A Dynamic Risk Score to Identify Increased Risk for Heart Failure Decompensation

Shantanu Sarkar*, *Member, IEEE*, and Jodi Koehler

***Abstract*—A method for combining heart failure (HF) diagnostic information in a Bayesian belief network (BBN) framework to improve the ability to identify when patients are at risk for HF hospitalization (HFH) is investigated in this paper. Implantable devices collect HF related diagnostics, such as intrathoracic impedance, atrial fibrillation (AF) burden, ventricular rate during AF, night heart rate, heart rate variability, and patient activity, on a daily basis. Features were extracted that encoded information regarding out of normal range values as well as temporal changes at weekly and monthly time scales. A BBN is used to combine the features to generate a risk score defined as the probability of a HFH given the diagnostic evidence. Patients with a very high risk score at follow-up are 15 times more likely to have a HFH in the next 30 days compared to patients with a low-risk score. The combined score has improved ability to identify patients at risk for HFH compared to the individual diagnostic parameters. A score of this nature allows clinicians to manage patients by exception; a patient with higher risk score needs more attention than a patient with lower risk score.**

***Index Terms*—Data fusion, heart failure (HF) prediction, implantable device diagnostics.**

I. Introduction

HEART failure (HF) is the most common cardiovascular disease that causes significant economic burden, morbidity, and mortality. In the U.S., more than 5.7 million have HF [1]. The primary cause of a significant proportion of hospitalizations is HF, with close to one million discharges for HF in the U.S. in 2007 [1]. The primary cause of HF hospitalization (HFH) is volume overload in which patients retain excess amount of fluid. The primary HF management strategy is to control excess fluid volume using diuretic therapy [2]. Further, ACE-Inhibitors, which control blood pressure, and β-blockers, which control heart rate, are known to reduce mortality in HF patients [2].

Implantable medical devices, such as pacemakers, implantable cardioverter defibrillator, cardiac resynchronization therapy defibrillator (CRT-D) and implantable loop recorders, provide daily measurements of several diagnostic parameters for possible evaluation of HF status in patients. Wireless transmission capabilities in these devices allow for automatic remote monitoring of the diagnostic parameters over the Internet without requiring patient compliance. Studies have shown that continuously monitored device diagnostics such as intrathoracic impedance (IMP) [3], atrial fibrillation (AF) burden and rate control information [4], and night heart rate (NHR), heart rate variability (HRV), and patient activity (ACT) [5] can identify HF patients at risk for future events and could potentially be used unilaterally to manage HF patients in an acute setting using remote monitoring alerts [6] or during a follow-up [7]. This paper investigates a method of combining the device diagnostic information in a Bayesian belief network (BBN) framework into a single risk score to improve the ability to identify patients at risk for HFH.

The HF syndrome is a complex interaction of multiple physiological processes with a high degree of uncertainty in the severity of the manifestation of the disease that may or may not require a HFH. The BBN approach [8], [9] allows for uncertain reasoning to estimate the probability of HF under a set of given diagnostic evidence. The BBN framework [8], [9] also allows for causal modeling, modular structuring, differential reasoning based on multiple diagnostic tests that are very similar in nature to a clinical decision making process. The BBN framework has been applied to other applications of biomedical informatics [10], [11].

II. Methods

A. Diagnostic Elements

Implanted medical devices monitor several clinical diagnostic parameters that may include IMP, AF burden, ventricular rate during atrial fibrillation (VRAF), ACT, NHR, and HRV. These parameters are monitored continuously and the device stores sample data points for each parameter daily. IMP is measured across the thorax between an electrode inside the heart and the device [3]. IMP is a surrogate measure for increasing fluid in the thorax, with an increase in fluid volume leading to a reduction in IMP. HRV is measured as the standard deviation of 5 min median of atrial (PP) intervals during a 24 h period. The HRV is a long-term measure of sympathetic tone changes, with reducing HRV implying increases in sympathetic tone. HRV is not measured during AF. NHR is measured as the average heart rate between midnight and 4 A.M. and is a measure for resting heart rate. ACT is measured as the number of minutes in a 24 h period the patient is active. A patient is considered active over a minute if the number of deflections of an accelerometer in the device exceeds a threshold. AF burden is measured as the total duration of the fast atrial rate during a 24 h period, as measured by PP intervals using a sensing electrode in the atrium, and with AV conduction ratio $\geq$ 2:1 [12]. In implantable loop recorder devices, AF is detected using incoherence of RR

