

Colon Surgery Outcome Prediction Using ACS NSQIP Data

Ankit Agrawal,
Reda Al-Bahrani
Dept. of Electrical Engg. and
Computer Science
Northwestern University
2145 Sheridan Rd
Evanston, IL 60201
USA
{ankitag,reda}
@eecs.northwestern.edu

Ryan Merkow,
Karl Bilimoria
Dept. of Surgery
Northwestern University
Chicago, IL 60611
USA
{r-merkow,k-bilimoria}
@northwestern.edu

Alok Choudhary
Dept. of Electrical Engg. and
Computer Science
Northwestern University
2145 Sheridan Rd
Evanston, IL 60201
USA
choudhar@eecs.northwestern.edu

ABSTRACT

We analyze colon surgery data from the ACS NSQIP program with the aim of developing accurate risk prediction models for post-operative adverse outcomes in colon surgery using data mining techniques. The data used in this study is de-identified and consists of 23 pre-operative risk factors, and 30-day postoperative mortality, serious morbidity, and overall morbidity outcomes for patients undergoing major colon surgical procedures in the year 2011. Our dataset had 27,011 such patient instances. Several data mining classification techniques were used on this data along with various data mining optimizations and validations to build predictive models for each of the three adverse outcomes, and were able to achieve a c-statistic of 0.905, 0.771, and 0.737 for 30-day mortality, serious morbidity, and overall morbidity respectively. Further, we also applied feature selection techniques to reduce the number of pre-operative risk factors in the model to 6, 5, and 5 for the three outcomes, while trying to have minimal degradation in c-statistic (0.88, 0.757, and 0.727 respectively).

Categories and Subject Descriptors

H.2.8 [Database Applications]: Data mining; J.3 [Life and Medical Sciences]: Medical information systems

Keywords

Biomedical informatics, Colon surgery, Decision making, Predictive modeling

1. INTRODUCTION

Accurate risk estimation for post-operative morbidity (complications) and mortality can improve both informed patient

consent by helping patients better understand risks and benefits, and also aid the physicians in surgical decision making by assessing the true patient-specific risks of a proposed procedure rather than relying on population-wide risk assessments [31]. It is estimated that more than 30 million surgical operations in the U.S. annually to remove deadly cancers, repair diseased organs and replace worn-out joints, resulting in more than 290,000 surgical-site infections each year, which cost about \$10 billion annually [5, 32, 6]. Thus, accurate risk estimation can potentially save thousands of complications and also reduce healthcare costs.

Colon cancer is the second most common cancer in women and third most common in men [24], and fourth most common cause of cancer death [3]. In the ACS NSQIP dataset used in this study, about 4% of patients did not survive more than 30 days after surgery, 26% patients developed serious morbidity, and 32% patients developed some kind of morbidity. Colon operations are relatively common and pose nontrivial morbidity and mortality risks [37], and there exist nearly half-a-dozen tools for its mortality risk assessment [11].

The American College of Surgeons (ACS) is a scientific and educational association of surgeons founded in 1913 to standardize surgical care, and its National Surgical Quality Improvement Program (ACS NSQIP) is the first nationally validated, risk-adjusted, outcomes-based program to measure and improve the quality of surgical care. It collects data on patient demographics, preoperative risk factors, lab-values, operative variables, and postoperative events using standardized definitions. Here, we use the ACS NSQIP data for colon surgery (both cancer and non-cancer) for the year 2011. The data has three binary variables for adverse outcomes - mortality, overall morbidity (any complication), and serious morbidity, all within 30 days of the surgical operation.

Applying data mining techniques to surgery data is useful to rank and link available attributes to the outcome. Here we use data mining techniques to estimate the risk of the three post-operative outcomes. Experiments with nearly 50 modeling techniques were conducted to find the best model for each of the three outcomes. Further, feature selection was used to find smaller attribute subsets in the data that could incur only minimal loss in accuracy, if at all.

The rest of the paper is organized as follows: Section 2

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

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

Copyright © 2013 ACM 978-1-4503-2174-7/13/08 ...\$15.00.

summarizes related work, followed by a brief description of the data mining techniques used in this study in Section 3. NSQIP data used in this work is described in Section 4. Experiments and results are presented in Section 5, followed by the conclusion and future work in Section 6.

