Exploring Preprocessing Techniques for Prediction of Risk of Readmission for Congestive Heart Failure Patients

Naren Meadem
Inst. of Technology, CWDS,
Univ. of Washington Tacoma
mnaren@uw.edu

Nele Verbiest
Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium
nele.verbiest@ugent.be

Kiyana Zolfaghar
Inst. of Technology, CWDS,
Univ. of Washington Tacoma
kiyana@u.washington.edu

Jayshree Agarwal
Inst. of Technology, CWDS,
Univ. of Washington Tacoma
jagarwal@u.washington.edu

Si-Chi Chin
Inst. of Technology, CWDS,
Univ. of Washington Tacoma
scchin@u.washington.edu

Senjuti Basu Roy
Inst. of Technology, CWDS,
Univ. of Washington Tacoma
senjutib@u.washington.edu

ABSTRACT

Congestive Heart Failure (CHF) is one of the leading causes of hospitalization, and studies show that many of these admissions are readmissions within a short window of time. Identifying CHF patients who are at a greater risk of hospitalization can guide the implementation of appropriate plans to prevent these readmissions. Developing predictive modeling solutions for such disease related risk of readmissions is extremely challenging in healthcare informatics. It involves integration of socio-demographic factors, health conditions, disease parameters, hospital care quality parameters, and a variety of variables specific to health care providers making the task immensely complex. This work, in collaboration with experts from Multicare Health Systems (MHS), describes a soon to be deployed prototype to predict risk of readmission within 30 days of discharge for CHF patients at MHS. We focus on data extraction and data preprocessing steps to improve prediction outcomes, including feature selection, missing value imputation and data balancing. We perform comprehensive empirical evaluations using the real-world health care data set provided by MHS. Our empirical evaluation demonstrates that we outperform one of the nearest competing previous results.

General Terms

Algorithms, Design, Experimentation

Keywords

Healthcare, Knowledge Discovery, Risk Prediction

1. INTRODUCTION

KDD-DMH'13, August 11, 2013, Chicago, Illinois, USA.

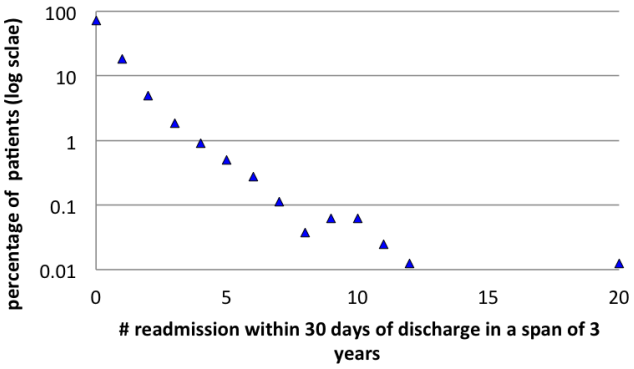

Figure 1: Y-axis shows the percentage of patients (in log scale) and X-axis shows the number of times patients have 30 day readmissions in a span of 3 years.

Congestive Heart Failure (CHF) has been identified as one of the leading causes of hospitalization, especially for adults older than 65 years of age [1]. Furthermore, studies show that CHF is one of the primary reasons behind readmission within a short time-span [17]. Based on the 2005 data of Medicare beneficiaries, it has been estimated that 12.5% of Medicare admissions due to CHF were followed by readmission within 15 days, accounting for about $590 million in health care costs [14]. The Center for Medicare and Medicaid Services (CMS) has started using the 30 day all cause heart failure readmission rate as a publicly reported efficiency metric. All cause 30 day readmission rates for patients with CHF have increased by 11% between 1992 and 2001 [14].

A variety of reasons could lead to readmissions - early discharge of patients, improper discharge planning, and poor care transitions, to name a few. Studies have shown that targeted interventions before or after discharge can reduce the probability of readmission, especially in elderly patients, and decrease the overall medical costs [4]. Proper pre-discharge planning [9] and post-discharge plans like home based follow up [8] and patient education [13, 11] can also reduce the readmission rates considerably and improve the health outcome of the patients.

