. The objective of our prediction model is to suppress false alarms while keeping true alarm detection rates high. In addition, by identifying alarms early, the response time of the medical personnel can be improved in the event of life-threatening arrhythmia alarms, and the alarm fatigue problem can be reduced among care providers. In Table I, we show the effectiveness (weighted accuracy in Equation 7) of our approach to suppress a large percentage of false alarms for two datasets as compared with the current state-of-the-art method.

Our Contribution: To the best of our knowledge this is the first reported application of time series early classification framework in the realm of suppressing false arrhythmia alarms. The contribution of this paper, summarized in Table II, is the following: 1) We characterize a classification

TABLE II

PROPERTIES (INTERPRETABILITY, EARLINESS, UNCERTAINTY, AND FLASE ALRAM SUPPRESSION (FAS)) USED TO CATEGORIZE THE METHODS

	Li et al. [2]	Behar et al. [3]	EDSC [9]	Our approach
Intrepretabilty	×	×	✓	✓
Earliness	×	×	✓	✓
Uncertainty	×	×	×	✓
FAS	✓	✓	×	✓

model to provide more *accurate* prediction (high true alarm detection and false alarm suppression) than the state-of-the-art methods on arrhythmia alarms. 2) We provide *interpretable* results in order to explain the rationale of the prediction, whereas all other published methods are black-box. 3) We provide *early* prediction before the alarm happens, which helps the practitioners to respond early to the alarm, whereas all other methods provide prediction at the time when alarms happen. 4) We provide a *cost-sensitive* model to achieve the desired level of false alarm suppression rates.

II. Background

A. Early Classification of Time Series

In the field of time series classification, early classification of time series has gained popularity [10], especially in application areas where critical time sensitive decision making is required, such as early warning of diseases [11]. The principal objective of early classification models for time series is to predict the label of the alarm as the ECG signal is progressively recorded and before the alarm happens. If the *observed* signal is insufficient to make an accurate prediction, more ECG signal data are used and the process is repeated until the time when the alarm happens. Early prediction of life-threatening cardiac alarms would allow care providers inside ICUs to be alert at the time of (or even before) true arrhythmia alarm events, and at the same time would automatically suppress false arrhythmia alarms.

B. Interpretable Early Classification Model

Medical experts tend to favor interpretable methods which provide visual clarification of prediction results rather than black-box methods. A method called early distinctive shapelet classification (EDSC) was proposed to provide interpretable early classification results [10]. The method extracts local discriminative patterns from the time series in order to characterize the target class locally. These discriminative local patterns, known as shapelets [12], are effective for early classification. An example of such shapelets is shown in Fig. 1. The patterns extracted from the two classes of the time series are discriminative, hence, a new signal can be classified *as soon as* a match between the signal and any of these extracted shapelets is found. In this way, the method is able to justify the prediction of the new signal as red (blue), because the new signal has a pattern that is similar to a pattern observed previously in the red (blue) class. For more details about EDSC, the reader is referred to [10].

In cases where signals from different classes are similar to each other, especially in the early phases of the signals, the shapelets extracted from these classes might exhibit similar patterns, which might mislead the prediction. For example, a true alarm signal might match a false alarm shapelet; in this case, the EDSC method would predict the signal as false alarm as soon as the match is found regardless of how *reliable* the match is. In other words, the EDSC method does not provide uncertainty estimates on the match between the signal and shapelet and depends only on the distance measurement (match) for the prediction of label of

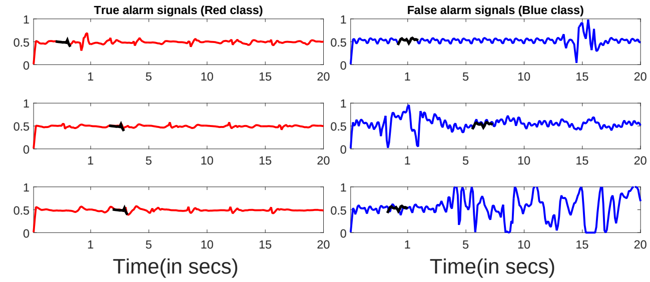

Fig. 1. Shapelets (in black) from true alarm (red) and false alarm (blue) classes

the time series. This drawback was addressed by Ghalwash et al. [13] where the EDSC method was extended to produce interpretable early classification of time series with uncertainty estimates, known as MEDSC-U. The uncertainty for the predicted label was used to decide the class membership of the time series signal. In this paper we investigate the MEDSC-U method for ECG signal classification and use the uncertainty estimates to decide the alarm class membership for ECG signals. The estimated uncertainties are used to develop a cost-sensitive decision algorithm for early alarm prediction using ECG signals.

