# Cardiovascular Risk Stratification with Heart Rate Topics

Alexander Van Esbroeck, Zeeshan Syed

University of Michigan, Ann Arbor, USA

## Abstract

*Recent work on heart rate motifs (HRM) has demonstrated that information in short heart rate patterns may be useful in identifying patients at elevated risk of cardiovascular death (CVD) following acute coronary syndrome. The information in HRM complements a variety of other clinical metrics including electrocardiographic (ECG) measures. While the HRM approach has value, it suffers from a focus on identifying and using information related to only a small number of discriminative patterns in heart rate time series, which loses valuable information among the full set of patterns in the data. We present a method based on topic models, an approach traditionally used in text analysis, to learn structure in the full set of short heart rate patterns in long-term ECG recordings. This model provides an interpretable representation of long-term ECG recordings and finds relationships amongst all short heart rate patterns across the entire patient population. When evaluated on data from 4,557 patients admitted with non-ST-elevation acute coronary syndrome, we show that heart rate topic models significantly improve risk stratification. This improvement is consistent even when considering information already available through the TIMI risk score and left ventricular ejection fraction, as well as several heart rate variability metrics.*

## 1. Introduction

The increasing prevalence of cardiovascular disease, combined with its high risk of mortality, has resulted in a variety of treatments (e.g., procedures, medications, devices) being developed to prevent future events and reduce an individual's risk of mortality. Despite the wide array of these treatment options, determining the appropriate level of therapy for an individual remains challenging. Accurate risk stratification is vital for matching patients to appropriate treatments, with the potential to both save lives and reduce health care costs.

A number of different biomarkers and clinical scores have been generated to quantify a patient's cardiac risk and to guide clinical decision making. Recent work has focused on improving patient assessment through cardiac biomarkers derived from long-term physiological signals such as the ECG. Of particular interest are changes in the heart rate over long periods of time, as a way of characterizing abnormal autonomic nervous regulation of the heart predisposing to fatal arrhythmias. Much of the initial work in this space (e.g., heart rate variability [1] and deceleration capacity [2]) relates either aggregate changes in heart rate or the presence of a specific heart rate pattern to an increase in cardiovascular risk. More recently, this work has been generalized to identify and integrate information in multiple heart rate motifs (i.e., short heart rate sequences discovered in a data-driven manner from long-term ECG) that are over- or underrepresented in patients experiencing cardiovascular outcomes [3].

While heart rate motifs have been found to be predictive of several cardiac outcomes [4], the number of motifs present in the data is overwhelming and requires the selection of a small set of predictive patterns. In this paper, we hypothesize that considering all of the patterns in the data at once can provide information useful for risk stratification. Our approach models the occurrence of all motifs appearing in a day-long ECG recording, automatically learning functional sets of patterns across a population. We leverage a method traditionally used in text analysis, topic models, to learn these sets of related heart rate motifs directly from the data. These data-derived topics provide a concise representation of the short patterns across a long-term recording, which can be used to better assess overall cardiac health.

The contributions of this paper are:

- we describe how topic models, a method for discovering latent structure in text documents, can be applied to the heart rate time series to assess cardiovascular risk;
- we evaluate the performance of a topic-based risk score alongside several popular risk measures on a real-world dataset consisting of day-long ECG recordings with nearly year-long patient follow-ups

## 2. Background

Most cardiac biomarkers derived from the ECG focus on the idea that an inability for the heart to adjust its rate to compensate for different physiological situations is indicative of a high risk for cardiovascular problems.

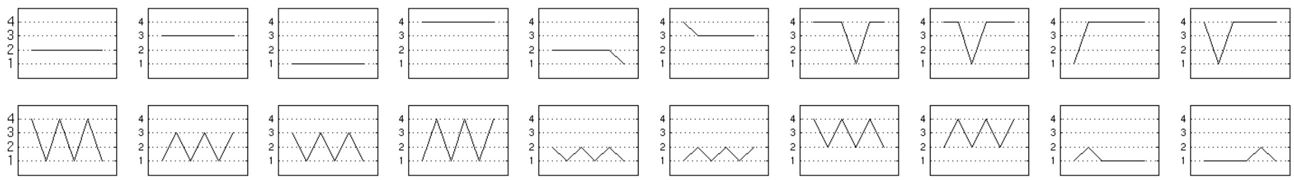

Figure 1. Two examples of heart rate topics. Each row depicts the 10 most representative heart rate patterns for a topic. In each pattern, a 1 indicates a low heart rate, and a 4 indicates a high heart rate.

