

Tree-based Regularization for Interpretable Readmission Prediction

Jialiang Jiang and Varun Chandola
Computer Science and Engineering
University at Buffalo

Sharon Hewner
School of Nursing
University at Buffalo

Abstract

Preventable hospital readmissions have been identified as one of the primary targets for reducing costs and improving healthcare delivery. However, most data driven studies for understanding readmissions have produced non-interpretable black boxes, which precludes them from being used effectively within the decision support systems in the hospitals. A novel strategy to improve the interpretability of a linear model by incorporating domain knowledge is proposed here. The central idea is to exploit the hierarchical relationships among the features (medical diagnosis codes, in this case) using a tree-structured sparsity-inducing regularization norm. The proposed method transforms the hierarchical relations among features into a graph and then applies graph-guided regularization during the model learning. Additionally, an evaluation metric is proposed to quantify the interpretability of a linear model with respect to the domain hierarchy. Results on two healthcare claims data sets are shown, where a model is learnt to predict a patient’s risk of readmission, based on the medical history and other relevant features. Results show that the proposed method is able to learn a model which can predict readmission risk with accuracies that are comparable to existing methods, but produces a highly interpretable output, which allows medical experts to draw clinically relevant insights and identify key factors associated with hospital readmissions. Some of these factors conform to existing beliefs, e.g., impact of surgical complications and infections during hospital stay. Other factors, such as the impact of mental disorder and substance abuse on readmission, provide empirical evidence for several pre-existing but unverified hypotheses. The findings of this study will be instrumental in designing the next generation decision support systems for preventing readmissions.

1 Introduction

Hospital readmissions are prevalent in the healthcare system and contribute significantly to avoidable costs. In United States, recent studies have shown that the 30-day readmission rate among the Medicare beneficiaries¹ is over 17%,

Copyright held by the author(s). In A. Martin, K. Hinkelmann, A. Gerber, D. Lenat, F. van Harmelen, P. Clark (Eds.), Proceedings of the AAAI 2019 Spring Symposium on Combining Machine Learning with Knowledge Engineering (AAAI-MAKE 2019). Stanford University, Palo Alto, California, USA, March 25-27, 2019.

¹A federally funded insurance program representing 47.2 % (\$182.7 billion) of total aggregate inpatient hospital costs in the

with close to 75% of these being avoidable (Mpa 2007), with an estimated cost of \$15 Billion in Medicare spending. Similar alarming statistics are reported for other private and public insurance systems in the US and other parts of the world. In fact, management of care transitions to avoid readmissions has become a priority for many acute care facilities as readmission rates are increasingly being used as a measure of quality (Conway and Berwick 2011).

Given that the rate of avoidable readmission has now become a key measure of the quality of care provided in a hospital, there have been increasingly large number of studies that use healthcare data for understanding readmissions. Most existing studies have focused on building models for predicting readmissions using a variety of available data, including patient demographic and social characteristics, hospital utilization, medications, procedures, existing conditions, and lab tests (Futoma, Morris, and Lucas 2015; Choudhry et al. 2013; Donze et al. 2013). Other methods use less detailed information such as insurance claim records (Yu et al. 2013; He et al. 2014). Many of these methods use machine learning methods, mainly Logistic Regression, to build classifiers and have reported consistent performance on a variety of clinical data sets. In fact, most papers about readmission prediction report AUC scores in the range of 0.65-0.75.

While the predictive models have shown promise, their moderate performance means that they are still not at a stage where hospitals could use them as “black-box” decision support tools. Moreover, such models are not easily interpretable to provide actionable insights to the decision makers. At the same time, beyond the selection of the initial set of features to learn from, these solutions do not explicitly utilize the rich information available in the medical domain.

In this paper, we explore incorporation of one such domain information, into the model learning process. Specifically, we utilize the hierarchical relationships among different medical diagnosis codes, available as a taxonomical tree (See Section 3 for details). The tree structure is utilized as a regularization penalty, to enforce the model (logistic regression) to learn a sparse solution such that the non-zero weights are localized within a few sub-trees. The key idea is that such a solution would be easier to interpret compared

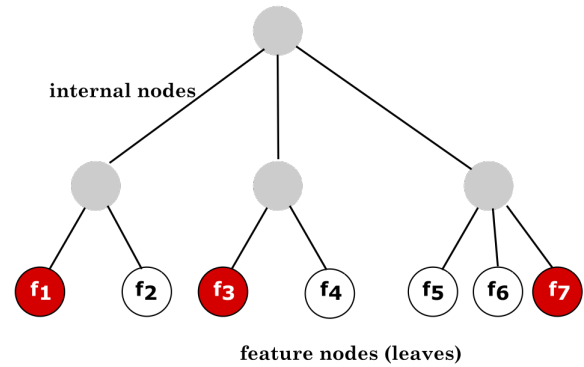
United States (Pfundner, Wier, and Steiner 2013).

to a solution in which the weights are “scattered” across. A graphical illustration is provided in Figure 1. The proposed method falls under the general class of *structured sparsity regularization* based machine learning models (Mosci et al. 2010), which consists of numerous schemes to exploit different types of relationships among features, including groups (Yuan and Lin 2006), sequential (Tibshirani et al. 2005), and graphs (Chen et al. 2010). However, regularization methods for scenarios where the features are related over a tree are sparse, and the existing ones provide an indirect way of capturing the tree structure (Zhao, Rocha, and Yu 2009), which, as observed later in the experiments, makes them inadequate for the target problem of readmission prediction.