Manuscript received February 28, 2012; revised June 13, 2012; accepted July 4, 2012. Date of publication July 20, 2012; date of current version December 14, 2012. *Asterisk indicates corresponding author.*

*S. Sarkar is with the CRDM Research Division, Medtronic, Inc., Moundsview, MN 55112 USA (e-mail: shantanu.sarkar@ medtronic.com).

J. Koehler is with the CRDM Research Division, Medtronic, Inc., Moundsview, MN 55112 USA (e-mail: jodi.koehler@medtronic.com).

Digital Object Identifier 10.1109/TBME.2012.2209646

interval time series over a period of time [13]. VRAF is the average ventricular rate during AF over a 24 h period.

B. Feature Extraction

Features are extracted for each diagnostic parameter to encode amplitude and temporal characteristics at monthly and weekly temporal scales. A large set of time series features was investigated for each parameter for both the time windows that included number of days outside a normal range of measured values, cumulative sum of difference between the raw measurement and an adaptive reference (CSAR) [3], cumulative sum of difference between the raw measurement and a fixed reference (CSFR), number of days CSAR or CSFR above a threshold, the slope of the raw measurements, and mean, median, minimum, and maximum measurement values.

For each parameter, the goal was to create four diagnostic categories based on a combination of features such that the HFH rate increases from one category to next. The category with the largest HFH rate was also designed to have the lowest occurrence rate.

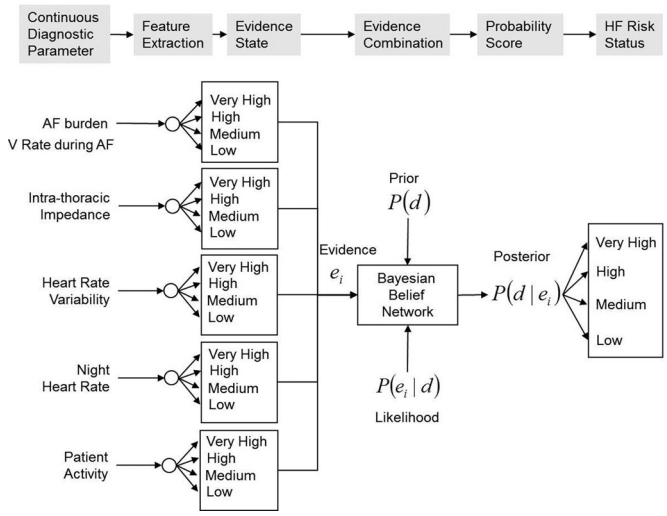

Fig. 1. Basic schematic for generation of the combined risk score.

C. Combined Risk Score

A BBN framework was used to combine the features for each diagnostic parameter. A six-node BBN framework with one parent node (d) and five child nodes (e_1–e_5) was used in this paper. To satisfy the assumption of conditional independence, each evidence node is a set of independent measurements. AF burden and the VRAF are combined together as one set of evidence (e_1). The other evidence nodes consist of IMP (e_2), HRV (e_3), NHR (e_4), and ACT (e_5). HRV and NHR were only calculated when patients were not in AF.

For the six-node BBN, the joint probability distribution can be expressed as

$$P(d, e_1, \ldots, e_5) = P(d) \cdot \prod_{i=1}^{5} P(e_i|d). \quad (1)$$

For discrete states of d, the posterior probability (risk score) for the occurrence of a HF event ($d = H$) given evidence from the diagnostic variables (e_1–e_5) is expressed as

$$P(d = H|e_1, \ldots, e_5) = \frac{P(d = H) \cdot \prod_{i=1}^{5} P(e_i|d = H)}{\sum_d P(d) \cdot \prod_{i=1}^{5} P(e_i|d)}. \quad (2)$$

The prior probability$P(d)$ and the conditional likelihood $P(e_i|d)$ are computed from existing data.