In their recent work on heart rate motifs, Chia and Syed explored identifying and integrating information from the frequencies with which different short-term heart rate patterns occur in long-term ECG to assess cardiac patients. The approach converted 24 hour heart rate time series into symbolic sequences, corresponding to low through high heart rate ranges, and discovered a set of over- or under-represented symbolic motifs in patients suffering cardiovascular death [3]. The information in these motifs was found to be complementary to other commonly used clinical variables including other ECG-derived risk scores.

However, there are several challenges with the use of heart rate motifs. For even a moderate choice of motif lengths, the number of motifs in the data grows unreasonably (with 4 symbols and motifs of length 8 there are over 60,000 patterns). The use of real-world long-term ECG datasets, which contain limited numbers of patients, necessitates the use of only a sparse subset of features to prevent overfitting. This has two effects: first, this prevents the leveraging of large numbers of patterns and their relationships for prediction. Second, this model describes only a small percentage of the original recordings, and in some cases an individual may have none of the predictive motifs present in their ECG. This provides a limited representation of the data, and can be challenging to interpret. The use of individual motif frequencies identifies no structure in the relationships between motifs, and cannot distinguish cases where one heart rate motif can appear in multiple contexts.

Understanding the higher-level structure of these motifs may provide novel insights into the data and correspondingly into cardiac physiology. We investigate the application of topic modeling to heart rate motifs. Topic models are used to describe collections of text documents by learning topics (sets of semantically related words). For example, the proceedings of a cardiology conference might contain topics such as electrophysiology, echocardiography, or atrial fibrillation. Each document is associated with a number of topics, and each topic is associated with a distribution over words.

Prior work has extended the use of topic models to ECG data, by first symbolizing and identifying "words" from long-term ECG recordings, and then identifying latent topics in these "words" across a patient population [5]. We consider heart rate motifs as words, constituents of a single document corresponding to a day-long ECG recording. After defining a document as the frequencies of all motifs occurring in a recording, we use this text-like representation to extract higher-level structure in the documents by learning topics over the motifs. By relating symbolically different but functionally similar motifs, all patterns in the data can be leveraged to identify useful predictive information. This allows the previously unwieldy motif frequency representation to be condensed into a mixture over a small set of topics, providing a similarly sparse representation but with the potential for greater predictive power. Two examples of heart rate topics are displayed in Fig. 1 with their 10 most probable patterns.

In this work we thoroughly evaluate this approach on a real-world dataset, comparing heart rate topics to a number of popular risk factors. We believe that topic models provide a better interpretation and understanding of heart rate motifs than the original approach, while maintaining and potentially expanding upon the predictive power of these features for cardiac assessment.

## 3. Methods

First, we extract the heart rate time series from the raw ECG recording. This involves identifying QRS complexes and then calculating the time between adjacent complexes. We use the open source QRS detection algorithms proposed by Hamilton et al. [6] and Zong et al. [7], and identify QRS complexes at time instants where both algorithms agree. The instantaneous heart rate was defined as the time between all pairs of normal sinus beats.

We then discretize the heart rate time series into a symbolic sequence, to transform it into a text-like representation that can be used to learn topic models. We use symbolic aggregate approximation (SAX) to achieve this representation [8]: given an alphabet size $A$, SAX divides the input values into $A$ equiprobable bins. It then assigns each value from the original series a symbol based on the bin in which it sits. In the hypothetical case of $A$ = 3, the symbols could be interpreted as denoting low, medium, and high heart rates. We apply SAX to each patient's heart rate time series separately, to account for individual variability in baseline heart rates.

Each patient's discretized heart rate time series can be abstractly represented as a text document, where we consider all substrings of a given length in this symbolic sequence as words. For a motif length $n$, we calculate the frequency of all $n$-length substrings in the symbolic sequence, allowing for overlaps. In our work, we consider all patients' day-long ECG recordings as analogous to documents, with each heart rate motif corresponding to a word. The topics, rather than representing sets of related words, reflect sets of related heart rate motifs.