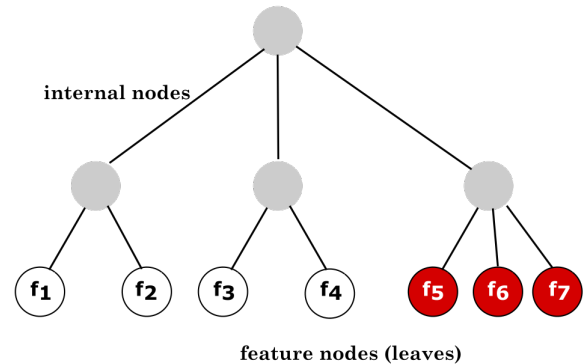
The proposed regularization scheme transforms the tree structure into a weighted graph that uses the “tree-distance” as the weight of the edge between the corresponding nodes in a graph, and then employs a graph based penalty to force the machine learning algorithms to favor solutions in which the non-zero weights are strongly linked in the graph. The regularizer is incorporated into a standard *logistic regression* classifier, using *truncated gradient descent* (Langford, Li, and Zhang 2008) for the optimization step. This is used to learn a readmission prediction model that uses diagnosis codes from a patient’s medical history to predict his or her readmission risk, as a binary label. Results on two data sets, extracted from: 1). New York State Medicaid records (MDW), and 2). MIMIC-III data set (a publicly available data set), show that the proposed model not only performs comparably, in terms of accuracy, to classical regularization schemes such as LASSO and existing tree-based regularizer (Zhao, Rocha, and Yu 2009), but learns a sparse model that is significantly better than others in terms of the interpretability. To this effect, we propose a quantitative metric to assess the interpretability of a model in which the features are arranged in a tree structure.

By analyzing the model trained on the MDW data, we infer several important insights to improve the understanding of readmissions. Some of our findings conform to existing beliefs, for example, the importance of bacterial infections during hospital stay. Other findings provide empirical evidence to support existing hypotheses amongst healthcare practitioners, for example, the effect of the type of insurance on readmissions (Hewner et al. 2014). Most interesting findings from our study reveal surprising connections between a patient’s non-disease background and the risk of readmission. These include behavioral patterns (mental disorders, substance abuse) and socio-economic background. For the result of the analysis of MIMIC-III data, it also has similar inference. For example, bacterial infections during hospital stay, chronic circulatory and respiratory system diseases are important factors to predict readmission. Moreover, profited by semantic refining ability of tree-based regularization, We can infer the significant disease classification to readmission straightforward. For example, class of Diseases Of The Circulatory System and Metabolic Disorders are highlighted for understanding of readmissions.

We believe that such findings can have a significant impact on how healthcare providers develop effective strate-



(a) No taxonomy-guided regularization



(b) Taxonomy-guided regularization

Figure 1: Two possible linear models for a hypothetical scenario with same sparsity and same model performance. The data set has seven features ($f_1 - f_7$) arranged as leaves of a tree. The red leaf nodes indicate the features that are non-zero in the learnt model. The first solution does not provide an interpretable solution as per the taxonomy. The second solution, on the other hand is more interpretable.

gies to reduce readmissions. At present, the healthcare efforts in this context have been twofold. First is the effort to improve the quality of care within the hospital and the second is to develop effective post-discharge strategies such as telephone outreach, community-based interventions, etc. The results from this study inform the domain experts on both fronts.

The rest of the paper is organized as follows. We review existing literature on readmission prediction in Section 2. We describe the data used for our experiments in Section 3 and formulate the machine learning problem in Section 4. We discuss the classification methodology in Section 5. We present the algorithm for measurement of model’s interpretability in Section 6. The results are presented in Section 7. We discuss the importance of interpretable outcome by including a real world case study in Section 8.

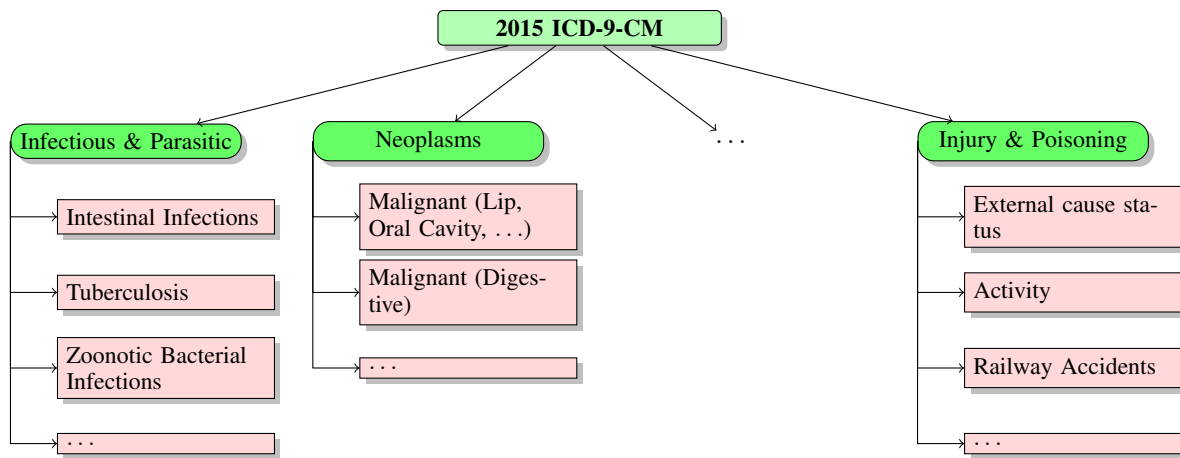


Figure 2: A sample portion of ICD9-CM classification. See [url²](#) for complete hierarchy.

2 Related Work

Coincident with the rising importance of readmissions in reducing healthcare costs, there have been several papers that use clinical and insurance claims information to build predictive models for readmissions (Kansagara et al. 2011) using different machine learning models including Deep Neural Networks (Jamei et al. 2017; Lin et al. 2018; Xiao et al. 2018), Logistic Regression (Futoma, Morris, and Lucas 2015; Choudhry et al. 2013; Donze et al. 2013; Niu 2013) and Support Vector Machines (Yu et al. 2013). However, most of these solutions have focused on improving the accuracy of the predictive model, and not necessarily on the interpretability of the model to improve the understanding of the readmission problem. Papers that focus on interpretability are limited to identifying the best features that predict readmission (Jiang et al. 2018) and have typically focused on a small set of patients or hospitals (Yu et al. 2013; Amarasingham et al. 2010). In this paper, we are focusing on a more direct approach that is scalable to any problem setting.