2. RELATED WORK

Optimizing risk adjustment methodologies is an important endeavor, particularly for quality improvement programs that seek accurate risk predictions and ultimately hospital comparisons. In the surgical arena, much work has focused on hierarchical modeling strategies [6, 12] as well as identifying the important predictors of postoperative outcomes [33, 32]. For example, in a recent article, Cohen et al [6] discuss the developmental history of ACS NSQIP modeling and the shift from logistic regression to generalized linear mixed models. Although the program recently moved to hierarchical models, it has not yet attempted a machine learning approach. A separate report by Syed and colleagues [39] implemented a computer based learning strategy to optimize risk adjustment. These investigators used a single method (i.e., support vector machines) to learn the relationships between CPT codes and morbidity and mortality. However, this report focused primarily on improving complexity adjustment and used a single method. In the present study, we sought to comprehensively examine nearly 50 different methods.

The concept that adequate risk adjustment can be performed on the basis of a limited number of predictors has been previously established using American College of Surgeons National Surgical Quality Improvement Program data. For example, in a 2010 report by Dimick et al [13], separate 5-predictor colectomy morbidity (ASA class, functional status, emergency surgery, albumin level and body mass index) and mortality (ASA class, functional status, emergency surgery, albumin level, dyspnea) models were highly correlated with the more complex models with respect to discrimination, calibration, and hospital-level performance. In another report, Merkow et al [32] identified 6 predictors (ASA class, procedural risk, functional status, emergency surgery and wound class) for the NQF endorsed death or serious morbidity colectomy model. They also found near equivalence in model discrimination and calibration. Nevertheless, these investigators did not use powerful computer based learning techniques.

3. DATA MINING TECHNIQUES

3.1 Modeling

We used 46 classification schemes in this study, including both direct classification models and their ensembles using various ensembling techniques. Due to space limitations, here we briefly describe only those classifiers whose results we present in the next section.

1. **Support vector machines:** SVMs are based on the Structural Risk Minimization (SRM) principle from statistical learning theory. A detailed description of SVMs and SRM is available in [40].
2. **Artificial neural networks:** ANNs are networks of interconnected artificial neurons, and are commonly used for non-linear statistical data modeling to model

complex relationships between inputs and outputs. Several good descriptions of neural networks are available [7, 14].

3. **Decision Table:** Decision table typically constructs rules involving different combinations of attributes, which are selected using an attribute selection search method. Simple decision table majority classifier [29] has been shown to sometimes outperform state-of-the-art classifiers.
4. **KStar:** KStar [10] is a lazy instance-based classifier, i.e., the class of a test instance is based upon the class of those training instances similar to it, as determined by some similarity function.
5. **J48 decision tree:** J48 (or C4.5) is a decision tree based classifier. While constructing the decision tree, the J48 algorithm [34] identifies the attribute that must be used to split the tree further based on the notion of information gain/gini impurity.
6. **Reduced error pruning tree:** Commonly known as REPTree [42], it is a implementation of a fast decision tree learner, which builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning.
7. **Random forest:** The Random Forest [9] classifier consists of multiple decision trees. The final class of an instance in a Random Forest is assigned by outputting the class that is the mode of the outputs of individual trees, which can produce robust and accurate classification, and ability to handle a very large number of input variables.
8. **Alternating decision tree:** ADTree [15] is decision tree classifier which supports only binary classification. It consists of two types of nodes: decision nodes (specifying a predicate condition, like 'age' > 45) and prediction nodes (containing a single real-value number). An instance is classified by following all paths for which all decision nodes are true and summing the values of any prediction nodes that are traversed.
9. **Decision stump:** A decision stump [42] is a weak tree-based machine learning model consisting of a single-level decision tree with a categorical or numeric class label. Decision stumps are usually used in ensemble machine learning techniques.
10. **M5 Model Trees:** M5 Model Trees [41] are a reconstruction of Quinlan's M5 algorithm [35] for inducing trees of regression models, which combines a conventional decision tree with the option of linear regression functions at the nodes.
11. **Naive Bayes:** The naive bayes classifier [18] is a simple probabilistic classifier that is based upon the Bayes theorem. This classifier makes strong assumptions about the independence of the input features, which may not always be true.
12. **Bayesian Network:** A Bayesian network is a graphical model that encodes probabilistic relationships among a set of variables, representing a set of random variables and their conditional dependencies via a directed acyclic graph (DAG).