The basic steps for the computation of the probability score described in (2) based on evidence from diagnostic parameters is shown in Fig. 1. In addition to the discrete states of d, evidence from each diagnostic parameter (e_1–e_5) has four discrete states.

D. Statistical Analysis

The dataset included data from HF patients with CRT-D devices available from four clinical studies with at least 90 days of follow-up data and evidence of device detected AF. Data at daily resolution from each diagnostic parameter are stored in the device for the last 425 days and can be remotely downloaded to a monitoring system. Patients with device detected AF were defined as patients who have at least 1 day of AF $>$ 5 min and their total AF burden during follow-up was $>$ 1 h. Patients with HF and AF were chosen as they are known to have a higher HFH event rate. The data cohort consisted of 591 patients with a total of 6521 monthly evaluations and 138 months with HFH events resulting in an overall HF event rate of 2.1% of monthly evaluations.

Performance evaluation of the features for individual diagnostic parameters and the combined probability score was performed by simulating a monthly follow-up which consisted of looking back at the diagnostics in the last 30 days and evaluating the occurrence of a HFH event in the following 30 days (see Fig. 2). The individual diagnostic parameter feature sets are grouped into four evidence states and the combined probability score is categorized into risk status groups as shown in Fig. 1. An Anderson–Gill model was used to compare survival free from HFH events based on the evidence states for each diagnostic parameter or the HF risk status groups. HFH were used as the endpoint in the data analysis. Each cardiovascular hospitalization collected during the clinical studies was carefully adjudicated for signs and symptoms of HF which included administration of IV or oral diuretic medication during the hospitalization.

A sensitivity and specificity analysis was performed for the combined risk score using the 30-day evaluation scheme shown in Fig. 2 with one data point every 30 days. Sensitivity (and specificity) is defined as the number of evaluations with score $\geq$ (or $<$) threshold and HFH (or no HFH) event in next 30 days divided by the total number of evaluations with HFH (without HFH) in next 30 days. The sensitivity and specificity computations are adjusted for multiple evaluations in patients using generalized estimating equation (GEE) with an exchangeable correlation structure. All statistical analyses were performed using SAS version 9.2 (SAS Institute Inc., Cary, NC).

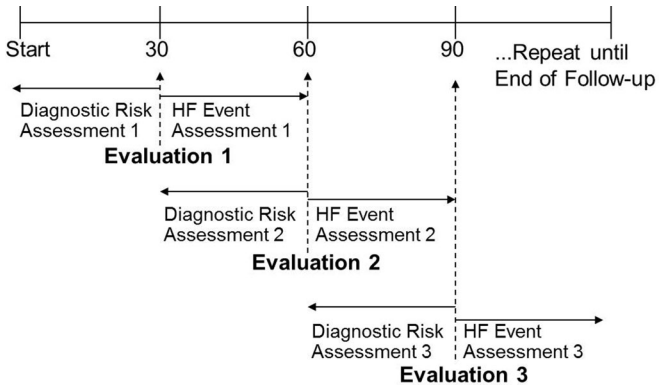

Fig. 2. Performance evaluation using a monthly evaluation scheme for monitoring patients with HF and AF.

III. Results

A. Feature Set

The selected feature set for the different parameters is detailed in Table I. The parameter set was chosen to have around 5% of the evaluations meeting the very high (VH) evidence state criteria with an event rate $> 5\%$. IMP met the criterion for a VH evidence state if the cumulative sum of the difference between adaptive reference and IMP in the last 7 days ($CSAR_7$) is > 45 Ω-days (Thr1). The high-evidence state chose the next 5–10% of the riskier evaluations. AF diagnostics met the high-evidence state criteria if there is at least 1 day with AF burden (AFB) $>$ 6 h and VRAF > 90 bpm on the same day in the last 30 days. The low-evidence group was chosen to include around 50% of the evaluations deemed to have low risk in terms of having event rates $< 2\%$. The higher risk evidence state is given priority if criteria are met for multiple evidence states in Table I.