Finally, the hierarchical relationship has never been exploited for building predictive models for readmission. Singh, et. al, (Singh et al. 2014) have presented a similar approach in the context of predicting disease progression, however, the authors focus on using the disease hierarchy to come up with new features that are fed into the classifier. Additionally, there is no standard of measurement for interpretability of prediction model especially for structured based, while we propose a general methodology to address the problem.

3 Data

For the experiments, we explored two different data sets that consist of healthcare insurance claims and electronic health records (EHR).

The first dataset is obtained from the *New York State Medicaid Data Warehouse* (MDW). Medicaid is a social health

²<http://www.icd9data.com/2015/Volume1/default.htm>

care program for families and individuals with low income and limited resources. We analyzed four years (2009–2012) of claims data from the MDW. The claims correspond to multiple types of health utilization including hospitalizations, outpatient visits, etc. While the raw data consisted of 4,073,189 claims for 352,716 patients, we only included the patients in the age range 18–65 with no obstetrics related hospitalizations. The number of patients with at least one hospitalization who satisfied these conditions were 11,774 and had 34,949 claims.

For each patient we have information of patient’s admission medical history extracted from four years of claims data represented as a binary vector that indicates if the patient was diagnosed with a certain disease in the last four years.

The second dataset is *Multi-parameter Intelligent Monitoring in Intensive Care* (MIMIC-III) public dataset (Johnson et al. 2016). This data is a large, freely-available database comprising de-identified health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012.

While the database includes information such as demographics, vital sign measurements made at the bedside (one data point per hour), laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality (both in and out of hospital), we focus on admission records to extract the medical codes as part of each patient’s history.

According to the dataset we retrieved from the MIMIC-III dataset, there are 46516 patients in total with 3996 of patients flagged as readmissions. The medical history for each patient consists of 6783 diagnosis codes.

Diagnosis Codes

Disease information is encoded in insurance claims and medical records using *diagnosis codes*. The *International Classification of Diseases* (ICD) is an international standard for classification of disease codes. The data used in this paper followed the ICD-9-CM classification which is a US adaptation of the ICD-9 classification. Conceptually, the ICD-9-CM codes are structured as a tree (See Figure 2 for a

sample) with 19 broad disease categories at level 1. The entire tree has 5 levels and has total of 14,567 diagnosis codes. While the primary purpose of ICD taxonomy has been to support the insurance billing process, it contains a wealth of domain knowledge about the different diseases.

Readmission Risk Flag

For each patient in the above described cohort, we assign a binary flag for readmission risk. The readmission risk flag is set to 1 if the patient had *at least* one pair of consecutive hospitalizations within 30 days of each other in a single calendar year, otherwise it is set to 0.

4 Problem Statement

Given a patient’s disease history, we are interested in predicting the *readmission risk* (binary flag) for the patient. The problem formulation is different from many existing studies (Futoma, Morris, and Lucas 2015), where the focus is on assigning a readmission risk to a single hospitalization event. Our focus is on understanding the impact of socioeconomic and behavioral factors on a readmission.

We denote each patient i as a vector \mathbf{x}_i consisting of 11,881 elements for the MDW dataset and 6,873 elements for the MIMIC-III admission dataset corresponding to the number of disease codes showed in data respectively. Note that while ICD-9-CM classification contains 14,567 codes, only 11,881 and 6,783 codes are observed in each data set used in this paper. We selected patients that age in between 18 and 65 and excluded pregnancy related diseases. All elements in the vector are binary. The readmission risk flag is denoted using $y_i \in \{0, 1\}$ where 1 indicates readmission risk.

From machine learning perspective our task is to learn a classifier from a training data set $\langle \mathbf{x}_i, y_i \rangle_{i=1}^N$ which can be used to assign the readmission risk flag to a new patient represented as \mathbf{x}_* . Note that the input vector \mathbf{x}_i is highly sparse. For example, in the NY Medicaid dataset, on average, there are only 36 non-zeros out of total 11,884 possible codes.

5 Methodology

We use a *logistic regression* (LR) model (Cox 1958) as the classifier, which, is the most widely used model in the context of readmission prediction (Futoma, Morris, and Lucas 2015). The LR model, for binary classification tasks, computes the probability of the target y to be 1 (readmission risk), given the input variables, \mathbf{x} as:

$$p(y = 1|\mathbf{x}) = \frac{1}{1 + \exp(-\beta^\top \mathbf{x})} \quad (1)$$

Where β is the LR model parameter (regression coefficients). We assume that \mathbf{x} includes a constant term corresponding to the intercept.

The model parameter β are learnt from a training data set $(\langle \mathbf{x}_i, y_i \rangle_{i=1}^N)$ by optimizing the following objective function:

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^N \log(1 + \exp(-y_i \beta^\top \mathbf{x}_i)) + \lambda \Omega(\beta) \quad (2)$$

where the first term refers to the training loss and the second terms is a regularization penalty imposed on the solution; λ being the regularization parameter.

Existing Regularization Schemes

L1 Regularizer Different forms of regularization penalties have been used in the past, including the widely used l_1 and l_2 norms (Tibshirani 1994). While l_2 norm ($\Omega(\beta) = \|\beta\|_2 = (\sum_j \beta_j^2)^{1/2}$) is typically used to ensure stable results, l_1 norm ($\Omega(\beta) = \|\beta\|_1 = \sum_j |\beta_j|$) is used to promote sparsity in the solution, i.e., most coefficients in β are 0.

