MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-40

Using EHRs and Machine Learning for Heart Failure Survival Analysis

Maryam Panahiazar[a], Vahid Taslimitehrani[b], Naveen Pereira[c], Jyotishman Pathak[a]

[a] *Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, USA*
[b] *Department of Computer Science and Engineering, Kno.e.sis Center, Wright State University, Dayton, OH, USA*
[c] *Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, USA*

Abstract

"Heart failure (HF) is a frequent health problem with high morbidity and mortality, increasing prevalence and escalating healthcare costs" [1]. By calculating a HF survival risk score based on patient-specific characteristics from Electronic Health Records (EHRs), we can identify high-risk patients and apply individualized treatment and healthy living choices to potentially reduce their mortality risk. The Seattle Heart Failure Model (SHFM) is one of the most popular models to calculate HF survival risk that uses multiple clinical variables to predict HF prognosis and also incorporates impact of HF therapy on patient outcomes. Although the SHFM has been validated across multiple cohorts [1-5], these studies were primarily done using clinical trials databases that do not reflect routine clinical care in the community. Further, the impact of contemporary therapeutic interventions, such as beta-blockers or defibrillators, was incorporated in SHFM by extrapolation from external trials. In this study, we assess the performance of SHFM using EHRs at Mayo Clinic, and sought to develop a risk prediction model using machine learning techiniques that applies routine clinical care data. Our results shows the models which were built using EHR data are more accurate (11% improvement in AUC) with the convenience of being more readily applicable in routine clinical care. Furthermore, we demonstrate that new predictive markers (such as co-morbidities) when incorporated into our models improve prognostic performance significantly (8% improvement in AUC).

Keywords:

Heart Failure; Survival Score; Electronic Health Records; Machine Learning.

Introduction

Heart failure (HF) is primarily caused by the inability of the heart to supply sufficient blood flow to the body. It has become one of the most deadly cardiovascular diseases in the 21st century [1]. Therefore, it is important to identify patients who are at a higher risk of mortality due to HF and assess the impact of HF therapy on their outcomes. Several studies have developed prognostic tools for HF, and one of the most commonly used tools is the Seattle Heart Failure Model (SHFM) [1]. SHFM was based on the PRAISE I clinical trial database and validated in five other cohorts.

While the derivation and validation of SHFM using such clinical trials databases provide a level of rigor in terms of data collected, they are typically limited to biased cohorts that tend to be homogenous. Further, in many cases, the environment in which the trials were conducted may not reflect the routine clinical care given to patients diagnosed with HF. Consequently, as EHRs become more ubiquitous and accessible for clinical research, it becomes imperative to investigate methods of predicting HF prognosis and the impact of HF therapy on important patient-related outcomes using EHRs–as opposed to data derived exclusively from clinical trials databases. The specific objectives of this study are to assess the performance of the SHFM using routinely collected EHR data in a community practice at Mayo Clinic, and to incorporate variables that were not part of the SHFM in our prognostic model (e.g., patient co-morbidities derived from the EHR) to assess improvement in the performance of survival analysis.

Our results suggest that (1) heart failure survival models built on EHRs are more accurate than the SHFM, (2) incorporating co-morbidities into the heart failure survival analysis prediction models improve the accuracy of our models, and (3) there are potential hidden interactions between diagnoses history of the patient, co-morbidities, and survival risk. We also build our models using multiple different machine learning algorithms and our results show that logistic regression and random forest return more accurate classifiers.

Background and Related Work

The SHFM was derived in a cohort of 1125 heart failure patients from the PRAISE I clinical trial with the use of a multivariate Cox model [1]. For variables such as medications (e.g., beta-blockers) and devices (e.g., defibrillators) that were not available in the derivation database, hazard ratios were estimated from published literature and "external" clinical trials. The model has been prospectively validated in 5 additional cohorts totaling 9942 heart failure patients and 17307 person-years of follow-up.

However, the SHFM has significant limitations for risk prediction in HF, particularly when used for routine clinical care. For instance, the hazard ratios for a subset of medications and devices variables in SHFM were estimated from prior published literature, and results from prior clinical trials may not be generalizable to a wider real-world population of HF patients. Limited patient-specific parameters (e.g., patient co-morbidities) have been used in the model to calculate survival score and can potentially lead to an improvement in the prediction of HF prognosis.