The HFH event rates for the different evidence states are shown in Table I. The comparison of VH to low evidence for each diagnostic variable has hazard ratios of 6.1, 4.4, 3.8, 5.4, and 3.9 for IMP, AF, NHR, HRV, and ACT, respectively ($p <$ 0.001 for all). Each of the diagnostic parameters has the ability to identify when patients are at risk for HFH.

The number of times two diagnostic variables had the same evidence state ranged from 5.9% (AF and ACT) to 16.1% (NHR and HRV) of all monthly evaluations when both evidence states were not low. A stepwise multivariate Anderson–Gill model showed that IMP, HRV, AF, and ACT all provided independent information ($p < 0.05$ for all). NHR did not provide independent information but was forced in the model as it is a vital sign with established treatment guidelines that physicians are familiar with in clinical practice.

B. Combined Risk Score Performance

The combined probability score is divided into four categories as shown in Fig. 1 and detailed in Table II. The performance results for the four groups in identifying patients at risk for a HFH within the next 30 days at monthly follow-up are shown in Table II. The VH risk evidence group consisted of evaluations with around the top 10% of risk scores. The next 10% consisted of the high-evidence group. The VH risk status is more likely to reflect evidence in last 7 days whereas the high-risk status indicates evidence in last 30 days and may have different therapeutic implications. The combined low-evidence group consisted of cases where each of the diagnostic parameters had low evidence. The results indicate that the patient with VH risk evidence at a follow-up is 15 times more likely to be hospitalized for HF in the next 30 days compared to patients with low-risk evidence at a follow-up. In comparison, the individual diagnostic parameters had a hazard ratio in the range from 3.8–6.1 for the comparison between VH and low-diagnostic evidence. Thus, combining the diagnostic information improves the ability to identify when the patients are at increased risk for HFH. Due to the dynamic nature of the probability score the same patient may have VH evidence at one follow-up and low evidence at another follow-up. The K–M plot showing time to HFH in the next 30 days for the risk groups derived from the combined risk score is shown in Fig. 3.

The receiver operating characteristic (ROC) curve, plotting the GEE estimates of sensitivity and specificity is shown in Fig. 4. The specificity increases and sensitivity decreases as the threshold for risk score value increases (going from low to V.High evidence category in Table II).

TABLE I
Feature Set for Each Diagnostic Parameter

Diagnostic Evidence		Feature Set	Event Rate
IMP	VH	$CSAR_7 \geq$ Thr1	7.5%
	H	$CSAR_{30} \geq$ Thr2	4.0%
	M	$CSAR_7 \geq$ Thr3 $\|\|$ $CSAR_{30} \geq$ Thr4	2.6%
	L	~('VH' $\|\|$ 'H' $\|\|$ 'M')	1.3%
AF	VH	ND(AFB_7>6hrs & $VRAF_7$>90bpm) $\geq$1	7.4%
	H	ND(AFB_{30}>6hrs & $VRAF_{30}$>90bpm)$\geq$1	3.7%
	M	ND(AFB_{30}>23hrs) < 30 & ND(AFB_{30}>6hrs) $\geq$1	3.0%
	L	~('VH' $\|\|$ 'H' $\|\|$ 'M')	1.8%
NHR	VH	ND($NHR_7\geq$85bpm) $\geq$5 $\|\|$ ND($CSAR_7\geq$Thr)=7	6.3%
	H	ND($NHR_{30}\geq$85bpm) $\geq$10 $\|\|$ $CSAR_7 \geq$ Thr	3.1%
	M	ND($NHR_{30}\geq$80bpm) $\geq$7 $\|\|$ $CSAR_{30} \geq$ Thr $\|\|$ $CSFR_7 \geq$ Thr	2.3%
	L	~('VH' $\|\|$ 'H' $\|\|$ 'M')	1.7%
HRV	VH	ND($HRV_7\leq$60ms) = 7	6.5%
	H	ND($HRV_{30}\leq$50ms) $\geq$14	4.3%
	M	ND($HRV_{30}\leq$60ms) $\geq$10 $\|\|$ $CSAR_{30} \geq$ Thr $\|\|$ ND($HRV_7\leq$60ms) $\geq$3 $\|\|$ $CSFR_7 \geq$ Thr	2.6%
	L	~('VH' $\|\|$ 'H' $\|\|$ 'M')	1.2%
ACT	VH	ND($ACT_7\leq$30min) = 7 $\|\|$ $CSAR_7 \geq$ Thr	5.7%
	H	ND($ACT_{30}\leq$30min) $\geq$10	4.3%
	M	ND($ACT_{30}\leq$60min) >10 $\|\|$ $CSAR_{30} \geq$ Thr $\|\|$ $CSAR_7 \geq$ Thr	2.4%
	L	~('VH' $\|\|$ 'H' $\|\|$ 'M')	1.5%