However, l_1 regularizer does not explicitly promote structural sparsity. Given that the features used in predicting readmission risk have a well-defined structure imposed by the ICD-9 standards, we explore regularizers that leverage this structure for model learning:

Sparse Group Regularizer This regularizer (also referred to as Sparse Group LASSO or SGL) assumes that the input features can be arranged into K groups (non-overlapping or overlapping) (Bach 2008). The SGL regularizer is given by:

$$\Omega(\beta) = \alpha \|\beta\|_1 + (1 - \alpha) \sum_{k=1}^K \|\beta_{G_k}\|_2 \quad (3)$$

where β_{G_k} are the coefficients corresponding to the group G_k . The above penalty function favors solutions which select only a few groups of features (group sparsity). For the task of readmission prediction, we divide the features corresponding to all numbers of diagnosis codes included into 19 non-overlapping groups, based on the top level groupings in the ICD-9-CM classification (See Table 1).

Tree Structured Group Regularizer This regularizer, also referred to as Tree Structured Group LASSO (TSGL), explicitly uses the hierarchical structure imposed on the features. The TSGL regularizer is given by:

$$\Omega(\beta) = \sum_{i=0}^D \sum_{j=1}^{N_i} \|\beta_{G_j^i}\|_1 \quad (4)$$

where G denotes the tree constructed using the hierarchy of the diagnosis codes. G_j^i denotes the j^{th} node in the tree at the i^{th} level. Thus G_1^0 denotes the root level, and so on.

A Novel Sparse Tree-Structure Regularizer

The TSGL regularizer, discussed above, treats the tree structure as a special overlapped group, which ignores the hidden relationship between nodes at different levels. To overcome this deficiency, we propose a different way to exploit the tree structure. The new regularization penalty is defined as:

$$\Omega(\beta) = \lambda_1 \sum_{i=1}^m \sum_{j=1}^m D_{ij} (\beta_i - \beta_j)^2 + \lambda_2 \|\beta\|_1 \quad (5)$$

Where the β_i and β_j are coefficients of features i and j , respectively and D_{ij} is the tree distance between features i and j and will be introduced in next subsection. The first penalty term ensures that the selected features are closer to each other in the taxonomy tree while the second term, $\|\beta\|_1$, ensures the overall sparsity of the solution.

1	Infectious And Parasitic Diseases
2	Neoplasms
3	Endocrine, Nutritional And Metabolic Diseases, And Immunity Disorders
4	Diseases Of The Blood And Blood-Forming Organs
5	Mental Disorders
6	Diseases Of The Nervous System And Sense Organs
7	Diseases Of The Circulatory System
8	Diseases Of The Respiratory System
9	Diseases Of The Digestive System
10	Diseases Of The Genitourinary System
11	Complications Of Pregnancy, Childbirth, And The Puerperium
12	Diseases Of The Skin And Subcutaneous Tissue
13	Diseases Of The Musculoskeletal System And Connective Tissue
14	Congenital Anomalies
15	Certain Conditions Originating In The Perinatal Period
16	Symptoms, Signs, And Ill-Defined Conditions
17	Injury And Poisoning
18	Supplementary Classification Of Factors Influencing Health Status And Contact With Health Services
19	Supplementary Classification Of External Causes Of Injury And Poisoning

Table 1: Top level disease groups in the ICD-9-CM classification

Distance Matrix for Tree The distance between any two nodes in the tree, D_{ij} is defined in terms of the length of the path between the two nodes. If the node i is an ‘‘ancestor’’ of node j , or vice-versa, the distance is defined as:

$$D_{ij} = (l_i - l_j)^2 \quad (6)$$

where l_i denotes the level or the number of steps from the root for node i . If the nodes i and j do not share any ancestral relationship, then the distance is defined as:

$$D_{ij} = ((l_i - l_c)^2 + (l_j - l_c)^2)^3 \quad (7)$$

where c is the node that is the *nearest common ancestor* for nodes i and j . The cubic power is used to sharply increase the weight with the number of levels to go up by to find the common ancestor. Thus, D_{ij} will be largest for two leaf nodes whose common ancestor is the root node.

The data matrix consisting of the distances between all pairs of leaf nodes (features) from the MIMIC-III data set is shown in Figure 3.

Optimization To solve the optimization problem in (2) using the regularization penalty defined in (5), we first convert the first penalty term (tree structure sparsity) into a graph constraint as:

$$\sum_{i=1}^m \sum_{j=1}^m D_{ij}(\beta_i - \beta_j)^2 = \beta^T \mathbf{L} \beta \quad (8)$$

where \mathbf{L} is a $m \times m$ matrix, such that:

$$L_{ij} = \begin{cases} 0 & \text{if } i = j \\ -D_{ij} & \text{otherwise} \end{cases} \quad (9)$$

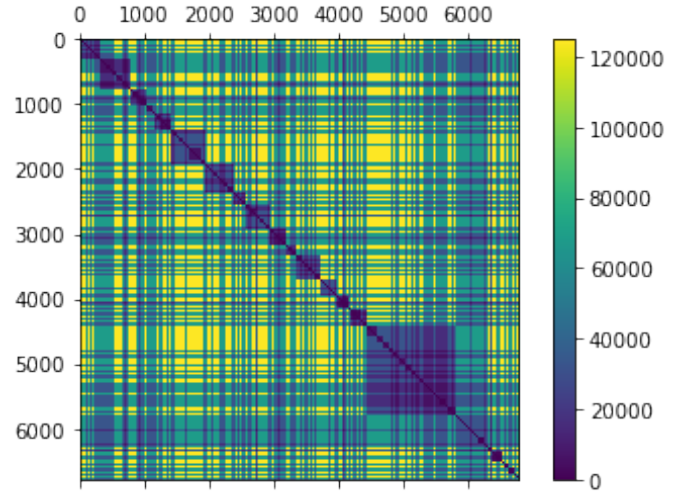


Figure 3: D_{ij} values for all feature pairs in MIMIC-III data.

