
Unsupervised Risk Stratification in Clinical Datasets: Identifying Patients at Risk of Rare Outcomes

Zeeshan Syed

University of Michigan, 2260 Hayward St., Ann Arbor, MI 48109 USA

ZHS@EECS.UMICH.EDU

Ilan Rubinfeld

Henry Ford Health System, 2799 West Grand Blvd., Detroit, MI 48202 USA

IRUBINF1@HFHS.ORG

Abstract

Most existing algorithms for clinical risk stratification rely on labeled training data. Collecting this data is challenging for clinical conditions where only a small percentage of patients experience adverse outcomes. We propose an unsupervised anomaly detection approach to risk stratify patients without the need of positively and negatively labeled training examples. High-risk patients are identified without any expert knowledge using a minimum enclosing ball to find cases that lie in sparse regions of the feature space. When evaluated on data from patients admitted with acute coronary syndrome and on patients undergoing inpatient surgical procedures, our approach successfully identified individuals at increased risk of adverse endpoints in both populations. In some cases, unsupervised anomaly detection outperformed other machine learning methods that used additional knowledge in the form of labeled examples.

1. Introduction

For many clinical conditions, patients experiencing adverse outcomes represent a small minority in the population. For example, the rate of cardiovascular mortality over a 90 day period following acute coronary syndrome (ACS) was found to be less than 2% in both the SYMPHONY and DISPERSE2 trials (Newby et al., 2003; Cannon et al., 2007). The corresponding rate of myocardial infarction in these trials was below 6%.

A similar case exists for patients undergoing surgical procedures. The rate of many important clinical complications, ranging from coma to bleeding, was well below 1% in the National Surgical Quality Improvements Program (NSQIP) data sampled at over 100 hospital sites (Khuri et al., 1998; Khuri, 2005). Less than 2% of the patients undergoing general surgery at these sites died in the 30 days following the procedure.

Identifying patients at risk of adverse outcomes in such a setting is challenging because most existing algorithms for clinical risk stratification rely on the availability of positively and negatively labeled training data. For outcomes that are rare, using these algorithms requires collecting data from a large number of patients to capture a sufficient number of positive examples for training. This process of collecting data from a large number of subjects is slow, expensive, and often burdens caregivers and patients. The costs and complexity of extensive, expertly-labeled data points have impeded the spread of even well validated and effective health care quality interventions (Schilling et al., 2008).

We address this issue by proposing an unsupervised anomaly detection-based approach to risk stratify patients without the need for labeled training examples. Our hypothesis is that patients at high risk of adverse outcomes can be detected as anomalies in a population. In the absence of any expert knowledge, we use a minimum enclosing ball (MEB) to find patients that lie in sparse regions of the feature space. We demonstrate the utility of this approach on data from over 4,000 patients admitted following ACS and from over 18,000 patients undergoing inpatient surgical procedures. In both cases, the MEB-based approach was able to successfully identify patients at increased risk of adverse events. In some cases, unsupervised anomaly detection even outperformed other machine learning methods using additional knowledge in the form of labeled

Appearing in *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel, 2010. Copyright 2010 by the author(s)/owner(s).

examples.

The main contributions of this paper are: (1) we present the hypothesis that patients at risk of many different future conditions can be discovered in a uniform manner as anomalies in a population, (2) we describe a risk stratification approach based on this hypothesis that is unsupervised and can be used to prognosticate patients for rare clinical outcomes, (3) we provide a detailed set of experimental results from two real-world applications and a variety of different adverse outcomes to rigorously evaluate the clinical utility of unsupervised anomaly detection, and (4) we contrast risk stratification for future adverse outcomes using anomaly detection with supervised machine learning methods applied to the same data.

2. Related Work

An extensive literature exists on the application of machine learning to prognosticate patients for adverse outcomes. Most of the existing work in this area relies on the availability of positively and negatively labeled examples for training. In contrast to these methods, which attempt to develop models for individual diseases, we focus on an unsupervised anomaly detection approach to identify high-risk patients. The hypothesis underlying our work is that patients who differ the most from other patients in a population are likely to be at an increased risk of adverse outcomes.