To train a topic model on the long-term ECG, we applied Latent Dirichlet Allocation (LDA) [9] to the collection of symbolized recordings. LDA is a generative model that assumes each topic underlying the collection can be characterized by a distribution over words in the vocabulary: a topic about atrial fibrillation may have high probabilities associated with words like "arrhythmia" or "atrial". Given only a collection of documents and their respective word frequencies, LDA provides a method to learn a set of topics that describe the collection, and the associations between each topic and individual words. It is then possible to associate the relationship of each document with these topics. For example, a document may be 70% related to atrial fibrillation, and 30% related to electrophysiology.

More concretely, in the context of heart rate analysis LDA defines an underlying set of $K$ topics, where each topic $k$ can be defined by a multinomial distribution over all of the heart rate patterns appearing in the data. Each ECG recording $r$ is itself generated by a multinomial distribution over topics, where the generative process assumes a two part process for each pattern $w_{ri}$ in the recording. First, a topic $z_{ri}$ is chosen by sampling a topic from that recording's topic distribution $\theta_r$. Then the pattern $w_{ri}$ is sampled from that topic's distribution over patterns $\beta_k$.

The model defines for a recording $r$ the prior distributions over observed patterns $w_{ri}$, their latent topics $z_{ri}$, and the recording distributions over topics $\theta_r$ as:

$$\theta_d \sim Dir(\alpha)$$

$$z_{di} \sim Multi(\theta_d)$$

$$w_{di} \sim Multi(\beta_{z_{di}})$$

Where $\alpha$ parameterizes a symmetric Dirichlet prior on topic distributions. We used variational inference estimation to learn these parameters from a collection of recordings.

## 4. Experiments

We trained the heart rate topic models, which do not take into account patient outcomes, on data from 4,557 patients admitted to hospitals with non-ST-elevation acute coronary syndrome. Prior work found an optimal choice of parameters with motif length of 6 and an alphabet size of 4 [5]. The Bayesian information criterion (BIC) was used to select the number of topics, which was set to 10.

In assessing the performance of heart rate topic models, however, we only focused on the set of patients with values recorded for LVEF ($n = 3,071$). The decision to focus on this subset for reporting performance statistics was because because patients within this group provided an opportunity to rigorously assess the improvement offered through our work relative to existing approaches for cardiovascular assessment. As a result, while we used all available data to train heart rate topic models in an unsupervised manner, we report performance consistently on patients with LVEF data available. We randomly divided these patients with LVEF measurements into evenly sized separate training and testing cohorts. Within the training cohort, we learned a classifier to translate the frequencies of the different topics learned in an unsupervised manner to a patient's estimate for CVD over follow-up. This classifier was evaluated on the testing cohort. The evaluation was done in two ways. First, we explored the ability of a topic-based risk score to provide supplemental information in a multivariate logistic regression model that did not factor in information related to the timings of events and the censoring of patients. This multivariate logistic regression model was adjusted for the TIMI risk core (TRS) [10] and left ventricular ejection fraction (LVEF). Second, we also explored the ability of a topic-based risk score to provide supplemental information in a multivariate Cox proportional hazards regression model, which incorporates timing and censoring information into the analysis. We compared our approach in both cases with different HRV metrics. The different variables were dichotomized as: TRS low (0-2), moderate (3-4) and high (5-7), LVEF over 40%, HRV SDNN below 70, HRV LFHF below 0.95, and the topic-based risk score (Topics) at the 50th percentile. In the logistic regression model, improvement by including the topic-based score with standard risk scores was measured using the area under the ROC curve (AUROC) and net reclassification improvement (NRI) as performance metrics [11]. In the Cox proportional hazards regression model, the different approaches were compared by studying the hazard ratios of the model.