In addition to SHFM, several other risk prediction models have been developed including SHOCKED, Frankenstein, PACE Risk Score, and HFSS [1]. These have been validated in independent cohorts along with SHFM: "The Heart Failure Survival Score (HFSS) was validated in 8 cohorts (2240 patients), showing poor-to-modest discrimination (c-statistic, 0.56–0.79), being lower in more recent cohorts. The Seattle Heart Failure Model was validated in 14 cohorts (16,057 patients), describing poor-to-acceptable discrimination (0.63–0.81), remaining relatively stable over time. Both models re-

ported adequate calibration, although overestimating survival in specific populations. The other 3 models were validated in a cohort each, reporting poor-to-modest discrimination (0.66–0.74) [1-5].

Furthermore, there are also studies that applied machine-learning algorithms to study risk factors and predict patient outcomes in HF. For example, Dai et al. [6] used boosting and support vector machine (SVM) schemes to build models to predict heart failure around six months before the actual diagnosis. Their results show that SVM has poor performance. Similarly, Austin et al. [7] used regression tree, bagging, Random Forest, boosting, SVM and logistic regression to classify HF patients with preserved Ejection Fraction (EF) from those patients with reduced EF. They concluded that logistic regression returns the most accurate models.

Methods

From a cohort of 119,749 Mayo Clinic patients between 1993-2013 with research authorization to access EHR data, we identified 5044 patients with a diagnosis of HF after applying specific criteria and excluding number of patients due to incomplete data (13.3%). These criteria included:

- A confirmed diagnosis of HF based on the ICD-9-CM code (428.x).
- An ejection fraction (EF) measurement ≤50% within two months of HF diagnoses.
- No prior diagnosis of coronary artery disease, myocarditis, infiltrative cardiomyopathy, and severe valvular disease.

We divided the EHR-derived dataset randomly into training (N=1560 patients) and test (N=3484 patients) datasets. In consultation with Mayo Clinic cardiologists who routinely treat HF patients, we identified the following features (variables) extracted from EHRs to calculate the survival score:

- Demographic variables including age, sex, race, ethnicity and survival status.
- Laboratory results including cholesterol, sodium, hemoglobin, lymphocyte count, and EF measurements.
- Medications including Angiotensin Converting Enzyme (ACE) inhibitors, Angiotensin Receptor Blockers (ARBs), β-adrenoceptor antagonists (β-blockers), Statins, and Calcium Channel Blocker (CCB).
- 26 major chronic conditions (ICD-9 code) as co-morbidities as defined by the U.S. Department of Health and Human Services [8].

Table 1 represents the characteristics of the study cohort with including training and test splits. The patients were 94% white 48% female. The average range of EF was 36±10.3. In terms of co-morbidities most of the patients (81.06%) had hypertension followed by hyperlipidemia (64.3%), chronic kidney disease (55.83%) and diabetes (37.4%).

As discussed earlier, our primary goal in this study is to assess the performance of SHFM using EHR data and propose a HF survival risk prediction model by adding new variables (e.g., patient co-morbidities) derived using the EHR data to improve prediction accuracy and performance of the model. To accomplish this goal, we designed two scenarios: In scenario A, we used COX proportional regression model [9] to predict the risk of survival in HF patients in the one, two and five years after the diagnosis of HF. In scenario B, we excluded all patients who died within one year after the first diagnosis of HF, and then based on the remainder of patients who survived after one year of HF diagnosis, we developed a series of models using different classifiers to classify these two groups of patients. Since most of the well known classification algorithms are developed for binary classification, we repeated scenario B to classify patients who died within two years and five years after the HF diagnosis. The following classification algorithms were used: random forest [10], logistic regression [11,12], support vector regression [13], decision tree [14] and ada boost [15].