Anomaly detection has been studied in the broad context of medicine in earlier work. Hauskrecht et al. (2007) described a probabilistic anomaly detection method to detect unusual patient-management patterns and identify decisions that are highly unusual with respect to patients with the same or similar conditions. More closely related to our research are efforts to use anomaly detection to evaluate patient data. For example, Tarassenko et al. (1995) applied novelty detection to the detection of masses in mammograms, Campbell et al. (2001) to blood samples, Roberts et al. (1994) to electroencephalographic signals, and Laurikkala et al. (2000) to vestibular data.

We supplement these efforts by developing and evaluating the use of unsupervised anomaly detection more broadly for clinical risk stratification. In contrast to previous work focusing on detecting existing disease, our research uses anomaly detection to identify patients at increased risk of adverse future outcomes. In addition, while most earlier research focuses on a specific clinical condition, we explore the more general idea of identifying patients at increased risk of many different adverse outcomes using a uniform approach

where patients are compared with the rest of the population.

We develop and evaluate our MEB-based approach using larger datasets than any of the earlier studies we found in the literature (both in the number of patients and the different clinical conditions studied). As part of this evaluation, we also include a comparison of our work to other machine learning methods that use additional information in the form of historical patient labels. This comparison is intended to provide insight into the extent to which unsupervised anomaly detection can achieve the accuracy possible with supervised approaches that require large volumes of training data to be collected with examples of rare events.

3. Methods

3.1. Finding the Minimum Enclosing Ball

Given the normalized feature vectors \mathbf{x}_i for patients $i = 1, \dots, n$, we identify anomalies by first learning the minimum volume hypersphere that encloses the data for all patients, i.e., the MEB (Tax & Duin, 2004). This task can be formulated as minimizing the error function:

$$F(R, \mathbf{a}, \xi_i) = R^2 + C \sum_i \xi_i \quad (1)$$

over R, \mathbf{a}, ξ_i subject to the constraints:

$$\|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0, \quad \forall i \quad (2)$$

where \mathbf{a} is the center and R is the radius of the MEB. The slack variables $\xi_i \geq 0$ account for errors corresponding to outliers in the data that do not fit within the radius R of the MEB. The parameter C controls the trade-off between the volume of the MEB and the number of errors.

The dual of the MEB problem is given by:

$$\begin{aligned} L(R, \mathbf{a}, \alpha_i, \gamma_i, \xi_i) = & R^2 + C \sum_i \xi_i \\ & + \sum_i \alpha_i (\|\mathbf{x}_i\|^2 - 2\mathbf{a} \cdot \mathbf{x}_i + \|\mathbf{a}\|^2) \\ & - \sum_i \alpha_i (R^2 + \xi_i) - \sum_i \gamma_i \xi_i \end{aligned} \quad (3)$$

where $\alpha_i \geq 0$ and $\gamma_i \geq 0$ correspond to the Lagrange multipliers. L is minimized with respect to R, \mathbf{a}, ξ_i and maximized with respect to α_i and γ_i . This can be simplified to:

$$L = \sum_i \alpha_i (\mathbf{x}_i \cdot \mathbf{x}_i) - \sum_{i,j} \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (4)$$

subject to the constraints:

$$0 \leq \alpha_i \leq C \quad (5)$$

The inner products in Equation 4 can be replaced by a kernel function to obtain a more flexible data description than a rigid hypersphere (Vapnik, 1998). In our work, we use the Gaussian kernel function $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2/2s^2)$, which is independent of the position of the dataset with respect to the origin, and only depends on the distances between objects.

