Prescriptive Cluster-Dependent Support Vector Machines with an Application to Reducing Hospital Readmissions *

Taiyao Wang and Ioannis Ch. Paschalidis, *Fellow, IEEE*

Abstract— We augment linear Support Vector Machine (SVM) classifiers by adding three important features: (i) we introduce a regularization constraint to induce a sparse classifier; (ii) we devise a method that partitions the positive class into clusters and selects a sparse SVM classifier for each cluster; and (iii) we develop a method to optimize the values of controllable variables in order to reduce the number of data points which are predicted to have an undesirable outcome, which, in our setting, coincides with being in the positive class. The latter feature leads to personalized prescriptions/recommendations. We apply our methods to the problem of predicting and preventing hospital readmissions within 30-days from discharge for patients that underwent a general surgical procedure. To that end, we leverage a large dataset containing over 2.28 million patients who had surgeries in the period 2011–2014 in the U.S. The dataset has been collected as part of the American College of Surgeons National Surgical Quality Improvement Program (NSQIP).

Index Terms— Support Vector Machines, clustering, sparsity, prescriptive analytics, medical informatics, machine learning.

I. Introduction

The *Support Vector Machine (SVM)* [1] is a binary classifier, widely used in practice due to its tractability for large scale problems. To obtain an SVM classifier, one needs to solve a convex quadratic problem with linear constraints, which can be done for large problem instances involving thousands of samples and hundreds of variables for each sample. The goal of this paper is to exploit the SVM framework to move beyond predictions and attempt to "control" future outcomes by appropriately modifying some of the key predictive variables. To that end, we develop a new method we call *Prescriptive Support Vector Machine (PSVM)*.

We will apply the new method to an important problem in health care; preventing hospital readmissions. The need for systematic, quantitative methods for addressing health care problems is compelling. An estimated $3 trillion is spent annually on health care in the U.S., a value that exceeds 17% of the U.S. Gross Domestic Product (GDP) – by far the largest among the 13 high-income Organization for Economic Cooperation and Development countries. The Centers for Medicare and Medicaid Services have identified

hospital readmissions, defined as an additional admission to address the same issue within 30 days after discharge, as an important and potentially preventable source of excessive resource utilization and increased cost of care [2].

An analysis of 2005 Medicare claims demonstrated that about 75% of 30-day readmissions, representing about $12 billion in Medicare spending, were potentially preventable [3]. As a result, through the enactment of the Readmissions Reduction Program section of the Affordable Care Act of 2012, readmissions have been increasingly used as a quality of care metric, and their reduction is mandated for certain diseases [2]. In this context, many surgical departments in the U.S. are establishing processes aimed at reducing 30-day readmissions. We refer to [4] for a general discussion of the benefits and some potential risks associated with the application of health analytics.

Several works exploit classical machine learning approaches, such as random forests, gradient tree boosting, logistic regression, linear and kernelized SVM, and related methods for predicting 30-day readmissions in patients with heart failure [5], [6], [7]. Recently, the authors developed an interpretable classification approach to predict chronic disease hospitalizations based on past Electronic Health Records (EHRs), establishing convergence, sample complexity and generalization guarantees [8], [9], [10]. Interpretability is indeed critical for medical and health informatics, as well as other areas, e.g., safety and security management. Without interpretable models, physicians may not use "black box" predictions even if they are highly accurate.

In this paper, we augment earlier SVM-based predictive analytics along three directions. First, we use a sparsity-inducing ℓ_1-norm-based constraint to obtain sparse classifiers which can generalize better out-of-sample and provide interpretability. Second, we leverage our work in [8], [10] to solve a joint clustering and classification problem and discover hidden clusters in the positive class and corresponding, per-cluster SVM-based classifiers. The third direction is the development of *prescriptive analytics*. In our setting, this consists of a method which leverages the SVM-based predictive model to devise personalized interventions with the potential to prevent a readmission by controlling/optimizing the value of some variables characterizing the patient.

There have only been very few works focusing on so-called *prescriptive analytics*. An example is [11], [12], which develop a data-driven framework to prescribe an optimal decision in a setting where the cost depends on uncertain problem parameters that need to be learned from data.

