©Copyright 2016
Aftab Hassan

Predictive Analytics and Decision Support for Heart Failure patients

Aftab Hassan

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Computer Science and Systems

University of Washington

2016

Reading Committee:

Senjuti Basu Roy, Chair

Ankur Teredesai

Martine De Cock

Program Authorized to Offer Degree:
Computer Science and Systems

University of Washington

Abstract

Predictive Analytics and Decision Support for Heart Failure patients

Aftab Hassan

Chair of the Supervisory Committee:
Assistant Professor Senjuti Basu Roy
Institute of Technology

In the last few years, legislations such as the Patient Protection and Affordable Care Act, also known as Obamacare, have emphasised the need for improving the quality of health care. Part of the programs introduced by this Act is the Hospital Readmissions Reduction Program (HRRP) which reduces payments to hospitals with excess readmissions. Hospitals are, therefore, constantly looking for ways to help reduce their readmission rate, and an idea of patients that are at a higher risk of getting readmitted is extremely beneficial. In this thesis, we first look at ways to predict the chance of a patient getting readmitted, and investigate predicting future healthcare costs, to understand how much money the patient would have to spend on hospital expenses during the readmission visit. We, then design a multi objective medication recommendation framework. We would like to echo the sentiments of the medical community here, that, while reducing readmissions is important, there are also other factors that go into ensuring a successful patient discharge experience, such as reducing mortality rate and the patient's length of stay at the hospital during the subsequent visit. We propose a novel framework that recommends personalized medications to patients by analyzing the complex interplay among a multitude of factors, such as, demographic factors, medical diagnoses, clinical factors, and how they contribute to the three objectives - thirty day readmission, subsequent length of stay and mortality rate. We then find the best medication combination to simultaneously reduce all three objectives by performing Multi

objective optimization. Our proposed framework is flexible enough to include or exclude additional factors, as well as layers, and can even obey constraints provided by the domain experts (i.e., doctors) in the design of this hierarchical network. We present a case study validated by a domain expert as well as comprehensive experimental results based on the proposed approach to demonstrate the effectiveness of our proposed framework. Our work is validated on the largest (1000+ bed) South Puget Sound healthcare system: MultiCare Health on their EMR data including medication transactions. While our primary effort is to design medication recommendations for heart failure patients, the proposed framework could be adapted for other diseases as well. We also develop a web based service called Pathway Finder with the objective of visually exploring and discovering clinical pathways that reduce the patients chance of readmission. We, therefore, try and play our small part in improving the quality of health care in the United States.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I am grateful to many people without whose help and contribution this work could not have been completed. First, I would like to express my gratitude to my advisor and committee chair Dr. Senjuti Basu Roy, who has been guiding my research, and giving me constant encouragement throughout. Next, I would like to appreciate Dr. Ankur Teredesai, who has given me valuable advice about my thesis. Our discussions have helped me to think of a problem from multiple perspectives, and in formulating a problem statement that is useful to the society. I would also like to thank other collaborators at the Center of Data Science, Dr. Martine De Cock, David Hazel, Rui Liu, Dr. Shanu Sushmita for their support in this research.

I would like to thank Stacey Newman for being such a wonderful friend, and showing me, by example, how there can always be more attention to detail, Harshavardhan Chitta for always being so supportive, and my friends at Sprinker Tennis and the University Y who have helped me stay healthy and focused.

On behalf of the Center for Data Science at the University of Washington, Tacoma, I would like to thank Edifecs Inc. and MultiCare Health System for providing the data sets used in this research, and Jenny He, PharmD, BCPS, Clinical Pharmacist, MultiCare Health System, for her clinical expertise in designing the medication recommendation framework.

Last but not least, I would like to express my deep appreciation to my family. Without support from my parents, I would never complete my Masters degree.

DEDICATION

To my grand father, Mr Peer Mohammed, who inspired me to pursue a Master's degree

Chapter 1

INTRODUCTION

In the last few years, legislations such as the Patient Protection and Affordable Care Act, also known as Obamacare, have emphasised the need for improving the quality of health care. A key component here, is patient care during his/her stay at the hospital and how it affects the overall discharge experience. In this thesis, we focus on improving the quality of discharge experience from various angles. Our first proposal is for a framework to predict 30-day hospital readmissions and the cost associated with that hospital visit. Next, while heart failure readmission is extremely important, there are also other factors that go into ensuring a successful patient discharge experience, such as reducing mortality rate and the patient's length of stay at the hospital during the subsequent visit. We propose a novel multi objective medication recommendation framework that recommends personalized medications to simultaneously reduce three objectives - thirty day readmission, subsequent length of stay and mortality rate. Finally, we develop a web based service called Pathway Finder with the objective of visually exploring and discovering clinical pathways that reduce the patient's chance of readmission.

In our readmission work, we evaluate state-of-the-art machine learning techniques in predicting whether a patient is at risk of being readmitted to the hospital within 30 days after discharge. This is a binary classification task, namely predicting whether the next admission of a given patient will be within 30 days or not. We find that the use of machine learning techniques allows us to achieve higher sensitivity (recall) without penalizing the specificity and precision too much. We provide comprehensive empirical results analysing the performance of the different machine learning algorithms we have used.

In addition to our readmission work, we also look at the problem of estimating the cost of

that hospital readmission. The ability to know both of these about a patient helps hospitals identify and plan for high risk patients, I.e, those that are highly likely to readmit and have a high cost associated with their readmission. We show that state of the art machine learning approaches out perform the statistical methods which were previously used in the cost prediction domain. In particular, a substantially lower mean absolute error (MAE) can be achieved with M5 model trees. Our research work in this domain is also useful to providers and insurers who wish to foresee the cost associated with a client.

Next, while reducing readmissions is important, there are also other factors that go into ensuring a successful patient discharge experience, such as reducing mortality rate and the patient's length of stay at the hospital during the subsequent visit. We propose a novel framework that recommends personalized medications to patients by analyzing the complex interplay among a multitude of factors, such as, demographic factors, medical diagnoses, clinical factors, and how they contribute to the three objectives - thirty day readmission, subsequent length of stay and mortality rate. We then find the best medication combination to simultaneously reduce all three objectives by performing Multi objective optimization. First, we propose to learn the structure and parameters of a hierarchical Bayesian network from the available patient data. We then perform Bayesian Inference to get conditional probabilities of each of the objectives for a given set of patient conditions such as demographic factors and diagnosis information for every valid medication combination, and then perform Multi Objective Optimization to optimize all of these objectives simultaneously. Our proposed framework is flexible enough to include or exclude additional factors, as well as layers, or can even obey constraints provided by the domain experts (i.e., doctors) in the design of this hierarchical network. We present a case study validated by a domain expert as well as comprehensive experimental results based on the proposed approach to demonstrate the effectiveness of our proposed framework. Our work is validated on the largest (1000+ bed) South Puget Sound healthcare system: MultiCare Health on their EMR data including medication transactions. While our primary effort is to design medication recommendations for heart failure patients, the proposed framework could be adapted for other diseases as

well. Our solution led to actionable insights for the chronic care management team, and is under consideration for deployment within the physician and pharmaceutical compliance workflows at KenSci Inc.

Finally, we develop Pathway-Finder, a novel interactive recommender system for clinical decision support. This is a cloud based web-service hosted on Microsoft Azure for Research platform with the objective of visually exploring and discovering clinical pathways, and understanding the effect of a care pathway on the patient's readmission. In doing so, we extend our earlier work of finding patients with a high chance of readmission, by making personalized intervention recommendations to minimize their 30-day HF readmission risk. Being hosted as cloud service enables the health providers to access the system without the need to deploy analytics infrastructure.

This thesis will detail the contributions made in these areas. Parts of this work have been published in conferences related to machine learning and healthcare analytics and I have been a co-author in these works, namely, Chapter 2, that was published at HIAI,2016 [60], Chapter 3 [25] that is pending approval from KDD,2016 and Chapter 4 that was published at ICDM,2014 [44]. Through all of this combined work, we play our small part in improving the discharge experience of a patient, as we join the initiative to improve and revolutionize the healthcare domain.

Chapter 2

PREDICTIVE ANALYTICS

2.1 Introduction

Patients with chronic conditions repeatedly get admitted to a hospital for treatment and care. They are often discharged when their condition stabilizes only to get readmitted again, many times within just a few days. This process is termed as *hospital readmissions*. The readmission problem in the U.S. is severe: currently one in five (20%) Medicare patients are readmitted to a hospital within 30 days of discharge. Three quarters of these readmissions (75%) are actually considered avoidable [31]. In addition to raising red flags about gaps in quality of care, hospital readmissions also place a huge financial burden on the health system. In 2011, there were approximately 3.3 million adult 30-day all-cause hospital readmissions in the United States, and they were associated with about $41.3 billion in hospital costs [28]. Avoidable readmissions account for around $17 billion a year [31]. In the U.S., the readmission rate of patients at a hospital is tracked as a proxy for measuring the overall quality of treatment a patient has received, and, under the Affordable Care Act, Medicare has started penalizing hospitals that have higher-than-expected rates of 30-day readmissions[1].

In this paper we tackle two related problems, namely *(1) predicting whether a patient is at risk of being readmitted to the hospital within 30 days after discharge*, and *(2) estimating the cost of that hospital readmission*. The ability to prioritize a care plan along both of these variables can enable hospital systems to more effectively allocate the limited human and budgetary resources available to the high-risk individuals (i.e., higher-cost, earlier readmissions). Potential care transition gaps and targeted interventions can be derived from such

[1]http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html, accessed on Oct 22, 2015

models with a more profound impact on overall population management.

Existing dedicated efforts for accurately predicting 30-day risk of readmission are mostly focused on a specific cohort[2], such as congestive heart failure patients [3], cancer patients [20], emergency readmissions [57], etc. While these models are very useful, there is a lot of value in having *all-cause* risk and cost of readmission models that are not tied to a specific disease. In addition to allowing to derive risk and cost scores for patients who do not belong to any of the well studied cohorts, these models can also be used for incoming patients for which we do not know (yet) which cohort they belong to. To the best of our knowledge, none of the recent efforts predict cost or risk for all-cause readmissions, which is a completely different medical and data mining problem involving large, heterogeneous patient population sizes compared to disease specific cohorts such as heart failure.

In this study, we evaluate state-of-the-art machine learning techniques for predicting 30-day risk and cost on admission data of patients provided by a large hospital chain in the Northwestern U.S. We treat the risk prediction problem as a binary classification task, namely predicting whether the next admission of a given patient will be within 30 days or not. The LACE index is often used in clinical practice for this purpose [67]. This index considers four numerical variables, namely length of stay (L), acuity level of admission (A), comorbidity condition (C), and use of emergency rooms (E). The LACE score of a patient is obtained by summing up the values of these four variables at the time of discharge. A threshold (usually ≥ 10) is then set to determine which patients are at "high" readmission risk [67]. We use LACE as a baseline to compare the performance of the machine learning algorithms we investigate in this paper. We find that the use of machine learning techniques allows to achieve higher sensitivity (recall) without penalizing the specificity and precision too much. On the cost prediction side, we find that the simple baseline strategy of forecasting that the next admission of a patient will cost as much as the average of his previous admissions works reasonably well. In addition, a substantially lower mean absolute error (MAE) can be

[2]A sub-group of a given population with similar characteristics (e.g., medical conditions), such as s group of diabetes patients.

achieved with M5 model trees.

The rest of the paper is organized as follows: after giving an overview of related work in Section 2.2, we formalize the risk and cost prediction problems in Section 2.3. The machine learning algorithms applied in this paper for risk and cost predictions are explained in Section 2.4. The dataset and features are described in Section 2.5. In Section 2.6 we discuss the performance of the algorithms. Finally, in Section 2.7 we conclude with our overall findings.

2.2 Related Work

In this section, we give a brief overview of research efforts done independently along each of the two dimensions: readmission risk prediction and healthcare cost prediction. To the best of our knowledge, there is no existing work that studies risk and cost prediction problems in a combined way.

2.2.1 Healthcare Cost Prediction

Previously proposed cost prediction models often used rule-based methods and linear regression models. A challenge with the rule-based methods (e.g. [38]) is that they require substantial domain knowledge which is not easily available and is often expensive. Linear regression models on the other hand are challenged by the skewed nature of healthcare data. Healthcare cost data typically features a spike at zero, and a strongly skewed distribution with a heavy right-hand tail [33]. As a result, the prediction models are posed with the challenge of an extreme value situation. This phenomenon is also observed in the dataset used in this study (see Figure 2.1). Consequently, several advanced statistical methods (in-sample estimation) have been proposed to overcome the skewness issue, such as General Linear Models (GLM) [45], mixture models [49], etc. For a comprehensive comparison of previously proposed statistical methods for healthcare cost prediction, we refer to the review paper [48]. The development of healthcare cost prediction models using machine learning methods has been more recent (e.g., [40, 61]). [40] investigate classification algorithms to predict whether an individual is going to incur higher or lower healthcare expenditure. [61]

use three machine learning algorithms for cost prediction – regression tree, M5 model tree and random forest, and observe improved performance when compared to traditional methods. In this paper, we also investigate these algorithms for the task of predicting cost of hospital readmission. To the best of our knowledge, their utility for predicting the costs of hospital readmissions specifically (as opposed to predicting general healthcare costs) has not been investigated before.

2.2.2 Hospital Readmission Prediction

In 2011, there were approximately 3.3 million adult 30-day all-cause hospital readmissions in the United States, and they were associated with about $41.3 billion in hospital costs [28]. Many of these hospitalizations are readmissions of the same patient within a short period of time. These readmissions act as a substantial contributor to rising healthcare costs [31]. Readmission rates are also used as a screening tool for monitoring the quality of service and efficiency of care provided by healthcare providers [3]. While predicting risk-of-readmission has been identified as one of the key problems for the healthcare domain, not many solutions are known to be effective [39, 52]. In fact, to improve the clinical process of heart failure patients for instance, healthcare organizations still leverage the proven best-practices, called "*Get With The Guidelines*" by the American Heart Association. In general, related work on risk-of-readmission prediction has primarily attempted to study cohort specific readmission risk, such as, heart failure, pneumonia, stroke, and asthma, but the effort of designing large scale machine learning algorithms for all-cause readmission is still at a rather rudimentary stage.

