Prediction of Invasive ICU Procedures from Electronic Health Record Data

Alex Kern (afkern), Jiwoo Lee (jiwoolee), Keaton Ollech (kollech)

Introduction

Intensive care unit (ICU) admissions remain the most costly and deadly type of hospital visit. In particular, ICU admissions often involve invasive procedures, such as catheterization, intubation, and mechanical ventilation that can significantly reduce quality-of-life and healthy-life-expectancy [1]. Thus, prediction of invasive ICU procedures and identification of features that are highly predictive of invasive ICU procedures may identify high-risk patients that can be closely monitored to prevent such procedures. We used the MIMIC-III database, which includes health-related data from ICU admissions for more than 40,000 patients, such as diagnoses, procedures, laboratory tests, and prescriptions [2]. We used machine learning approaches, including lasso regression, gradient boosting, and support vector machines. The input to our algorithm was electronic health record information, including diagnoses, procedures, laboratory tests, and prescriptions. We used lasso regression, gradient boosting, and support vector machines to output a prediction of invasive ICU procedures. Taken together, we used machine learning approaches on the MIMIC-III database to predict the invasive ICU procedures in the hopes that earlier identification of high-risk patients may prompt preventative interventions.

Related Work

Several papers have examined prediction of ICU mortality [3-5]. Zhao et al. used demographics, comorbidities, and laboratory tests at the initial presentation in a logistic regression model and obtained an AUROC of 0.74 and 0.83 for ICU admission and ICU mortality, respectively [3]. Given that patient conditions rapidly evolve during ICU stays, features obtained at initial presentation are limited. Johnson et al. used physiologic and laboratory measurements sampled at random timepoints in a gradient boosting model and obtained an AUROC of 0.92 [4]. While this study acquired high accuracy, using measurements sampled at random timepoints does not accurately represent ICU stays. Sinuff et al. used data from 12 different observational studies in a rule-based model and obtained an AUROC of 0.85 [5]. Even though this study used manual validation against physician scores, the use of 12 different observational studies may have introduced additional bias.

Furthermore, several papers have examined prediction of ICU readmission [6-8]. Notably, such papers have used machine learning approaches on the MIMIC-III database as well. For example, Rojas et al. used patient characteristics, nursing assessments, prior admissions, medications, ICU interventions, diagnostic tests, vital signs, and laboratory results in a gradient-boosted machine model and obtained an AUROC of 0.76 [6]. Hammer et al. used demographical factors, surgical factors, physiological parameters, ICU treatment, and the acuity of illness in a logistic regression model and obtained an AUROC of 0.78 [7]. Xue et al. used physiological measurements and medications in a graph-based model and obtained an AUROC of 0.636 and 0.637 in a baseline and grouping model, respectively [8]. These papers employ a comprehensive, diverse set of features, yet only focus on single machine learning approaches. To our knowledge, there is no study to date that examines prediction of invasive ICU procedures. Given that invasive ICU procedures are a critical factor in ICU readmission, prediction of invasive ICU procedures may play a critical role in prediction of ICU readmission. In other words, our work with prediction of invasive ICU procedures may identify high-risk patients that can be closely monitored to prevent such procedures.

Dataset and Features

First, we downloaded the following tables from the MIMIC-III database [2]: ICUSTAYS, DIAGNOSES_ICD, PROCEDURES_ICD, MICROBIOLOGYEVENTS, PRESCRIPTIONS, D_ICD_DIAGNOSES, and D_ICD_PROCEDURES. ICUSTAYS contains general information about a patient's ICU stay, including the check-in and check-out time. DIAGNOSES_ICD and PROCEDURES_ICD contains information about a patient's diagnoses and procedures, and D_ICD_DIAGNOSES and D_ICD_PROCEDURES were used to map ICD-9 codes to human-readable features. MICROBIOLOGYEVENTS and PRESCRIPTIONS contain the results of laboratory tests and prescriptions, respectively. Every feature (e.g., diagnoses, procedures, laboratory tests, and prescriptions) has an associated chart-time and is coded by an indicator variable of whether a patient has a particular feature.