Therefore, during the initial hospitalization (either during admission or discharge) of a patient, if her Risk of Readmission (RoR) within a given time frame (such as, within 30 days or 60 days) could be calculated, that may in turn lead to developing improved post-discharge planning for the patient. Furthermore, such insights may guide healthcare providers to develop programs to improve the quality of care and administer targeted interventions - thus reducing the readmission rate and the cost incurred in these readmissions. This can also facilitate proper resource utilization by the hospitals.

While actionable insights [12, 10] could be gathered by predicting the RoR, the task itself is very complex. First and foremost, one has to understand *the domain-specific factors or attributes* that cause readmissions. For example, the cohort identification of RoR could be based on coded ICD9 diagnoses[1] and clinical measures such as ejection fraction. Furthermore, the dataset is noisy, inconsistent, skewed, and has a significant amount of missing values. As an example, we observe that 63% of the records in our given dataset have no value for the important attribute ejection fraction.

Not many solutions [14, 15] are known to be effective. In fact, health care organizations still leverage the proven best-practices called *Get With The Guidelines* written by the American Heart Association to improve the clinical process of CHF patients.

Our primary contributions in this work can be summarized as follows:

- Partnering with Multicare Health Systems (MHS) (a leading health care provider in the state of Washington), we embark on the task of identifying CHF patients who are likely to get readmitted within 30 days of discharge for any cause[2]. We formalize the problem and study the attributes pertinent to cause readmission for CHF patients[3].
- We overcome the three main factors that make the data for the RoR prediction task complex: We propose attribute selection to deal with the high-dimensionality of the data, we use data imputation to overcome the missing value problem and use class balancing techniques to solve the problem of skewed data.
- We perform an extensive experimental study using a real world dataset (provided by MHS) that demonstrates that our proposed method outperforms our nearest competitor [14].

While our discussions primarily focus upon CHF, the methods and issues we outline are generic and applicable to various other diseases as well.

The remainder of the paper is organized as follows: in Section 2, we discuss our approach to perform RoR prediction for CHF patients. Section 3 summarizes the experimental evaluations. Related work is studied in Section 4. Finally, we conclude in Section 5 and propose future research directions.

[1]ICD-9 is the International Classification of Diseases, 9th Revision, Clinical Modification Codes.

[2]30 days is chosen as the readmission window, because it is a clinically meaningful time-frame for hospitals and medical communities to take action to reduce the probability of readmission [14]. Furthermore, 30 day readmission rates are also used by CMS as a potential efficiency measure for hospitals [14]. All cause readmission is considered as a publicly reported efficiency metric by CMS.

[3]Our developed prototype *Risk-O-Meter* has been accepted to KDD 2013, demo track.

2. PROPOSED METHOD

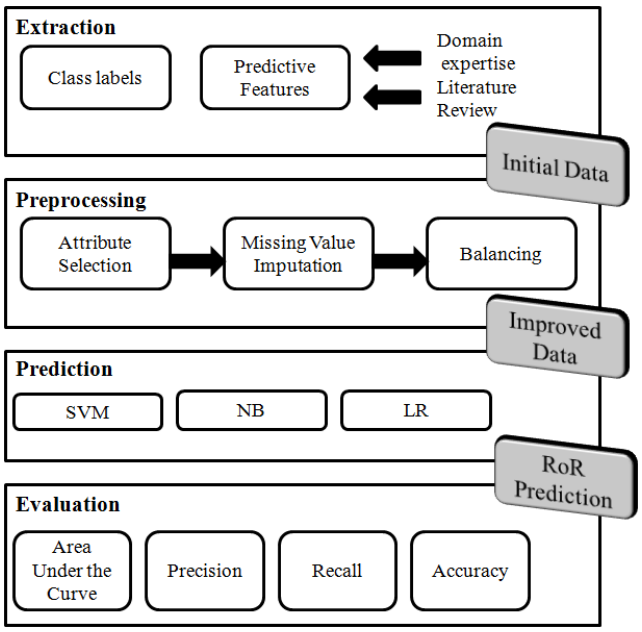

Figure 2: The overall architecture for the RoR prediction process.

In this section we discuss our methodology to deal with the RoR within 30 days of discharge prediction problem. Formally, the problem is formulated as a supervised learning problem, more specifically as a binary classification task. The class of a patient is *Readmission* if the elapsed period between the last discharge and next admission is smaller or equal to 30, and *No Readmission* else.