We apply our methods to a data set containing over 2.28

Department of Electrical and Computer Engineering, Division of Systems Engineering, and Center for Information and Systems Engineering, Boston University, Boston, MA 02215, USA. E-mail: {wty, yannisp}@bu.edu, url: http://sites.bu.edu/paschalidis/.

million patients who had surgeries in 2011–2014 in the U.S. The data are collected as part of American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP) [13]. Earlier work studied risk factors for 30-day readmissions for categories of surgical patients, e.g., orthopaedic trauma injuries [14], knee and hip arthroplasty [15], and ventral hernia repair [16]. A simple readmission score using few variables was developed based on 2011 NSQIP data only in [17]. To the best of our knowledge, our work is the first to develop analytics for 30-day readmissions after general surgery using millions of NSQIP records.

The remainder of this paper is organized as follows. Sec. II reviews SVM-based classification and the joint clustering and classification method [10]. These are key building blocks for the prescriptive method which is presented in Sec. III. The data and pre-processing steps are outlined in Sec. IV. Experimental results are in Sec. V and conclusions in Sec. VI.

Notation: All vectors are column vectors and are denoted by bold lowercase letters. For economy of space, we write $\mathbf{x} = (x_1, \ldots, x_{\dim(\mathbf{x})})$ to denote the column vector $\mathbf{x}$, where $\dim(\mathbf{x})$ is the dimension of $\mathbf{x}$. We use prime to denote the transpose of a vector. Unless otherwise specified, $\|\cdot\|$ denotes the ℓ_2 norm and $\|\cdot\|_1$ the ℓ_1 norm. We will use $\|\mathbf{x}\|_p = (\sum_{i=1}^{\dim(\mathbf{x})} |x_i|^p)^{1/p}$ to denote the ℓ_p norm, where $p \geq 1$. We will also use the notation $[N]$ for the set $\{1, \ldots, N\}$.

II. SVM BASED PREDICTIVE ANALYTICS

The SVM algorithm [1] seeks a separating hyperplane in the variable space, so that data samples from the two different classes reside on two different sides of the hyperplane. The minimum over all the distances from the input data samples to the hyperplane is called *margin*. The goal of SVM is to find the optimal hyperplane that has the maximum margin. In cases where data samples are neither linearly nor perfectly separable, the *soft-margin* SVM tolerates misclassification errors and can leverage kernel functions to map the features into a higher dimensional space where linear separability is possible (*kernelized SVMs*) [1].

A. SLSVM: Sparse Linear SVM

Following [8], [10] and our interest in interpretable classifiers, we formulate a Sparse version of Linear SVM (SLSVM) as follows. We are given training data $\mathbf{x}_i \in \mathbb{R}^D$ and labels $y_i \in \{-1, 1\}$, $i = 1, \ldots, n$, where $\mathbf{x}_i$ is the vector of variables characterizing the ith patient and $y_i = 1$ (resp., $y_i = -1$) indicates that the patient is (resp., is not) readmitted. We will refer to the class with labels equal to 1 as the positive class and the other class as the negative class.

We seek to find a hyperplane orthogonal to some vector $\boldsymbol{\beta} \in \mathbb{R}^D$ that passes from $-\beta_0 \in \mathbb{R}$, which can be done by solving the following quadratic programming problem:

$$\min_{\boldsymbol{\beta}, \beta_0, \xi_i} \quad \frac{1}{2} \|\boldsymbol{\beta}\|^2 + C \sum_{i=1}^{n} \xi_i \tag{1}$$
$$\text{s.t.} \quad \xi_i \geq 0, \quad \forall i,$$
$$y_i(\mathbf{x}_i' \boldsymbol{\beta} + \beta_0) \geq 1 - \xi_i, \quad \forall i,$$