3.2. Identifying Anomalies in the Population

To test if patient k is an anomaly in the population, we compare the distance from the feature vector \mathbf{x}_k to the center of the hypersphere \mathbf{a} . The patient is declared to be an anomaly in the population if \mathbf{x}_k has a distance that is greater than the radius R of the MEB (Tax & Duin, 2004):

$$\begin{aligned} \|\mathbf{x}_k - \mathbf{a}\| &= (\mathbf{x}_k \cdot \mathbf{x}_k) - 2 \sum_i \alpha_i (\mathbf{x}_k \cdot \mathbf{x}_i) \\ &+ \sum_{i,j} \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j) > R^2 \end{aligned} \quad (6)$$

Consistent with the discussion in Section 3.1, we replace the inner product in Equation 6 with the Gaussian kernel function. This leads to the patient k being declared an anomaly if:

$$\sum_i \alpha_i \exp\left(\frac{-\|x_k - x_i\|^2}{2s^2}\right) \geq -R^2/2 + C \quad (7)$$

While Equation 7 can be used to assign binary labels to patients by categorizing them as anomalies or non-anomalies, the distance of each patient’s feature vector from the center of the MEB can also be used as a continuous anomaly score. In our work, we favor this approach, since it allows for a more fine-grained assessment of patient risk.

4. Evaluation

We evaluated our MEB-based unsupervised anomaly detection scheme on two separate datasets.

We used data from 4,189 patients in the MERLIN trial (Scirica et al., 2007), who were admitted to a hospital with non-ST-elevation ACS. The endpoint of sudden cardiac death was adjudicated by a blinded Clinical Events Committee for a median follow-up of 348 days following the index event. During the follow-up period 77 sudden cardiac deaths occurred.

We also used data from the American College of Surgery NSQIP (Khuri et al., 1998; Khuri, 2005) for 18,248 patients undergoing inpatient surgical procedures in 2005. This data was sampled at over 100 different hospital sites and consisted of patients undergoing both general and vascular surgery. Patients were followed for 30 days post-surgery for mortality and various morbidities. We studied the ten most rare morbidity outcomes: coma > 24 hours (21 events); peripheral nerve injury (25); myocardial infarction (41); stroke or cerebrovascular accident (56); pulmonary embolism (76); failure of extracardiac graft or prosthesis (80); renal insufficiency (83); cardiac arrest (94); renal failure (127); and bleeding requiring transfusion (150). We also studied the endpoint of mortality (369).

In both these datasets, we used baseline clinical characteristics as features. This corresponded to 20 variables collected at randomization for the MERLIN trial, and 45 variables collected preoperatively for the NSQIP data. The MEB was learned using the Statistical Pattern Recognition Toolbox (Czech Technical University). We experimented with different values of the Gaussian kernel parameter ($\log_2 s = 0, \dots, 10$) while fixing $C=1$. Each patient’s anomaly score was defined to be the distance of the patient’s feature vector from the center of the MEB. We assessed the predictive ability of unsupervised anomaly detection by calculating the area under the receiver operating characteristic curve (AUROC) for these scores relative to the different endpoints. The AUROC is widely used in medicine, and is generally considered the standard for evaluating risk stratification methods (Altman, 1991). The results of this process were interpreted as follows: an AUROC less than 0.6 has no clinical utility, 0.6 to 0.7 limited clinical utility, 0.7 to 0.8 modest clinical utility, and 0.8 or higher genuine clinical utility (Ohman et al., 2000).

We compared unsupervised anomaly detection with different supervised learning methods in an attempt to study the extent to which unsupervised anomaly detection can achieve results similar to other methods requiring the collection of large volumes of training data with positive and negative examples. For each dataset and endpoint, we trained and evaluated the accuracy of a logistic regression (LR) model using five-

fold cross-validation (Hosmer & Lemeshow, 2000). We also studied the accuracy of support vector classification (SVC) and support vector regression (SVR) using five-fold cross-validation (Gunn, 1998). The results of unsupervised anomaly detection were compared with the best results obtained for LRV, SVC and SVR. For both SVC and SVR, we experimented with different choices of the Gaussian kernel parameter as described earlier. For SVR, we additionally varied ϵ between 0.1 and 0.5. LR training was carried out using Matlab (Mathworks Inc.) while both SVC and SVR were trained using LIBSVM (National Taiwan University).