Despite several years of continued research efforts in modeling risk of readmission and healthcare cost, a dual predictive tool that utilizes healthcare data to predict risk and cost of hospital readmission has not been explored before. This study makes the first step in that research direction.

2.3 Problem Description

The goal of this study is to predict a patient's 30-day **risk** of hospital readmission and the associated **cost** of that readmission. We assume that the learning task at hand is a combination of a supervised classification problem (risk prediction) and a regression problem for predicting the cost (in dollars) of the readmission. The feature vector $X_i = (x_{i1}, x_{i2}, ..., x_{iM})$ of an instance i includes information about general demographics such as age and gender of the patient, as well as specific clinical and cost information at the time of discharge from the hospital. The goal is to produce an output vector $Y_i = (y_{i1}, y_{i2})$ consisting of a label y_{i1} that indicates whether the next admission of the patient will be in 30 days ("yes") or not ("no"), and the cost y_{i2} of the next admission. Let us use $\mathcal{X}$ to denote the set of all instances (feature vectors), and let $\mathcal{Y} = \{\text{yes, no}\} \times \mathbb{R}^+$ be the set of all dual labels. Given training examples of the form (X_i, Y_i) with $X_i \in \mathcal{X}$ and $Y_i \in \mathcal{Y}$, the aim is to learn a model $\mathcal{H} : \mathcal{X} \rightarrow \mathcal{Y}$ that can label new, unseen instances from $\mathcal{X}$ with a dual label from $\mathcal{Y}$ in an accurate way. We address this multi-label prediction learning problem in a manner similar to binary relevance [63], by learning a model for each label:

- **Risk of 30-Day Readmission – Classification Task:**
  For given training examples of the form (X_i, y_{i1}), where $X_i \in \mathcal{X}$ and $y_{i1} \in \{\text{yes, no}\}$, the goal is to learn a model $\mathcal{H}_1 : \mathcal{X} \rightarrow \{\text{yes, no}\}$ that accurately predicts whether the next admission of a patient will be within 30 days.

- **Cost of Readmission – Regression Task:**
  For given training examples of the form (X_i, y_{i2}), where $X_i \in \mathcal{X}$ and $y_{i2} \in \mathbb{R}^+$ (cost in dollars), the goal is to learn a model $\mathcal{H}_2 : \mathcal{X} \rightarrow \mathbb{R}^+$ that accurately predicts the cost of the next admission.

The combined model is then obtained as $\mathcal{H}(X) = (\mathcal{H}_1(X), \mathcal{H}_2(X))$, for X in $\mathcal{X}$.

We also tried other techniques for multi-label prediction like – Label Powerset [11] and Chain Classifier [54], but initial evaluation results with these methods were not as good as

those obtained with the binary relevance approach, so we omit them from this paper.

X (Input Vector)		Y (Output Vector)	
Admission		Next Admission	
Unique Identifier	Features (X)	Risk (y_1)	Cost (y_2)
id_1	$x_{11}, x_{12}, ...x_{1M}$	yes	\$45,132
id_1	$x_{21}, x_{22}, ...x_{2M}$	yes	\$41,305
id_1	$x_{31}, x_{32}, ...x_{3M}$	no	\$17,809
id_2	$x_{41}, x_{42}, ...x_{4M}$	yes	\$21,305
id_2	$x_{51}, x_{52}, ...x_{5M}$	no	\$55,809
id_3 (**test case**)	$x_{61}, x_{62}, ...x_{6M}$	?	?

Table 2.1: Example input and output scenario for the risk and cost prediction task. Here, the first column indicates a unique identifier for each patient in the dataset.

Table 2.1 illustrates the input and output representations for the risk and cost prediction problem. Let us assume that the patient with id_1 was admitted to the hospital four times (say on Jan 14, Feb 2, Feb 28 and Apr 15). That means that he had three *re*admissions, two within 30 days (high risk) with cost being \$45,132 and \$41,305 respectively, and one after 30 days (low risk) with cost being \$17,809. These response features are constructed based on attributes from the original, raw data (shown in Table 2.2). During the training phase, data of patient id_1 and id_2 will be used to train binary classifiers (to predict risk) and regression models (to predict cost); during the test phase the models will be used to predict the risk and the cost of patient id_3's next encounter.

We evaluate the accuracy of the learned models in several ways. Accuracy is traditionally measured as the percentage of instances that are classified correctly. It has been emphasized that the use of accuracy as an evaluation measure for data where there is an imbalance

between positive and negative classes can yield misleading conclusions [19, 26]. Readmission data is typically imbalanced. As Table 2.2 shows, approximately 27% of the admissions in our study are within 30 days (i.e. 27% positive instances), while the remaining 73% happen after 30 days (i.e., 73% negative instances). In addition to accuracy, we therefore also evaluate the binary classification models in terms of sensitivity (recall), specificity (true negative rate), and precision. Recall that sensitivity is TP/(TP + FN), specificity is TN/(TN+FP), and precision is TP/(TP+FP), with TP, FP, TN, and FN respectively denoting the number of true positives, false positives, true negatives and false negatives. The performance of the cost prediction algorithms is evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), with a lower error indicating a better performance.

2.4 Methods

In this section we give an overview of the machine learning algorithms used in this study. For risk-of-readmission prediction we used state-of-the-art classification techniques, while for cost prediction we used regression techniques.

2.4.1 Risk Prediction Methods

Support Vector Machine: A Support Vector Machine (SVM) is a statistical learning method for training classifiers based on different types of kernel functions – polynomial functions, radial basis functions, etc. An SVM learns a linear separating hyperplane by maximizing the margin between the classes [16]. The decision boundary is maximised with respect to the data points from each class (known as support vectors) that are closest to the decision boundary. For this study, we tested SVM with linear and radial kernel, and we report the results for radial in Table 2.3 because its performance was better than the linear kernel settings. We also tested for different regularization parameters ($C = 1, 5, 10, 15$), but the overall results did not change.

Logistic Regression: Logistic Regression is an example of a discriminative classifier that models the posterior $p(y_1|X)$ directly given the input features. That is, it learns to

map the input (X) vector directly to the output class label y_1 (risk in our case). When the response is a binary (dichotomous) variable, logistic regression fits a logistic curve to the relationship between X and y_1 [51]. The class decision for the given probability is then made based on a threshold value. The threshold is often set to 0.5, i.e. if $p(y_1|X) \geq 0.5$, then we predict that the next readmission of the patient will be within 30 days, and otherwise not. We tested with multiple threshold values to make the class decision.

Decision Trees: An alternative approach to linear classification is to partition the space into smaller regions, where the interactions are more manageable. Like for the other methods described in this section, the goal of a classification tree is to predict a response y_1 (risk in our case) from inputs X. This is done by growing a binary tree. At each internal node in the tree, a test is applied to one of the inputs, and depending on the outcome, the left or the right sub-branch of the tree is then selected. Eventually a leaf node is reached, where the prediction is made. For this study, we used an implementation of the classification and regression tree algorithm (CART) [7] in R. We tested the performance of classification trees using different complexity parameters (cp = 0.01, 0.001, 0.0005). In Table 2.3 we report the results of the best performing tree with cp set to 0.01 value.

Random Forest: Random forest regression is an ensemble learning method that operates by constructing a multitude of regression trees at training time and outputting the mean prediction of the individual trees for new observations. Each tree is constructed using a random sample of the observation and feature space from the original dataset. This has the effect of correcting the tendency of individual regression trees to overfit the training data [6].

Generalised Boosted Modeling (GBM): Boosting is an approach to machine learning based on the idea of creating a highly accurate predictive model ensemble by combining many relatively weak and inaccurate models [21]. In other words, boosting is an optimization technique that minimizes the loss function by adding, at each step, a new model that best reduces the loss function. It is often used to grow an ensemble of classification trees, like we do in this paper. In this study we use the gbm implementation of AdaBoost in R, which is an

implementation of extensions to Freund and Schapire's AdaBoost algorithm and Friedman's gradient boosting machine[3].

All the models are trained and tested using R[4]. Additionally, we also set the output of each model to be the class probability (prob= TRUE), instead of the class labels. This was done in order to test for different decision threshold values $(0.0 - 1.0)$ to find the optimal balance between different evaluation measures. We report results for thresholds between $0.2 - 0.52$ in Figures 2.2-2.5.

2.4.2 Cost Prediction Methods

Linear Regression: Linear regression is used extensively in the literature on healthcare cost prediction, so, even though it has its limitations, it can not be ignored in this study. We use a linear regression model to predict cost using an M-dimensional vector of predictive variables (see Table 2.2).

M5 Model Tree: M5 model trees are a generalization of the CART model [7]. The structure of an M5 model tree follows that of a decision tree, but has multiple linear regression models at the leaf nodes, making the model a combination of piecewise linear functions. The algorithm for the training of a model tree breaks the input space of the training data through a recursive partitioning process similar to the one used in CART. After partitioning, linear regression models can be fit on the leaf nodes, making the resulting regression model locally accurate.

In addition to the linear regression and M5 Model Tree methods, **decision trees** and **generalised boosted modeling (GBM)** as described for risk prediction task were also used for predicting the cost.

---

[3]http://cran.r-project.org/web/packages/gbm/gbm.pdf

[4]http://www.r-project.org

2.5 Dataset and Features

The study in this paper includes admission data of patients provided by a large hospital chain in the Northwestern United States. Each admission record includes demographic information (e.g., gender, ethnic group), clinical information (e.g., primary diagnosis), care provider details, administrative data (e.g., length of stay) and billing information (e.g., charge amount). First, we performed data filtering as part of data pre-processing. Of the available ∼221K admission records, we excluded instances of admissions for which the patient died before discharge, or was transferred to another acute care facility within the hospital chain, or left against medical advice. Additionally, we excluded records where the next admission date is unknown, since they cannot be used to evaluate the correctness of cost and risk of readmission prediction. We also excluded hospitalizations with unspecified primary diagnosis.

Next, we performed several feature engineering steps. There were 214 features in the raw data. We used a forward stepwise regression approach [14] to select a subset of this feature set. This subset is shown in Table 2.2. Most of the features from Table 2.2 correspond directly to features from the raw data; others have been constructed based on previous history of the patient. That is, most of the features are drawn from individual admission records, but some are aggregated across multiple admission records of the same patient. The features from the latter category are:

- **Number of Comorbidities:** this is the total number of unique comorbidities[5] that were registered for a patient up to the time of discharge. We used the Elixhauser comorbidity [18] information of a patient to identify all comorbidities associated to that patient. Comorbidity is associated with worse health outcomes, increased healthcare costs and is known to impact prediction of risk of readmission [15]. Therefore, we use it as one of the predictor variables.

- **Number of Existing conditions:** this is is the total number of unique diagnoses

[5]Two or more coexisting medical conditions or disease processes that are additional to an initial diagnosis.

registered for this patient up to this point, including during previous admissions. The list of existing conditions of a patient is represented using ICD9-CM codes in the raw data (∼4K distinct values). We grouped the ICD9-CM codes using Clinical Classification Software (CCS)[6], and included the count of distinct CCS codes per patient as a feature.

- **Number of Previous Admissions:** this is the total number of hospital admissions registered for this patient up to this point. Here, the assumption is that a patient with a history of several hospital admissions is more likely to be readmitted again.

In this paper we use frequency counts (e.g. Number of Comorbidities) to overcome the limitation of significant sparseness in this dataset, for future research, we aim to explore statistical methods to overcome this limitation. Finally, we randomly sampled ∼10K instances with the feature set shown in Table 2.2 to train and test our models.

Table 2.2: Overview of the feature set used in the prediction of "risk" and "cost" of readmission

Feature	**Type**	**Distribution**
Gender	Categorical	Female (5,818)
		Male (4,176)
Adult	Boolean	Yes (9,792)
		No (202)
Age $\geq$ 65	Boolean	Yes (4,801)
		No (5,193)
Ethnic Group	Categorical	Caucasian (8,303)

Continued on next page

[6]https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp

Table 2.2 continued: Overview of the feature set used in the prediction of "risk" and "cost" of readmission

Feature	**Type**	**Distribution**
		African American (669)
		Hispanic/Latino (257)
		American Indian (185)
		Asian (172)
		Pacific Islander (157)
		Multi-Racial (83)
		Non-Hispanic (25)
		Middle Eastern (18)
		Eskimo (4)
		Other (121)
Marital Status	Categorical	Single (2,387)
		Married (4,148)
		Widowed (1,939)
		Divorced (1,084)
		Separated (211)
		Significant Other (192)
		Legally Separated (23)
		Domestic Partner (4)
		Other (2)
		Unknown (4)
Admit Type	Categorical	Emergency (7,350)
		Elective (2,519)

Continued on next page

Table 2.2 continued: Overview of the feature set used in the prediction of "risk" and "cost" of readmission

Feature	**Type**	**Distribution**
		Urgent (86)
		Trauma Center (39)
Financial Class	Categorical	Medicare (5,595)
		Medicaid (924)
		Self-pay (319)
		Other (3,156)
Care Type	Categorical	Acute (9,975)
		Geropsychiatric (19)
No. of Comorbidities	Numeric	Mean: 6.43 (SD = 4.25)
No. of Existing Conditions	Numeric	Mean: 10.67 (SD = 8.05)
Length of Stay (Days)	Numeric	Mean: 4.28 (SD = 4.41)
Same Day Discharge	Boolean	Yes (105)
		No (9,889)
Blood Pressure at Discharge	Categorical	80-89 or 120-139 (4,189)
		<80 and < 120 (3,529)
		90-99 or 140-159 (1,557)
		> 99 or > 159 (719)
No. of Previous Admissions	Numeric	Mean: 1.55 (SD = 2.87)
Next Admission < 30 Days	Categorical	Yes (2,697), 27%
(Response Variable)		No (7,297), 73%

Continued on next page

Table 2.2 continued: Overview of the feature set used in the prediction of "risk" and "cost" of readmission

Feature	Type	Distribution
Additional Features for Cost Prediction		
Current Admit Cost ($)	Numeric	Mean: 53,530 (SD = 72,888)
Current Bed Charge ($)	Numeric	Mean: 10,120 (SD = 14,458)
Cost of Next Readmission ($) **(Response Variable)**	Numeric	Mean: 54,140 (SD = 74,400)

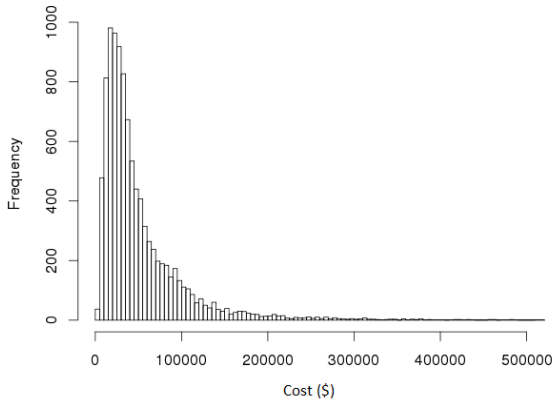


Figure 2.1: Distribution of hospital readmission cost in the readmission dataset

2.6 Result Analysis

We evaluated the machine learning methods from Section 2.4 for the risk and cost of hospital readmission prediction problems described in Section 2.3 on the dataset described in Section 2.5. In this section we present the results and discuss the key observations.