$$\|\boldsymbol{\beta}\|_1 \leq T.$$

In the above formulation, the first term is proportional to one over the minimum distance between a hyperplane that passes from $-\beta_0 - 1$ and a hyperplane that passes from $-\beta_0 + 1$, i.e., one over the thickness of a band (margin) in which we would like to avoid placing any data points so as to increase the separability between the two classes. The parameter C is a tunable parameter and ξ_i is a misclassification penalty for each data point i. The constraint on $\|\boldsymbol{\beta}\|_1$ imposes sparsity in the variable vector $\boldsymbol{\beta}$, thus, allowing only a sparse subset of features to be selected for the classification decision. The parameter T is also tunable and controls the level of sparseness. There exist close connections to previous work, such as elastic net regularization [18], ℓ_1-norm SVM [19], and a robust optimization approach for obtaining appropriate regularizers to learning problems [20]. A drawback of the formulation (1) is that it is difficult to kernelize, while kernelized elastic net has also been proposed in [21].

B. JCC: Joint Clustering and Classification

In [8], [10] the authors have proposed a Joint Clustering and Classification (JCC) problem based on the Sparse Linear Support Vector Machine (SLSVM) framework. The SLSVM method we saw in Sec. II-A can in fact be seen as a special case of JCC where only one cluster is being used. The classification problem under consideration satisfies the following assumptions. (i) The negative class samples are assumed to be i.i.d. and drawn from a single cluster with distribution P_0. (ii) The positive class samples belong to L clusters, with distributions $P_1^1, \ldots, P_1^L$. (iii) Different positive clusters have different features that separate them from the negative samples (see Fig. 1 for an example).
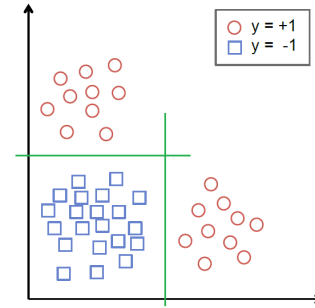

Fig. 1: The positive class contains two clusters and each cluster is linearly separable from the negative class.

Let $\mathbf{x}_i^+$ and $\mathbf{x}_j^-$ be the D dimensional positive and negative samples, y_i^+, y_j^- the corresponding labels, where $i \in [N^+]$ and $j \in [N^-]$ and $y_i^+ = 1$, $\forall i$ and $y_j^- = -1$, $\forall j$. Assuming L hidden clusters in the positive class, we try to discover: (a) the L hidden clusters (denoted by a mapping function $l(i) = l$, $l \in [L]$) and (b) L classifiers ($\boldsymbol{\beta}^l, \beta_0^l$), as the solution to the following Joint Clustering and Classification (JCC) problem:

$$\min_{\boldsymbol{\beta}^l, \beta_0^l, l(i)} \sum_{l=1}^{L} \left(\frac{1}{2} \|\boldsymbol{\beta}^l\|^2 + \lambda^+ \sum_{i:l(i)=l} \xi_i^{l(i)} + \lambda^- \sum_{j=1}^{N^-} \zeta_j^l \right) \tag{2}$$

$$\text{s.t.} \sum_{d=1}^{D} |\beta_d^l| \leq T^l, \quad \forall l \in [L],$$

$$\xi_i^{l(i)} \geq 1 - y_i^+ \beta_0^{l(i)} - \sum_{d=1}^{D} y_i^+ \beta_d^{l(i)} x_{i,d}^+, \forall i \in [N^+],$$

$$\zeta_j^l \geq 1 - y_j^- \beta_0^l - \sum_{d=1}^{D} y_j^- \beta_d^l x_{j,d}^-, \forall j \in [N^-], \forall l \in [L],$$

$$\xi_i^{l(i)}, \; \zeta_j^l \geq 0, \quad \forall i \in [N^+], \; \forall j \in [N^-], \forall l \in [L],$$

where T^l is a parameter controlling the sparsity of the classifier in cluster l.

In formulation (2), we have introduced different misclassification penalties, $\xi_i^{l(i)}$, ζ_j^l for positive and negative samples. In fact, the misclassification costs of the negative samples are counted L times, since these samples are drawn from a single distribution and are not clustered but simply copied into each cluster. The parameters λ^- and λ^+ control the relative weight of these misclassification costs from negative and positive samples and should be appropriately selected to negate the overcounting, specifically, we set $\lambda^+ = L\lambda^-$. The constraint $\sum_{d=1}^{D} |\beta_d^l| \leq T^l$ is an ℓ_1-relaxation of the sparsity requirement to the local classifiers, which is essential to align the formulation with the problem assumptions and to estimate more robust local classifiers. The selection of the tuning parameters is discussed in more detail in [8], [10].