We conducted two additional analyses. First, we measured the AUROC when the MEB was trained for each endpoint using data from patients known to be event free. This experiment attempted to study the extent to which anomaly detection could identify patients at risk of future rare outcomes using data only from patients known to be free of these outcomes. Second, we also studied the binary categorization of patients by the MEB into anomaly and non-anomaly groups. This was done by means of Kaplan-Meier survival analysis (Efron, 1988) to compare event rates for patients declared to be anomalies and non-anomalies. Hazard ratios (HR) and 95% confidence intervals (CI) were estimated using a Cox proportional hazards regression model (Cox, 1972). Consistent with the medical literature, findings were considered to be significant for a p-value less than 0.05. In contrast to the AUROC, Kaplan-Meier survival analysis does not provide an overall measure of the discriminative ability of a risk variable. Instead, it typically measures relative differences in the rate of events across different patient groups once the variable has been dichotomized. It accounts for both the timing of events (i.e., whether events occur at the start of the study or late into the study) as well as censoring (i.e., patients dropping out before the study is complete). This experiment therefore focused on supplementing the AUROC results by studying whether patients categorized as high or low risk had different rates of adverse outcomes.

5. Results

5.1. ACS and NSQIP Data

Figure 1 presents the AUROC for unsupervised anomaly detection as the kernel parameter s was varied. MEB-based discrimination of patients at risk of sudden cardiac death was maximized for $\log_2 s = 1$. This corresponded to an AUROC of 0.67.

Figure 2 presents the AUROC for the different morbidity endpoints and mortality in the NSQIP data. Sim-

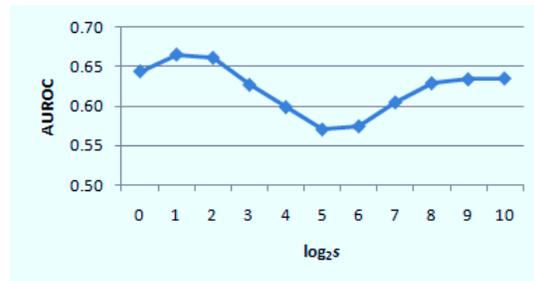


Figure 1. MEB AUROC for sudden cardiac death following ACS as the Gaussian kernel parameter s is varied.

ilar to the ACS case, performance was generally maximized at $\log_2 s = 1$. For eight of the ten morbidity endpoints studied, this choice of the kernel parameter resulted in an AUROC higher than 0.7 (in two of these cases the AUROC was greater than 0.8). We obtained analogous results for the endpoint of mortality. The AUROC was maximized for $\log_2 s = 1$, corresponding to a value of 0.86.

5.2. Comparison with LR, SVC and SVR

Figure 3 presents a comparison of the AUROC for the MEB approach (with $\log_2 s = 1$) to the best results obtained for LR, SVC and SVR.

For the ACS dataset, all three supervised methods had a higher AUROC than unsupervised anomaly detection. This difference was significant for LR and SVC (both 0.72), but only marginal for SVR (0.68 vs. 0.67).

For the NSQIP dataset, the predictive ability of unsupervised anomaly detection was comparable to the results obtained with LR, SVC and SVR. The MEB AUROC was greater than or equal to 0.7 in every case where one or more of the supervised approaches showed moderate or higher clinical utility. In some cases (e.g., coma > 24 hours, myocardial infarction, stroke or cerebrovascular accident, graft failure, cardiac arrest and bleeding), the MEB AUROC was higher than any of the other methods. Overall, the average MEB AUROC across all morbidities (0.74) was higher than the corresponding average AUROC for LR (0.72) and SVC (0.70), and marginally less than the average AUROC for SVR (0.75).

5.3. Training on Negative Examples

Training the MEB only on patients known to be event free improved performance for all endpoints in both datasets ($\log_2 s = 1$). The AUROC for the sudden cardiac death population increased from 0.67 to 0.69 when the MEB was trained only on negative examples. Table 1 presents the corresponding changes in AUROC for the NSQIP population. For six of the NSQIP endpoints, the resulting AUROC exceeded 0.8.