13. **Logistic Regression:** Logistic Regression [22] is used for prediction of the probability of occurrence of an event by fitting data to a sigmoidal S-shaped logistic curve. Logistic regression is often used with ridge estimators [30] to improve the parameter estimates and to reduce the error made by further predictions.
14. **AdaBoost:** AdaBoost [16] is a commonly used ensemble technique for boosting a nominal class classifier. In general, boosting can be used to significantly reduce the error of any weak learning algorithm that consistently generates classifiers which need only be a little bit better than random guessing.
15. **LogitBoost:** The LogitBoost algorithm is an ensemble technique implementation of additive logistic regression which performs classification using a regression scheme as the base learner, and can handle multi-class problems. In [17], the authors explain the theoretical connection between Boosting and additive models.
16. **Bagging:** Bagging [8] is a meta-algorithm to improve the stability of classification and regression algorithms by reducing variance. Bagging is usually applied to decision tree models to boost their performance.
17. **Random subspace:** The Random Subspace classifier [21] constructs a decision tree based classifier consisting of multiple trees, which are constructed systematically by pseudo-randomly selecting subsets of features, trying to achieve a balance between overfitting and achieving maximum accuracy.
18. **Rotation Forest:** Rotation forest [36] is a method for generating classifier ensembles based on feature extraction, which can work both with classification and regression base learners. The training data for a the base classifier is created by applying Principal Component Analysis (PCA) [25] to K subsets of the feature set, followed by K axis rotations to form the new features for the base learner, to encourage simultaneously individual accuracy and diversity within the ensemble.

3.2 Feature Selection

We used 2 feature selection techniques in this study - first to find a subset of features from the available feature set, and then to evaluate the predictive potential of the each of the attribute in the resulting subset of features.

1. **Correlation Feature Selection (CFS):** CFS is used to identify a subset of features highly correlated with the class variable and weakly correlated amongst them [19]. CFS was used in conjunction with a greedy step-wise search to find a subset with best average merit.
2. **Information Gain:** This is used to assess the relative predictive power of the predictor attributes, which evaluates the worth of an attribute by measuring the information gain with respect to the outcome status: $IG(Class, Attrib) = H(Class) - H(Class|Attrib)$, where $H(.)$ denotes the information entropy.

4. ACS NSQIP DATA

The developmental history and current details of ACS NSQIP, including sampling strategy, data abstraction procedures, variables collected, outcomes, and structure are described elsewhere [2, 1, 6, 26, 27, 28, 23]. In brief, hospitals collect standardized and audited data on patient demographics, preoperative risk factors, laboratory values, operative variables, and postoperative complications. Trained Surgical Clinical Reviewers (SCR) using standard ACS NSQIP tools and definitions gather data based on established timelines. Patients are followed for postoperative outcomes for 30 days after the index operation irrespective of whether the patient is an inpatient, has been discharged to their home or another facility, or has been readmitted to another hospital. Data definitions are rigorous and standardized across all participating institutions [38].

Postoperative outcomes assessed in this study were mortality, serious morbidity, and overall morbidity. Serious morbidity was defined as the occurrence of any one of the following surgical or medical complications: stroke or cerebrovascular accident, coma (lasting greater than 24 hours), peripheral nerve injury, myocardial infarction, cardiac arrest, pneumonia, ventilation dependence (greater than 48 hours), reintubation, acute renal insufficiency or failure, venous thromboembolism, sepsis or septic shock, organ space/deep surgical site infection (SSI), wound dehiscence, graft failure or postoperative bleeding requiring a blood transfusion. Overall morbidity was defined as the occurrence of any of the above-mentioned adverse events with the addition of superficial SSI or urinary tract infection. Patients were precluded from being categorized as having the following complications if the condition was documented preoperatively: SSI, pneumonia, ventilator dependence, reintubation or renal insufficiency/failure.

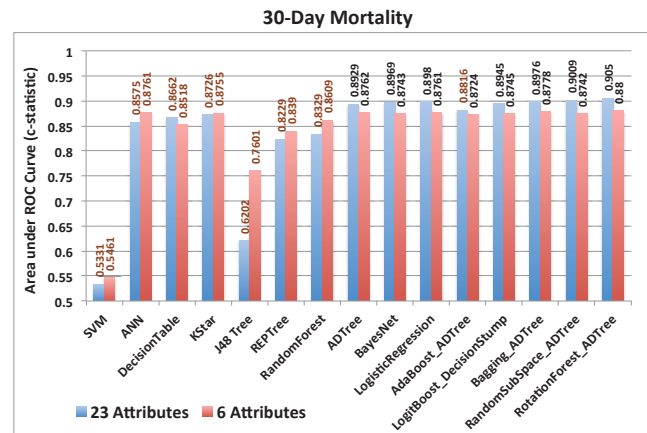


Figure 1: Prediction performance comparison for 30-day mortality in terms of area under the ROC curve (c-statistic).