The revised objective function for the regularized LR model can be written as:

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^N \log(1 + \exp(-y_i \beta^T \mathbf{x}_i)) + \lambda_1 \beta^T \mathbf{L} \beta + \lambda_2 \|\beta\|_1 \quad (10)$$

Truncated Gradient Descent Note that the objective function and the tree penalty term have a convex form such that one can calculate the gradient of the objective function with respect to the weight vector, β , and use that within a gradient descent algorithm. However, due to the presence of the l_1 term ($\|\beta\|_1$), a direct gradient descent formulation is not possible. We employ *Truncated Gradient Descent* (Langford, Li, and Zhang 2008) which has been shown to be effective in learning solutions under l_1 regularization penalties.

The idea behind truncated gradient descent is to ignore the l_1 term when calculating the gradient, and round small coefficients (that are not larger than a small threshold) to zero after every k online steps, i.e., at every k^{th} step:

$$\beta^{(t)} = T_0(\beta^{(t-1)} - \eta \nabla J) \quad (11)$$

Where ∇J is the gradient without the l_1 penalty term and T_0 defined by:

$$T_0(\beta_j, \theta) = \begin{cases} 0 & \text{if } |\beta_j| < \theta \\ \beta_j & \text{otherwise} \end{cases} \quad (12)$$

That is, we first apply the standard gradient descent rule, and then round small coefficients to zero to enforce the l_1 sparsity.

6 A Quantitative Measure for Interpretability

In classical sparsity inducing models, sparsity is measured using the number of non-zero coefficients or weights. While this is reasonable for settings with “flat” structure, e.g., LASSO or Group LASSO, this does not reveal the true interpretiveness of a solution, in the context of a tree structure. For instance, Figure 1 illustrates how two solutions with same number of non-zero coefficients can have different interpretability.

We propose a novel measure to assess the interpretiveness of a solution. The proposed measure is calculated in a *bottom-up* fashion, starting from the leaf nodes. For the i^{th} node in the tree, we define the purification noise, P_i as:

$$P_i = E_i + \sum_{j \in C_i} P_j \quad (13)$$

where C_i is the set of non-leaf children of node i and E_i is the *Shannon Entropy* of the current i^{th} node by measuring the information loss of all leaf child nodes, i.e., children that are actual features:

$$E_i = -p_i \log_2(p_i) \quad (14)$$

where p_i is the fraction of leaf children of node i with a non-zero coefficient. Starting from the parents of the leaf nodes, the purification noise is recursively computed, and finally the purification noise for the root node, i.e., P_{root} is treated as the overall purification noise for the entire solution.

7 Results

In this section we present our findings by applying logistic regression classifier for the task of readmission prediction on the MDW data and MIMIC-III data described earlier. We first compare the performance of the different regularization strategies to the classification task using the area under the ROC-curve (AUC) for each classifier as our evaluation metric due to the imbalance of data. We also compare the different strategies to report the purification noise (interpretability score) value for each solution. For each strategy, we run 10 experiments with random 80-20 splits for training and test data, respectively. The optimal values for the regularization parameters for each strategy are chosen using cross-validation. We use the MATLAB package, SLEP (Liu, Ji, and Ye 2009), for the Tree Structured Group Regularization experiments. The proposed regularization method was developed in Python.

7.1 Comparing Different Regularization Strategies

Here we compare the performance of different regularization methods discussed in Section 5. The results are summarized in Table 2 and Figure 4.

For the MIMIC-III data set, the best performance, in terms of AUC is obtained using the classical, $l1$ and $l2$ regularizations. However, the interpretability is highest for the proposed tree-structured measure, followed by the earlier published TSGL algorithm. On the other hand, the results for

Table 2: Comparison of Different Regularization Strategies

	MIMIC-III			
	LR ($l2$)	LR ($l1$)	TSGL	SparseTree-based
AUC	0.87	0.88	0.73	0.71
Purification Noise	194.85	194.86	67.40	30.12
	MDW			
	LR ($l2$)	LR ($l1$)	TSGL	SparseTree-based
AUC	0.67	0.68	0.70	0.62
Purification Noise	716.48	289.49	134.63	21.69

the MDW data set show that the tree based regularizations perform on par with the classical methods. However, the interpretability is significantly higher for the proposed tree-based regularization scheme. The $l2$ regularizer, for obvious reasons, does not produce a sparse solution (194.85 of MDW and 716.48 of MIMIC-III), while the other three regularizers induce significant sparsity. However, the structured regularizers are able to achieve significantly low structured sparsity (30.12 of MDW and 21.69 of MIMIC-III) which is consistent with the ICD-9-CM hierarchy.

Effect of Regularization Parameter, λ_1 The role of the regularization parameter, λ_1 , in (10) is to control the penalty on the tree-structure of the solution. Figure 4 shows how the AUC score and the interpretability score vary with λ_1 . By increasing λ_1 , we note significant improvement of performance by leveraging more prior hierarchical information as well as outstanding decrease on purification noise, which indicates highly interpretability.