Unsupervised Risk Stratification in Clinical Datasets

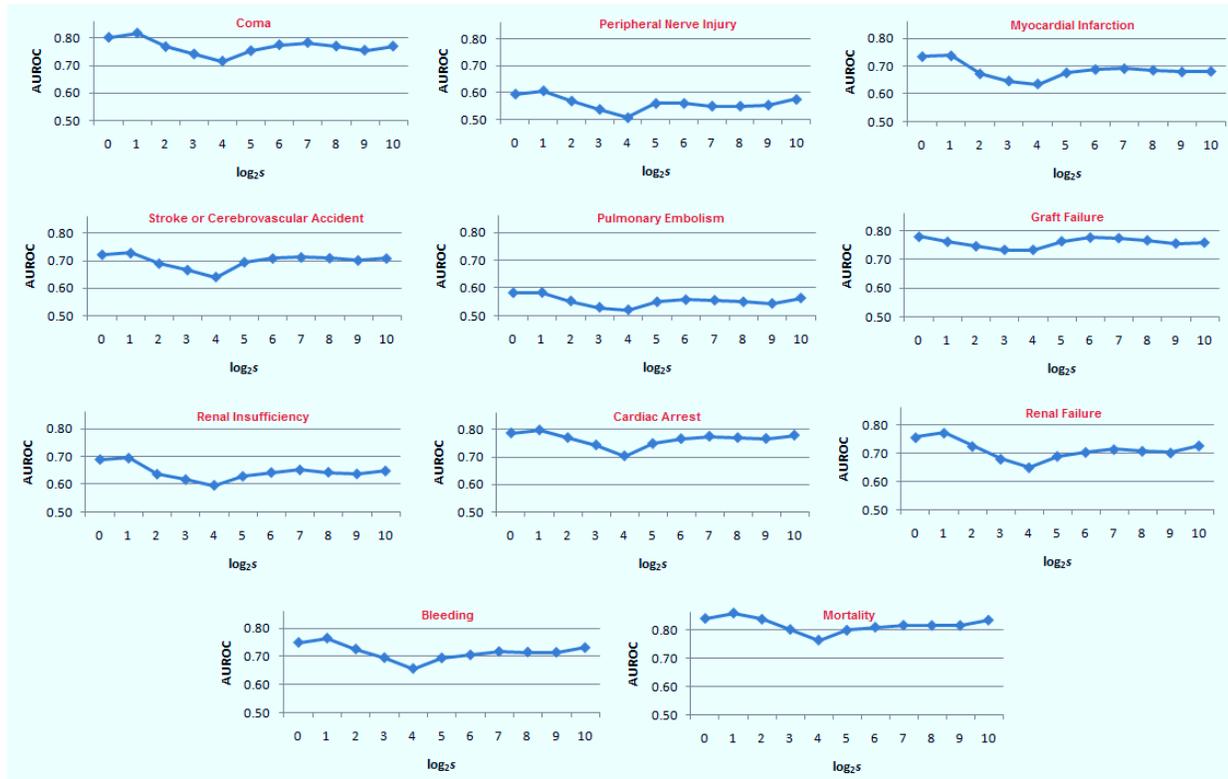


Figure 2. MEB AUROC for inpatient surgical morbidities and mortality as the Gaussian kernel parameter s is varied.