III. Model Description

We begin by briefly describing the modified early distinctive shapelet classification method for uncertainty estimation (MEDSC-U) [13]. Given a time series dataset D, where each time series example is an ECG signal of 20 seconds prior to the alarm event, each signal is associated with a label (true or false alarm). The task is to correctly classify the ECG signal as early as possible. The MEDSC-U extracts all shapelets of different lengths for early classification. For each shapelet a distance threshold is learned such that the shapelet discriminates between classes. Then, MEDSC-U ranks the shapelets using a utility function that incorporates earliness and accuracy of the shapelet. The shapelets are then pruned by selecting the top performing shapelets that cover the entire dataset and finally the method classifies unknown time series based on the most *confident* matching shapelets. In Section III-C, we show how to characterize this method in order to suppress a large fraction of false alarms while keeping near-perfect true alarm detection rates.

A. Learning Phase

The MEDSC-U method has three steps to extract all discriminative shapelets for early classification.

1. Shapelet Extraction: The shapelet is defined as $S = (s, l, \delta, c)$ where s is a time series subsequence of length l, c is the alarm label of the shapelet (true or false alarm), and δ is a distance threshold which needs to be learned. The distance threshold is estimated by computing the distances between the subsequence s and all time series in the dataset. To compute the distance between a subsequence s of length l and a time series T of length L (where $l \leq L$), a window of length l is slide over the time series T to extract all subsequences $\{h_1, h_2, ..., h_{L-l+1}\}$ of length l. Then, the distance can be computed as

$$dist(s,T) = \min_{\forall i \in \{1,2,\ldots,L-l+1\}} dist(s,h_i) \quad (1)$$

where $dist(s,h_i)$ is the Z-normalized Euclidean distance, which is computed as

$$dist(s,h_i) = \sqrt{\frac{1}{l}\sum_{j=1}^{l}\left(\frac{s_j-\mu_s}{\sigma_s} - \frac{h_{ij}-\mu_{h_i}}{\sigma_{h_i}}\right)^2} \quad (2)$$

where μ and σ are the mean and standard deviation of the subsequence. In both [10] and [13] the distance was computed using Euclidean distance without Z-normalization, however, we use the Z-normalized Euclidean distance due to different variances of the ECG signal examples. The distance threshold δ is computed such that the shapelet discriminates between the two alarm classes. Then, the MEDSC-U method iterates over all time series in D to extract all subsequences of length *l*, where *l* is the length of the potential shapelet. The method varies *l* between *minL* and *maxL* which are user-defined parameters.

2. Ranking shapelets: MEDSC-U assigns a score to each shapelet that incorporates both the earliness and the accuracy. The earliness defines how early, on average, the shapelet matches the target time series (the shapelet *S* matches the time series *T* if $dist(s,T) \le \delta$). Then, the shapelets are sorted in descending order based on their scores.

3. Prunning Phase: The process begins from the highest ranked shapelet S. The MEDSC-U method removes all time series from the dataset that are covered by the shapelet S. This shapelet is stored along with all other shapelets that have the same score as S (equal-performance shapelets as S). Then, the next ranked shapelet is considered. If the shapelet covers any of the remaining time series, the shapelet and all other equal-performance shapelets are added to the extracted list and all covered time series are removed. The method iteratively does so until all time series in the dataset are covered.