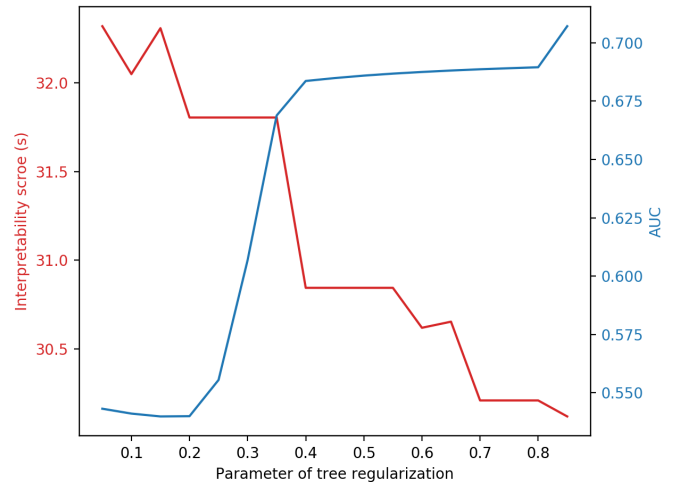


Figure 4: Performance as a function of the regularization parameter λ_1 in MIMIC-III dataset

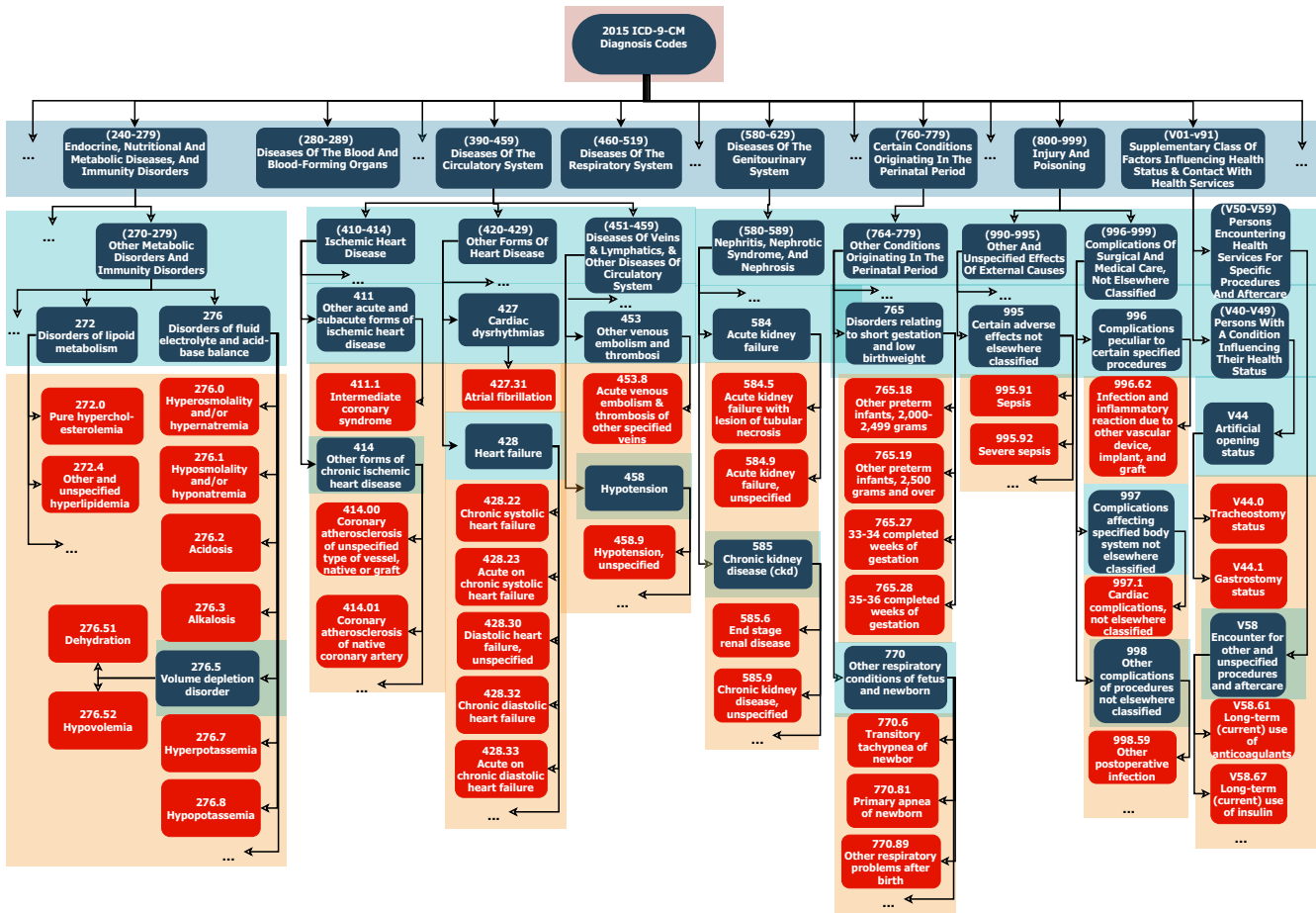


Figure 5: Non-zero features learnt using the proposed regularization from MIMIC-III data set

7.2 Qualitative Interpretation of the Solution

Figure 5 provides a graphical illustration of the non-zero features learnt by the proposed model. The red nodes denote the actual diagnosis codes that had non-zero weights. The blue nodes are the ancestors of the selected leaf nodes. We first note that the model consists of only 87 (out of 11567) features or disease codes. Additionally, the 87 codes are concentrated within a few higher level disease categories. Since the sub-trees containing the non-zero features are relatively dense, it is easy to summarize them with a higher order disease category. For instance, one can determine that Disorders of fluid electrolyte and acidbase balance is an important factor in determining readmissions, which has been confirmed in actual clinical studies (Badawi and Breslow 2012). Similarly, pregnancy related complications are an important factor in determining readmissions.

On the other hand, a similar visualization for $l1$ regularization shows a highly non-interpretable result, as seen in Figure 6. For instance, consider the feature - *Isolated tracheal or bronchial tuberculosis ...*, which has a non-zero weight. It is unclear why only that

diagnosis code is selected whereas the other 35 siblings (other forms of tuberculosis) are ignored.

8 Discussion

Section 7 shows that leveraging the hierarchical information in the ICD-9CM classification improves the predictive capability of logistic regression while gaining better interpretability. In this section, we demonstrate the necessity of high level interpretation to a medical record prediction model from the healthcare perspective. The focus is to show that an interpretable decision support tool for readmission risk prediction can be effective, as shown in the following real world case study.