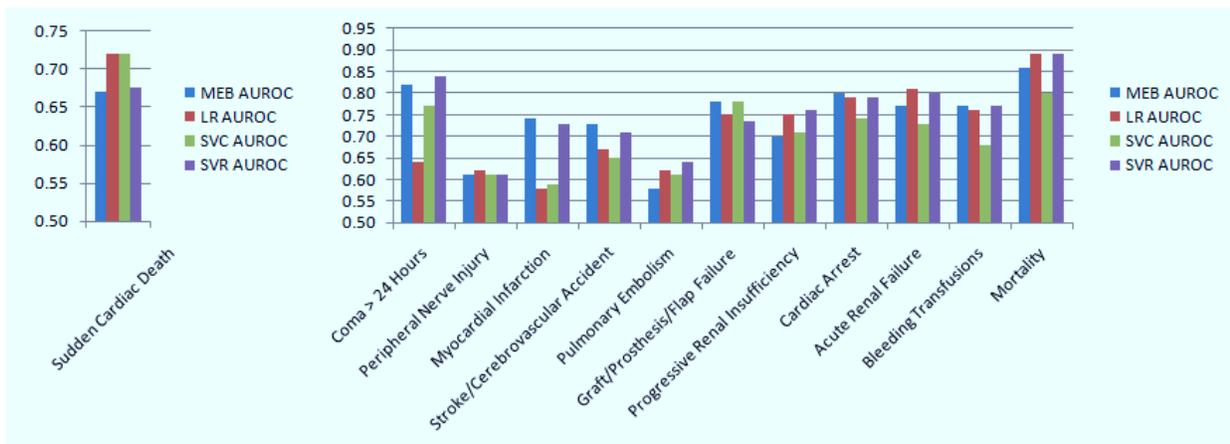


Figure 3. Comparison of AUROC for MEB, LR, SVC and SVR.

Table 1. AUROC for MEB trained on all patients (ALL) and MEB trained on negative examples (NEG).

EVENT	AUROC MEB (ALL)	AUROC MEB (NEG)
COMA	0.82	0.84
PERIPHERAL NERVE INJURY	0.61	0.66
MYOCARDIAL INFARCTION	0.74	0.79
STROKE OR CEREBROVASCULAR ACCIDENT	0.73	0.78
PULMONARY EMBOLISM	0.58	0.64
GRAFT FAILURE	0.78	0.85
RENAL INSUFFICIENCY	0.70	0.75
CARDIAC ARREST	0.80	0.85
RENAL FAILURE	0.77	0.83
BLEEDING	0.77	0.82
MORTALITY	0.86	0.91

5.4. Survival Analysis

Tables 2 and 3 present the hazard ratios of patients categorized as being anomalies by the MEB approach relative to patients categorized as being non-anomalies ($\log_2 s = 1$).

For the ACS population, patients who were classified as anomalies showed a statistically significant ($p < 0.05$) increase in their risk of sudden cardiac death. These patients experienced more than a 50% increase in their risk of adverse outcomes.

Our results on the NSQIP data paralleled these findings. Patients identified as anomalies by the MEB showed an elevated risk for all but two of the morbidity endpoints, and for the outcome of death within 30 days postoperatively. For most morbidities, there was a two- to three-fold increased risk of adverse outcomes. This increase in risk was even higher in patients outside the MEB for the endpoint of death.

6. Summary and Discussion

In this paper, we developed and evaluated the hypothesis that patients who are anomalies in a population are at an increased risk of adverse future outcomes. We present this as a way to risk stratify patients without using extensive *a priori* information or requiring large volumes of training data with both positive and negative examples. Collecting such data is often slow, expensive, and difficult when only a minority of patients in a population experience events.

We proposed an MEB-based unsupervised anomaly detection approach to identify potentially high-risk patients in a population. Our results on over 22,000 patients showed that patients who lie in sparse regions of the feature space are at an increased risk of sudden cardiac death following ACS and for both morbidity

and mortality following inpatient surgical procedures. In some cases, this approach of risk stratifying patients based on their anomaly score outperformed supervised machine learning methods that used additional knowledge in the form of both positively and negatively labeled examples. These cases can be attributed to supervised methods being unable to generalize for complex, multi-factorial clinical events when only a small number of patients in a large training population experience events.

For the NSQIP data, unsupervised anomaly detection achieved an AUROC of 0.8 or higher for three outcomes in NSQIP (mortality, coma, and cardiac arrest), and between 0.7 and 0.8 for six other outcomes (myocardial infarction, stroke or cerebrovascular accident, graft failure, renal insufficiency, renal failure, and bleeding). We believe that these results (moderate or genuine clinical utility) are quite encouraging as post-operative morbidity outcomes are generally difficult to predict. These results were reinforced by Kaplan-Meier survival analysis. For mortality and most morbidity outcomes, patients outside the MEB exhibited a two- to three-fold increased risk of adverse endpoints.