VH = Very High, H = High, M = Medium; L = Low, ND = Number of days, CSAR= Cumulative Sum Adaptive Reference, CSFR = Cumulative Sum Fixed Reference, Thr = Threshold; Subscripts indicate look-back window size

TABLE II
Performance Results Using the Anderson–Gill Model

Risk Group	Risk Score Ranges	Monthly followup	HFH Rate	Hazard Ratio (95% CI)
V.High	≥0.25	9.4%	7.5%	15.1 (8.1-28.4)*
High	[0.14-0.25)	9.8%	4.9%	9.6 (4.9-18.8)*
Med	[0.04-0.14)	42.2%	1.7%	3.4 (1.9-6.2)*
Low	<0.04	38.6%	0.5%	Reference

* p-value < 0.001

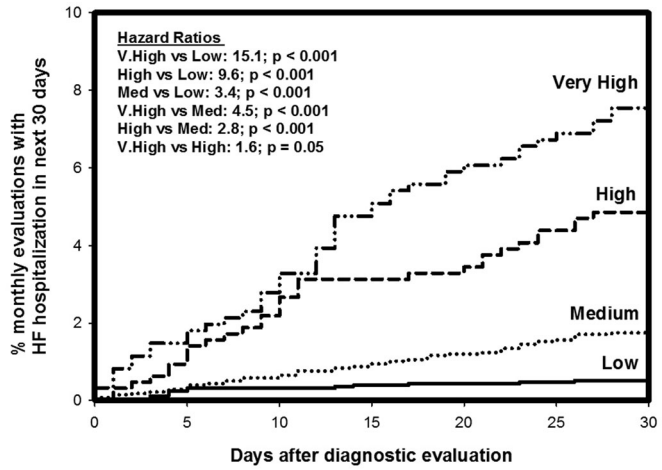

Fig. 3. Kaplan–Meier curves showing time to HFH for the different risk score groups.

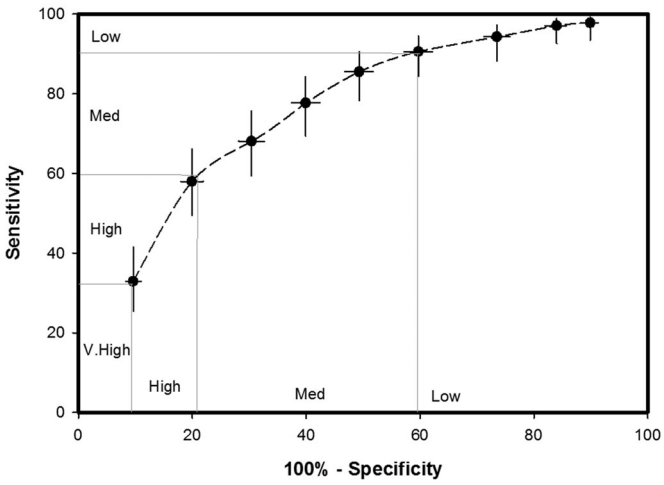

Fig. 4. ROC curve for GEE estimates of sensitivity and specificity. The error bars represent the 95% confidence interval of the GEE estimate. The thresholds for the various risk groups are indicted in the ROC curve.