B. Testing Phase

When an ECG signal T with unknown label (true or false alarm) is encountered, the distance between the *observed* signal and all extracted discriminative shapelets is computed. When the shapelet $S=(s,l,\delta,c)$ matches T (i.e. the distance $dist(s,T)$ between T and S is less than or equal to δ) then T is classified as class c. Since ECG signals from bedside monitors are often contaminated with artifacts and noise which cause false alarms. The distance between T and S contains uncertainty in itself. To account for that uncertainty, the distance is defined as a random variable d

$$d = dist(s,T) + \varepsilon \quad (3)$$

where ε is some random variable with 0 mean and standard deviation equal to σ.

If the shapelet S matches T, the confidence C_S^c of classifying T as class c can be estimated by computing two components: 1) confidence in the fact that d is less than a threshold δ and 2) confidence in the ability of shapelet S to accurately classify time series T. Therefore, C_S^c is defined as

$$C_S^c = C_S(d<\delta|S\ matches\ T)C_S(class(T)=c|S\ matches\ T)$$

The first component is defined as

$$C_S(d<\delta|S\ matches\ T) = \frac{(\delta - dist(s,T))^2}{\sigma^2 + (\delta - dist(s,T))^2} \quad (4)$$

The closer $dist(s,T)$ is to δ, the lower the confidence is. Also, larger σ means lower confidence. More details about the derivation of Equation 4 can be found in [13]. The second component is computed as

$$C_S(class(T)=c|S\ matches\ T) = Precision(S) \quad (5)$$

where *Precision* is the fraction of the matched time series that are from class c [13]. Thus the lower bound of the class confidence estimate of the prediction C_S^c is calculated as

$$C_S^c \ge \frac{(\delta - dist(s,T))^2}{\sigma^2 + (\delta - dist(s,T))^2} \times Precision(S) \quad (6)$$

Since both terms in this product take value between 0 and 1, the highest value of the C_S^c is 1.

Equation 6 computes the confidence of predicting the time series T as class c using the shapelet S. So, for any time series T, the distance $dist(s,T)$ between the time series and the shapelet is computed. If the distance is less than or equal to the threshold, then the confidence C_S^c is computed using Equation 6. If the distance is greater than the threshold, the confidence is not computed, Hence, the confidence is computed only when the shapelet matches the time series.

When multiple shapelets match T over time, the overall confidence of the prediction increases as more evidences are gathered for the particular time series. For more details regarding computing the class confidence when multiple shapelets match, the readers are encouraged to read [13].

C. One-Sided MEDSC-U (1-MEDSCU)

We describe how to adapt MEDSC-U for the task of suppressing false alarm while keeping high true alarm detection rates. Since missing true alarm could lead to fatal consequences and risking patients' lives, the naive method is to predict every alarm as a true alarm. In this case, the true alarm detection (sensitivity) is 100% but false alarm suppression (specificity) is 0%.

To ensure that no true alarms is missed, we provide cost-sensitive alarm detection by comparing the computed C_S^c to a predefined confidence threshold value. In particular, we set the confidence threshold for predicting true alarm as very low and for false alarm as very high (99%). Therefore, when a true alarm shapelet (shapelet extracted from true alarm signals) matches the time series, the signal is classified immediately as a true alarm. On the other hand, when a false alarm shapelet (shapelet extracted from false alarm signals) matches the time series, we check the estimated confidence at that particular time point. If the estimated confidence is less than the predefined confidence threshold (no *strong* evidence yet that the signal is a false alarm), we delay our prediction task and look at a larger signal in the hope that the confidence estimate will increase with access to more data. If at the end of the time series the conditions failed to satisfy (no confident true or false alarm shapelets match so far), we classify the ECG signal as a true alarm.

Setting high confidence threshold for false alarm prediction ensures that a signal can be predicted as false alarm *only* if the confidence in our prediction is more than 99%, thus ensuring high true alarm detection rates. This approach in decision making ensures we do not miss any true alarm. On the other hand, we predict the signal as a true alarm as soon as a match is found so that we provide an early alert for every true alarm. Our approach could be viewed as a hybrid approach between MEDSC-U and EDSC methods, where it utilizes high confidence level for predicting false alarm and predicts a true alarm as soon as a match is found. We call our approach One Sided MEDSC-U (1-MEDSCU).