5. EXPERIMENTS AND RESULTS

In our experiments, we used the WEKA toolkit 3.6.7 for data mining [20]. 3-fold cross-validation was used for evaluation. Area under the ROC curve, or c-statistic was used

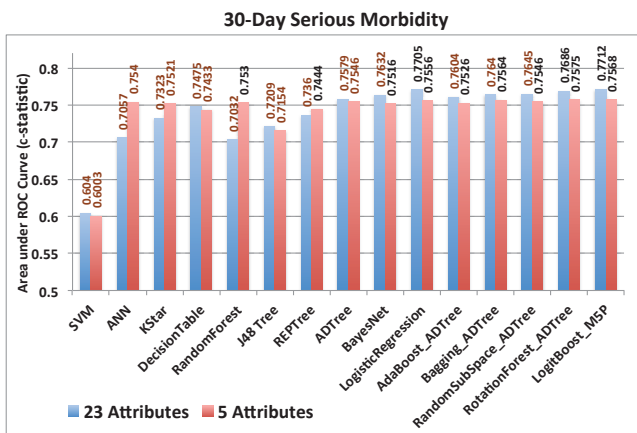


Figure 2: Prediction performance comparison for 30-day serious morbidity in terms of area under the ROC curve (c-statistic).

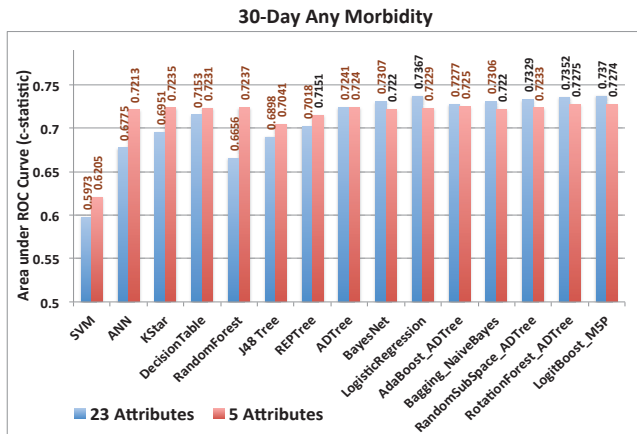


Figure 3: Prediction performance comparison for 30-day overall morbidity in terms of area under the ROC curve (c-statistic).

as the metric for evaluation.

Figures 1, 2, and 3 present the results on 15 classification schemes for the three post-operative adverse outcomes of 30-day mortality, 30-day serious morbidity, and 30-day overall morbidity respectively. The 15 classification schemes shown here vary for the three outcomes and consist of most of the popular classifiers. For each of the ensembling techniques, many underlying classifiers were tried but only the one with the best c-statistic is shown in these figures. Blue bars represent the results with the entire set of 23 attributes, and red bars represent the results with the reduced set after feature selection. As described earlier, the original dataset consisted of 23 attributes. Using CFS technique for feature selection yielded a subset of 6, 5, and 5 features for the post-operative outcomes of mortality, serious morbidity, and overall morbidity respectively.

In each of the figures, the technique that resulted in the best c-statistic is placed at extreme right. The number on

top of the each bar is the corresponding c-statistic. The numbers in brown represent that the c-statistic is significantly lower than the best model at $p=0.05$. Other numbers in black indicate that the performance is not statistically distinguishable from the best model at $p=0.05$. As is evident from these figures, there are many classification schemes that perform comparably well for these outcomes. For 30-day mortality, Rotation Forest with ADTrees as the underlying classifier gives the best c-statistic of 0.905 with 23 attributes, and of 0.88 with 6 attributes. For 30-day serious morbidity, LogitBoost with M5 model tree as the underlying classifier gives the best c-statistic of 0.771 with 23 attributes, and of 0.757 with 5 attributes. For 30-day overall morbidity, again LogitBoost with M5 model tree as the underlying classifier gives the best c-statistic of 0.737 with 23 attributes, and of 0.727 with 5 attributes.