As shown in Figure 5, the model is well-informed by the ICD-9-CM hierarchy. Interpretable learning grants the model the ability to conclude high-level important category that are more understandable to the healthcare providers in the medical facilities. For example, the result of proposed model suggests 428-Heart Failure as a important disease category results in readmission instead of extremely specific ICD-9-CM disease code like 428.33-Acute on Chronic Diastolic Heart Failure. The high level important concept ab-

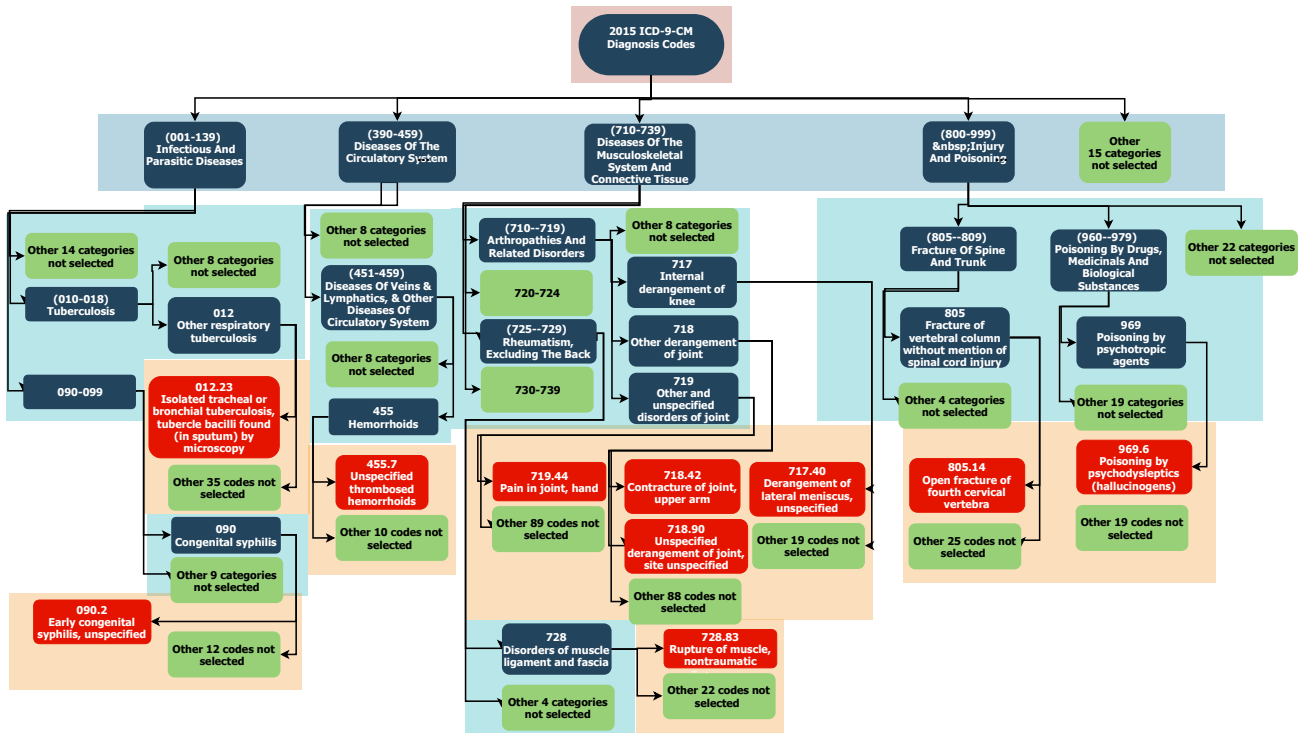


Figure 6: Non-zero features learnt using the l_1 regularization from MIMIC-III data set

straction makes it easier to focus on important aspects while removing excessive attention from specific disease codes, that may distract the healthcare staff from the key factors during the post-discharge phase.

Case Study: *M.J. was a 55-years-old white male with medical history of **Hypertension**, **Coronary Artery Disease** and **High Cholesterol**, and came to the emergency department with left flank pain on 7/1/2018. He was admitted to the hospital for left kidney stone and treated with intravenous fluid and pain medication. M.J. was also found of left **hydro nephrosis** due to obstructing kidney stone and M.J. underwent urology surgery to have the stone removed on 7/4/2018. However, his hospital course was complicated with **sepsis** (urinary tract infection) and **electrolyte imbalance** (hyperkalemia) due to acute renal injury. M.J. finished the course of antibiotics for urinary tract infection and his renal function was back to the baseline. After a 2-week hospital stay, the patient was debilitated so he was discharged to a skilled nursing facility on 7/14/2018 for rehabilitation.*

On 7/21/2018, M.J. was found unresponsive and

*pulse-less by the staff in the skilled nursing facility. He was resuscitated by emergency medical services and re-admitted to the intensive care unit for cardiac arrest due to **electrolyte imbalance** (hyperkalemia). Staff reported that he was not eating or drinking since being discharged to the skilled nursing facility on 7/14/2018. During the ICU, M.J. developed multi-organ failure and died in the intensive care unit on 7/31/2018.*

Given the patient’s medical history in the above case study, the proposed model would assign a readmission risk to the patient. But at the same time, the model would provide additional factors that could be true for the patient. For instance, **electrolyte imbalance** would be a factor for readmission, along with the other chronic diseases. The discharge staff would include that in the notes to ensure that it is monitored in rehabilitation phase and possibly save the patient’s life.

9 Conclusions

In the last decade, there have been numerous studies that link factors pertaining to a patient’s hospital stay to the risk of readmission. However, most studies have been on a focused cohort, limited to one or few hospitals. We show here

that similar results can be achieved using claims data, which has fewer elements but provides a large population coverage; the entire state of New York for this study and the wide population coverage of MIMIC-III. Even with the large volume of data, the predictive algorithms are not accurate enough to be used as decision making tools. However, model interpretation can reveal insights which can inform the strategies for reducing and/or eliminating readmissions.