Second, we engineered our outcome, which represents undergoing invasive ICU procedures, using a string-matching technique and a list of common invasive ICU procedures (e.g., catheterization, intubation, and mechanical ventilation). The outcome captured 33 features related to invasive ICU procedures (e.g., venous cath NEC, non-invasive mech vent, resp tract intubat NEC, etc.). The 33 features used in the outcome were removed from features. Our final dataset included 6,557,399 unique patient-feature items capturing 46,476 unique patients, 57,784 unique ICU stays, and 13,578 features. There were a total of 21,283 (45.79%) unique patients and 25,099 (43.44%) unique ICU stays associated with invasive ICU procedures. Some of the most common features included potassium chloride (3.46%), insulin (2.59%), and D5W (2.57%).

Third, we excluded features based on the chart-times of features to evaluate the earliest time that risk of invasive ICU procedures can be predicted with high predictive performance. The consideration of different cut-off times is informative, earliest identification of risk may prompt preventative interventions. The average length of an ICU stay can range between 2 to 10 days [9]. The mean and median ICU stay in our data was 2.72 and 6.77 days, respectively. Thus, we used three cut-off times of one hour, 12 hours, and 24 hours. In other words, at each cut-off time, we only considered features documented at one hour, 12 hours, and 24 hours into an ICU stay.

Methods

Before beginning our analysis, we wanted to examine if there were any uninformative features. We used the function nearZeroVar() on our data, because a feature with near zero variance would likely lack prediction power and be uninformative. This method proved useful in reducing our feature space since our dataset contained many near zero variance features. We then chose to implement three different methods for predicting invasive ICU procedures: cross-validated lasso regression, cross-validated gradient boosting, and a support vector machine (SVM).

Despite excluding features with near zero variance, we anticipated that not all of the remaining features would be informative in prediction of invasive ICU procedures. Lasso regression is a shrinkage method, so it sets some of the coefficients in the model to zero in order to decrease the variance and reduce overfitting. While the decrease in variance comes at the cost of an increase in bias, we primarily wanted to avoid overfitting, and thus using a shrinkage method seemed like a good choice. Lasso regression relies on a linear model, but uses an alternate fitting procedure to linear regression that is more restrictive. We also used cross-validation, which resamples the training data many times in order to reveal flaws in the model, and tunes the model's hyperparameters in an attempt to minimize the test error.

We chose gradient boosting because we wanted to try implementing a non-linear method. In general, non-linear methods tend to be more complex and less interpretable. We also used cross-validation for this method to tune hyperparameters and minimize test error. As with lasso regression, gradient boosting makes use of shrinkage, which helps the model remain resistant to overfitting. Like other supervised methods, gradient boosting still defines an objective function and optimizes it. Gradient boosting learns slowly by first using samples that are easy to predict, then sequentially builds more trees using the previously grown trees. Its resistance to overfitting also derives from the fact that it proceeds slowly and gradually improves its predictions in places where it does not perform well.