Figure 4 presents the relative predictive power of the resulting smaller subset of attributes identified by CFS for each of the three post-operative adverse outcomes. There are 7 distinct attributes across the final attributes used in the three models: American Society of Anesthesiology (ASA) class (1 or 2, 3, 4 or 5), sepsis/septic shock, emergency procedure, ventilator dependence, functional status (independent, partially dependent, totally dependent), acute renal failure and the procedural group. The procedural groups were defined based on the principal Current Procedural Terminology [4] (CPT) code, and categorized into the following clinically meaningful groups: partial laparoscopy with anastomosis, partial laparoscopy with ostomy, partial open with anastomosis, partial open with ostomy, total laparoscopy with ostomy, total open with ostomy. These findings are important given the interest in determining a small set of attributes that are important for multiple outcomes.

6. CONCLUSION AND FUTURE WORK

In this workshop paper, we present our preliminary results of data mining on ACS NSQIP data on colon surgical outcomes. We evaluated nearly 50 classification schemes for each of the three post-operative outcomes - mortality, serious morbidity, and overall morbidity, all within 30 days of surgery. c-statistic of up to 0.905, 0.771, and 0.737 were achieved for the three outcomes respectively. Further, feature selection techniques were able to significantly reduce the number of attributes in the model, incurring a minimal cost in c-statistic (0.88, 0.757, and 0.727 respectively). Given the prediction quality, we believe that the resulting models can be very useful to not only accurately estimate risk of post-operative adverse outcomes, but also aid doctors in decision making and improve informed patient consent by providing a better understanding of the risks involved in a particular treatment procedure, based on patient-specific attributes. Accurate risk prediction can potentially also save valuable resources by avoiding high risk procedures that may not be necessary for a particular patient.

Future work includes developing more complex models for the studied outcomes, and also exploring conditional outcome models using some intra-operative and/or post-operative outcomes (e.g. risk of 30-day mortality/morbidity, given that the patient has (not) suffered serious/overall morbidity within 5 days of surgery), and exploring the use of undersampling/oversampling to deal with unbalanced data. We also plan to do similar analysis for other types of surgical operations using both ACS NSQIP data and other