## 5. Results

The topic-based risk score showed low correlation with all of the risk factors ($r < 0.20$), with the exception of LFHF ($r = 0.40$). In the logistic regression framework for evaluation, the addition of the topic-based risk score significantly improved discrimination and net reclassification relative to the TIMI risk score and LVEF (Table 1). The AUROC improved from 0.782 to 0.788 when the HRV metrics were added to the TIMI risk score and LVEF, while the NRI was 0.47. A similar improvement was seen when

Table 1. AUROC and NRI with corresponding p value indicating the prediction improvement gained by adding risk factors (to the right of the parentheses) to the model including only the variables in the parentheses. HRV includes both SDNN and LFHF.

| Variables | AUROC | NRI | p |
|---|---|---|---|
| TRS, LVEF | 0.782 | n/a | n/a |
| (TRS, LVEF), HRV | 0.788 | 0.47 | $< 0.001$ |
| (TRS, LVEF), Topics | 0.790 | 0.32 | 0.006 |
| (TRS, LVEF, HRV), Topics | 0.791 | 0.30 | 0.010 |

Table 2. Hazard ratios of risk factors with 90-day cardiovascular death times under a Cox proportional hazards model

| Factor | Hazard Ratio | p |
|---|---|---|
| TRS | 10.3 | $< 0.001$ |
| LVEF | 3.76 | $< 0.001$ |
| Topics | 2.75 | 0.002 |
| SDNN | 1.20 | n.s. |
| LFHF | 1.81 | 0.026 |

the topic-based risk score was added to a model containing TRS, LVEF, and both HRV metrics (AUROC improved from 0.788 to 0.791 while NRI was 0.30).

When timing and censoring information was factored into the evaluation using the Cox proportional hazards regression approach, both the TIMI risk score and LVEF had a higher hazard ratio than the topic-based risk score. However, the topic-based risk score achieved a higher hazard ratio than HRV and also provided information that was complementary to both the TIMI risk score and LVEF demonstrating the incremental improvement possible through this method.

## 6. Conclusions

A risk-score derived from a heart rate topic model, which models the full set of short heart rate patterns appearing in a long-term ECG recording, significantly improved cardiovascular risk stratification when combined with TRS, LVEF, and several HRV metrics.

## References


[1] Malik M. Heart rate variability. Annals of Noninvasive Electrocardiology 1996;1(2):151–181.

[2] Bauer A, Kantelhardt J, Barthel P, Schneider R, Mäkikallio T, Ulm K, Hnatkova K, Schömig A, Huikuri H, Bunde A, et al. Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study. The Lancet 2006;367(9523):1674–1681.

[3] Chia C, Syed Z. Computationally generated cardiac biomarkers: Heart rate patterns to predict death following coronary attacks. In SIAM International Conference on Data Mining. 2011.

[4] Syed Z, Stultz C, Scirica B, Guttag J. Computationally generated cardiac biomarkers for risk stratification after acute coronary syndrome. Science Translational Medicine 2011; 3(102):102ra95–102ra95.

[5] Van Esbroeck A, Chia CC, Syed Z. Heart rate topic models. In Association for the Advancement of Artificial Intelligence. 2012.

[6] Hamilton P, Tompkins W. Quantitative investigation of qrs detection rules using the mit/bih arrhythmia database. Biomedical Engineering IEEE Transactions on 1986; (12):1157–1165.

[7] Zong W, Moody G, Jiang D. A robust open-source algorithm to detect onset and duration of qrs complexes. In Computers in Cardiology. Ieee, 2003; 737–740.

[8] Lin J, Keogh E, Lonardi S, Chiu B. A symbolic representation of time series, with implications for streaming algorithms. In Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. ACM, 2003; 2–11.

[9] Blei D, Ng A, Jordan M. Latent dirichlet allocation. The Journal of Machine Learning Research 2003;3:993–1022.

[10] Antman E, Cohen M, Bernink P, McCabe C, Horacek T, Papuchis G, Mautner B, Corbalan R, Radley D, Braunwald E. The timi risk score for unstable angina/non–st elevation mi. JAMA the journal of the American Medical Association 2000;284(7):835.

[11] Pencina M, D'Agostino Sr R, D'Agostino Jr R, Vasan R. Evaluating the added predictive ability of a new marker: from area under the roc curve to reclassification and beyond. Statistics in medicine 2008;27(2):157–172.



Address for correspondence:

Alexander Van Esbroeck
2260 Hayward Street, Ann Arbor, MI 48109, USA
alexve@umich.edu