The overall framework is depicted in Figure 2. We extract a dataset that describes meaningful features of previously seen patients and their class labels. Next, we improve the predictive ability of the dataset by applying attribute selection, missing value imputation and class balancing techniques. Then, we apply predictive models to this improved dataset to obtain RoR predictions. The outcomes are evaluated with respect to several evaluation measures. In the following we describe the aspects of the overall process in more detail.

2.1 Data Extraction, Integration and Exploration

Getting familiar with the data is overwhelming challenging, albeit extremely important for this problem, where the challenges are unique and very specific to the domain. Real world clinical data is noisy and heterogeneous in nature, severely skewed, and contains hundreds of pertinent attributes. It contains information on patients' socio-demographical characteristics, marital status, ethnicity, diagnosis, discharge information, comorbidity factors[4], other cost related factors pertaining to a particular hospital admission and many more.

[4]Comorbidities are specific patient conditions that are secondary

After extraction, we resort to a systematic and comprehensive data exploration process through visualization to gather our first hand knowledge on the domain and identify the pertinent factors correlated to CHF related admissions[5].

2.1.1 Data Selection

Hospital encounters of patients with discharge diagnosis of CHF (either primary or secondary) are identified as the potential index for CHF related admissions. We consider patients with a discharge diagnosis of ICD9-CM[6] for this purpose, as listed in Table 1. Our entity of observation is each CHF hospital encounter[7]. Once all the admission instances related to CHF have been identified, we exclude the admissions with in-hospital deaths from the analyses, because we are interested in predicting readmissions. All inter-hospital transfers are also regarded as readmissions.

ICD-9 CM codes	Description
402.01	Malignant hypertensive heart disease with CHF
402.11	Benign hypertensive heart disease with CHF
402.91	Unspecified hypertensive heart disease with CHF
404.01	Malignant hypertensive heart and kidney disease with CHF, with chronic kidney disease stage I through stage IV, or unspecified
404.03	Malignant hypertensive heart and kidney disease with CHF, and chronic kidney disease stage V, or end stage renal disease
404.11	Benign hypertensive heart and kidney disease with CHF and with chronic kidney disease stage I through stage IV, or unspecified
404.13	Benign hypertensive heart and kidney disease with CHF and chronic kidney disease stage V or end stage renal disease
404.91	Unspecified hypertensive heart and kidney disease with CHF and with chronic kidney disease stage I through stage IV, or unspecified
404.93	Unspecified hypertensive heart and kidney disease with CHF and chronic kidney disease stage V or end stage renal disease
428.XX	CHF codes

Table 1: ICD-9 CM codes for CHF

to the patient's principal diagnosis and that require treatment during the stay.

[5]Data preprocessing using discretization and data visualization are detailed in a separate manuscript [3] and are beyond the scope of this paper.

[6]ICD-9 CM is the the International Classification of Diseases, 9th Revision, Clinical Modification Codes.

[7]Multiple hospital encounters of the same patient may be considered separately in this process.

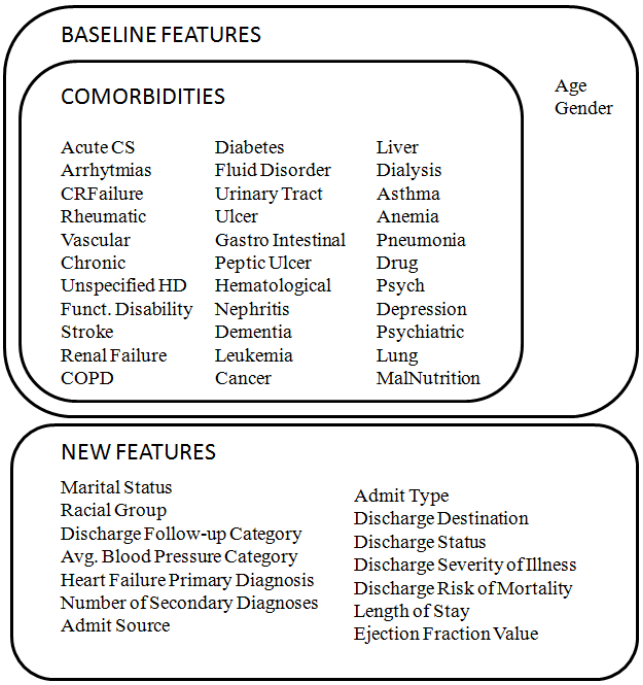

Figure 3: Features used in the RoR prediction process.