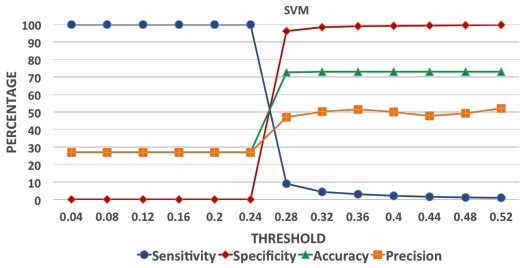


Figure 2.2: Risk prediction performance results of SVM

2.6.1 Risk Prediction

The results of the five machine learning algorithms as well as the LACE baseline, are presented in Figures 2.2-2.5 and Table 2.3. Among the existing risk prediction tools, the LACE index is regularly used in hospitals [67]. This index considers four numerical variables, namely length of stay (L), acuity level of admission (A), comorbidity condition (C), and use of emergency rooms (E). A LACE score is obtained by summing up the values of these four variables. A threshold (usually ≥ 10) is then set to determine patients with "high" readmission risk [67]. We use LACE as a baseline to compare the performance of the machine learning algorithms we investigate in this paper.

We evaluated the models developed with all five machine learning methods using 10-fold cross-validation across different threshold values (see Figures 2.2-2.5). This was done so that a threshold value that would give the highest possible sensitivity, but at the same time also have comparable specificity to that of the LACE tool can be identified. It should be noted that for the 30-day risk of readmission prediction task, higher sensitivity is more desirable. This is because correctly identifying the "high risk" patients who are likely to be readmitted within 30 days is more crucial than correctly identifying low risk patients (discussed in Section 2.1). Overall results corresponding to the best threshold values for all the models are shown in Table 2.3, and Figure 2.6 shows the trade-off between sensitivity and specificity for all the models.

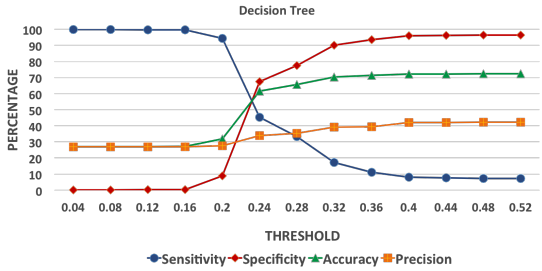


Figure 2.3: Risk prediction performance results of Decision Trees

Algorithm	**Sensitivity (%)**	**Specificity (%)**	**Precision (%)**
LACE	76.42	38.95	31.63
SVM	98.11	1.84	26.98
Decision Trees	94.07	9.04	27.65
Random Forest	84.76	25.60	29.63
Logistic Regression	92.47	13.24	28.26
GBM	90.43	18.24	29.02

Table 2.3: Performance comparison of different machine learning methods for the task of predicting whether the next hospital admission of a patient will be within 30 days. The results are based on 10-fold cross-validation.

There are three key observations to be made from these results. First, the results in Table 2.3 suggest that most machine learning methods show promising results when compared to the baseline LACE method. Not only was it possible to achieve higher sensitivity than LACE, but this was done without penalizing the specificity and precision too much. More in detail, with 3 out of the 5 machine learning methods we achieved a sensitivity of over 80%, while the results for specificity and precision remained comparable to that of LACE (sensitivity = 76%). The sensitivity results for the other two methods, namely SVM and decision tree,

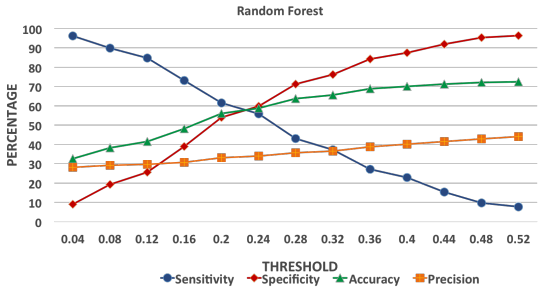


Figure 2.4: Risk prediction performance results of Random Forest

were also very high (sensitivity $\geq$ 94%) and the precision score was comparable, but the proportion of "low risk" instances which were correctly identified was very low (specificity $\leq$ 9%).

Second, the rate of change in sensitivity and specificity slightly differs across different machine learning methods. For instance, as the threshold values increase, the sensitivity and specificity in the decision trees, logistic regression, and generalised boosted models exhibit sigmoid curves (see Figure 2.3, and 2.5), characterized by a small progression in the beginning and then accelerating and converging over larger threshold values. For random forest models, the change is almost linear (Figure 2.4). For SVM a steep drop in sensitivity and rise in specificity is observed between 0.24 to 0.28 threshold values (Figure 2.2). Further investigation of the SVM results showed that there was big increase in the number of true negatives and false negatives around these threshold values, illustrating that SVM is less robust than the other methods, and that its good performance depends on fine tuning of the cutoff threshold.

The third key observation is that, across all methods, 50-60% is the maximum score that can be achieved when a "perfect" balance across all measures (sensitivity, specificity, accuracy and precision) is desired. This is a meaningful result because it shows that the machine learning methods can give good performance for the majority ($>$ 50%) of instances across all measures. It is interesting to observe in Figures 2.2-2.5 that this optimal point of

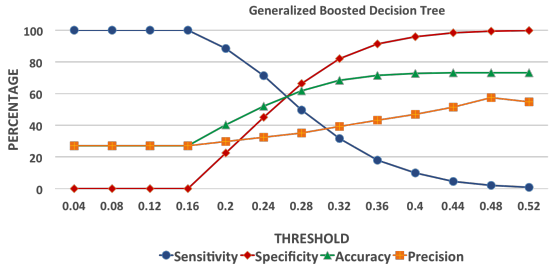


Figure 2.5: Risk prediction performance results of GBM

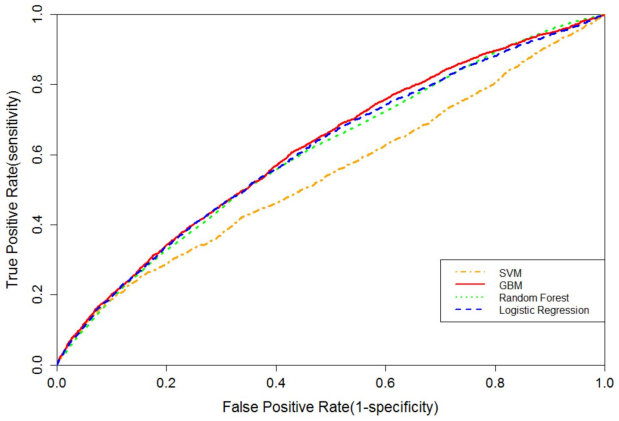


Figure 2.6: ROC curve comparing risk prediction performance results. It shows the trade-off between sensitivity and specificity. It can be seen that GBM is the best classifer, while SVM is the worst.

balance emerges within the same small range of threshold values across all different machine learning methods.

Overall, for the risk prediction task, the results for most machine learning methods for any type of readmission ("all-cause") are promising when compared to a standardized risk prediction tool (LACE). It was possible to achieve higher sensitivity (recall) without penalizing the specificity and precision too much. Improving the precision and specificity further will be a task to explore in future.

2.6.2 Cost Prediction

We measured the performance of the methods for cost prediction using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). A lower error indicates that the predicted dollar amount is closer to the actual cost. As for risk prediction, we evaluated all models using 10-fold cross-validation. An overview of MAE and RMSE results is presented in Table 2.4. We compared the results of four machine learning methods, namely linear regression, M5 model tree, generalised boosted model and decision tree, with those of two baseline methods:

- **Average Baseline (AB):** the Average Baseline (AB) measure is the overall mean cost μ of individual average encounter costs for all the beneficiaries within the training set prior to the current encounter for which we are predicting the cost. This mean (μ) score is then used as the baseline predicted cost for all patient-encounter pairs in the test set.

- **Current Admit Cost (CAC):** the Current Admit Cost (CAC) baseline model is a linear regression model fitted using only the current admission cost during the training period as a predictor variable, with next readmission cost being the response variable. Note that the difference between this current admit cost baseline and the competing linear regression model is that all features (as shown Table 2.2) from the readmission dataset were used to train the linear regression model, while only the 'current admit cost' variable was used in the CAC baseline model.

As can be seen in Table 2.4, for all-cause readmissions, our data mining models exhibit lower prediction error compared to the Average Baseline (AB) method in terms of both MAE and RMSE. Within that, M5 model tree has the lowest prediction error. The errors for the Current Admit Cost (CAC) baseline method were interestingly enough comparable to the errors of several of the more sophisticated methods.

Overall, two key observations can be made from the performance results shown in Table 2.4. First, current admit cost and average cost are strong baseline models, and therefore

Table 2.4: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) in dollars for the cost prediction task

Algorithm	**MAE ($)**	**RMSE ($)**
Average Baseline (AB)	21,609	27,176
Current Admit Cost (CAC)	20,882	26,458
Linear Regression (LM)	20,232	26,124
M5 Model Tree (M5)	18,263	24,824
Generalised Boosted Model (GBM)	20,065	26,388
Decision Tree (DT) (cp=0.01)	20,512	26,328

current hospital admission cost or average cost alone can be a good indicator for the next readmission cost provided it is available to the care provider. Second, among all machine learning algorithms, M5 model tree performed best and achieved a substantially lower MAE than strong baseline models for predicting next readmission cost.

The knowledge that the cost distribution in our dataset is highly skewed (see Figure 2.1), as is known to be the case with healthcare costs in general, inspired us to delve deeper into investigating for which fraction of the patients our models can predict costs with error margins that are reasonably bounded. To this end, we divided the population into 11 different cost buckets, shown on the horizontal axis in Figure 2.7. The cost buckets range from the 5% lowest cost patients (subpopulation 0-5%) to the 10% highest cost patients (subpopulation 90-100%). Next, for each of our cost prediction methods, we measured the average MAE over all patients within each subpopulation. The results are shown in Figure 2.7.

It is interesting to observe that all methods display a similar behavior: the predictions across all models are most accurate within the middle of the range, i.e. for moderate cost patients (40-70%). For the low cost patients (0-40%), the machine learning techniques clearly outperform the baseline models. This is especially the case for the M5 model tree. For the

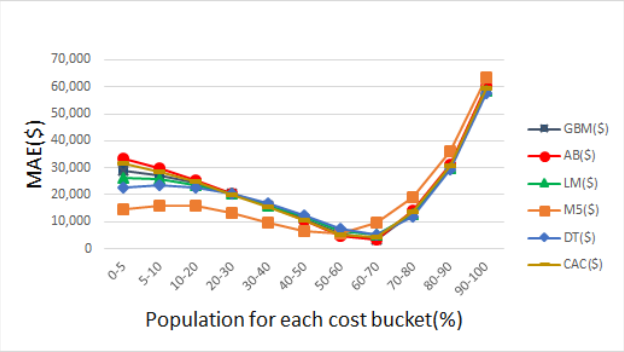


Figure 2.7: Comparison of average Mean Absolute Error (MAE) across different cost buckets.

high-cost patients (70-100%) it is interestingly enough the other way around, although the difference in error between the different techniques is relatively small compared to the size of the actual healthcare costs in this case. Still, the results in Figure 2.7 indicate that it might be beneficial to train a hierarchical model that first predicts a cost bucket and then uses a model trained specifically for that cost bucket to arrive at a final prediction in dollars.

2.7 Summary

The rate of hospital readmissions of patients is a key measure that is tracked for numerous reasons. Consequently, risk stratification of a population and readmission models are becoming increasingly popular. Recent data mining efforts either predict healthcare costs or risk of hospital readmission, but not both. The goal of this study was a dual predictive modeling effort that utilizes healthcare data to predict the risk and cost of any hospital readmission ("all-cause"). For this purpose, we explored machine learning algorithms to do accurate predictions for risk and cost of 30-day readmission. For the task of risk prediction, results for most machine learning methods for any type of readmission ("all-cause") were promising when compared to a standardized risk prediction tool (LACE). It was possible to achieve higher sensitivity (recall) without penalizing the specificity and precision too much. On the cost prediction side, two key observations were made from the performance results

of the machine learning methods. First, average admission cost and current admission cost are strong predictors, and therefore they alone can be a good indicator for the next readmission cost. Second, among the four machine learning algorithms, M5 model tree consistently performed better and achieved a substantially lower MAE than strong baseline models for predicting next readmission cost.