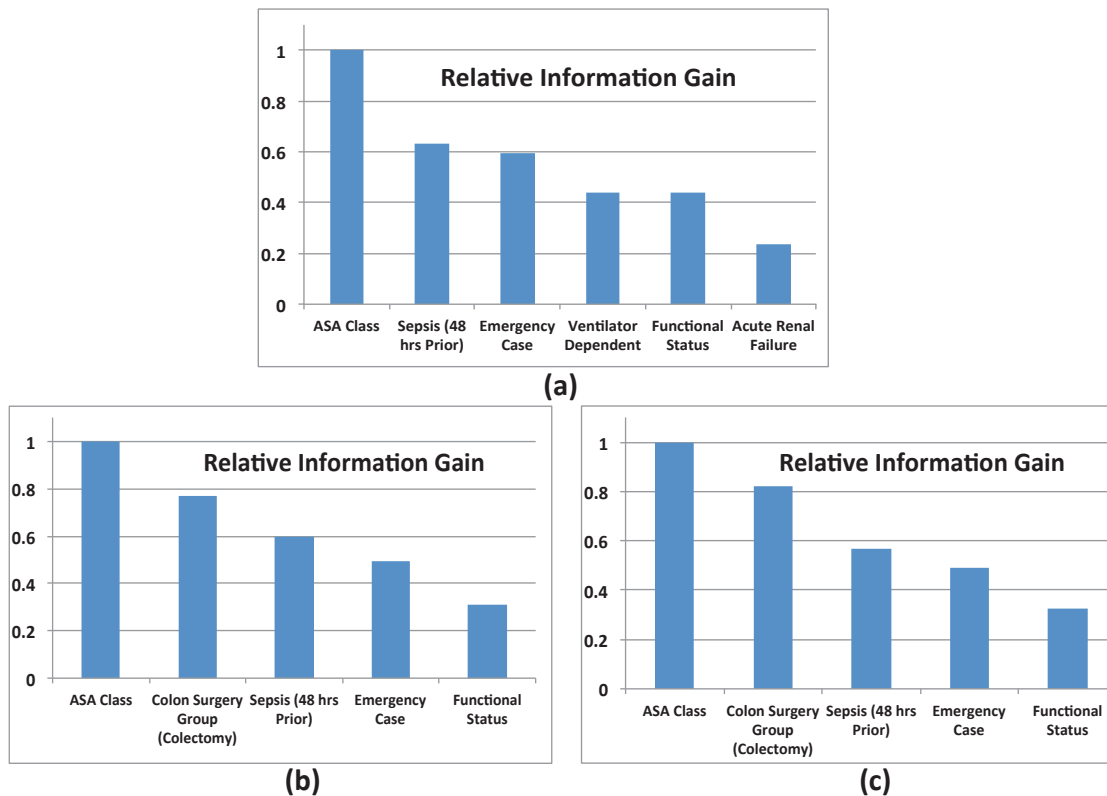


Figure 4: Relative information gain of features resulting from the CFS technique for (a) 30-day mortality, (b) 30-day serious morbidity, and (c) 30-day overall morbidity

available data. Finally, we would also like to integrate the current and future work into healthcare and clinical decision making in practice. One possible way to do so is to develop risk calculators for different types of outcomes.

7. ACKNOWLEDGMENTS

This work is supported in part by the following grants: NSF awards CCF-0833131, CNS-0830927, IIS-0905205, CCF-0938000, CCF-1029166, and OCI-1144061; DOE awards DE-FG02-08ER25848, DE-SC0001283, DE-SC0005309, DESC0005340, and DESC0007456; AFOSR award FA9550-12-1-0458.

8. REFERENCES

- [1] American college of surgeons - national surgical quality improvement program. available at <http://www.facs.org/acsnqip.html>. accessed january 1st, 2012.
- [2] American college of surgeons - national surgical quality improvement program: Semiannual report july 2012. chicago: American college of surgeons.
- [3] Cancer, world health organization, jan. 2013, <http://www.who.int/mediacentre/factsheets/fs297/en/> accessed may 12, 2013.
- [4] Cpt 2013: Current procedural terminology. chicago: American medical association, 2013.
- [5] Url: New ways to calculate the risks of surgery, wall street journal, feb. 2010, <http://online.wsj.com/article/SB10001424052748703422904575039110166900210.html>, accessed may 12, 2013.
- [6] Optimizing {ACS} {NSQIP} modeling for evaluation of surgical quality and risk: Patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. *Journal of the American College of Surgeons*, (0):-, 2013.
- [7] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford: University Press, 1995.
- [8] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123-140, 1996.
- [9] L. Breiman. Random forests. *Machine learning*, 45(1):5-32, 2001.
- [10] J. G. Cleary and L. E. Trigg. K*: An instance-based learner using an entropic distance measure. In *In Proceedings of the 12th International Conference on Machine Learning*, pages 108-114. Morgan Kaufmann, 1995.
- [11] M. E. Cohen, K. Y. Bilimoria, C. Y. Ko, and B. L. Hall. Development of an american college of surgeons national surgery quality improvement program: morbidity and mortality risk calculator for colorectal surgery. *Journal of the American College of Surgeons*, 208(6):1009-1016, 2009.
- [12] J. Dimick, A. Ghaferi, N. Osborne, C. Ko, and B. Hall. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg.*, 255(4):703-707, 2012.