2.1.2 Feature Extraction

Critical factors influencing early recurrent admissions are identified through data exploration, review of related studies and help of domain experts. In Figure 3, we list the features that we use in our models. The first set of 35 features, referred to as baseline features in this paper, consists of the predictor variables used by researchers at Yale University, who have tried to solve a similar problem [14]. It consists of two demographic variables namely *age* and *gender* and comorbidities related to the diagnosis information derived from the primary and secondary diagnosis of the patient. The new set of 14 features that we add to these baseline features, consists of other socio-demographic, clinical and administrative data, based on inputs from domain experts and recent studies [6, 5].

2.2 Pre-processing of the Dataset

Once the initial domain knowledge has been amassed and initial important factors pertinent to CHF related admissions are identified, the next task ahead is to pre-process the data to make it amenable for building predictive models.

Feature Selection: Because of the complexity and uniqueness of the domain, hospital readmission due to CHF is a complex phenomenon governed by multiple features. One of our major challenges before the classification task is to determine the subset of attributes that have a significant impact on readmission of patients from the myriad of attributes present in the data set. We consider two state-of-the-art feature selection techniques: Pearson's Chi-square test and Stepwise regression [7].

Missing Value Imputation: We observe that some of the important attributes in the dataset have no value for certain patients. These missing values not only impede the

actual prediction task, but may also lead to biased results.

We use a simple but effective clustering based technique for imputing missing values. The dataset (including instances with missing values) is first divided into a set of clusters using the K-modes clustering method. Then each instance with missing-values is assigned to a cluster that is most similar to it. Finally, missing values of an instance are patched up with the plausible values generated from its respective cluster.

Reducing Class-imbalance: Once the data is integrated, it is observed that the labeled dataset is highly skewed - i.e., the number of instances with *No Readmission* label significantly outnumbers the number of instances with class label *Readmission*. Such imbalance introduces bias in the actual predictive model, as the model with such skewed class distribution would inevitably predict the majority class far more frequently than the alternative class. To circumvent that problem, we use both over- and undersampling. These techniques alter the class distribution of the training data such that both classes are well represented. Oversampling works by re-sampling the rare class records [7], while undersampling decreases the number of records belonging to the majority class by randomly eliminating tuples.

2.3 Predictive Model Building

For the final classification task we use three predictive models. The first one is Logistic Regression (LR), a method that models the outcomes (class labels) as a so-called *logit* function of the predictive variables. We also use Naive Bayes (NB), a well-known simple classifier that assumes independence within the features. The last algorithm is Support Vector Machines (SVM), which separates the data as well as possible, using kernels to model irregular borders. In our case, we use the well-established RBF kernel. Many other classifiers are available, but as the main goal of this work is to showcase the effect of preprocessing techniques and not to compare classifier performances, we limit ourselves to these three classifiers.

3. EXPERIMENTAL EVALUATION

In this section we provide a description of the data set and the used attributes in conjunction. The experiments are conducted using the infrastructure provided by MHS. We use R-studio and R version 2.15.1 to develop the models and use the SQL server 2008 for the database.

The Cardiovascular datamart is of our primary interest. This data mart was developed in 2011 to support an internal clinical process improvement initiative. It is a robust analytics environment, holding approximately 8,600 patients diagnosed with CHF and servicing over 16,800 hospital encounters since January 1st, 2009. The driver of the clinical process improvement is to reduce the high rate of readmission of CHF patients. A detailed description of the views in this datamart is listed below.

- View 1 (3 attributes): patient characteristics that are specific to heart failure systolic anterior motion.
- View 2 (12 attributes): ejection fraction value results.
- View 3 (39 attributes): support follow-up and other inpatient related metrics based on coded diagnosis for inpatient only.
- View 4 (19 attributes): CHF patients (based on clinical results) ejection fraction values less than 40%.
- View 5 (5 attributes): heart failure ICD9 diagnosis codes.
- View 6 (10 attributes): CHF list of medication.
- View 7 (2 attributes): list of diagnosis for arterial fibrillation.
- View 8 (26 attributes): arterial fibrillation patients (based on primary or secondary diagnosis) from the CHF cohort.
- View 9 (26 attributes): active and inactive medications for CHF patients.