Chapter 3

MULTI-OBJECTIVE CONGESTIVE HEART FAILURE MEDICATION RECOMMENDATION

3.1 Introduction

Heart failure is one of the most prominent health care challenges facing the world today. Across the U.S. approximately six hundred and ten thousand (610,000) patients die from it, five hundred and fifty thousand (550,000) new cases are diagnosed each year, and more importantly, over five million (5M) americans currently live with a specific type of heart failure: Congestive Heart Failure (henceforth referred to as CHF). Heart Failure alone is responsible for 11 million physician visits each year in the U.S., which is more than all forms of cancer combined. However, studies indicate that majority of these are readmissions within a short window of time [50, 55]. Reasons for readmissions vary from early discharge of patients, improper discharge planning, and poor care transitions. Factors that could be externally controlled (or administered) are construed as *interventions* and are applicable at different phases of a patients life cycle. In particular, studies[50, 55] have shown that targeted interventions during pre and post discharge [56] phases such as home based follow up [50], patient education [36], and administering appropriate procedures during the hospital stay, can reduce the readmission rates considerably, and improve patient health significantly.

Machine learning approaches for care management of heart failure patients have been catered towards predicting the 30-day readmission risk for CHF patients [70, 47, 68, 69]. A related effort recommends procedural interventions for CHF patients to reduce 30-day hospital readmissions [42] but does not cover medications. Also, while reducing readmission is important, there are other equally or even more important measures for value based accountable care such as *the effect of recommended medication on the patientś mortality*

rate, and subsequent length of stay at the hospital. However, to the best of our knowledge, no prior work exists for medication recommendation, or for simultaneous optimization of multiple objectives.

Individual optimization of these objectives is not ideal because focusing on the optimization of one objective at a time can sometimes have a contrasting effect on other objectives. A strategy that optimizes multiple such objectives significantly aids the physician as well as the care line managers who want to move towards value based care. Recent studies investigating the relationship between length of stay, mortality rate, and readmission rate for heart failure patients from an acute care hospital setting have shown that a greedy approach of reducing each objective is not an ideal solution. We summarize some of these findings below:

- Shorter length of stay (LOS) has been associated with higher risk of readmission [64, 9].
- Increased Risk of Mortality is found among patients cared for by physicians with short Length of Stay Tendencies [58].
- Average LOS has dropped from 9.3 days in 1988 to 4.9 days in 2006. However, financial pressures to reduce Length of stay has led to poorer patient outcomes, namely higher risk of readmission and mortality [58].

Consequently, we argue that the ultimate goal of reducing LOS, mortality rate and readmissions is achievable only if one takes a principled approach to all of the objectives.

Example 1. Consider a male, Caucasian patient, aged 66. He suffers from Chronic Obstructive Pulmonary Disease (CO-PD), Peptic Ulcer (PU) and Iron deficiency (ID). Consider the following three medications are available: LISINOPRIL-HCTZ TABS 20-25 MG OR (MED.28364), FUROSEMIDE TABS 20 MG OR (MED.8864), METOLAZONE TABS 10 MG OR (MED.28682). These nine factors are predictors and we wish to learn how they relate to the likelihood of the three response variables, being, 30 day Heart Failure readmission, Seven day Length of Stay and Twelve month Mortality. Our objective is to find

the ideal medication combination from a pool of valid medication combinations which take a principled approach to all three objectives.

Problem Definition 1. *Create a Multi-Objective medication recommendation framework that makes personalized medication recommendations for CHF patients to simultaneously minimize* 30*-day readmission, 12-month Mortality and 7-day Length of Stay at the next hospital visit.*[1].

To recommend medications, we consider a multitude of factors, such as, socio-demographic factors, co-morbidities or chronic conditions, and medications.

Our framework relies on the following four steps: (1) Since we deal with very high dimensional data involving several hundreds of factors, we first attempt to *learn the structure of the network automatically* from the data itself. For structure learning, our designed solution appropriately adapts Constraint Based Bayesian Network Learning algorithm [53, 10], *Score Based Learning Algorithm* [32] and *Hybrid Algorithm* [62] that combines both Constrained based and Score based approaches. (2) Once the structure is defined, we use parameter learning [27, 5] techniques to compute the probability of the directed edges. (3) We then perform Bayesian Inference to compute the probability of 30-day Readmission, 7-day subsequent Length of Stay and 12-month mortality for all possible medication combinations from within our pool of medications. In doing so, we get probability scores for each response variable, for each condition. (4) Finally, since our framework tries to optimize multiple objectives simultaneously, we perform Multi-Objective Optimization[13] to ensure that the recommended medications takes a principled approach to optimize all three objectives.

A sample scenario for a patient is explained in Example 1 below, for whom the structure may look like the one presented in Figure 3.1.

---

[1]The durations of 30-day for Readmission, 7 day for Length of Stay and 12 Month for Mortality were chosen after careful discussion and advice from medical experts. This is considered *clinically meaningful* by different healthcare services and standards [1].

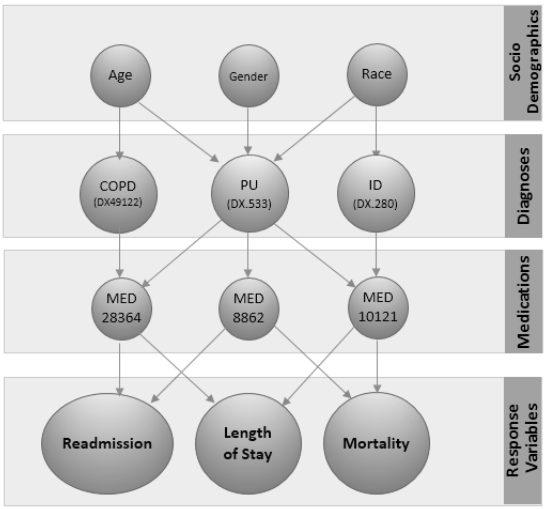


Figure 3.1: A learned network structure considering Example 1

Our framework is flexible in that it is possible to specify constraints coming from a domain expert or medical practitioner. For example, if a physician feels that a connection between Gender and presence or absence of Pneumonia is not meaningful, such relations can be blacklisted. Also, since we are working on recommending medications, it is possible that certain medication combinations taken together can cause side-effects or have a bad patient outcome. Such medication combinations should therefore be discarded. The medication pool considered in our study comprises of the most important medications prescribed to CHF patients, as recommended by a clinical pharmacist from MultiCare Health System, a not-for-profit health care organization based in Tacoma, Washington. In addition, we also study an alternate approach where medications considered are the ones that show highest co-occurrence with CHF when queried against an online repository of biomedical literature called Pubmed.[2]

Finally, we present comprehensive experimental results to validate the effectiveness of our proposed framework and present a case study validated by a domain expert. We use the

[2]http://www.ncbi.nlm.nih.gov/pubmed

data from Multicare as well, to validate the effectiveness of our proposed framework.

The contribution of our work may be summarized as:

1. We initiate the study of recommending medications that take a principled approach towards minimizing 30-day readmission, twelve-month mortality and seven-day length of stay for CHF patients, which has not been studied previously to the best of our knowledge.

2. We convert the problem to a hierarchical Bayesian Network learning task and perform Pareto Optimization [46] to enable the Multi-Objective decision support.

3. We present comprehensive experimental results and verify the correctness of these claims by talking to a domain expert and present these results as a case study.

The rest of the paper is organized as follows: Section 3.2 describes the proposed medication recommendation framework. Section 3.4 discusses the experimental setting, empirical evaluation, and case study results. We summarize related works in Section 3.5. Section 3.6 concludes the paper and indicates directions for future work.

3.2 Solution

A hierarchical Bayesian network effectively depicts the relationships between the patient attributes and how their interplay affects the response variables (Readmission, Length of Stay and Mortality). Thus, we model the medication recommendation as a network learning task using the Bayesian network learning principles. After that, we consider all valid medication combinations from our medication pool and use Bayesian Inference to compute the conditional probability of each of the response variables being true, given this medication combination and other patient attributes. By doing so, we get an individual probability value between 0 and 1 for each of the response variables, which we call *Risk scores*. After that, we perform Multi-Objective Optimization to find an optimal balance between these risk scores from above, and recommend the medication combination(s) that forms the pareto

set [46]. We thereby succeed in recommending a medication combination that simultaneously optimizes all three response variables, ensuring the patient a successful discharge experience.

Figure 3.2 describes the high level design of the framework.

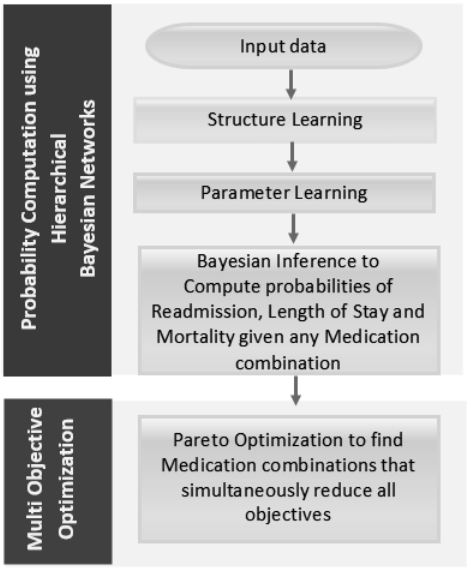


Figure 3.2: Design of the proposed framework

3.2.1 Bayesian Network and notations

A Bayesian Network represents the relationship between a set of random variables in a system as a directed acyclic graph, along with the probability density functions (PDFs) for each node in the network. The structure of a Bayesian Network implies that the value of a particular node is conditional only on the values of its parent nodes, and so together with the network structure, the PDFs are sufficient to represent the joint probability distribution of the domain

$$Pr(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{i=n} Pr(X_i | P_a^{X_i}) \tag{3.1}$$

3.2.2 Structure Learning

Our structure learning solution relies upon the *Causal Sufficiency Assumption* and the *Markov Assumption*[59]. We use *Constraint Based, Score-Based, and Hybrid* methods to learn the structure of the network. We describe them briefly one by one next.

3.2.2.1 Constraint Based Methods

These methods make use of the conditional independence tests using statistical tests on the data set. We use a computationally efficient algorithm, *Grow and Shrink* [8] which relies on detecting the *Markov Blanket* [32] of the variables to induce the network structure. Markov blanket for a node X in a Bayesian network is the set of nodes composed of X's parents, it's children, and it's children's other parents. It operates by identifying the local neighborhood of each variable in the Bayesian network as a preprocessing step, in order to facilitate the recovery of the exact structure around each variable in subsequent steps. This has the added advantage of being easier to verify and possibly correct a *posteriori* by an outside expert. The overall algorithm runs in two phases : in the *grow phase*, given a node X, it starts with an empty set S and it adds variables to S as long as they are dependent (i.e., Markov Blanket property is violated) with X given the current contents of S. In this process, however, there may be some variables that were added to S that were really outside the blanket. In the *shrinking phase*, those variables are identified and removed.

3.2.2.2 Score-Based Methods

Constraint-based algorithms suffer from poor "robustness", i.e., large effects on the output of the algorithm is observed, for small changes of the input i.e. single errors in the independence tests. To overcome that shortcoming, we apply Score-based approaches that uses hill climbing heuristics.

Score-based methods creates several Bayesian Network and assigns a score to each candidate of them, typically one that measures how well that Bayesian network describes the

data set $\mathcal{D}$. Assuming a structure $\mathcal{R}$, its score is,

$$Score(\mathcal{R}, \mathcal{D}) = Pr(\mathcal{R}|\mathcal{D}) = \frac{Pr(\mathcal{D}|\mathcal{R}) \times Pr(\mathcal{R})}{Pr(\mathcal{D})}$$

As Score-based algorithms attempt to maximize this posterior probability, returning the structure $\mathcal{R}$ that maximizes it is prohibitively expensive. Since the search space of all possible structures is exponential to the number of variables n, this posses tremendous computational challenges. In our solution, we apply hill climbing based greedy heuristics and use Bayesian information criterion (BIC) [41] to approximate $Pr(\mathcal{R}|\mathcal{D})$. Our solution performs a local search and does not need the brute-force computations. The search is started from either an empty, full, or possibly random network. After that, it iteratively attempts to add, remove, or reverse every possible single-edge and consider that change which increases the score most. The process stops when there is no single-edge change that increases the score.

3.2.2.3 Hybrid Approach

We finally apply a hybrid approach to learn the network structure, namely the *max-min hill climbing algorithm* [62] which combines ideas from both Score-based and Constraint-based approaches. It first reconstructs the skeleton of a Bayesian network and then performs a Bayesian-scoring greedy hill-climbing search to orient the edges. The latter phase does not provide any theoretical guarantees. However, this algorithm appears effective in many high dimensional real world problem (such as ours) and tackles the limitations posed by the other algorithms.

The resultant network structure generated by these algorithms satisfy any constraint that is already specified, it will include the edges that are pre-specified to exist, and vice-versa.

3.2.2.4 Additional Constraints

However, there are also additional constraints that are employed on the network which come from domain knowledge. We segregate the attributes used in our data set into categories/layers as shown in Figure 3.1. Connections are only allowed between socio-demographic

factors and diagnosis, between diagnosis and medication, and between medication and the response variables.

In our implementation based on medication list from NIH (explained in 3.4.2.2), we have an additional constraint to remove edges between two nodes in the network that show zero co-occurrence when queried against Pubmed.