A patient's disease history is typically expressed using diagnosis codes, which can take as many as 18000 possible values, with many more possibilities in the next generation ICD-10 disease classification. With so many possible features, ensuring model interpretability is a challenge. However, using structured sparsity inducing models, such as the one proposed here, one can ensure that the truly important factors can be identified within the hierarchy.

10 Acknowledgment

This material is based upon work partially supported by a seed grant from the University at Buffalo Germination Space Program. We thank Chiahui Chen, MS, RN, FNP-BC, School of Nursing, UB for developing the case study.

References

Amarasingham, R.; Moore, B. J.; Tabak, Y. P.; Drazner, M. H.; Clark, C. A.; Zhang, S.; Reed, W. G.; Swanson, T. S.; Ma, Y.; and Halm, E. A. 2010. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. *Medical Care* 48(11):981–989.

Bach, F. R. 2008. Consistency of the group lasso and multiple kernel learning. *Journal of Machine Learning Research* 9:1179–1225.

Badawi, O., and Breslow, M. J. 2012. Readmissions and death after icu discharge: development and validation of two predictive models. *PloS one* 7(11):e48758; e48758–e48758.

Chen, X.; Kim, S.; Lin, Q.; Carbonell, J. G.; and Xing, E. P. 2010. Graph-structured multi-task regression and an efficient optimization method for general fused lasso. *CoRR* abs/1005.3579.

Choudhry, S. A.; Li, J.; Davis, D.; Erdmann, C.; Sikka, R.; and Sutariya, B. 2013. A public-private partnership develops and externally validates a 30-day hospital readmission risk prediction model. *Online J Public Health Inform.* 5(2).

Conway, P. H., and Berwick, D. M. 2011. Improving the rules for hospital participation in medicare and medicaid. *JAMA* 306(20):2256–2257.

Cox, D. R. 1958. The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)* 215–242.

Donze, J.; Aujesky, D.; Williams, D.; and Schnipper, J. L. 2013. Potentially avoidable 30-day hospital readmissions in medical patients: Derivation and validation of a prediction model. *JAMA Internal Medicine* 173(8):632–638.

Futoma, J.; Morris, J.; and Lucas, J. 2015. A comparison of models for predicting early hospital readmissions. *Journal of Biomedical Informatics* 56:229 – 238.

He, D.; Mathews, S. C.; Kallou, A. N.; and Hutfless, S. 2014. Mining high-dimensional administrative claims data to predict early hospital readmissions. *J Am Med Inform Assoc.* 21(2):272–279.

Hewner, S.; Seo, J. Y.; Gothard, S. E.; and Johnson, B. 2014. Aligning population-based care management with chronic disease complexity. *Nursing Outlook* 62(4).

Jamei, M.; Nisnevich, A.; Wetchler, E.; Sudat, S.; and Liu, E. 2017. Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. *PLoS ONE* 12:e0181173.

Jiang, S.; Chin, K.-S.; Qu, G.; and Tsui, K. L. 2018. An integrated machine learning framework for hospital readmission prediction. *Knowledge-Based Systems* 146:73 – 90.

Johnson, A. E. W.; Pollard, T. J.; Shen, L.; Lehman, L.-w. H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; and Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data* 3:160035 EP –.

Kansagara, D.; Englander, H.; Salanitro, A.; Kagen, D.; Theobald, C.; Freeman, M.; and Kripalani, S. 2011. Risk prediction models for hospital readmission: a systematic review. *Journal of American Medical Association* 306(15).

Langford, J.; Li, L.; and Zhang, T. 2008. Sparse online learning via truncated gradient. *Journal of Machine Learning Research* 10:777–801.

Lin, Y.-W.; Zhou, Y.; Faghri, F.; Shaw, M. J.; and Campbell, R. H. 2018. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *bioRxiv*.

Liu, J.; Ji, S.; and Ye, J. 2009. *SLEP: Sparse Learning with Efficient Projections*. Arizona State University.

Mosci, S.; Rosasco, L.; Santoro, M.; Verri, A.; and Villa, S. 2010. Solving structured sparsity regularization with proximal methods. In *Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases: Part II, ECML PKDD'10*, 418–433.

2007. Promoting greater efficiency in medicare. MPA Committee Report to Congress.

Niu, Y. 2013. Regression Models for Readmission Prediction Using Electronic Medical Records. Master's thesis, Wayne State University, Detroit, Michigan.

Pfuntner, A.; Wier, L. M.; and Steiner, C. 2013. Costs for hospital stays in the united states, 2011. *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs* 168.

Singh, A.; Nadkarni, G.; Gutttag, J.; and Bottinger, E. 2014. Leveraging hierarchy in medical codes for predictive modeling. In *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB '14*, 96–103.

Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J.; and Knight, K. 2005. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society Series B* 91–108.

Tibshirani, R. 1994. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58:267–288.

- Xiao, C.; Ma, T.; Dieng, A. B.; Blei, D. M.; and Wang, F. 2018. Readmission prediction via deep contextual embedding of clinical concepts. *PloS one* 13(4):e0195024.
- Yu, S.; Esbroeck, A. v.; Farooq, F.; Fung, G.; Anand, V.; and Krishnapuram, B. 2013. Predicting readmission risk with institution specific prediction models. In *Proceedings of the 2013 IEEE International Conference on Healthcare Informatics*, 415–420.
- Yuan, M., and Lin, Y. 2006. Model selection and estimation in regression with grouped variables. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B* 68:49–67.
- Zhao, P.; Rocha, G.; and Yu, B. 2009. The composite absolute penalties family for grouped and hierarchical variable selection. *The Annals of Statistics* 37(6A):3468–3497.