Table 1 – Patient Characteristics for HF Study Cohort

	Variables	Value
Demographic	Age (years)	78±10
	Sex (male)	52%
	Race (White)	94%
	Ethnicity (Not Hispanic or Latino)	84%
Laboratory	BMI	28.7±11.25
	Systolic Blood Pressure (mm/Hg)	120±25
	Ejection Fraction (EF)	36±10.3
	Hemoglobin (g/dL)	11.8±1.2
	Cholesterol (mg/dL)	144±35
	Uric Acid (g/dL)	7.1±2.5
	Sodium (mEq/L)	128±4.2
	Lymphocytes (x10(9)/L)	1.32±0.7
Medications	ACE inhibitors	55.7%
	Beta blockers	48.6%
	Angiotension Receptor Blockers	12.8%
	Calcium Channel Blockers	4.1%
	Statins	43.2%
	Diuretics	68.7%
	Allopurinol	18.5%
	Aldosterone Blockers	18.5%
Comorbidities	Hypothyroidism	21.2%
	Acute myocardial infarction	16.3%
	Alzheimers	11.9%
	Anemia	5 3.01%
	Asthma	10.72%
	Atrial fibrillation	48.56%
	Benign prostatic hyperplasia	9.5%
	Cataract	31.4%
	Chronic Kidney Disease	55.83%
	Pulmonary disease	30.4%
	Depression	25.5%
	Diabetes	37.4%
	Glaucoma	9.4%
	Hip/pelvic fracture	4.3%
	Hyperlipidemia	64.3%
	Hypertension	81.06%
	Ischemic heart disease	70.2%
	Osteoporosis	18.3%
	Rheumatoid Arthritis	39.2%
	Stroke	12.4%
	Breast cancer	2.2%
	Colorectal cancer	1.58%
	Prostate cancer	4.5%
	Lung cancer	2.45%
	Endometrial cancer	0.00%

To investigate the effect of variables extracted from the EHR data on the performance of our models, we designed two sets of predictor variables. In the first set called *baseline* (*BL*), we applied the same variables used in the SHFM. Since our EHR

derived dataset does not have information about patients' NYHA class, QRS duration, or device implantations (e.g., defibrillators), we did not include them in our models. In the next variable set called *extended (EX)*, we added the following predictor variables to the *BL* model: race, ethnicity, BMI, calcium channel blocker (CCB) and 26 different co-morbidities. Then we compared the performance of our *BL* and *EX* models with the SHFM.

Results

This section reports a systematic validation of our HF survival risk prediction model(s) in both scenarios A and B. As we mentioned earlier, to minimize the effect of overfitting and increase generalizability of our models, we separated our cohort randomly into training (N=1560 patients) and test (N=3484 patients) datasets. To validate models which are developed in scenario A (survival models), we designed two approaches. In the first approach (A1), we divided our predictor variables into two parts: variables that were common between both the SHFM model and our models (e.g., EF measurement), and variables used by just our model (e.g., patient co-morbidities). For variables that were common we used the hazard ratios defined by SHFM. For variables that were not used in SHFM we used the hazard ratios extracted from our analysis. Table 2 presents the performance of our model (A1) in the form of Area Under Curve (AUC). In the second approach (A2) we used the hazard ratio developed by our model to calculate the AUCs (also shown in Table 2).

Table 2 shows that accuracy drops on average by 9% in the AUC when using a mixture of hazard ratios from our model and SHFM. Although it is not surprising, one plausible reason might be that SHFM was developed using a clinical trials database, whereas our model has been developed using an EHR database in a real community practice. Consequently, combining hazard ratios derived from two different datasets can lead to a degradation in the accuracy of the prediction model.

Table 2 – Perfomance (Area Under Curve) of Scenario A (Survival Analysis) for both baseline and extended variable sets

	1 year		**2 years**		**5 years**	
	BL	EX	BL	EX	BL	EX
A1	0.71	0.80	0.68	0.75	0.66	0.71
A2	0.77	0.83	0.77	0.82	0.75	0.80

Figure 1 also represents the Receiver Operating Characteristics (ROC) of the models developed in scenario A on the BL variable set. From Table 2 and Figure 1, we can observe that the performance of our models drops when we want to predict the risk of heart failure two and five years after the first heart failure event. We hypothesize that the main reason of observing a drop in the accuracy is because the dataset for two and five year models are imbalanced and usually the classifiers used in this study do not work well on imbalanced datasets. Figure 2 shows the ROC curve of the scenario A models on EX variable set.