IV. Data description and Pre-processing

Two different critical alarm datasets were extracted from PhysioNet's MIMIC II version 3 repository [14] [15]. The database is a multiparameter ICU repository containing patient records of up to eight signals from bedside monitors in ICUs. The signals are sampled at 125 Hz. The extracted datasets contains the time stamps and human-annotated true and false asystole and ventricular tachycardia alarms. We extracted a subset of patient's records which contained only signal from lead ECG II, because it was identified as the sensor which contained the least number of missing values across the patients. For each alarm a 20-second window prior to the alarm event was extracted similar to [4]. Few alarm events contained missing values, which we ignored in this study. Finally, we ended up with 261 asystol (ASYS) alarms and 629 ventricular tachycardia (VTACH) alarms. Details about distribution of true and false alarms in the individual datasets are explained in Table III.

The raw signals extracted from MIMIC II was very noisy with high frequency signal components. In order to obtain a smooth signal, the ECG signal was passed through a low pass filter to remove the white noise. A 20-second analysis window prior to the alarm event was considered in our algorithm. However, each 20-second ECG signal contained 2500 data points in the time series, which increased the computational cost in our pattern extraction algorithm. Thus, we down-sampled each ECG signal from 125 Hz to 12.5 Hz, resulting in 250 temporal points in each signal.

V. Experimental setup

Assume that the number of true alarms is N. We partition the true alarm dataset into four distinct partitions, hence, each partition has $N/4$ true alarms. For each partition, we randomly select $N/4$ false alarms from the false alarm dataset to ensure balanced training data. We train our method (and the baseline methods) using the training data (of size $N/2$) and test them on the remaining examples. In addition, we repeat the entire process 20 times (each time has 4 distinct partitions on true alarm) which results in 80 different combination of training data.

TABLE III

Datasets description

Dataset	Total alarms	True Alarms (%)	False Alarms (%)
ASYS	261	40 (15.3%)	221 (84.7%)
VTACH	629	227 (36.09%)	402 (63.91%)

We used 4 evaluation measures: True alarm detection (TAD) rate, which is sensitivity; False alarm suppression (FAS) rate, which is specificity; and Earliness, which is the fraction of the time points used for classification. However, since missing true alarm (positive class) is more severe than missing false alarm (negative class), different errors incur different weights. The balanced accuracy (the average between sensitivity and specificity) considers similar weights for different errors. To account for this, false negative is penalized more than false positive by $1+\beta^2$, where higher β penalizes false negative more than false positive. Therefore, the weighted balanced accuracy ($WAcc$) is computed as:

$$WAcc = (WSens + Spec)/2 \tag{7}$$

where

$$WSens = TP/(TP + (1+\beta^2)FN)$$
$$Spec = TN/(TN + FP)$$

where TP, TN, FP, FN is the number of true positives, true negatives, false positives, and false negatives, respectively.

We compared our method to three baseline models. (1) BeharRaw: Behar et al. [3] was applied on the raw ECG signals. (2) BeharFiltered: Behar et al. [3] was applied on the filtered ECG signals. (3) EDSC [10] with the Z-normalized version of distance measuring (Equation 2). The original EDSC method resulted in 0 sensitivity, thus we did not include it as a baseline method.

VI. Discussion

A. Accuracy performance

The evaluation of each method is shown in Table IV. Clearly, our method has near optimal TAD rate, while all other methods have much less TAD rate. For example, on ASYS dataset, BeharFiltered has comparable TAD rate (92.37%) to our method (99.12%), however, it has lower FAS rate (18.97%) compared to 1-MEDSCU (34.29%). EDSC has better FAS rate (74.16%) than ours but on the cost of TAD rate (83.62%). This shows that our method has moderate FAS rates while keeping high TAD rate, which is an extremely challenging task. The same conclusion applies on VTACH. However, we claim that we can obtain the desired level of FAS by adjusting the confidence threshold of the method, which will reduce TAD rate but will still be comparable to other methods. This is explained in the next section.

In addition to TAD and FAS, it is clear that our method has better weighted accuracy $WAcc$ than all other methods. There is a statistically significant difference (pvalue is shown in the last column of Table IV) between our method and all other methods using significance level 0.05, except for EDSC on ASYS at $\beta = 2$.