Two different approaches have been proposed for (2) [8], [10]. The first, transforms the problem into a Mixed Integer Programming (MIP) problem but can only solve small-scale problems. The second approach is an alternating optimization approach which applies to large-scale problems and also gives rise to theoretical performance guarantees. It is shown in [8], [10] that it is better to perform joint clustering and classification instead of separating the two tasks.

III. SVM BASED PRESCRIPTIVE ANALYTICS

Prescriptive Support Vector Machines (PSVM) is a prescriptive method we introduce in this paper that builds on top of SLSVM and JCC.

Suppose we have generated the per-cluster optimal predictive hyperplanes using the JCC approach we described in Sec. II-B. Let $\mathscr{C}$ be an index set of variables for each patient we can control/modulate by applying certain interventions/therapies. For each patient i in the positive class, with variable vector $\mathbf{x}_i$, we are interested in optimizing the value of the controllable variables $x_{i,d}$, for $d \in \mathscr{C}$, so that the patient is predicted to belong to the negative class.

There is, however, a cost for large changes to the value of the controllable variables, which introduces a trade-off between "flipping" the patient to the negative class and implementing interventions that lead to large changes in the controllable variables (see Fig. 2). The following formulation optimizes a linear combination of the corresponding two terms in the objective. Specifically, consider a patient i in cluster l, where $i \in [N^+]$, $\mathbf{x}_i$ is vector of variables characterizing the patient, and $\mathbf{y}_i$ is the patient's variables after applying the prescription/intervention. Let $(\boldsymbol{\beta}^l, \beta_0^l)$ be the coefficients associated with the predictive hyperplane discovered by JCC
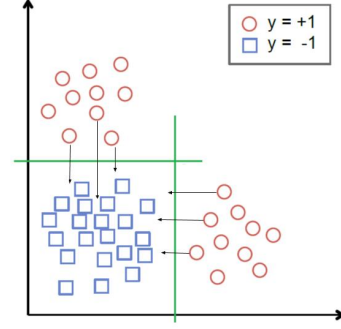

Fig. 2: The readmitted patients moved from the readmitted side of the prediction hyperplane to the non-readmitted side.

in the l-th cluster. To determine $\mathbf{y}_i$ we solve the following convex optimization problem:

$$\min_{\mathbf{y}_i, \xi_i} \quad \lambda \xi_i + \|\mathbf{y}_i - \mathbf{x}_i\|_p^p \tag{3}$$

$$\text{s.t.} \quad \beta_0^l + (\boldsymbol{\beta}^l)' \mathbf{y}_i \leq \xi_i - 1,$$
$$y_{i,d} = x_{i,d}, \; \forall d \notin \mathscr{C},$$
$$\xi_i \geq 0,$$
$$L_{i,d} \leq y_{i,d} \leq U_{i,d}, \; \forall d \in C,$$

where $L_{i,d}$ and $U_{i,d}$ are bounds on the controllable variables for each patient i. The parameter λ trades-off the failure to flip the patient to the negative side of the hyperplane with the required change in the patient's characteristics measured by the term $\|\mathbf{y}_i - \mathbf{x}_i\|_p^p$. The higher the value of λ, the more attention is given to the goal of preventing a readmission. To select an appropriate value for λ we can use cross-validation, based on some cost function that accounts for the cost of reducing readmissions and the cost of prescriptions.

Notice that problem (3) can be solved independently for each patient who is predicted to belong to the positive class (readmitted). Thus, it is naturally distributed and can obtain a prescription for each at-risk patient with only local computations. The form of the problem (3) depends on the selection of the ℓ_p norm; for instance, when $p = 2$, we have a quadratic programming problem and when $p = 1$, we have a linear programming problem.

IV. DATA AND PREPROCESSING

In this section we describe the data set we use to test and validate our methods.