For the ACS data, the results of AUROC and Kaplan-Meier survival analysis were mixed. While Kaplan-Meier survival analysis found that patients outside the MEB had more than a 50% increase in their risk of sudden cardiac death over time, the AUROC suggested only limited clinical utility. This difference may be attributed to the AUROC being affected by censoring. In NSQIP, this was not an issue since patients were followed for all 30 days unless they died. In MERLIN, however, patients had a maximum follow-up of two years but a median follow-up just under one year due to censoring. The results of Kaplan-Meier survival analysis, which accounts for the timing of events censoring, may therefore be more meaningful for this

Table 2. Kaplan-Meier survival analysis of unsupervised anomaly detection for sudden cardiac death following acute coronary syndrome.

EVENT	# OUTSIDE MEB	# INSIDE MEB	HR	CI	P
SUDDEN CARDIAC DEATH	34/922	43/3267	1.66	1.32-2.07	<0.001

Table 3. Kaplan-Meier survival analysis of unsupervised anomaly detection for morbidity and mortality following inpatient surgical procedures.

EVENT	# OUTSIDE MEB	# INSIDE MEB	HR	CI	P
COMA	15/5584	6/12664	2.38	1.48-3.83	<0.001
PERIPHERAL NERVE INJURY	11/5584	14/12664	1.34	0.90-1.98	0.152
MYOCARDIAL INFARCTION	28/5584	13/12664	2.12	1.59-3.07	<0.001
STROKE OR CEREBROVASCULAR ACCIDENT	32/5584	34/12664	1.74	1.34-2.27	<0.001
PULMONARY EMBOLISM	31/5584	45/12664	1.25	0.99-1.57	0.056
GRAFT FAILURE	63/5584	17/12664	2.91	2.22-3.80	<0.001
RENAL INSUFFICIENCY	53/5584	30/12664	2.01	1.60-2.51	<0.001
CARDIAC ARREST	71/5584	23/12664	2.65	2.10-3.36	<0.001
RENAL FAILURE	93/5584	34/12664	2.50	2.05-3.04	<0.001
BLEEDING	103/5584	47/12664	2.24	1.88-2.66	<0.001
MORTALITY	311/5584	58/12664	3.53	3.07-4.07	<0.001

specific population. We present the AUROC largely to maintain consistency between our evaluation of the NSQIP and MERLIN datasets, and also to avoid concealing the low AUROC in MERLIN.

In both of the patient populations studied, we found that the performance of our unsupervised anomaly detection approach could be improved by training the MEB only on negative examples. This is likely due to the ability of the MEB to better characterize normal patients when positive examples are excluded from training. We believe that in applications where known negative examples are available, this additional information can allow for a better anomaly detection-based method for risk stratification.

Our study does have limitations. For both the MERLIN and NSQIP data, most of the features available to us corresponded to binary variables (e.g., male or female, hypertensive or non-hypertensive). For some of these variables, the dichotomization process removed information that may have been useful to identify anomalies (e.g., systolic blood pressure). We also limited our choice of features to baseline clinical characteristics, and did not use data such as imaging results and biomarker laboratory values for patients. Prior work has shown these data to have significant prognostic value. However, these features were missing in a majority of the patients in both study populations. Finally, we note that while we evaluated the use of unsupervised anomaly detection in two datasets, ad-

ditional studies are needed on a wider set of different clinical conditions to fully explore and validate the relationship of our risk stratification approach with adverse outcomes.

Acknowledgment

We would like to thank the reviewers for their comments and suggestions on improving the presentation of this work. We also thank Sarah Whitehouse for editorial assistance preparing this manuscript.