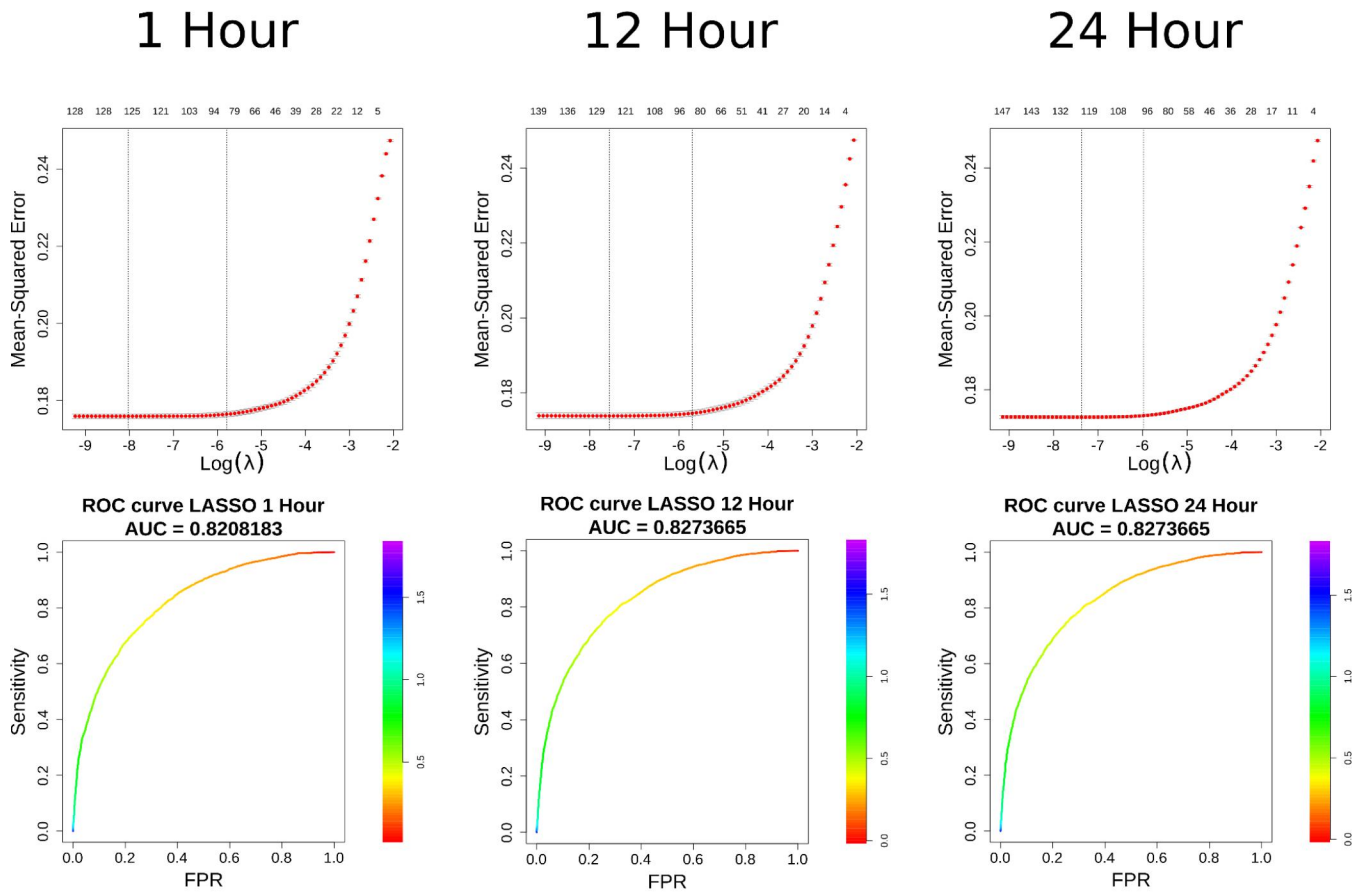
Finally, we used an SVM. We wanted to test how this "off-the-shelf" method could compare to our carefully cross-validated previous methods. SVMs create a separating hyperplane to label the classes of the test data. In order to choose the best hyperplane, this method maximizes the distance, or margin, between the points closest to the hyperplane, called support vectors. We used a radial kernel since our data is non-linear, meaning we tried to solve the equation $K(x, y) = \exp(-\gamma\Sigma_{j=1}^{p}(x_{ij} - y_{ij})^2)$, where γis a hyperparameter for the smoothness of the decision boundary.

Experiments/Results/Discussion

In order to develop accurate, computationally efficient, and relatively easily interpretable models to predict invasive ICU procedures from the relatively large number of features, we developed models using only high-variance variables in the three different time point datasets. 130 variables were identified as being high variance in the 1 hour dataset, 144 in the 12 hour dataset, and 150 in the 24 hour dataset. Notably, these features overlap between the datasets, such that the 12 hour dataset is a superset of the 1 hour, and the 24 hour is a superset of the 12 hour. This high variance filtering thus removes a majority of the predictors, likely for events or diagnoses which occurred in very few ICU stays. This filtering allows our model to avoid overfitting to rare events which might be 100% predictive of invasive ICU procedures (or lack thereof) in our dataset but not necessarily in the broader world.