A. NSQIP Dataset Description

The ACS-NSQIP was created to improve surgical techniques and outcomes and catalogs over 300 variables on comorbidities, intra-operative events, and 30-day outcomes using prospective random sampling [13]. It contains no protected health information.

The NSQIP dataset contains variables such as:

- Demographic and health care status characteristics, such as age, gender, race, body mass index, smoking, diabetes, hypertension requiring medication, and admittance from the emergency room.
- Procedure information (e.g., CPT code [22], ICD9 code [23]), the American Society of Anesthesiologists (ASA) classification, and wound classification.
- Pre-operative, intra-operative, and post-operative variables, including hospital length of stay information, superficial/deep/organ space surgical site infections, and existence/description of complications (e.g., pneumonia, infections, bleeding, thromboembolic events, etc.).
- Laboratories, pre-operative and post-operative values.

After data pre-processing steps we describe below, there were a total of $2,288,938$ de-identified patients, $135,293$ of whom were readmitted within 30 days, resulting in a readmission rate of 5.91%. A total of 230 variables were available for analysis, most of which were binary and integer with the remaining being continuous.

B. NSQIP Dataset Preprocessing

Data pre-processing steps we applied were as follows:
- Patients who died within 30 days from discharge were not included in the total of $2,288,938$ patients, as these events compete with readmission.
- Categorical variables (e.g., race, discharge, destination, insurance type, CPT code, ICD9 code) were numerically encoded by what is typically referred to as *one hot encoding*, which amounts to introducing a new indicator variable for each category.
- Missing values of categorical variables were treated as new categories and missing values of numerical variables were replaced by *k*-nearest-neighbors imputation.
- Features with small standard deviation (< 0.005) were removed.
- One of every two features which were highly linearly correlated (absolute value of correlation > 0.8) was removed.
- Feature scaling was applied for all features to bring all values into the $[0,1]$ range, specifically, all variables were normalized by subtracting the minimum and dividing by the range.

The variables were further separated into two classes: pre-operative variables and post-operative variables. Pre-operative variables are those that can be known before or during the main surgical procedure while post-operative variables, including complications, can only be determined after the surgery has been completed. The reason for considering these two classes of variables is that some post-operative variables may be affected by the controllable variables which may be modulated using our prescriptive method.

C. Controllable Variables

We consider three types of controllable variables on which to intervene using prescriptive analytics:
- Pre-operative lab tests: sodium, Blood Urea Nitrogen (BUN), serum creatinine, serum albumin, bilirubin,

SGOT (Serum Glutamic-Oxaloacetic Transaminase), alkaline phosphatase, White Blood Cell count (WBC), hematocrit (HCT), platelet count, Partial Thromboplastin Time (PTT), Prothrombin Time (PT), and International Normalized Ratio (INR) of PT values.
- Length of stay at the hospital: total length of stay, days from admission to operation, days from operation to discharge.
- SSI (Surgical Site Infection) or Infection: occurrences of deep incisional SSI, occurrences of organ space SSI, and post-operative occurrences of Urinary Tract Infection (UTI).

Pre-operative lab values could be altered through appropriate medications and treatment before the operation to bring them closer to levels not associated with readmission. The length of stay at the hospital could be to shortened, or lengthened as appropriate. Recommendations can also target the tightening of infection control measures that affect the variables described in the third item above. In the work we report in this paper we focus on the pre-operative hematocrit (HCT), as it is a variable that can be directly impacted (increased) through blood transfusion. The predictive models also suggest that pre-operative hematocrit (HCT) is one of the most important controllable variables.

V. PERFORMANCE EVALUATION AND EXPERIMENTAL RESULTS

A. Prediction Results

1) Prediction Accuracy: In the readmission prediction problem, one typically considers two distinct performance metrics computed out-of-sample, i.e., over a test set not seen during training. These metrics are the false positive rate (or false alarm rate, or one minus the specificity of the test) and the detection rate (or the true positive rate, or sensitivity of the test). A Receiver Operating Characteristic (ROC) curve, is a curve that evaluates the performance of a binary classifier as the decision threshold is varied, created by plotting the true positive rate against the false positive rate at different threshold settings. To have a single metric to compare different ROC curves, we will consider the Area Under the ROC Curve (AUC). An ideal prediction model has an AUC equal to 1, whereas a random prediction would yield an AUC of 0.5. Anything with an AUC greater than 0.7 is considered a moderately good predictive model.