References

- Altman, D.G. *Practical Statistics for Medical Research*. Chapman & Hall, 1991.
- Campbell, C. and Bennett, K.P. A linear programming approach to novelty detection. *Advances in Neural Information Processing Systems*, pp. 395–401, 2001.
- Cannon, C.P., Husted, S., Harrington, R.A., Scirica, B.M., Emanuelsson, H., Peters, G., and Storey, R.F. Safety, Tolerability, and Initial Efficacy of AZD6140, the First Reversible Oral Adenosine Diphosphate Receptor Antagonist, Compared With Clopidogrel, in Patients With Non-ST-Segment Elevation Acute Coronary Syndrome Primary Results of the DISPERSE-2 Trial. *Journal of the American College of Cardiology*, 50(19):1844–1851, 2007.

- Cox, DR. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 187–220, 1972.
- Efron, B. Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of the American Statistical Association*, 83(402):414–425, 1988.
- Gunn, S.R. Support vector machines for classification and regression. *ISIS Technical Report*, 14, 1998.
- Hauskrecht, M., Valko, M., Kveton, B., Visweswaran, S., and Cooper, G.F. Evidence-based Anomaly Detection in Clinical Domains. In *AMIA Annual Symposium Proceedings*, volume 2007, pp. 319–323. American Medical Informatics Association, 2007.
- Hosmer, D.W. and Lemeshow, S. *Applied Logistic Regression*. Wiley-Interscience, 2000.
- Khuri, S.F. The NSQIP: a new frontier in surgery. *Surgery*, 138(5):837–843, 2005.
- Khuri, SF, Daley, J., Henderson, W., Hur, K., Demakis, J., Aust, JB, Chong, V., Fabri, PJ, Gibbs, JO, Grover, F., et al. 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. *Annals of Surgery*, 228(4):491–507, 1998.
- Laurikkala, J., Juhola, M., and Kentala, E. Informal identification of outliers in medical data. In *The Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology*. Citeseer, 2000.
- Newby, LK, Bhapkar, MV, White, HD, Topol, EJ, Dougherty, FC, Harrington, RA, Smith, MC, Asarch, LF, Califf, RM, et al. Predictors of 90-day outcome in patients stabilized after acute coronary syndromes. *European Heart Journal*, 24(2):172–182, 2003.
- Ohman, E.M., Granger, C.B., Harrington, R.A., and Lee, K.L. Risk stratification and therapeutic decision making in acute coronary syndromes. *Jama*, 284(7):876–878, 2000.
- Roberts, S. and Tarassenko, L. A probabilistic resource allocating network for novelty detection. *Neural Computation*, 6(2):270–284, 1994.
- Schilling, P.L., Dimick, J.B., and Birkmeyer, J.D. Prioritizing quality improvement in general surgery. *Journal of the American College of Surgeons*, 207(5):698–704, 2008.
- Scirica, B.M., Morrow, D.A., Hod, H., Murphy, S.A., Belardinelli, L., Hedgepeth, C.M., Molhoek, P., Verheugt, F.W.A., Gersh, B.J., McCabe, C.H., et al. Effect of ranolazine, an antianginal agent with novel electrophysiological properties, on the incidence of arrhythmias in patients with non ST-segment elevation acute coronary syndrome: results from the Metabolic Efficiency With Ranolazine for Less Ischemia in Non ST-Elevation Acute Coronary Syndrome Thrombolysis in Myocardial Infarction 36 (MERLIN-TIMI 36) randomized controlled trial. *Circulation*, 116(15):1647–1652, 2007.
- Tarassenko, L., Hayton, P., Cerneaz, N., and Brady, M. Novelty detection for the identification of masses in mammograms. In *Artificial Neural Networks, 1995., Fourth International Conference on*, pp. 442–447, 1995.
- Tax, D.M.J. and Duin, R.P.W. Support vector data description. *Machine Learning*, 54(1):45–66, 2004.
- Vapnik, V. *Statistical Learning Theory*. Wiley, New York, 1998.

NSQIP Disclaimer

The American College of Surgeons National Surgical Quality Improvement Program and the hospitals participating in it represent the source of the data used herein; they have not verified and are not responsible for the statistical validity of the data analysis or for the conclusions derived by the authors.