After further exploration of the dataset we identify a total of 49 attributes to be related to CHF admission.

We follow a 10 fold cross-fold validation procedure to test the model. We consider four evaluation metrics to assess the quality of the different models:

- **Area Under the Curve (AUC)**: this measure evaluates the trade-of between the rate of patients that are correctly classified as *Readmission* and the rate of patients incorrectly classified as *Readmission*.
- **Precision:** This is the probability that a patient that is predicted as *Readmission* indeed belongs to the class *Readmission*.
- **Recall:** This measures the probability that a patient that truly belongs to the class *Readmission* is also predicted to be *Readmission*.
- **Accuracy:** This is the rate of correctly classified patients.

Depending on the final goal of the RoR prediction, the different evaluation measures are less or more appropriate. The AUC measure is typically interesting when the problem is imbalanced. The precision is important if there is a high cost related to falsely predicting patients to belong to the class *Readmission*. Recall is relevant if the detection of patients that belong to *Readmission* is the main goal. The accuracy is the traditional evaluation measure that gives a global insight in the performance of the model.

We compare all results with the Yale baseline method [14]. In this work, a hierarchical logistic regression model was developed to calculate hospital risk-standardized 30 day all-cause readmission rates for patients hospitalized with CHF. The Yale model primarily focused on cardiovascular and comorbidity variables; moreover, the RoR prediction was limited to patients older than 65 years at the time of an index admission. As our problem at hand is different from the one considered in the Yale Model, our comparison primarily relies on the basis of the attributes suggested by the former model, that is, the Yale model that we refer to here uses the baseline features as described in Figure 3 and applies the predictive models LR, NB and SVM directly to the corresponding dataset, without any data preprocessing.

The results are presented in Table 2. For brevity, we only show the models that are in the top-3 with respect to performance for one of the considered evaluation metrics. Each line in the table represents a model, and it is indicated which

Baseline							
Feature Selection	*Missing Value Imputation*	*Sampling*	*Predictive Model*	*AUC*	*Precision*	*Recall*	*Accuracy*
-	-	-	NB	0.59	0.36	0.04	0.77
-	-	-	SVM	0.58	0.27	0.45	0.61
-	-	-	LR	0.59	0.40	0.01	0.78
Top-3: AUC							
Feature Selection	*Missing Value Imputation*	*Sampling*	*Predictive Model*	*AUC*	*Precision*	*Recall*	*Accuracy*
Stepwise	Clustering	Over sampling	SVM	0.64	0.33	0.51	0.66
Stepwise	Clustering	-	NB	0.64	0.58	0.16	0.79
Chi-Square	Clustering	Over sampling	SVM	0.64	0.32	0.50	0.66
Top-3: Recall							
Feature Selection	*Missing Value Imputation*	*Sampling*	*Predictive Model*	*AUC*	*Precision*	*Recall*	*Accuracy*
Chi-Square	Clustering	Under sampling	LR	0.60	0.28	0.56	0.58
Stepwise	Clustering	Over sampling	LR	0.59	0.28	0.56	0.58
Stepwise	Clustering	Under sampling	LR	0.59	0.27	0.56	0.58
Top-3: Precision							
Feature Selection	*Missing Value Imputation*	*Sampling*	*Predictive Model*	*AUC*	*Precision*	*Recall*	*Accuracy*
Stepwise	Clustering	-	NB	0.64	0.58	0.16	0.79
Chi-Square	Clustering	-	NB	0.63	0.37	0.32	0.73
Stepwise	Clustering	Over sampling	SVM	0.64	0.33	0.51	0.66
Top-3: Accuracy							
Feature Selection	*Missing Value Imputation*	*Sampling*	*Predictive Model*	*AUC*	*Precision*	*Recall*	*Accuracy*
Stepwise	Clustering	-	NB	0.64	0.58	0.16	0.79
Chi-Square	Clustering	-	NB	0.63	0.37	0.32	0.73
Stepwise	Clustering	Over sampling	SVM	0.64	0.33	0.51	0.66

Table 2: Evaluation of the baseline method and the newly proposed techniques. In the first three columns, it is indicated which preprocessing techniques are applied. In the fourth column, the predictive model is shown and the last four columns show the resulting evaluation metrics.

preprocessing algorithms are applied. A minus sign means that none of the corresponding preprocessing techniques was carried out.