We randomly chose 60% and 20% of the patients in the dataset to form the training and validation set and keep the remaining 20% of the patients as a test set.

We compared the methods we presented in Sec. II with some standard machine learning methods, namely, Random Forest (RF) [24] and Logistic Regression (LR) [25]. The Random forest [24] is a large collection of decision trees and it classifies by averaging the decisions of each tree. Logistic regression is widely used as a base for comparison in medical machine learning studies. In this work, a logistic regression model was fitted with an additional regularization term: an ℓ_2-norm term (similar to ridge regression) [25].

In Table I, we compare the performance of the various classification methods: Random Forests (RF) [24], SLSVM, ℓ_2-regularized logistic regression (L2LR) with pre-operative variables and post-operative variables [25]. JCC was also applied to the problem but resulted into a single positive cluster, which is identical to SLSVM. In Table I, the 2nd column reports AUC using only pre-operative variables and the 3rd column lists the corresponding AUC using all (pre-operative and post-operative) variables.

Methods were implemented in Python (Python Software Foundation, https://www.python.org/) [26] and Matlab (MathWorks, Natick, MA). For random forests, the number of trees grown was 500. Cross-validation was used to tune parameters of all methods, e.g., the number of variables randomly sampled as candidates at each split for RF, regularization strength for SLSVM and L2LR.

TABLE I: Performance of the various classification methods.

Method	pre-op AUC	post-op AUC
L2LR	72.32%	83.53%
RF	73.11%	84.91%
SLSVM	72.28%	83.48%

Based on the results of Table I, using post-operative variables results into substantially better performance. AUCs of all the methods were similar, perhaps because the NSQIP dataset contained a large amount of data and a sufficient number of highly predictive features. It is interesting that we can predict with such a high accuracy 30-day readmissions. In fact, just this information can be extremely useful as the health care system can target at-risk patients and monitor them post-discharge to reduce the risk of readmission.

2) Important Variables: For each variable, we computed a two-tailed p-value using Welch's t-test, where the null hypothesis was that the two cohorts (readmitted and non-readmitted patients) have equal means. We found 177 variables with a p-value less than 10^{-6}.

Using this analysis, the variables with the most statistically significant values in the two patient cohorts (readmitted and non-readmitted), were: (*i*) return to Operating Room (OR) after the main surgery and before discharge, (*ii*) length of stay, (*iii*) occurrences of Surgical Site Infection (SSI, either organ/space SSI, superficial SSI, deep incisional SSI), (*iv*) occurrences of urinary tract infection, (*v*) occurrences Deep Vein Thrombosis (DVT)/thrombophlebitis, (*vi*) occurrences of pulmonary embolism, (*vii*) pneumonia occurrences, (*viii*) estimated probability of morbidity, (*ix*) occurrences of sepsis, (*x*) occurrences myocardial infarction, (*xi*) occurrences of progressive renal insufficiency, (*xii*) stroke with neurological deficit, (*xiii*) disseminated cancer, (*xiv*) patient currently on dialysis (pre-op), (*xv*) pre-operative HCT, (*xvi*) total operation time (in minutes), and (*xvii*) Body Mass Index (BMI).

B. Prescriptive Results

In this section we evaluate the effectiveness of prescriptions obtained by solving problem (3) for each patient. We focus on optimizing the patient's pre-operative hematocrit (HCT) using a blood transfusion. Transfusions infuse blood into the patient bloodstream and, typically, a discrete number of bags, each containing 100cc of blood, gets prescribed. We will limit the number of bags of blood given to a patient to 3, corresponding to 300cc of blood, which can be considered as a safe upper limit for blood transfusion. Each bag of blood given to patient increases HCT by roughly 3%. We thus define 4 possible treatments as follows:

- Treatment 1: No transfusion.
- Treatment 2: 1 bag of blood transfusion.
- Treatment 3: 2 bags of blood transfusion.
- Treatment 4: 3 bags of blood transfusion.