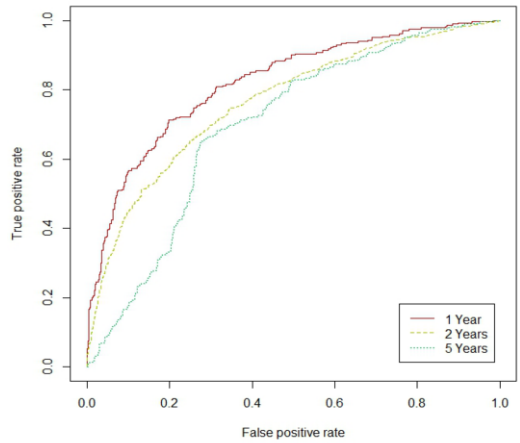


Figure 1 - ROC curve for Scenario A on baseline variable sets

As we discussed earlier, one of our goals in this study was to investigate the effect of adding more predictor variables to our models, and potentially improve model performance. To this end, we designed two variable sets called *BL* and *EX*. In the *BL* set, we have 16 predictor variables and in the *EX* set, we consider 45 variables to develop our models. Table 2 shows that the AUC for models developed using the *EX* variable set increased by 7.7%, 6.5% and 6.6% compared to the *BL* variable sets for 1-, 2- and 5-years models, respectively. In the next part of our results, we show the performance of the classification models (scenario B). In this scenario, we excluded all patients who died within 1-, 2- and 5-years after the first diagnosis of HF, and then developed models to classify them separately from patients who did survive after the HF diagnosis.

Table 3 – Perfomance of Senario B (Classification) for both baseline and extended variable sets

	1 year		2 years		5 years	
	BL	EX	BL	EX	BL	EX
Decision Tree	0.60	0.66	0.50	0.50	0.50	0.50
Random Forest	0.62	**0.80**	0.65	0.72	0.62	0.72
Ada Boost	0.59	0.74	0.66	0.71	0.61	0.68
SVM	0.56	0.46	0.61	0.52	0.55	0.38
Logistic Regression	**0.68**	**0.81**	0.7	0.74	0.61	0.73

Table 3 represents the AUC of different classifiers for both BL and EX variable sets. Much like scenario A, we make the same observation in scenario B where inclusion of additional variables derived from the EHR to the models significantly improves classifier performance.

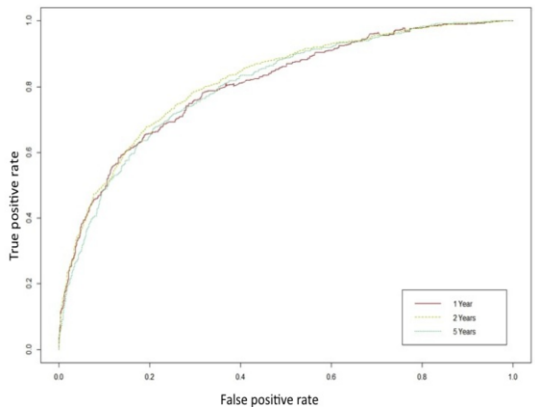
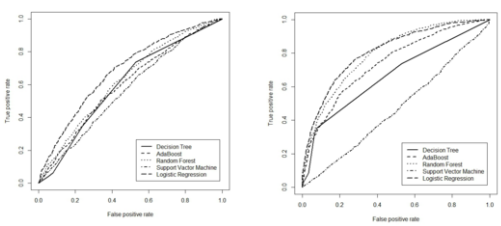
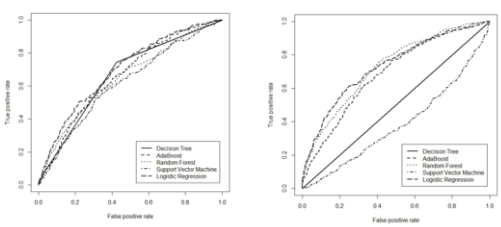


Figure 2 - ROC curve for Scenario A on exdended variable sets

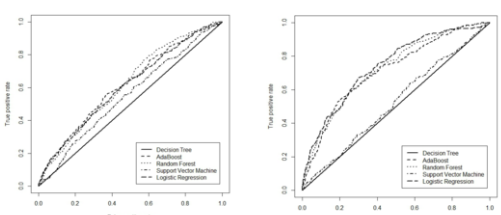
Figure 3 also represents the ROC of the models developed in scenario B on the BL and EX variable sets for one, two, and five years for all classifiers shown in Table 2, including Decision Tree, Random Forest, AdaBoost, SVM, and Logistic Regression. There are a number of reasons why SVM may have been less accurate in developing this prediction model. First, SVM is not an appropriate method for handling both continuous and categorical variables in the same model. Second, our data suggest that SVM may be more strongly affected by classification imbalance in the data than either Boosting or logistic regression.