TABLE IV

EVALUATION OF THE MODELS IN TERMS OF TRUE ALRM DETECTION RATE (TAD), FALSE ALARM SUPPRESSION RATE (FAS), EARLINESS (100 - EARLINESS) AND WEIGHTED ACCURACY (WACC). LARGER VALUE HAS BETTER PERFORMANCE. PVLAUE IS COMPUTED BETWEEN OUR METHOD AND THE BEST BASELINE METHOD ON THE CORRESPONDING EVALUATION MEASURE

		Behar (Raw)	Behar (Filtered)	EDSC	1-MEDSCU	pvalue
ASYS	TAD	84.62±11.46	92.37±11.27	83.62±15.19	99.12± 3.25	
	FAS	35.03±7.64	18.97±7.24	74.16±9.41	34.29±12.36	
	100-Earliness	0	0	62.8±6.27	38.39±9.05	
	$WAcc$ ($\beta = 2$)	47.11±11.13	49.16±10.84	**66.12±11.55**	65.68±6.32	0.74
	$WAcc$ ($\beta = 3$)	41.9±13.7	45.96±14.33	59.75±13.39	**65.05±7.55**	1.20e-03
VTACH	TAD	86.22±8.69	52.60±25.27	64.78±23.16	**95.67±8.81**	
	FAS	31.18±5.7	51.07±24.75	65.07±14.77	20.32±13.43	
	100-Earliness	0	0	59.9±11.71	39.96±9.34	
	$WAcc$ ($\beta = 2$)	44.49±3.11	37.20±5.99	48.65±2.47	**52.85±5.52**	4.72e-09
	$WAcc$ ($\beta = 3$)	36.34±2.91	32.80±8.32	42.58±3.39	**48.56±7.41**	3.13e-09

B. Controlling False Alarm Suppression Rate

Our method has advantage over other methods in controlling the balance between TAD and FAS. In other words, the false alarm confidence threshold used in 1-MEDSCU controls the sensitivity of the model to predict true and false alarm. When a test ECG signal matches a false alarm shapelet (blue shapelet as in Figure 1), the method computes the confidence of the match. If the estimated confidence is greater than the false alarm confidence threshold, then 1-MEDSCU predicts the signal as a false alarm. Increasing the false alarm confidence threshold *guarantees* that no true alarm is incorrectly predicted as a false alarm but at the same time decreases the false alarm suppression rate. In the previous results, we used 99% false alarm confidence threshold to ensure near-optimal TAD rates. By varying the confidence level we can obtain FAS rate comparable to other method but still with higher (but not near-optimal) TAD rates. The results of varying the false alarm confidence threshold is shown in Fig. 2.

The blue dotted (red dashed) line represents the varying FAS (TAD) rates for different false alarm confidence thresholds (x-axis), respectively. The blue marks (star, circle, and diamond) indicate the FAS rates, while the red marks show the TAD rates achieved by the three baseline models. It is clear that, our model can achieve similar FAS rates as the baseline methods with comparable or even higher TAD rate. For example, in order to achieve FAS rate similar to EDSC (blue star) we can achieve a TAD rate of 83 (the vertical line that touches the blue star, touches the red dashed line at 83%) by setting false alarm confidence threshold to 4.9%. So, we achieve comparable TAD rates as to EDSC (red star). Comparing to BeharRaw, it has 35% FAS (blue diamond) and 85% TAD (red circle), while 1-MEDSCU can achieve 99% TAD at 35% FAS, significantly outperforming BeharRaw.

Therefore, by varying the confidence threshold, we can achieve the desired level of FAS with comparable or even better TAD rates.