None of the proposed models are winners with respect to all evaluation metrics. All proposed models outperform the baseline methods with respect to AUC, and for each of the considered baseline methods, there is a new model that outperforms all the baseline models. The baseline methods using NB or LR as predictive model are outperformed for all evaluation metrics by the new method that uses stepwise feature selection, clustering for data imputation and NB as predictive model. The baseline method using SVM as predictive model is outperformed by the new method that uses stepwise feature selection, clustering for data imputation, oversampling and NB as predictive model. These conclusions indicate that it is indeed useful to use the newly introduced features and to apply preprocessing techniques to the data.

4. RELATED WORK

Preventing hospitalization is a prominent factor to reduce patient morbidity, improve patient outcomes, and curb health care costs. An increasing body of literature attempts to develop predictive models for hospital readmission risks. These studies range from all-cause readmissions to readmission for specific diseases such as heart failure, pneumonia, stroke, and asthma. Each of these models exploits various predictor variables assessed at various times related to index hospitalization (admission, discharge, first follow-up visit, etc.). One of the significant research results for predicting RoR for CHF patients was proposed by the University of Yale [14] (considering a different problem definition). The attributes proposed by the Yale Model are considered as a baseline in this work.

In another research study [16], a real-time predictive model was developed to identify CHF patients at high risk for readmission within the 30-day timeframe. In this model, some clinical and social factors available within hours of hospital presentation are used in order to have a real-time prediction model. Although the model demonstrated good discrimination for 30-day readmission (AUC 0.72), the dataset size is very small (1372 HF patients).

One of the recent studies for predicting 30-day RoR for heart failure hospitalization is done in [2]. In this work, administrative claim data is used to build a regression model on 24,163 patients from 307 hospitals. However, like the Yale Model [14], this work has only focused on patients more than 65 years old from CHF registry in the general US population from 2004 to 2006 and its best performance had a AUC under 0.60.

5. CONCLUSION

Partnering with MHS we study the problem of RoR of CHF patients within 30 days of discharge. The problem is formalized as a binary classification problem and different prediction models are developed and validated. We consider a complex, very high dimensional clinical dataset provided by MHS towards identifying the risk factors related to readmission of patients discharged with diagnosis of CHF, within 30 days of discharge. We propose a framework to solve this task that focuses on extracting predictive features and preprocessing of the data. Our comprehensive experimental study exhibits the benefit of the additional features and confirms that preprocessing improves the predictive models.

Our ongoing work involves investigation of additional feature and classification techniques to improve the quality of prediction. Additionally, we continue to investigate the deployment of the developed technique inside the MHS domain to make it available to be used by physicians [18]. Next, we are in the process of integrating our preprocessing and modeling component in the patient care pipeline to improve the quality of care. We are also in the process of predictive modeling for risk of readmission for other diseases particularly chronic autoimmune ones where very little domain knowledge exists for identifying major risk factors.

6. ACKNOWLEDGMENTS

This work is supported by MHS (grant no: A73191). Additionally, we are thankful to the data architects and the clinicians at MHS for their valuable time and insightful discussions during the initial stage of the study.

7. ADDITIONAL AUTHORS

Ankur Teredesai (Inst. of Technology, CWDS, Univ. of Washington Tacoma, email:
ankurt@uw.edu) and David Hazel (Inst. of Technology, CWDS, Univ. of Washington Tacoma, email:
dhazel@uw.edu) and Paul Amoroso (Multicare Healthcare Systems, email:
Paul.Amoroso@multicare.org) and Lester Reed (Multicare Healthcare Systems, email:
Lester.Reed@multicare.org)

8. REFERENCES

[1] K. F. Adams, G. C. Fonarow, C. L. Emerman, T. H. LeJemtel, M. R. Costanzo, W. T. Abraham, R. L. Berkowitz, M. Galvao, and D. P. Horton. Characteristics and outcomes of patients hospitalized for heart failure in the united states: Rationale, design, and preliminary observations from the first 100,000 cases in the acute decompensated heart failure national registry (ADHERE). *American Heart Journal*, 149(2):209–216, Feb. 2005.