Since we do not have in the NSQIP data information on whether a blood transfusion has been performed, we assume a baseline treatment depending on the patient's HCT as follows:

- For female patients, if HCT<37, 0 bags of blood are assumed to have been given; if 37<HCT<40, 1 bag of blood is assumed to have been given; if 40<HCT<43, 2 bags of blood are assumed to have been given; and, if HCT>43, 3 bags of blood are assumed to have been given.
- For male patients, if HCT<41, 0 bags of blood are assumed to have been given; if 41<HCT<44, 1 bag of blood is assumed to have been given; if 44<HCT<47, 2 bags of blood are assumed to have been given; and, if HCT>47, 3 bags of blood are assumed to have been given.

We then use formulation (3) to obtain a prescription for each patient in a test dataset, using the ℓ_2 norm ($p = 2$) in the penalty associated with the prescribed change in the patient's HCT. In the absence of ground truth, we then evaluate the effect of the prescription using a variety of prescriptive methods. We will compare the readmission rate when prescriptions are being implemented with a baseline rate set to be equal to the actual readmission rate of 5.85% for patients in the test set. To be able to compare the readmission rate with or without the prescriptions, we calibrate each predictive model by selecting a decision threshold (i.e., a point on the ROC curve corresponding to the model) so that the model yields the same readmission rate of 5.85% in the absence of any prescriptions.

Table II reports the results. The first column lists the predictive model used to evaluate the effects of the prescriptions. The second column lists the readmission rate after the optimal prescription is applied to each patient in the test set. The third column lists the readmission rate in the absence of any prescriptions. The average reduction of the readmission rate across the three predictive models is 1.24%, which implies a 21% decrease compared to the baseline readmission rate of 5.85%. The average percentage change of HCT due to the prescriptions is equal to 4.40%.

VI. Conclusions

We developed a new framework to decide prescriptions or other interventions that reduce the rate of an undesirable

TABLE II: Effect of optimal prescriptions.

Method	Prescriptive rate	Baseline rate
L2LR	4.95%	5.85%
RF	4.70%	5.85%
SLSVM	4.18%	5.85%

event. We build this prescriptive capability based on an SVM-based predictive model. Decisions can be decomposed for each subject (patient).

We applied this new framework to a large dataset of 2.28 million patients tracked by the ACS-NSQIP over a four year period (2011–2014). We considered personalized decisions to potentially increase the pre-operative HCT for each patient through a blood transfusion. The objective of the prescriptive method is to prevent 30-day readmissions and reduce the corresponding readmission rate.

Our results show that our prescriptive SVM approach reduces the readmission rate by an average of 1.24% from the readmission rate of 5.85% in the absence of prescriptions. This amounts to a relative percentage decrease of 21%. Considering that more than $12 billion were spent in 2005 by Medicare on potentially preventable readmissions, these types of readmission rate reductions can lead to dramatic savings on an annual basis. Future work will consider kernelized methods and speeding them up them for large-scale datasets.