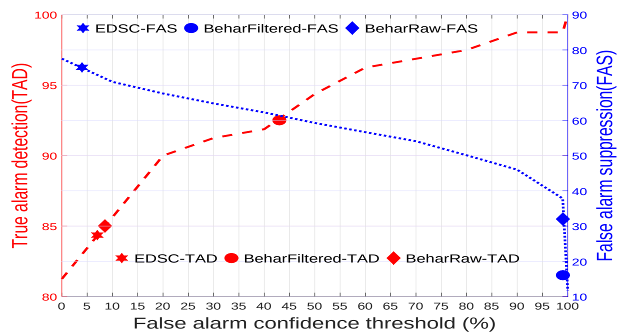

Fig. 2. Varying false alarm confidence threshold for ASYS. The red line shows increasing true alarm detection with increasing false alarm confidence threshold. The blue line show decreasing false alarm suppression with increasing false alarm confidence threshold. The red and blue marks indicate TAD and FAS respectively achieved by the baseline methods

C. Earliness

The results of earliness of the methods are shown in Table IV. 1-MEDSCU not only provide accurate results (as shown in the previous sections) but also provide these results early. So, the prediction takes place even before the actual alarm alerts, whereas all other methods, except EDSC, provide results at the time when the alarm happens. The prediction of our method is provided, on average, using around 60% of the time points (8 seconds before the actual alarm) at a false alarm confidence threshold of 99%. Although, it is evident that EDSC has better earliness performance than 1-MEDSCU, our method outperforms EDSC in terms of TAD as in Table IV. By varying the false alarm confidence threshold, the earliness of 1-MEDSCU improves as shown in Fig. 3. At 4.9% confidence threshold, the prediction of our method were provided using only 40% of time series' length, comparable to EDSC.

From Fig. 2 and 3, we conclude that by lowering the false alarm confidence threshold we obtain earlier predictions and

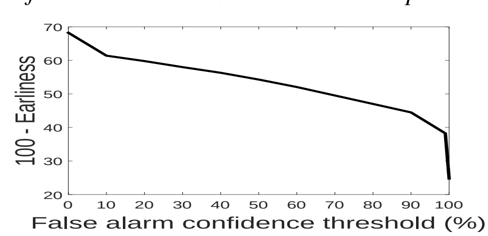

Fig. 3. Trend of Earliness with varying false alarm confidence threshold (100 - Earliness) on ASYS. Larger the value, the earlier the prediction

higher FAS rates but on the cost of reducing TAD rates. Therefore, a proper trade-off has to decided by hospital administrators between earliness, FAS and TAD.

D. Interpretability: Case Study

We present an example to show the effectiveness of our proposed interpretable method that utilizes the confidence levels to produce more accurate results. In Figure 4, a true alarm signal matches a false alarm shapelet (solid blue segment) with confidence 1% at time point 4 (so 16 seconds before the alert). EDSC would classify that example at that time as false alarm. However, 1-MEDSCU does not classify the signal at that time because the confidence is less than the false alarm confidence threshold (99%), therefore, delays the decision. At time 4.9 second, another false alarm shapelet (dotted blue) matches the signal resulting in confidence 8%. 1-MEDSCU continues until time 16 where the signal matches a true alarm shapelet (red shapelet). The method immediately classifies the signal correctly as a true alarm, because the method does not require confidence to predict the signal as a true alarm. It is clear that the signal can be mistakenly classified as a false alarm because two evidences (two shapelets) were found in the early phases of the signal. However, since the evidences are not strong enough the method continues until either a very strong evidence to classify as a false alarm is found or any evidence to classify it as a true alarm is found, thus ensuring high TAD rates.

VII. Conclusion And Future Work

In this paper, we address the problem of suppressing high cardiac false alarms using univariate ECG signals. The objective of this paper is to reduce false alarms as much as possible without compromising TAD performance. We have achieved this objective in the proposed (1-MEDSCU) model by keeping high confidence threshold for false alarm predictions to ensure high TAD. We were able to suppress a moderate percentage of FAS while keeping high rate of early TAD predictions. We show that the proposed early alarm detection approach has outperformed the state-of-the-art methods on both datasets in terms of weighted accuracy. In addition, we showed that we can control the FAS rate on the cost of TAD rate, nevertheless, the method achieved better suppression rate than other methods keeping comparable TAD rate. In addition, we showed that our method provides not only accurate results but also provides interpretable results early.