[2] H. BG, C. LH, and F. GC. Incremental value of clinical data beyond claims data in predicting 30-day outcomes after heart failure hospitalization. *Circ Cardiovasc Qual Outcomes*, 4(4):60–67, 2011.

[3] S. Chin, K. Zolfaghar, J. Agarwal, S. B. Roy, A. Teredesai, D. Hazel, P. Amoroso, and L. Reed. Data exploration using discretization techniques, an application for hospital readmission analysis. *Under Review*, 2013.

[4] P. C. Coleman EA. The care transitions intervention: Results of a randomized controlled trial. *Archives of Internal Medicine*, 166(17):1822–1828, Sept. 2006.

[5] A. D. DonzÃl' J. Potentially avoidable 30-day hospital readmissions in medical patients: Derivation and validation of a prediction model. *JAMA Internal Medicine*, 173(8):632–638, Apr. 2013.

[6] C. Franchi, A. Nobili, D. Mari, M. Tettamanti, C. D. Djade, L. Pasina, F. Salerno, S. Corrao, A. Marengoni, A. Iorio, M. Marcucci, and P. M. Mannucci. Risk factors for hospital readmission of elderly patients. *European Journal of Internal Medicine*, 24(1):45–51, Jan. 2013.

[7] J. Han and M. Kamber. *Data mining: concepts and techniques*. Morgan Kaufmann, 2006.

[8] P. Harrison, P. Hara, J. Pope, M. Young, and E. Rula. The impact of postdischarge telephonic follow-up on hospital readmissions. *Popul Health Manag*, 14:27–32, 2011.

[9] T. Hunter, J. Nelson, and J. Birmingham. Preventing readmissions through comprhensive discharge planning. *Prof Case Manag.*, 18:56–63, 2013.

[10] H. Kaur and S. K. Wasan. Empirical study on applications of data mining techniques in healthcare. *Journal of Computer Science*, 2(2):194–200, 2006.

[11] T. M. Koelling, M. L. Johnson, R. J. Cody, and K. D. Aaronson. Discharge education improves clinical outcomes in patients with chronic heart failure. *Circulation*, 111(2):179–185, Jan. 2005.

[12] H. C. Koh and G. Tan. Data mining applications in healthcare. *Journal of Healthcare Information Management Vol*, 19(2):65, 2011.

[13] H. M. Krumholz, J. Amatruda, G. L. Smith, J. A. Mattera, S. A. Roumanis, M. J. Radford, P. Crombie, and V. Vaccarino. Randomized trial of an education and support intervention to preventreadmission of patients with heart failure. *Journal of the American College of Cardiology*, 39(1):83–89, Jan. 2002.

[14] H. M. Krumholz, S. L. T. Normand, P. S. Keenan, Z. Q. Lin, E. E. Drye, K. R. Bhat, Y. F. Wang, J. S. Ross, J. D. Schuur, and B. D. Stauffer. *Hospital 30-day heart failure readmission measure methodology. Report prepared for the Centers for Medicare & Medicaid Services.*

[15] K. Ottenbacher, P. Smith, S. Illig, R. Linn, R. Fiedler, and C. Granger. Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke. *Journal of clinical epidemiology*, 54(11):1159–1165, 2001.

[16] A. R, M. BJ, and T. YP. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. *Journal of Medical Care*, 10:981–988, Feb. 2010.

[17] J. Ross, J. Chen, Z. Lin, H. Bueno, J. Curtis, P. Keenan, S. Normand, G. Schreiner, J. Spertus, M. Vidan, Y. Wang, Y. Wang, and H. Krumholz. Recent national trends in readmission rates after heart failure hospitalization. *Circ Heart Fail*, 3:97–103, 2010.

[18] K. Zolfaghar, J. Agarwal, D. Sistla, S. Chin, S. B. Roy, N. Verbiest, A. Teredesai, D. Hazel, P. Amoroso, and L. Reed. Risk-o-meter: An intelligent clinical risk calculator. In *Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2013.