IV. Discussion

The risk score derived using a combination of HF device diagnostics available in implantable devices was able to identify when a patient is at a higher risk of a HF event. Presently, these devices wirelessly transmit data at regular intervals (e.g., monthly) to a remote monitoring system where the information can be combined to generate the risk score. Clinicians can use the risk score as a triage tool to manage patients by exception. A high or VH risk score would indicate that a patient needs more attention, a low-risk score would indicate the patient needs less attention. For example, the contributing factors for a VH risk score can be IMP and AF and high VRAF. IMP corroborated by signs and symptoms may suggest volume overload, which would suggest fluid management using diuretic therapy in the short term. Once fluid status is stabilized in a patient, then AF and high VRAF can be treated with rate control medications such as β-blockers to prevent future HF events caused by AF and high VRAF in that patient. Thus, the risk score and the contributing factors for the risk will provide the opportunity to clinicians to modify, personalize, and optimize therapeutic options for HF patients in a timely proactive manner with the intention to improve clinical outcomes such as reduction in HFH events.

The BBN framework also provides the capability of adding more variables assuming that the measurements provide orthogonal information. Additional clinical variables, such as B-type natriuretic peptide, measures of renal dysfunction, blood pressure, and respiration rate, which are known to identify HF, may be added to the model.

The data presented in this paper were used to develop the BBN model. The performance of the model needs further evaluation using an independent validation dataset. Prospective clinical trials are being performed to validate the performance of the risk score.

References

[1] V. L. Roger *et al.*, "Heart disease and stroke statistics–2011 update: A report from the American heart association," *Circulation*, vol. 123, no. 4, pp. e18–e209, Feb. 2011.

[2] S. A. Hunt *et al.*, "2009 focused update incorporated into the ACC/AHA 2005 guidelines for the diagnosis and management of heart failure in adults," *Circulation*, vol. 119, no. 14, pp. e391–e479, Apr. 2009.

[3] C. M. Yu *et al.*, "Intrathoracic impedance monitoring in patients with heart failure: Correlation with fluid status and feasibility of early warning preceding hospitalization," *Circulation*, vol. 112, no. 6, pp. 841–848, 2005.

[4] S. Sarkar *et al.*, "Burden of atrial fibrillation and poor rate control detected by continuous monitoring via implanted devices identifies when a patient is at risk for heart failure hospitalization," *J. Am. Coll. Cardiol.*, vol. 57, p. E107, Apr. 2011.

[5] P. B. Adamson *et al.*, "Continuous autonomic assessment in patients with symptomatic heart failure: Prognostic value of heart rate variability measured by an implanted cardiac resynchronization device," *Circulation*, vol. 110, no. 16, pp. 2389–2394, Oct. 2004.

[6] G. H. Crossley *et al.*, "The CONNECT (clinical evaluation of remote notification to reduce time to clinical decision) trial: The value of wireless remote monitoring with automatic clinician alerts," *J. Am. Coll. Cardiol.*, vol. 57, no. 10, pp. 1181–1189, Mar. 2011.

[7] D. J. Whellan *et al.*, "Combined heart failure device diagnostics identify patients at higher risk of subsequent heart failure hospitalizations: Results from PARTNERS HF study," *J. Am. Coll. Cardiol.*, vol. 55, pp. 1803–1810, 2010.

[8] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufmann, 1988.

[9] F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*, 2nd ed. New York: Springer, 2007.

[10] W. Long, "Temporal reasoning for diagnosis in a causal probabilistic knowledge base," *Artif. Intell. Med.*, vol. 8, pp. 193–215, 1996.

[11] P. J. F. Lucas *et al.*, "Bayesian networks in biomedicine and health-care," *Artif. Intell. Med.*, vol. 30, pp. 201–214, 2004.

[12] H. Purerfellner *et al.*, "Accuracy of atrial tachyarrhythmia detection in implantable devices with arrhythmia therapies," *Pacing Clin. Electrophysiol.*, vol. 27, no. 7, pp. 983–992, Jul. 2004.

[13] S. Sarkar *et al.*, "A detector for a chronic implantable atrial tachyarrhythmia monitor," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1219–1224, Mar. 2008.