3.2.3 Parameter learning

3.2.3.1 Background

After the structure of the network is constructed, the next step is to learn the parameters of the network, given the structure. Using Example 1, this step is analogous to creating PDFs to each node in the constructed network to create the conditional probability table at each node. Using the sample network of Example 1, this step will compute all the following probabilities at node MED.28364.

$$\Pr(\text{MED.28364= i— DX 533 =0 \& DX 280=0})$$
$$\Pr(\text{MED.28364= i— DX 533 =0 \& DX 280=1})$$
$$\Pr(\text{MED.28364= i— DX 533 =1 \& DX 280=0})$$
$$\Pr(\text{MED.28364= i— DX 533 =1 \& DX 280=1}), \forall_{i=0,1}$$

Typically, for parameter learning, a prior distribution is assumed over the parameters of the local PDFs before the data is used (for example, this can be uniform), or it could be estimated using the given data itself. The distribution of a node X conditional upon its parents may have any form. The *conjugacy* of this prior distribution is also desirable, meaning that the posterior belongs to the same family as the prior, albeit with different parameters.

In our implementation, we use Bayesian Parameter Estimation [37] to learn the parameter θ. In this method, the prior distribution over θ (i.e., $Pr(\theta)$) is known. Now the posterior distribution of θ is calculated according to Bayes rule:

$$P(\theta|D) = \frac{Pr(D|\theta)Pr(\theta)}{\int Pr(D|\theta)Pr(\theta)d\theta}$$

Our objective is to calculate the Maximum A Posteriori (MAP in short), i.e.,

$$\hat{\theta}_{MAP} = argmax_{\theta}P(\theta|D) = argmax_{\theta}Pr(D|\theta)Pr(\theta)$$

The prior $Pr(\theta)$ is calculated using a Beta-Distribution for binary variables which gives rise to a posterior which is also a Beta distribution.

3.2.4 Bayesian Inference

Once the network structure has been learnt, Bayesian Inference is applied to find conditional probabilities of Readmission, Length of Stay and Mortality given a medication combination. From within the medication pool, we generate all triplets of medications, since a set of 3 is the average medication size, as told by the clinical pharmacist. Since our medication pool consists of 57 (explained in Section 3.4.2.2) medications, this gives us ${}^{57}C_3$ valid medications combinations (29260) that can be used to apply Inference on a given patient. However, as earlier mentioned, not all of these medications can be used in combination with each other. Upon excluding these combinations, we get a total of 15829 valid three-medication combinations.

We then apply Bayesian Inference for the patient details in Example 1

- P(**Readmit** — Age=66, Gender=Male, Race = Caucasian, DX 280=1, DX 533=1, DX 49122=1, MED.28364=1, MED.8862=1, MED.10121=1)
- P(**LOS** — Age=66, Gender=Male, Race = Caucasian, DX 280=1, DX 533=1, DX 49122=1, MED.28364=1, MED.8862=1, MED.10121=1)
- P(**Mortality** — Age=66, Gender=Male, Race = Caucasian, DX 280=1, DX 533=1, DX 49122=1, MED.28364=1, MED.8862=1, MED.10121=1)

As the output of the Inference stage, we get risk scores (probabilities) of each of the response variables, given each available medication combination, for a given patient. In the final step of Multi-Objective Optimization [46], we find the medication combinations which give rise to the Pareto Set.

3.3 Multi-Objective optimization

Example 2. Consider the patient mentioned in Example 1. Imagine that 6 different medication combinations are possible to be administered to this patient which might have the following risk scores (refer to Table 3.1).

Consider only the first two combinations of medications (not all 6) and it is not obvious how these two combinations should be compared. Combination-two has a mean of 22 percent, which is less than Combination-one's mean of 30. However, we cannot say with certainty that the latter combination provides a better treatment plan. This is because, here, Combination-one has a smaller LOS score than that of the other combination, which simply makes these two combinations incomparable with each other. Thus, both of these combinations need to be part of the solution, since neither of these two *dominates* the other in all three objectives.

To further formalize the aforementioned observation, we are dealing with multiple objectives which are potentially conflicting in nature. Hence, the problem is studied as a

SLN	Medication Combination	Risk Scores		
		Readmit Score	LOS Score	Mortality Score
1	MED.28364, MED.8864, MED.20106	25	30	34
2	MED. 40140, MED.39800, MED.28542	15	31	21
3	MED.28364, MED.8864, MED.28682	25	30	10
4	MED.13448, MED.17700, MED.8865	10	40	21
5	MED.6501, MED.6478, MED.25477	30	30	34
6	MED.28689, MED.28364, MED.25476	9	20	15

Table 3.1: Different Risk Score Combination for Example 1

Multi-Objective Optimization [13] problem. For a nontrivial Multi-Objective Optimization problem such as ours, there may not exist a single solution that simultaneously optimizes each objective (Refer to Example 2 above). In that case, the objective functions are said to be conflicting, and there exists a number of Pareto optimal solutions. A solution is called *Pareto optimal* [46], if none of the objective functions can be improved in value without degrading some of the other objective values. Without additional information, all Pareto optimal solutions are considered equally good (as vectors cannot be ordered completely). Our goal, therefore is to find a representative set of Pareto optimal solutions.

3.3.0.1 Preliminary

Multi-Objective Optimization, which is also called multicriteria optimization or vector optimization, has been defined as finding a vector of decision variables satisfying constraints to give acceptable values to all objective functions.

In general, it can be mathematically defined as: find the vector $X^* = [x_1^*, x_2 * *, \ldots, x_m^*]^T$ to optimize

$$F(X) = [f_1(X), f2(X), \ldots, f_k(X)]^T$$

where $X^* \in \mathfrak{R}^n$ is the vector of decision or design variables (risk scores), and $F(X) \in \mathfrak{R}^k$ is the vector of objective functions, which must each be either minimized or maximized.

We formally define some important concepts related to pareto optimization first.

Definition 1. **Pareto Dominance:** A vector $U = [u_1, u_2,$
$\ldots, u_k] \in \mathfrak{R}^k$ is dominant to vector $V = [v_1, v_2, \ldots, v_k] \in \mathfrak{R}^k$ (denoted by $U \prec V$) if and only if $\forall i \in \{1, 2, \ldots, k\}$, $u_i \leq v_i \wedge \exists j \in \{1, 2, \ldots, k\} : \ u_j < v_j$. In other words, there is at least one u_j which is smaller or equal to corresponding v's.

Definition 2. **Pareto Optimality:** A point $X^* \in \Omega$ (Ω is a feasible region in $\mathfrak{R}^n$) is said to be Pareto optimal (minimal) with respect to all $X \in \Omega$ if and only if $F(X^*) \prec F(X)$. Alternatively, it can be readily restated as $\forall i \in \{1, 2, \ldots, k\}$, $\forall X \in \Omega - \{X^*\} f_i(X^*) \leq f_i(X) \wedge \exists j \in \{1, 2, \ldots, k\} : \ f_j(X^*) < f_j(X)$. In other words, the solution $X*$ is said to be Pareto optimal (minimal) if no other solution can be found to dominate X^* using the definition of Pareto dominance.

Definition 3. **Pareto Set:** For a given MOP, a Pareto set $\mathcal{P}^*$ is a set in the decision variable space consisting of all the Pareto optimal vectors $\mathcal{P}^* = \{X \in \Omega | \nexists X' \in \Omega : \ F(X') \prec F(X)\}$. In other words, there is no other X' as a vector of decision variables in Ω that dominates any $X \in \mathcal{P}^*$.

3.3.0.2 Obtaining Pareto Optimal Set

Once we get the conditional probabilities from the previous step of Bayesian Inference, the Pareto Optimization Algorithm finds the medication combination(s) which constitute the Pareto Set.

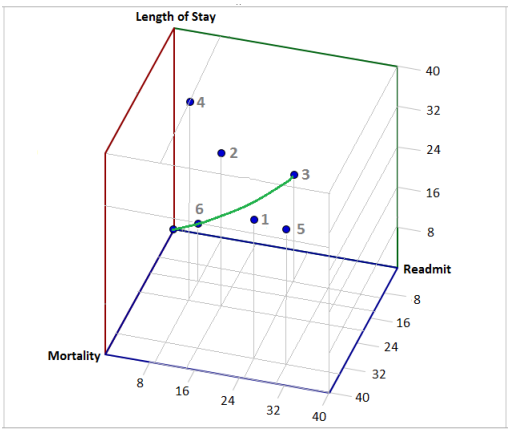


Figure 3.3: Pareto Set for Example 2

The pseudo-code can be found in Algorithm 1. Initially the first input combination is assumed to be part of the Pareto Solution Pool (referred to as *Pool*, hereafter) (line 1). We start traversing from the second input combination and see if it is a Pareto Dominant Solution (lines 7-10). If yes, the non-dominant solution from the *Pool* is replaced by this Pareto Dominant Solution. If no, we check whether it is at least Pareto Optimal (lines 12-15). If found to be Pareto Optimal (not dominated by anything else), this solution is added to the *Pool*. We thus form a Pareto Solution Pool comprising of the Pareto Dominant and Pareto Optimal solutions, which our framework demands(lines 18-20).

Considering Example 2, the following are the observations based on the algorithm (Figure 3.3 illustrates them visually).

(a) The combination that gives risk scores (25,30,10) dominates (25,30,34) and (30,30,34). (b) The combination that gives risk scores (9,20,15) dominates (15,31,21) and (10,40,21). (c) Neither (25,30,10) nor (9,20,15) dominates each other, so both are retained in the Pareto

Set [46]. Finally, *3* and *6* are selected, and the framework recommends both of these combinations.

Complexity Analysis: If there are k possible combinations of medications and m different objective functions, the worst case complexity of Algorithm 1 is $O(k^2 \times m)$.

Algorithm 1 Pareto Optimization Algorithm

```
1:  ParetoSolutionPool[1] = input[1];
2:  for (i in 2 :input) do
3:      boolean ParetoDominant = false;
4:      boolean ParetoOptimal = true;
5:
6:      for (j in 1:ParetoSolutionPool) do
7:          if (input[i] ParetoDominates ParetoSolutionPool[j]) then
8:              remove ParetoSolutionPool[j];
9:              ParetoDominant = True;
10:
11:         if (ParetoDominant == false and input[i] isParetoDominatedBy ParetoSolutionPool[j]) then
12:             ParetoOptimal = false;
13:             break;
14:
15:     if (ParetoDominant == True —— ParetoOptimal == true) then
16:         add input[i] to ParetoSolutionPool;
```

3.4 Experiment

Our empirical analyses are on an 8 core Windows Server with each core having 16GB of RAM. The development environment comprises of R and Java.

3.4.1 Dataset

We use data from MultiCare hospital system for the heart failure cohort. This is dynamic data that is constantly populated, as new patients make a hospital visit. Our dataset contains patient records from the year 2010 until present.

Dataset Preparation The processed data set consists of $74,009$ patient encounter records with 149 attributes including the response variables. Each admission record consists

of socio-demographic factors (eg Age, Gender, Ethnicity), comorbidities/chronic conditions (Diabetes, Renal failure, Pneumonia), and medication information. Table 3.2 summarizes the dataset used in the study.

Group	**Category**	**Description**
1	Demographics	Age, Gender, Race, Marital Status (4 attributes)
2	Clinical Readings	16 Clinical Readings (Hemoglobin level, Creatinine level etc)
3	Medical Diagnosis	37 diagnosis (Arrhythmia, Peptic Ulcer, Fluid Disorder etc)
4	Medications	57 Medications used by CHF patients
5	Others	Length of stay, # secondary diagnosis (2 attributes)
6	Response Variables	30-day Readmission flag, 7-day Length Of Stay on next visit flag, 12-Month Mortality flag(3 attributes)

Table 3.2: Attribute summary for heart failure cohort

3.4.2 Implemented algorithms

For our Bayesian Network, we implement 3 different structure learning algorithms, namely Hill Climbing (HC), Grow-Shrink (GS) (default), Hybrid (HY) and compare them with a baseline implementation, described below.

Baseline Algorithm For a given patient for whom we would like to recommend medications, our baseline method consists of finding the closest match to a previous patient who had a successful treatment plan, using Jaccard index [2], and comparing his actual medications with that of the matched patient.

Our dataset is randomly partitioned as train data and test data in 80% and 20% split. In our train dataset, we find all instances of the patients who had an ideal outcome (i.e., who

did not have a 30-day Readmission, 7-day Length of Stay on next hospital visit, and death in the 12-Months following discharge) upon using the medications that were recommended. We call these instances the *Ideal Outcome Instances*.

For each test tuple in the test data, we identify the closest input vector in the *Ideal Outcome Instances* set using Jaccard index [2] computed as:

$$J(\mathcal{A}, \mathcal{B}) = \frac{|\mathcal{A} \cap \mathcal{B}|}{|\mathcal{A} \cup \mathcal{B}|}$$

where A denotes the vector comprising of attributes in Group 1, 2, 3 and 5 described in Table 3.2 for the *Ideal Outcome Instances* set and B denotes the input vector (comprising the same set of attributes ad A) observed in the test data.

3.4.2.1 Medications List

To come up with a comprehensive list of valid medication combinations, we collaborate with a pharmacist at Multicare. We reduce the 8000 medications that have been historically prescribed to heart failure patients at MultiCare to a set of the 57 most important medications. The rest of our discussion considers only these medications set. What is challenging though is that these medications may appear together in many possible combinations that the framework must account for during recommendation. Note that we also treat the two different dosages of the same medication as two different medications, as that is most clinically meaningful.

Moreover, not all of these medications can be used in combination with each other. With the help of our domain expert collaborator, certain medication combinations are blacklisted, as using them together could cause adverse side effects. The blacklisted medication combinations are shown in Table 3.4.