ACKNOWLEDGMENTS

REFERENCES

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
[2] "Readmissions reduction program," https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program.html, accessed: 2018-02-22.
[3] J. James, *Medicare Hospital Readmissions Reduction Program: To Improve Care and Lower Costs, Medicare Imposes a Financial Penalty on Hospitals with Excess Readmissions.* Project HOPE, 2013.
[4] I. C. Paschalidis, "How machine learning is helping us predict heart disease and diabetes," *Harvard Business Review*, May 30 2017. [Online]. Available: https://hbr.org/2017/05/how-machine-learning-is-helping-us-predict-heart-disease-and-diabetes
[5] R. Amarasingham, B. J. Moore, Y. P. Tabak, M. H. Drazner, C. A. Clark, S. Zhang, W. G. Reed, T. S. Swanson, Y. Ma, and E. A. Halm, "An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data," *Medical care*, vol. 48, no. 11, pp. 981–988, 2010.
[6] J. D. Frizzell, L. Liang, P. J. Schulte, C. W. Yancy, P. A. Heidenreich, A. F. Hernandez, D. L. Bhatt, G. C. Fonarow, and W. K. Laskey, "Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches," *JAMA cardiology*, vol. 2, no. 2, pp. 204–209, 2017.
[7] S. B. Golas, T. Shibahara, S. Agboola, H. Otaki, J. Sato, T. Nakae, T. Hisamitsu, G. Kojima, J. Felsted, S. Kakarmath *et al.*, "A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data," *BMC medical informatics and decision making*, vol. 18, no. 1, p. 44, 2018.
[8] T. S. Brisimi, T. Xu, T. Wang, W. Dai, W. G. Adams, and I. C. Paschalidis, "Predicting chronic disease hospitalizations from electronic health records: An interpretable classification approach," *Proceedings of the IEEE*, vol. 106, no. 4, pp. 690–707, 2018.
[9] T. S. Brisimi, T. Xu, T. Wang, W. Dai, and I. C. Paschalidis, "Predicting diabetes-related hospitalizations based on electronic health records," *Statistical methods in medical research*, p. 0962280218810911, 2018.
[10] T. Xu, T. S. Brisimi, T. Wang, W. Dai, and I. C. Paschalidis, "A joint sparse clustering and classification approach with applications to hospitalization prediction," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*, 2016, pp. 4566–4571.
[11] D. Bertsimas and N. Kallus, "From predictive to prescriptive analytics," *arXiv preprint arXiv:1402.5481*, 2014.
[12] D. Bertsimas and B. Van Parys, "Bootstrap robust prescriptive analytics," *arXiv preprint arXiv:1711.09974*, 2017.
[13] A. M. Ingraham, K. E. Richards, B. L. Hall, and C. Y. Ko, "Quality improvement in surgery: the american college of surgeons national surgical quality improvement program approach," *Advances in surgery*, vol. 44, no. 1, pp. 251–267, 2010.
[14] V. Sathiyakumar, C. S. Molina, R. V. Thakore, W. T. Obremskey, and M. K. Sethi, "Asa score as a predictor of 30-day perioperative readmission in patients with orthopaedic trauma injuries: an nsqip analysis," *Journal of orthopaedic trauma*, vol. 29, no. 3, pp. e127–e132, 2015.
[15] A. J. Pugely, J. J. Callaghan, C. T. Martin, P. Cram, and Y. Gao, "Incidence of and risk factors for 30-day readmission following elective primary total joint arthroplasty: analysis from the acs-nsqip," *The Journal of arthroplasty*, vol. 28, no. 9, pp. 1499–1504, 2013.
[16] F. Lovecchio, R. Farmer, J. Souza, N. Khavanin, G. A. Dumanian, and J. Y. Kim, "Risk factors for 30-day readmission in patients undergoing ventral hernia repair," *Surgery*, vol. 155, no. 4, pp. 702–710, 2014.
[17] D. J. Lucas, A. Haider, E. Haut, R. Dodson, C. L. Wolfgang, N. Ahuja, J. Sweeney, and T. M. Pawlik, "Assessing readmission after general, vascular, and thoracic surgery using acs-nsqip," *Annals of surgery*, vol. 258, no. 3, p. 430, 2013.
[18] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
[19] J. Zhu, S. Rosset, R. Tibshirani, and T. J. Hastie, "1-norm support vector machines," in *Advances in neural information processing systems*, 2004, pp. 49–56.
[20] R. Chen and I. C. Paschalidis, "A robust learning approach for regression models based on distributionally robust optimization," *Journal of Machine Learning Research*, vol. 19, no. 13, 2018. [Online]. Available: http://jmlr.org/papers/v19/17-295.html
[21] Y. Feng, S.-G. Lv, H. Hang, and J. A. Suykens, "Kernelized elastic net regularization: Generalization bounds, and sparse recovery," *Neural computation*, vol. 28, no. 3, pp. 525–562, 2016.
[22] "Current Procedural Terminology (CPT)," American Medical Association, 2014. [Online]. Available: www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt.page
[23] "International classification of diseases, 9th revision (ICD9)," World Health Organization, 1999. [Online]. Available: www.who.int/classifications/icd/en/
[24] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
[25] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning.* Springer series in statistics Springer, Berlin, 2001, vol. 1.
[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.