Using these reduced sets of variables, we then used LASSO regression, gradient boosted tree (GBM), and support vector machine (SVM) models to predict the probability of an invasive ICU procedure. We used cross-validation to pick model hyperparameters for these models, with the exception of SVM, where the extensive tuning of parameters was computationally infeasible.

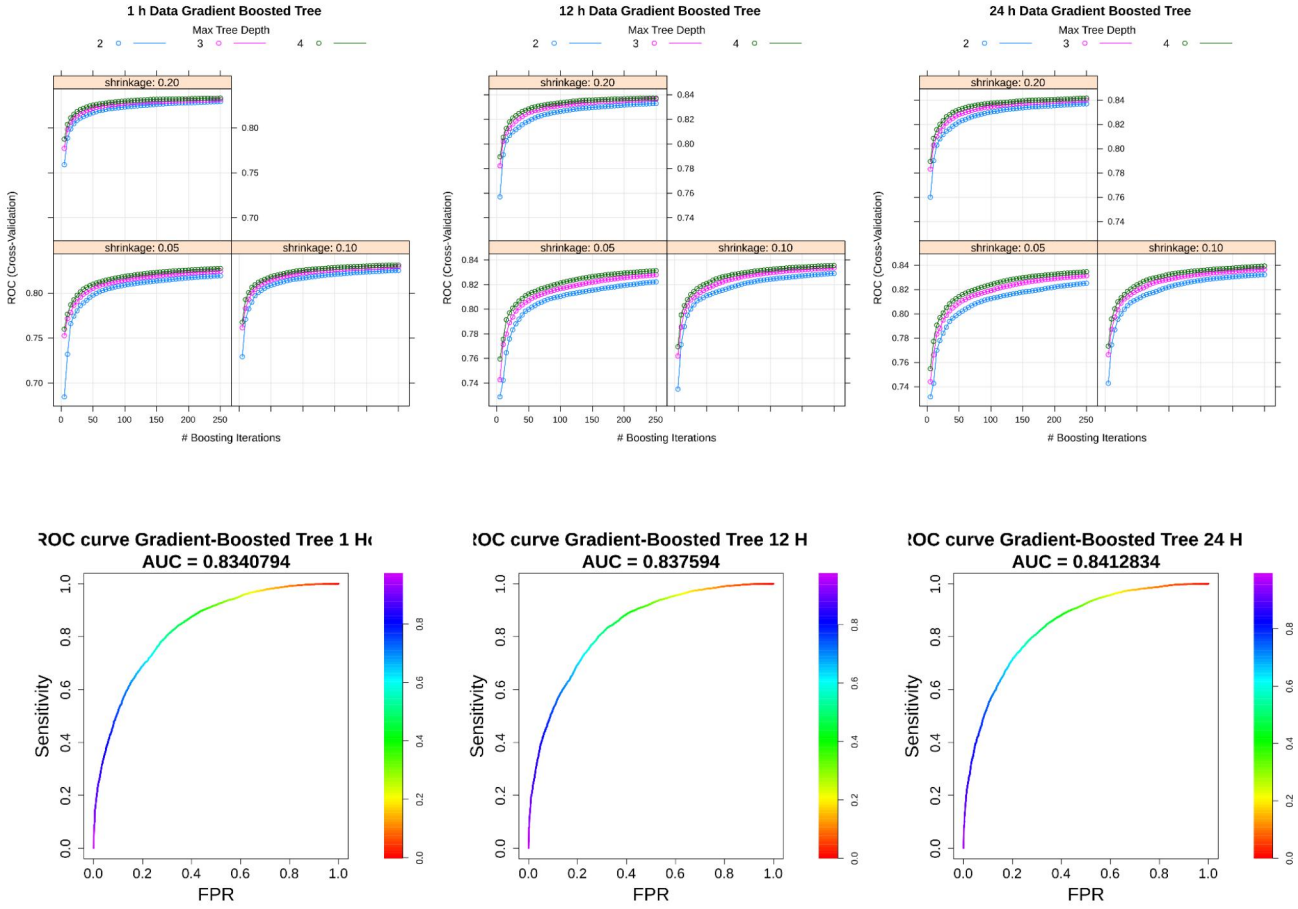
Figure 1 LASSO Model



For LASSO regression, we used five-fold cross-validation to identify the value for λ one standard error above λ_{min} to balance performance and model interpretability. This led to values of 0.003046284, 0.003343293, and 0.002529079 for λ for the 1 hour, 12 hour, and 24 hour models (Figure 1 Top) . These values shrunk relatively few variables down to zero, leading to 86, 89, and 101 nonzero predictors for these models. This is perhaps not surprising due to the relatively low number of high variance variables used in these models. All three models (1 hour, 12 hour, and 24 hour) performed relatively well (Figure 1 Bottom) with slight improvements in performance as more data was included, with the 24 hour model having an area under the receiver-operator curve (AUROC) of 0.827 and the 1 hour model having 0.821.

Figure 2 Gradient Boosted Tree Model

1 Hour 12 Hour 24 Hour



We next used GBM with parameters 250 boosting iterations, 3 minimum observations per node, a maximum interaction depth of 4, and shrinkage of 0.2, identified via five-fold cross-validation (Figure 2 Top). The final cross-validation AUROCs were similar to the performance on the test set, and the performance relative to the number of boosting iterations showed a similar trend for the three time point datasets, with the 1 hour and 12 hour models flattening out to lower overall performance than the 24 hour model, and all models performing better with more boosting iterations on the cross-validation ROC.

These models performed better at all time points than the LASSO models, with AUROC = 0.834, 0.838, 0.841 for 1 hour, 12 hour, and 24 hour models, respectively (Figure 2 Bottom). Once again, the models with more predictors based on a longer ICU stay performed slightly better.