- [13] J. Dimick, N. Osborne, B. Hall, C. Ko, and J. Birkmeyer. Risk adjustment for comparing hospital quality with surgery: how many variables are needed? *Journal of the American College of Surgeons*, 210(4):503–508, 2010.
- [14] L. Fausett. *Fundamentals of Neural Networks*. New York, Prentice Hall, 1994.
- [15] Y. Freund and L. Mason. The alternating decision tree learning algorithm. In *Proceeding of the Sixteenth International Conference on Machine Learning*, pages 124–133. Citeseer, 1999.
- [16] Y. Freund and R. E. Schapire. Experiments with a new boosting algorithm. 1996.
- [17] J. Friedman, T. Hastie, and R. Tibshirani. Special invited paper. additive logistic regression: A statistical view of boosting. *Annals of statistics*, 28(2):337–374, 2000.
- [18] H. George. John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345, 1995.
- [19] M. Hall. *Correlation-based feature selection for machine learning*. PhD thesis, Citeseer, 1999.
- [20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: An update. *SIGKDD Explorations*, 11(1), 2009.
- [21] T. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.
- [22] D. Hosmer and S. Lemeshow. *Applied Logistic Regression*. John Wiley and Sons, Inc., 1989.
- [23] A. Ingraham, K. Richards, B. Hall, and C. Ko. Quality improvement in surgery: the american college of surgeons national surgical quality improvement program approach. *Adv Surg.*, 44:251–67, 2010.
- [24] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman. Global cancer statistics. *CA: A Cancer Journal for Clinicians*, 61(2):69–90, 2011.
- [25] I. T. Jolliffe. *Principal Component Analysis*. Springer, second edition, 2002.
- [26] S. Khuri, J. Daley, W. Henderson, K. Hur, J. Demakis, J. Aust, V. Chong, P. Fabri, J. Gibbs, F. Grover, K. Hammermeister, G. Irvin, G. McDonald, E. Passaro, L. Phillips, F. Scamman, J. Spencer, and J. Stremple. The department of veterans affairs’ nsqip: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. national va surgical quality improvement program. *Ann Surg.*, 228(4):491–507, 1998.
- [27] S. Khuri, W. Henderson, J. Daley, O. Jonasson, R. Jones, D. Campbell, A. Fink, R. Mentzer, L. Neumayer, K. Hammermeister, C. Mosca, and N. Healey. Successful implementation of the department of veterans affairs’ national surgical quality improvement program in the private sector: the patient safety in surgery study. *Ann Surg.*, 248(2):326–36, 2008.
- [28] S. Khuri, W. Henderson, J. Daley, O. Jonasson, R. Jones, D. Campbell, A. Fink, R. Mentzer, and J. Steeger. The patient safety in surgery study: background, study design, and patient populations. *J Am Coll Surg.*, 204(6):1089–102, 2007.
- [29] R. Kohavi. The power of decision tables. In *Proceedings of the 8th European Conference on Machine Learning*, ECML ’95, pages 174–189, 1995.
- [30] S. le Cessie and J. van Houwelingen. Ridge estimators in logistic regression. *Applied Statistics*, 41(1):191–201, 1992.
- [31] M. M. Data-based risk calculators becoming more sophisticated and more popular. *JAMA*, 302(7):730–731, 2009.
- [32] R. Merkow, B. Hall, M. Cohen, X. Wang, J. Adams, W. Chow, E. Lawson, K. Bilimoria, K. Richards, and C. Ko. Validity and feasibility of the american college of surgeons colectomy composite outcome quality measure. *Annals of surgery*, 257(3):483–489, 2013.
- [33] R. Merkow, T. Kmiecik, D. Bentrem, D. Winchester, A. Stewart, C. Ko, and K. Bilimoria. Effect of including cancer-specific variables on models examining short-term outcomes. *Cancer*, 119(7):1412–9, 2012.
- [34] J. Quinlan. *C4. 5: programs for machine learning*. Morgan Kaufmann, 1993.
- [35] J. R. Quinlan. Learning with continuous classes. pages 343–348. World Scientific, 1992.
- [36] J. Rodriguez, L. Kuncheva, and C. Alonso. Rotation forest: A new classifier ensemble method. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(10):1619–1630, oct. 2006.
- [37] P. Schilling, J. Dimick, and J. Birkmeyer. Prioritizing quality improvement in general surgery. *J Am Coll Surg.*, 207(5):698–704, 2008.
- [38] M. Shiloach, S. Frencher, J. Steeger, K. Rowell, K. Bartzokis, M. Tomeh, K. Richards, C. Ko, and B. Hall. Toward robust information: data quality and inter-rater reliability in the american college of surgeons national surgical quality improvement program. *J Am Coll Surg.*, 201(1):6–16, 2010.
- [39] Z. Syed, I. Rubinfeld, J. Patton, J. Ritz, J. Jordan, A. Doud, and V. Velanovich. Using procedural codes to supplement risk adjustment: a nonparametric learning approach. *Journal of the American College of Surgeons*, 212(6):1086–1093, 2011.
- [40] V. N. Vapnik. The nature of statistical learning theory. *Springer*, 1995.
- [41] Y. Wang and I. Witten. Induction of model trees for predicting continuous classes. In *Proc European Conference on Machine Learning Poster Papers*, pages 128–137, Prague, Czech Republic, 1997.
- [42] I. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Pub, 2005.