Medication Name	**Dosages**
Lisinopril oral	2.5-40 mg
Losartan oral	25-150 mg
Metoprolol tartrate oral	12.5-400 mg
Metoprolol succinate oral	12.5-400 mg
Carvedilol oral	3.125-25 mg
Furosemide oral or IV	20-160 mg
Bumetanide oral or IV	0.5-10 mg
Torsemide oral	5-200 mg
Potassium chloride oral	10-80 mEq
Spironolactone oral	12.5-200 mg
Digoxin oral	0.0625-0.5 mg
Hydralazine oral	25-300 mg
Isosorbide mononitrate oral	40-120 mg daily
Isosorbide dinitrate oral	20-120 mg
Dobutamine IV infusion	-
Milrinone IV infusion	-

Table 3.3: Medications and Dosages

Invalid Combinations	
Lisinopril	Losartan
Metoprolol Succinate	Carvedilol
Metoprolol Tartrate	Carvedilol
Furosemide	Torsemide
Furosemide	Bumetanide
Torsemide	Bumetanide
Isosorbide Mononitrate	Isosorbide Dinitrate

Table 3.4: Blacklisted Medication Combinations

3.4.2.2 Medication List from NIH

We consider an alternate approach to find the most important set of medications that are relevant to Congestive Heart Failure and the response variables, namely, 30-day Readmission, 7-day Length Of Stay and 12-Month Mortality, by querying Pubmed. We make use of a Web API [3] which counts the number of citations for each medication from our entire medication pool, and its co-occurrence with each response variables. The medications are then ranked in descending order of citations, and the top 50 fifty medications are used in the study for Approach 2.

3.4.3 Case Study

We compare the medications recommended by our framework with the actual medications that are prescribed to five randomly selected patients who had an ideal outcome, i.e. they

[3]Base URL : http://www.ncbi.nlm.nih.gov/pubmed/?term=

Medication Name	Number of citations			
	HEART FAILURE	**READMISSION**	**LENGTH OF STAY**	**MORTALITY**
DIGOXIN	3420	39	76	1148
DOPAMINE	1339	8	134	1632
DOBUTAMINE	1730	23	86	988
AMIODARONE	1247	9	120	1277
CARVEDILOL	1444	16	12	639
METOPROLOL	1037	10	28	740
LOSARTAN	749	2	2	505
MILRINONE	631	13	36	234
SPIRONOLACTONE	1592	33	24	672
FUROSEMIDE	1752	51	124	606

Table 3.5: Medications and Citations from Pubmed

did not have a 30-day readmission, a 12-month mortality or a length of stay more than 7 days on the next admission.

3.4.3.1 Results

A survey of questions regarding the usefulness of the medication recommendations is used. In Approach 1, the medication pool consisted of the most frequent CHF medications recommended by the pharmacist(as explained in 3.4.2.1). In Approach 2, the medication pool consisted of those medications that returned the highest number of citations when queried against Pubmed (as explained in 3.4.2.2). For both Approach 1 and 2, the implementation involved Bayesian Structure Learning (Grow-and-Shrink), Parameter Learning, Bayesian In-

ference and Pareto Optimization. In Approach 3, the medication pool consisted of those recommended by the pharmacist using the baseline model(as explained in 3.4.2).

The pharmacist gives scores in a scale of 1 to 5 (low to high). The results in Table 3.7 demonstrate that our proposed framework is effective. Also, the medications suggested by the pharmacist are more relevant to heart failure than those discovered by querying Pubmed.

3.4.4 Empirical Study

As an alternate approach, we use the constructed Bayesian network to generate a set of recommendation rules. Our experiments used 59207 patients (80 % of the dataset) for training and 8203 patients (Ideal Outcome Instances from within 20 % of the dataset) for testing.

Without loss of generality, let us assume that a total of $|\mathcal{X}'|$ of $|\mathcal{X}|$ factors are medications, and the remaining set $\{\mathcal{X}\} - \{\mathcal{X}'\}$ of factors are other attributes describing the patient condition.

For each patient record d whose actual class label is 0 (i.e.,$Readmission = 0$), we use only $|\{\mathcal{X}\} - \{\mathcal{X}'\}|$ attributes of record d (denoted as $d(\mathcal{X} - \mathcal{X}')$) and feed it through the constructed network to obtain the inference probability p_1. Then, we use the entire patient record (with both medication and non-medication attributes), and use that to make a second inference probability p_2.

$$p_1 = Pr(Readmit = 0|d(\mathcal{X} - \mathcal{X}'),\ p_2 = Pr(Readmit = 0|d)$$

If $p_2 > p_1$ (which indicates that our constructed model infers that the set of medications associated with the patient input is effective in further bringing down her readmission risk), we store the set of medications $\{\mathcal{X}'\}$ associated with d as the generated recommendation, given the values for the non-medication attributes.Using Example 1, a recommendation rule in our case may look as follows:

- Rule-Example: if Gender = Male & Age = 66 & Race = Caucasian & diagnosis=

COPD & diagnosis= PU & diagnosis= ID & Readmit=0, recommended medications are MED.28364 & MED.8864 & MED.28682.

Similar check is performed to generate rules for patient records associated with actual class label of 1.

3.4.4.1 Evaluation Measures

We use four metrics to evaluate our experiment results: 1) the number of exact matches from the rules of the test data (HIT); 2) accuracy of the recommendations (ACCY); 3) True positive rate (TPR) or Recall [4]; 4) True negative rate (TNR) or Specificity. For each pair of a set of recommended medications and a set of observed procedures, we define a true positive (tp) case, if a recommended medication appears in the observed medication set. We define a true negative (tn) case, if a non-recommended medication does not appear in the observed medication set. False positive (fp) occurs when the recommended medication does not appear in the observed set. False negative (fn) occurs when the non-recommended medication actually appears in the observed medication set. Therefore, we compute ACCY, TPR and TNR as follows:

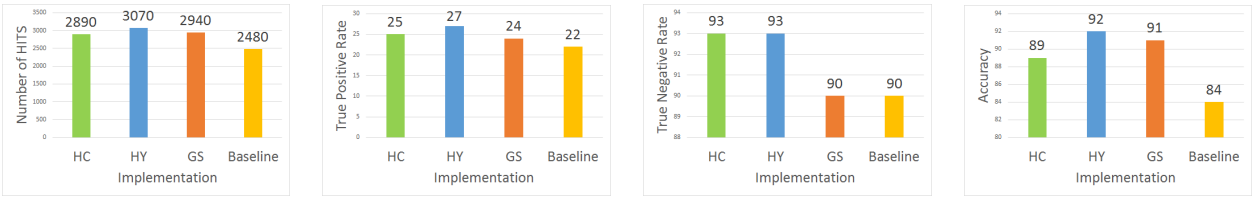
$$ACCY = \frac{tp + tn}{tp + tn + fp + fn}$$
$$TPR = \frac{tp}{tp + fn}$$
$$TNR = \frac{tn}{tn + fp}$$

3.4.4.2 Results

The experimental results are presented in Figure 3.4. Understandably the HIT values are in the lower side for all the algorithms, while the other three measures (especially Accuracy) are high and demonstrates the effectiveness of our proposed methods.

[4]Note that Precision is not relevant in our settings, as it captures the actually correct recommendations out of all recommendation.

Figure 3.4 and Table 3.8 show our experiment results. The results indicate that HY outperforms all other implementations in three of the four metrics (HITS, TPR and ACCY). For TNR, HC and HY show the same percentage rate. The results also demonstrate that all of the implementations outperform the Baseline for all metrics. These results corroborate the effectiveness of different algorithm for the task of Multi-Objective medication recommendation.



(a) Hits (b) True Positive Rate (c) True Negative Rate (d) Accuracy

Figure 3.4: Effectiveness of different algorithms for Medication recommendation

3.5 Related Work

To the best of our knowledge, no prior work has taken a Multi-Objective approach towards medication recommendation. We compare and contrast relevant existing research here.

Bayesian Network in Healthcare:

In one of our previous studies [42], we look into recommending procedures to CHF patients with the objective of reducing 30-day hospital readmission. However, being a single objective problem, Multi-Objective Optimization [13], which is the focus of the current work, was not required. Also, the current problem we are solving focuses on medications which bring along with it a lot of constraints. For example, certain medication combinations can be harmful when used together, and the structure learning component of our Bayesian Networks should be flexible enough to incorporate these constraints. In the proposed framework, we also try and understand the impact of all valid medication combinations on patient outcome, which adds to the computation process.

Medication Recommendation System: There have also been a few attempts to build

medication recommendation systems. The purpose of one such study [34] was to develop a new alerting and recommender system for preventing medication errors. However, this is a not a recommender system in the true sense, since it still requires doctors/physicians to make recommendations. It merely compares the recommendation with statistical data for that same drug, compiled from past records in the database, and sends an onscreen alert to the physician if the alert engine finds that the entered data exceed the threshold from historical data.

Another very recent effort [66] to develop a medication recommendation system uses the semantic web to find relationships between medications and classify them. However, no machine learning approaches are used, and also the evaluation is based only on user ratings, which can sometimes be misleading.

There have also been efforts to create decision support for generating treatment plan for other diseases, such as, such as, coronary diseases [4], ulcers [12], sepsis [23], and depression [35]. Unlike us, none of these work deal with the problem of high dimensionality, scale, or multi-layer modeling.

Application of Multi-Objective Otimization : While Multi-Objective Optimization is traditionally studied as a theoretical research topic [46], we describe a few attempts of how they are handled in *practice* here.

A recent effort in Multi-Objective Optimization [13] looks at how to improve the chance of success of a *public-private partnership (PPP) scheme* [65] where the Multi-Objectivity comes from having different stake holders, and different critical success factors for each stake holder. A pool of solutions is created based on findings from a Bayesian Network. However, the final Multi-Objective solutions are computed based on a weighted average of scores of all the objectives, and choosing the ones that are above a pre-defined threshold. We, however, believe that setting a hard threshold is a slightly naive approach and go a step further in letting the Pareto Optimization Algorithm determine the best solutions. Also, this study is conducted mainly on data coming from user surveys and is on a very small scale.

Another *recent work* [17] uses the Pareto optimization technique to create new keyboard

layouts with the objective of minimizing finger travel distance and spelling errors. However, the scores for these metrics are computed by calculating the Euclidean distance between keyboard keys, and through a spell checking algorithm. No machine learning concepts are used here. Since the healthcare domain we are working on is very dynamic, it is important that we use a machine learning approach so as to discover new causal relationships and trends from underlying data.

In view of the above, it is easy to observe that our effort is orthogonal to these related work.

3.6 Summary

We investigate the problem of recommending medications to simultaneously reduce 30 day readmission, twelve month mortality and seven day length of stay, taking a principled approach to all three objectives. We consider a multitude of factors ranging from socio-demographic factors such as age and gender, diagnosis information, clinical factors and medications used. Our proposed solution that relies on learning the structure of a hierarchical Bayesian Network from such multi-dimensional data, followed by Parameter Learning to compute the conditional probabilities between these factors. Finally, we perform Multi-Objective Optimization(Pareto Optimization) to find the medication combination(s) that optimize all three objectives simultaneously. Our experimental results as well as case studies demonstrate the effectiveness of the proposed framework. Exploration of medication recommendation to minimize other objectives such as *cost* incurred during hospital stay, and expanding the framework to scale for large number of related conditions are fertile grounds for future work. Our solution led to actionable insights for the chronic care management team, and is under consideration for deployment within the physician and pharmaceutical compliance workflows at KenSci Inc.

SLN	Recommended	Actual	HITS	TP	FP	FN	TN	JACCARD
1	• FUROSEMIDE TABS 20 MG OR • METOLAZONE TABS 10 MG OR • LISINOPRIL-HCTZ TABS 20-25 MG OR	• FUROSEMIDE 20 MG OR TABS • METOLAZONE TABS 10 MG OR • HYDRALAZINE HCL TABS 25 MG OR	2	67 %	33 %	33 %	98 %	50 %
2	• LISINOPRIL 20 MG OR TABS • CARVEDILOL TABS 3.125 MG OR TABS • HYDRALAZINE HCL TABS 25 MG OR	• LISINOPRIL 20 MG OR TABS • CARVEDILOL 3.125 MG OR TABS • ISOSORBIDE MONONITRATE TBCR 60 MG OR	2	67 %	33 %	33 %	98 %	50 %
3	• LISINOPRIL-HCTZ TABS 20-25 MG OR • ISOSORBIDE MONONITRATE TBCR 60 MG OR • BUMETANIDE TABS 1 MG OR	• BUMETANIDE TABS 1 MG OR • TORSEMIDE 20 MG OR TABS • ARVEDILOL 3.125 MG OR TABS	1	33 %	67 %	67 %	96 %	20 %
4	• LISINOPRIL-HCTZ TABS 20-25 MG OR • METOPROLOL TARTRATE TABS 50 MG OR • HYDRALAZINE HCL 20 MG/ML IJ SOLN	• LISINOPRIL-HCTZ TABS 20-25 MG OR • METOPROLOL TARTRATE TABS 50 MG OR • HYDRALAZINE HCL 20 MG/ML IJ SOLN	3	100 %	0 %	0 %	100 %	100 %
5	• LISINOPRIL-HCTZ TABS 20-25 MG OR • FUROSEMIDE SOLN 10 MG/ML OR • HYDRALAZINE HCL 20 MG/ML IJ SOLN	• LISINOPRIL-HCTZ TABS 20-25 MG OR • FUROSEMIDE 40 MG OR TABS • SPIRONOLACTONE 25 MG OR TABS	1	33 %	67 %	67 %	98 %	20 %

Table 3.6: Medication Recommendation for five patients obtained using Pareto Optimization

Question	**Response**
1. Average rating for Approach 1 over 5 patients	3/5
2. Average rating for Approach 2 over 5 patients	1/5
3. Average rating for Approach 3 over 5 patients	2/5
4. Which approach complies the most with medical guidelines?	Approach 1
5. Which approach recommends medications that are most relevant to heart failure?	Approach 1

Table 3.7: Survey Response

	HC	HY	GS	Baseline
HITS	2,890	3,070	2,940	2,480
TNR	93	92	90	90
TPR	25	27	24	22
ACCY	89	92	91	84

Table 3.8: Performance of Different Bayesian Network Algorithms

Chapter 4

PATHWAY-FINDER: SYSTEM DEVELOPMENT OF AN INTERACTIVE RECOMMENDER SYSTEM FOR SUPPORTING PERSONALIZED CARE PATHWAYS

4.1 Introduction

Clinical pathways are widely used by hospitals for managing the patient treatment process. Effective clinical pathway analysis captures clinical best practices that are shown to contribute to targeted outcome, such as the optimal length of stay for each patient [29, 30]. While the initial development of clinical pathways is a time-consuming process, requiring the collaborations among hospital physicians, nurses and staff, it is also critical that these pathways continue to evolve, incorporating continued feedback as to how chosen care pathways are impacting quality of care for the patient. The challenge for care providers is that, given the limited visit time with each patient[1], how do they optimize that time, informing treatment plans and providing quality of care, not only for that patient, but also by collected feedback on outcomes to support continued evolution of the care pathways.