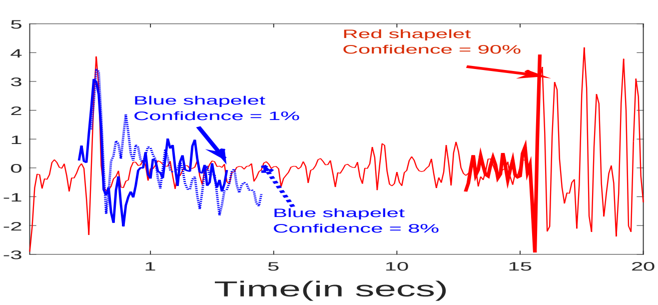

Fig. 4. True alarm example wrongly classified as false alarm by EDSC at time 4, however, correctly classified as true alarm by 1-MEDSCU at time 16

Currently, 1-MEDSCU works for univariate time series. In future we will extend it for multivariate time series [16] in order to improve the performance of the model. We will also investigate the problem of false alarm suppression in other medical domains.

Acknowledgment

The authors would like to thank Dušan Ramljak for valuable discussions.

References

[1] A. Aboukhalil, L. Nielsen, M. Saeed, R. G. Mark, and G. D. Clifford, "Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform," *Journal of Biomedical Informatics*, vol. 41, no. 3, pp. 442 – 451, 2008.

[2] Q. Li and G. D. Clifford, "Signal quality and data fusion for false alarm reduction in the intensive care unit," *Journal of Electrocardiology*, vol. 45, no. 6, pp. 596 – 603, 2012.

[3] J. Behar, J. Oster, Q. Li, and G. Clifford, "ECG signal quality during arrhythmia and its application to false alarm reduction," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 6, pp. 1660–1666, 2013.

[4] R. Salas-Boni, Y. Bai, P. R. E. Harris, B. J. Drew, and X. Hu, "False ventricular tachycardia alarm suppression in the ICU based on the discrete wavelet transform in the ECG signal," *Journal of Electrocardiology*, vol. 47, no. 6, pp. 775 – 780, 2014.

[5] F. Scalzo, D. S. Liebeskind, and X. Hu, "Reducing false intracranial pressure alarms using morphological waveform features." *IEEE Trans. Biomedical Engineering*, vol. 60, no. 1, pp. 235–239, 2013.

[6] J. P. K. Jr., "Clinical alarm hazards: a top ten health technology safety concern," *Journal of Electrocardiology*, vol. 45, no. 6, pp. 588 – 591, 2012.

[7] B. J. Drew, P. Harris, J. K. Zgre-Hemsey, T. Mammone, D. Schindler, R. Salas-Boni, Y. Bai, A. Tinoco, Q. Ding, and X. Hu, "Insights into the problem of alarm fatigue with physiologic monitor devices: A comprehensive observational study of consecutive intensive care unit patients," *PLoS ONE*, vol. 9, no. 10, 10 2014.

[8] O. Sayadi and M. B. Shamsollahi, "Life-Threatening Arrhythmia Verification in ICU Patients Using the Joint Cardiovascular Dynamical Model and a Bayesian Filter," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 2748–2757, 2011.

[9] Z. Xing, J. Pei, and P. S. Yu, "Early prediction on time series: A nearest neighbor approach." in *International Joint Conferences on Artificial Intelligence*, 2009, pp. 1297–1302.

[10] Z. Xing, J. Pei, P. S. Yu, and K. Wang, "Extracting interpretable features for early classification on time series." in *SIAM International Conference on Data Mining*, 2011, pp. 247–258.

[11] M. Ghalwash, V. Radosavljevic, and Z. Obradovic, "Early diagnosis and its benefits in sepsis blood purification treatment," in *IEEE International Conference on Healthcare Informatics, ICHI 2013, 9-11 September, 2013, Philadelphia, PA, USA*, 2013, pp. 523–528.

[12] L. Ye and E. Keogh, "Time series shapelets: a new primitive for data mining," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2009, pp. 947–956.

[13] M. Ghalwash, V. Radosavljevic, and Z. Obradovic, "Utilizing temporal patterns for estimating uncertainty in interpretable early decision making," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014, pp. 402–411.

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

[15] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13).

[16] M. Ghalwash, V. Radosavljevic, and Z. Obradovic, "Extraction of interpretable multivariate patterns for early diagnostics," in *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, Dec 2013, pp. 201–210.