BL model, 1 year *EX model, 1 year*



BL model, 2 years *EX model, 2 years*



BL model, 5 years *EX model, 5 years*

Figure 3 - ROC curve for Scenario B on BL & EX variable sets(x=False positive rate, y=True positive rate)

Discussion

In this study we explored how to improve SHFM by considering routine clinical care data. Since models that are built based on EHRs are more accurate (11% improvement in AUC) and are applicable in standard routine care, it is imperative to leverage EHR data for survival analysis and prediction modeling in HF and other chronic conditions. We also showed that incorporating new predictive markers (co-morbidities) in our models improved the performance significantly (8% improvement in AUC) and gives us insights about the pathophysiology of HF. Another highlight of this study is calculating the Hazard Ratio (HR) based on real-world EHR data, whereas other studies, including SHFM, have used the HR from literature and extrapolation of results from clinical trials which may not reflect routine care for HF patients [1,5]. Finally, we observe that there are potential hidden interactions between diagnoses, history of the patient, co-morbidities, and survival risk that warrant further research. Note that since we calculate the model output based on individual patient chracteristics, it is plausible to incorporate the output derived from these predictive models within EHRs and facilitate clinical decision making for managing HF patients with better treatment options–an area for future research. In summary, our results suggest that heart failure survival models built on EHRs are more accurate, and incorporating co-morbidities into the HF models significantly improves the accuracy of our models.

Conclusion

In this study we assessed the performance of SHFM using Mayo Clinic's EHR dataset. Our results demonstrate an improvement in accuracy as compared to the standard SHFM and also suggest the ready applicability of our model to standard clincal care in the community. We also incorporated additional predictor variables that included 26 co-morbidities into our models that lead to further improvement in the prognostic predictive accuracy. Finally, we built a heart failure risk prediction model using a series of machine learning techniques and observed that logistic regression and random forest return more accurate models compared to other classifiers.

Acknowledgments

This project was funded in part by support from AHRQ (R01 HS023077).

References

[1] Alba AC, Agoritsas T, Jankowski M, et al. Risk Prediction Models for Mortality in Ambulatory Patients with Heart Failure: A Systematic Review. Circulation. Heart failure, vol. 6, no. 5, pp. 881–9, Sep. 2013.

[2] Cabassi A, Champlain J, Maggiore U, et al. Prealbumin. Improves Death Risk Prediction of BNP-added Seattle Heart Failure Model: Results from a Pilot Study in Elderly Chronic Heart Failure Patients. International Journal of Cardiology, vol. 168, no. 4, pp. 3334–9, Oct. 2013.

[3] Hussain S, Kayani AM, Munir R, et al. Validation of the Seattle Heart Failure Model (SHFM) in Heart Failure Population. Journal of the College of Physicians and Surgeons–Pakistan: JCPSP, vol. 24, no. 3, pp. 153–6, Mar. 2014.

[4] Levy WC, Mozaffarian D, Linker DT, et al. The Seattle

Heart Failure Model: Prediction of Survival in Heart Failure. Circulation, vol. 113, no. 11, pp. 1424–33, Mar. 2006.

[5] Prasad H, Sra J, Levy AC, and Stapleton DD. Influence of Predictive Modeling in Implementing Optimal Heart Failure Therapy. The American Journal of the Medical Sciences, vol. 341, no. 3, pp. 185–90, Mar. 2011.

[6] Dai W, Brisimi T S, Adams W G, Mela T, Saligrama V, and Paschalidis L C. Prediction of hospitalization due to heart diseases by supervised learning methods. International Journal of Medical Informatics. 2014.

[7] Austin P C, Tu J V, Ho J E, Levy D, and Lee D S. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. Journal of Clinical Epidemiology, 66.4: 398-407, 2013.

[8] U.S. Department of Health & Human Services http://www.hhs.gov/ash/initiatives/mcc/

[9] Fox J. Cox Proportional-Hazards Regression for Survival Data, Appendix to An R and S-PLUS Companion to Applied Regression, February, 2002.

[10] Liaw A, and Wiener M, Classification and Regression by Random Forest. R news 2.3: 18-22, 2002.

[11]Hosmer D, Lemeshow S, and Sturdivant RX. Introduction to the logistic regression model. John Wiley & Sons, Inc., 2000.

[12]Taslimitehrani V, Dong G. A new clinical prediction method using contrast pattern aided logistic regression with application on traumatic brain injury. In IEEE International Conference on BioInformatics and BioEngineering (BIBE), Nov 2014. Best Student Paper Award.

[13]Furey TS, et al. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 16.10: 906-914, 2000.

[14]Safavian SR, and Landgrebe D. A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics 21.3 1991, 660-674, 2000.

[15]Collins M, Schapire RE, and Singer Y. Logistic regression, AdaBoost and Bregman distances. Machine Learning 48.1-3, 253-285, 2002.