We develop `Pathway-Finder`, a novel interactive recommender system for clinical decision support. Through clinical pathways analysis, we identify and gather what is known about the patient (through EMR records and intake form) as well as recommend appropriate interventions that can lead to improved care quality, and report on the efficacy of those interventions.

The proposed interactive system surfaces patient information relevant to that encounter, as well as supporting identification of new factors, which can then be visualized to show the connections between patient demographic characteristics, disease conditions(comorbidities

[1]The average visit duration with physicians less than 20 minutes [22].

or diagnoses), possible interventions, and targeted clinical outcomes. In this demonstration, we use a heart failure cohort provided in [43] to minimize the 30-day risk of readmission for Heart Failure (the targeted outcome). However, the system is flexible enough to support clinical decision support for a variety of disease conditions and targeted outcomes.

The contributions of `Pathway-Finder` are as follows:

- First, the system provides interactive discovery and exploration of clinical pathways analysis;
- Second, the system iteratively collects necessary patient information that drive the development of treatment plan;
- Third, the system visualizes the trace and predicted outcome of a patient, supporting personalized intervention recommendation;
- Last, the system implements a key-value structure on Microsoft Azure for Research platform, supporting real-time interactive visualization.

The modeling and the solution relies on learning the structure and the probability distribution of a *Bayesian Network* [24] from the available patient data. As the number of attributes (or attribute values) increases, the Bayesian Network grows in size, resulting in an exponential number of look ups to perform in order to recommend interventions. To recommend interventions real time, the system, therefore, makes use of a novel representation of Bayesian network which is hosted on a cloud-based infrastructure like Windows Azure for Research.

Section 4.2 describes the technical specifications of our proposed system. `Pathway-Finder` uses Bayesian Network learning for offline computation (Section 4.2.1) and provides a scalable key-value structure to store exponential numbers of conditional probabilities to support real-time factor retrieval for the visualization (Section 4.2.2). Section 4.2.3 demonstrates the four stages of `Pathway-Finder` with a basic use case. We conclude the paper in Section 4.7.

4.2 Technical Specifications: System Overview

`Pathway-Finder` is a cloud based web-service hosted on Microsoft Azure for Research platform. The objective is to interactively discover more about the user health conditions and adaptively recommend care-pathways to minimize her 30-day readmission risk for heart failure. The majority of the proposed system components in this demonstration are precomputed and stored to increase the speed of the application. Figure-4.1 provides the overview of the system that comprises of offline and online layer. The UI enlists simple socio-demographic factors and the user (i.e. clinician) selects respective values for those from the drop-down. After that, the system alternatively suggests a set of diagnoses and interventions (utilities and procedures) to the user and then she/he selects some of them. In Section 4.2.1 we describe the offline computations, and Section 4.2.2 is used to describe the computations that take place, once the user starts interacting with the system.

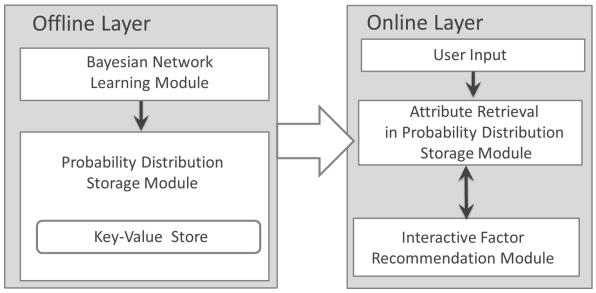


Figure 4.1: Architectural Overview of `Pathway-Finder`

4.2.1 Offline Computation

The offline computation is performed using two modules. (a) Bayesian Network Learning Module, and (b) Key Value Storage Module. We describe these two modules next.

(a) Bayesian Network Learning Module: At the heart of the system `Pathway-Finder`, there exists an intervention recommendation module which is mostly designed offline. The

objective of this module is "discover" the causal relationship between different factors (or attributes) related to heart failure readmission and how their interplay impacts the readmission risk. We briefly describe our solutions next.

A hierarchical Bayesian network effectively depicts the causal relationships between the factors and how their interplay relates to lowering the heart failure readmission risk. Thus, we model the intervention recommendation as a network learning task using the Bayesian network learning principles. For structure learning, our designed solution appropriately adapts Constraint Based Bayesian Network Learning algorithm [53, 10], *Score Based Learning Algorithm* [32] and *Hybrid Algorithm* [62] that combines both Constrained based and Score based approaches. Once the structure is defined, we use parameter learning [27, 5] techniques to learn the probability distribution at each node. In our implementation, we use Bayesian Parameter Estimation [37] to learn the parameter θ. The created probability distributions will go through the Probability Storage Module after that.

(b) Probability Distribution Storage Module: In this section, we present methods to transform the exponential number of conditional probabilities learned from Bayesian Networks (Section 4.2.1) to create a scalable storage module for efficient retrieval.

For the simplicity of exposition, imagine that a Bayesian Network has learned that "Gender" (M/F) and "Race" (imagine "Caucasian" and "Pacific Islanders" are two possible race values) has *causal relationship* with "Heart Failure (HF)". The task is to store the probability distribution function (Pdf in short) of each of these three nodes in a *key-value store*. The keys are multiple set of composite keys consisting of the minimum number of combination of the attributes required to maintain the uniqueness and have a cascading relationship with each other. The values are the probability of diagnoses, given key and the rest of the attributes. For our example above, the first set of keys are the two "Gender" values with rest of the attribute combinations as the value (such as, $Pr(HF|M, Caucasian), Pr(HF|M, PacificIslander)$ (similarly for females). The second set of keys will be for different ("Gender", "Race") combinations and the values are the probabilities of HF given the keys.

4.2.2 Online Computation

The online layer is designed with three modules that are described next.

(a) User Input: The interface accepts simple socio-demographic attribute values from the user as it is shown in stage-1 of Figure 4.3a. After that, the two other modules, iteratively interact with the user.

(b) Probability Distribution Lookup Module: This module is invoked multiple times to do look up either for the diagnoses or for the interventions. Based on the user input, the search goes inside the Probability Distribution Lookup Module to retrieve either the diagnoses or the interventions that the user is most likely to have.

The look up from the key-value store checks if the key is present in the store or not. Based on the user input, various combinations of keys are formed. Once the conditional probability for all possible diagnoses as entered by user have been looked up, the intervention will be recommended for the diagnosis with the highest probability. For example, if a user enters "Gender=M",and "Race=Caucasian", then the first look up is on "Gender". The second look up is on "Gender" and "Race", both. The lookup continues until the conditional probability for the diagnoses are retrieved as shown in the Figure-4.2. If the key is not present in the store, then based on the user's input the most similar key-value pair will be retrieved.

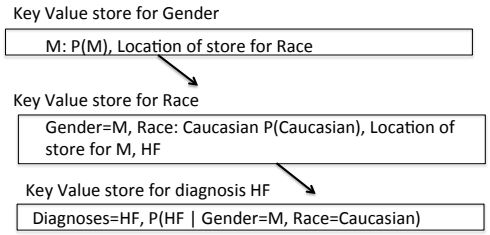


Figure 4.2: Key-Value Lookup Example

For the above example, imagine there are 5 diagnoses/comorbidities (DX1,DX2,DX3,DX4,DX5)

that the user may have and `Pathway-Finder` wants to suggest the top-3 likely ones to the user. For this, the call comes to this module to compute the following 5 probabilities.

$$Pr(DX1|Gender = "M" \, and \, Race = "Caucasian")$$
$$Pr(DX2|Gender = "M" \, and \, Race = "Caucasian")$$
$$Pr(DX3|Gender = "M" \, and \, Race = "Caucasian")$$
$$Pr(DX4|Gender = "M" \, and \, Race = "Caucasian")$$
$$Pr(DX5|Gender = "M" \, and \, Race = "Caucasian")$$

The module ranks all the diagnoses/comorbidities based on individual probabilities and suggests the top 3 highest diagnoses as most likely to investigate as described next.

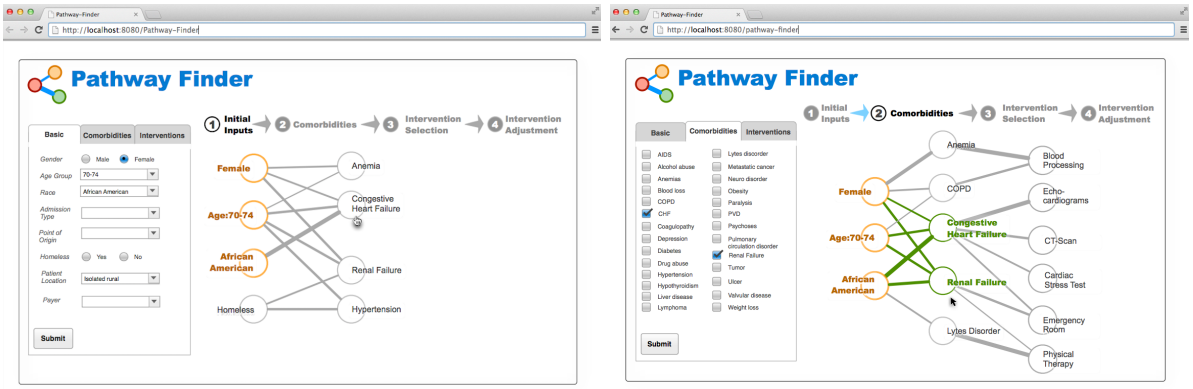
(c) Interactive Exploration/Recommendation module: There is an iterative interaction between this module and the lookup module. Based on the user selected values so far, the system iteratively suggests more interventions or discovers more diagnoses for her, until one of the stopping conditions is reached. For the example scenario, out of three diagnoses (imagine those are DX2, DX4, DX5) that are suggested to the user, she only selects DX4. Then, based on that the most appropriate set of interventions (let us say top-3) is selected to her (by making a call to the look up module) to retrieve the interventions that minimizes her readmission risk the most.

4.2.3 System Demonstration

In this section we demonstrate the four stages of `Pathway-Finder`: 1) Initial input collection; 2) Comorbidities exploration; 3) Intervention recommendation; 4) Outcome prediction and intervention adjustment loop. Figure 4.3 presents the progress of each stage.

The system is showcased using State Inpatient Databases (SID)[2] of Washington State of year 2010 and 2011. The HF cohort contains 3,908 distinct diagnosis codes and 2,049 procedure codes and total of 119,988 patient records.

[2]http://www.hcup-us.ahrq.gov/sidoverview.jsp



(a) Stage 1: Initial Input Collection

(b) Stage 2: Comorbidities Recommendation

(c) Stage 3: Intervention Recommendation

(d) Stage 4: Intervention Adjustment

Figure 4.3: Screen Shots of `Pathway-Finder`

4.3 Initial Input Collection

At the first stage, we collect preliminary information from user's input in an interactive way. We collect the patients' basic information first. `Pathway-Finder` proceeds with the limited information provided by the user, displaying associated comorbidities from the learned Bayesian Networks described in Section 4.2. For example, as shown in Figure 4.3a, a user provides her age(=70-74), gender(=Female), ethnic group(=African American) which are orange circles, leaving the others blank. Provided with the preliminary inputs, the network is expanded with a set of corresponding diagnosis/comorbidities, which are Anemia, Con-

gestive Heart Failure, Renal Failure, and Hypertension. The thickness of the links between factor nodes implies the probabilities of one factor leading to the other. The thicker they are, the stronger the connections. The user can interact with the form or directly interact with the network by clicking on the nodes. After clicking the "Submit" button, we enter the stage 2 of the system.

4.4 Comorbidities Recommendations

At the second stage, the user can add diagnosis/comorbidities information. For example, as shown in Figure 4.3b, she clicks on the two circles that turn green: Congestive Heart Failure and Renal Failure. Once the user clicks "Submit" at this page, our system will show her the appropriate interventions based on all the information collected and other diagnosis or comorbidities that could also appear for this specific patient. In this case, the interventions we recommended are Blood Processing, Echocardiograms, CT Scan, Cardiac Stress Test, Emergency Room, and Physical Therapy (gray circles in Figure 4.3b). Meanwhile we encourage more inputs from users by suggesting that the patient could also have Chronic Obstructive Pulmonary Disease (COPD) and Fluid and Electrotype (Lytes) Disorder. The additional suggestions were learned from our data and can be used to improve the prediction accuracy of our system. This stage is to assist clinicians to identify the appropriate interventions and other diseases the patient of interest might have.

4.5 Intervention Recommendations

At the third stage, a clinician can continue filling information about the patient and select appropriate interventions. In our case, the clinician selects Echocardiograms, Cardiac Stress Test, and Physical Therapy (the green circles in the third layer in Figure 4.3c). On the other hand, she finds the patient actually has Fluid and Electrotype (Lytes) Disorder, so she adds this new diagnosis to the network, which is the green circle in the second layer. Based on all the user inputs from the first three stages, our system predicts the Readmission Risk of the patient. As shown in Figure 4.3c, `Pathway-Finder` estimates 30% probability that the

patient of interest will be readmitted.

4.6 Intervention Adjustment

At the last stage, `Pathway-Finder` allows clinicians to adjust their intervention strategies to preview the impact of the new treatment plan to the targeted outcome. In our example shown in Figure 4.3d, the clinician adds CT Scan and observes a reduced readmission risks to 17%. This stage provides an iterative process to support a loop of discovery between interventions and the targeted outcome.

4.7 Summary

We propose an *interactive system* called `Pathway-Finder`, with the objective to visually explore, discover, and recommend clinical pathways for health conditions. We demonstrate `Pathway-Finder` to interactively recommend interventions to minimize the readmission risk due to Heart Failure (HF). At the heart of `Pathway-Finder`, there exists a Bayesian Network that learns the causal relationship among different factors and how that contributes to HF readmission risk. Further novelty of the system includes a key-value based representation of Bayesian Network, which enables us to perform real time lookup to interactively recommend interventions. Our demonstration also involves a high dimensional real patient dataset with hundreds and thousands of records and several hundreds of factors. To the best of our knowledge, `Pathway-Finder` is the first system that is empowered with the ability to interactively recommend and explore pathways for different clinical conditions.

Chapter 5

CONCLUSION

In this thesis, we propose solutions and tools to improve the quality of hospital discharge experience.

In our readmission work, we evaluate state-of-the-art machine learning techniques in predicting whether a patient is at risk of being readmitted to the hospital within 30 days after discharge. We also look at the problem of estimating the cost of that hospital readmission and show how state of the art machine learning approaches out perform the statistical methods which were previously used in the cost prediction domain.

Keeping in view with the sentiments of the medical community, that along with thirty day hospital readmission, there are other factors that determine a successful hospital discharge experience, we propose a novel framework that recommends personalized medications to patients by analyzing the complex interplay between a multitude of factors, such as, demographic factors, medical diagnoses, clinical factors, and how they contribute to the three objectives - thirty day readmission, subsequent length of stay and mortality rate. We then find the best medication combination to simultaneously reduce all three objectives by performing Multi objective optimization. We perform Structure learning and Parameter learning to learn the structure and parameters of the underlying hierarchical Bayesian network from the available patient data. We then perform Bayesian Inference to get conditional probabilities of each of the objectives for a given set of patient conditions such as demographic factors and diagnosis information for every valid medication combination, and then perform Multi Objective Optimization to optimize all of these objectives simultaneously. Our proposed framework is flexible enough to include or exclude additional factors, as well as layers, or can even obey constraints provided by the domain experts (i.e., doctors) in the design of

this hierarchical network. We present a case study validated by a domain expert as well as comprehensive experimental results based on the proposed approach to demonstrate the effectiveness of our proposed framework. Our experimental results as well as case studies demonstrate the effectiveness of the proposed framework. In our proposed multi objective medication recommendation framework, further work, however, would be needed in the exploration of medication recommendation to minimize other objectives such as cost incurred during hospital stay, and expanding the framework to scale for large number of related conditions. Besides heart failure, we are also investigating other diseases, such as COPD and diabetes, to construct a medication recommendation framework to simultaneously minimize multiple objectives.

We also develop a web based service called Pathway Finder with the objective of visually exploring and discovering clinical pathways that reduce the patient's chance of readmission. To the best of our knowledge, Pathway-Finder is the first system that is empowered with the ability to interactively recommend and explore pathways for different clinical conditions.

Through all of this combined work, we contribute to Health Analytics research and play our small part in the improvement of lives and health of patients worldwide.

BIBLIOGRAPHY

[1] Ruben Amarasingham et al. 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*, 2010.

[2] Baeza-Yates et al. *Modern information retrieval*, volume 463. ACM press New York, 1999.

[3] Uri Balla, Stephen Malnick, and Ami Schattner. Early readmissions to the department of medicine as a screening tool for monitoring quality of care problems. *Medicine*, 87(5):294–300, September 2008.

[4] John Bittl et al. Bayesian methods affirm the use of percutaneous coronary intervention to improve survival in patients with unprotected left main coronary artery disease. *Circulation*, 2013.

[5] George EP Box and George C Tiao. *Bayesian inference in statistical analysis*, volume 40. John Wiley & Sons, 2011.

[6] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[7] Leo Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth Publishing Company, 1984.

[8] Facundo Bromberg et al. Efficient markov network structure discovery using independence tests. In *SDM*, 2006.

[9] H Bueno et al. Trends in length of stay and short-term outcomes among medicare patients hospitalized for heart failure, 1993-2006. *JAMA*, pages 2141–2147, 2010.

[10] Jie Cheng and Russell Greiner. Comparing bayesian network classifiers. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 101–108. Morgan Kaufmann Publishers Inc., 1999.

[11] Everton Alvares Cherman, Maria Carolina Monard, and Jean Metz. Multi-label problem transformation methods: a case study. *CLEI Electron. J.*, 14(1), 2011.

[12] Insook Cho et al. Using ehr data to predict hospital-acquired pressure ulcers: A prospective study of a bayesian network model. *International journal of medical informatics*, 2013.

[13] Kalyanmoy Deb. Multi-objective optimization. In *Search methodologies*, pages 403–449. Springer, 2014.

[14] Shelley Derksen and H. J. Keselman. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, 45(2):265–282, 1992.

[15] J. Donze, D. Aujesky, D. Williams, and J. L. Schnipper. 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.

[16] Harris Drucker, Christopher J. C. Burges, Linda Kaufman, Alexander J. Smola, and Vladimir Vapnik. Support vector regression machines. In *Advances in Neural Information Processing Systems 9, Proceedings of the 1996 NIPS conference*, pages 155–161, 1996.

[17] Mark Dunlop et al. Multidimensional pareto optimization of touchscreen keyboards for speed, familiarity and improved spell checking. *CHI '12*, pages 2669–2678, 2012.

[18] Anne Elixhauser, Claudia Steiner, D Robert Harris, and Rosanna M Coffey. Comorbidity measures for use with administrative data. *Medical care*, 36(1):8–27, 1998.

[19] Mehrdad Fatourechi, Rabab K. Ward, Steven G. Mason, Jane Huggins, Alois Schlögl, and Gary E. Birch. Comparison of evaluation metrics in classification applications with imbalanced datasets. In *7th International Conference on Machine Learning and Applications (ICMLA)*, pages 777–782, 2008.

[20] N. K. Francis, J. Mason, E. Salib, L. Allanby, D. Messenger, A. S. Allison, N. J. Smart, and J. B. Ockrim. Factors predicting 30-day readmission after laparoscopic colorectal cancer surgery within an enhanced recovery programme. *Colorectal Disease*, 17(7):148–154, Jul 2015.

[21] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer System Science*, 55(1):119–139, 1997.

[22] Andrew Gottschalk and Susan A. Flocke. Time spent in face-to-face patient care and work outside the examination room. *Annals of Family Medicine*, 3(6):488–493, November 2005.

[23] Eren Gultepe et al. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. *JAMA*, 2013.

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

[25] Aftab Hassan et al. Multi-objective congestive heart failure medication recommendation. In *KDD*. IEEE, 2016.

[26] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, September 2009.

[27] David Heckerman, Dan Geiger, and David M Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3):197–243, 1995.

[28] Anika L. Hines, Marguerite L. Barrett, Joanna Jiang, and Claudia A. Steiner. Conditions with the largest number of adult hospital readmissions by payer. Technical report, Healthcare Cost and Utilization Project (HCUP), 2011.

[29] Zhengxing Huang, Wei Dong, Lei Ji, Chenxi Gan, Xudong Lu, and Huilong Duan. Discovery of clinical pathway patterns from event logs using probabilistic topic models. *Journal of Biomedical Informatics*, 47:39–57, February 2014.

[30] Haruko Iwata, Shoji Hirano, and Shusaku Tsumoto. Construction of clinical pathway based on similarity-based mining in hospital information system. *Procedia Computer Science*, 31:1107–1115, 2014.

[31] Stephen F. Jencks, Mark V. Williams, and Eric A. Coleman. Rehospitalizations among patients in the Medicare fee-for-service program. *New England Journal of Medicine*, 360(14):1418–1428, 2009.

[32] Finn Verner Jensen and Thomas Dyhre Nielsen. *Bayesian networks and decision graphs*. Springer, 2007.

[33] A.M Jones. Models For Health Care. Technical report, HEDG, c/o Department of Economics, University of York, January 2010.

[34] Miyo K et al. Development of case-based medication alerting and recommender system: A new approach to prevention for medication error. *Medinfo*, pages 871–874, 2007.

[35] Michel CA Klein and Gabriele Modena. Estimating mental states of a depressed person with bayesian networks. In *Contemporary Challenges and Solutions in Applied Artificial Intelligence*, pages 163–168. Springer, 2013.

[36] Todd M. Koelling et al. Discharge education improves clinical outcomes in patients with chronic heart failure. *Circulation*, 111(2):179–185, January 2005.

[37] Stuart C Kramer and Harold W Sorenson. Bayesian parameter estimation. *Automatic Control, IEEE Transactions on*, 1988.

[38] Richard Kronick, Todd P. Gilmer, Tony Dreyfus, and Theodore G. Ganiats. CDPS-Medicare: The chronic illness and disability payment system modified to predict expenditures for Medicare beneficiaries. Technical report, 2002.

[39] 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.*

[40] Cha Bibudh Lahiri and Nitin Agarwal. Predicting healthcare expenditure increase for an individual from Medicare data. In *Proceedings of the ACM SIGKDD Workshop on Health Informatics*, 2014.

[41] Wentian Li and Dale R Nyholt. Marker selection by akaike information criterion and bayesian information criterion. *Genetic Epidemiology*, 21:S272–7, 2000.

[42] Rui Liu et al. A framework to recommend interventions for 30-day heart failure readmission risk. *IEEE 2014*, pages 911–916, 2014.

[43] Rui Liu et al. A framework to recommend interventions for 30-day heart failure readmission risk. In *ICDM*. IEEE, 2014.

[44] Rui Liu et al. Pathway-finder: An interactive recommender system for supporting personalized care pathways. In *ICDM*. IEEE, 2014.

[45] Willard G. Manning, Anirban Basu, and John Mullahy. Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics*, 24(3):465–488, 2005.

[46] R Timothy Marler et al. Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*, 26(6):369–395, 2004.

[47] Naren Meadem et al. Exploring preprocessing techniques for prediction of risk of readmission for congestive heart failure patients. In *DMH-KDD*, 2013.

[48] Borislava Mihaylova, Andrew Briggs, Anthony O'Hagan, and Simon G. Thompson. Review of statistical methods for analysing healthcare resources and costs. *Health Economics*, 20(8):897–916, 2011.

[49] John Mullahy. Heterogeneity, excess zeros, and the structure of count data models. *Journal of Applied Econometrics*, 12(3):337–50, 1997.

[50] Mary D Naylor et al. Comprehensive discharge planning and home follow-up of hospitalized elders: A randomized clinical trial. *JAMA*, 1999.

[51] Andrew Y. Ng and Michael I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In *Advances in Neural Information Processing Systems 14, Proceedings of the 2001 NIPS conference*, pages 841–848. MIT Press, 2001.

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

[53] Judea Pearl and Thomas S Verma. A theory of inferred causation. *Studies in Logic and the Foundations of Mathematics*, 1995.

[54] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier chains for multi-label classification. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases: Part II*, pages 254–269, 2009.

[55] Michael W. Rich et al. A multidisciplinary intervention to prevent the readmission of elderly patients with congestive heart failure. *New England Journal of Medicine*, 1995.

[56] JK Schneider et al. A medication discharge planning program measuring the effect on readmissions. *Clinical Nursing Research*, 1993.

[57] E. Shadmi, N. Flaks-Manov, M. Hoshen, O. Goldman, H. Bitterman, and R. D. Balicer. Predicting 30-day readmissions with preadmission electronic health record data. *Med Care*, 53(3):283–289, Mar 2015.

[58] William N. Southern et al. Increased risk of mortality among patients cared for by physicians with short length-of-stay tendencies. *JGIM*, pages 712–718, 2015.

[59] Peter Spirtes and Christopher Meek. Learning bayesian networks with discrete variables from data. In *KDD*, 1995.

[60] Shanu Sushmita et al. Predicting 30-day risk and cost of all-cause hospital readmissions. In *Proceedings of ACM Digital Health 2015 (5th International Conference on Digital Health)*, 2016.

[61] Shanu Sushmita, Stacey Newman, James Marquardt, Prabhu Ram, Viren Prasad, Martine De Cock, and Ankur Teredesai. Population cost prediction on public healthcare datasets. In *Proceedings of ACM Digital Health 2015 (5th International Conference on Digital Health)*, pages 87–94, 2015.

[62] Ioannis Tsamardinos, Laura E Brown, and Constantin F Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65(1):31–78, 2006.

[63] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, 3(1):1–13, 2007.

[64] J. M. Vinson, M. W. Rich, J. C. Sperry, A. S. Shah, and T. McNamara. Early readmission of elderly patients with congestive heart failure. *Journal of the American Geriatrics Society*, 38(12):1290, 1990.

[65] Jingzhu Xie1 et al. Multiobjective bayesian network model for public-private partnership decision support. *Inquiry*, pages 1069–1081, 2013.

[66] Burasakorn Yoosooka et al. Medication recommender system. *TICST 2015*, pages 313–317, 2015.

[67] Bichen Zheng, Jinghe Zhang, Sang Won Yoon, Sarah S. Lam, Mohammad Khasawneh, and Srikanth Poranki. Predictive modeling of hospital readmissions using metaheuristics and data mining. *Expert Systems with Applications*, 42(20):7110–7120, 2015.

[68] Kiyana Zolfaghar et al. Big data solutions for predicting risk-of-readmission for congestive heart failure patients. In *Big Data, 2013 IEEE International Conference on*, pages 64–71. IEEE, 2013.

[69] Kiyana Zolfaghar et al. Predicting risk-of-readmission for congestive heart failure patients: A multi-layer approach. *CoRR*, 2013.

[70] Kiyana Zolfaghar et al. Risk-o-meter: an intelligent clinical risk calculator. In *KDD*, pages 1518–1521, 2013.