Finally, we utilized SVM models with a radial basis function kernel, with gamma = 1/(number of features), cost = 1, and epsilon = 0.1. These models outperformed the gradient boosted tree models at all time points, with the 1 hour model having sensitivity = 0.8386 and specificity = 0.6732 while the 24 hour gradient-boosted tree model had sensitivity = 0.8364 and specificity = 0.6658 on the test set. The 12 hour SVM had even better performance at sensitivity = 0.8409 and specificity = 0.6804, with the 24 hour SVM model performing the best of all at sensitivity = 0.8386 and specificity = 0.6927, albeit with slightly lower sensitivity than the 12 hour model.

Table 1

Model Type (Timing)	LASSO (1 hour)	LASSO (12 hour)	LASSO (24 hour)	GBM (1 hour)	GBM (12 hour)	GBM (24 hour)	SVM (1 hour)	SVM (12 hour)	SVM (24 hour)
Sensitivity	0.8616	0.8616	0.8609	0.8325	0.8322	0.8364	0.8386	0.8409	0.8386
Specificity	0.5873	0.5939	0.6015	0.6530	0.6645	0.6658	0.6732	0.6804	0.6927
Accuracy	0.7381	0.7411	0.7441	0.7517	0.7567	0.7595	0.7641	0.7686	0.7729

Overall, all three classes of models predicted invasive ICU procedures relatively well, with minimum AUROCs of 0.821 and 0.834 for LASSO and gradient-boosted trees with data available at 1 hour, and sensitivity = 0.8386 and specificity = 0.6732 for the SVM. Accuracy, sensitivity, and specificity metrics for all models are in Table 1. The SVM models performed the best, followed by the gradient-boosted trees and then the LASSO model. However, our SVM may be prone to overfitting, but we would need to use cross-validation or additional data sources to verify this theory. Adding additional data up to 24 hours into the ICU stay improved performance mildly for both classes of models. The performance increases for the models from the additional 23 hours worth of data were relatively modest. This perhaps reflects that much of the data collected later in the ICU stay is more specific to the conditions of individual patients and would not make it into the limited set of high variance features we used. This can also be seen from the relatively low numbers of high variance features added to the 12 and 24 hour models. However, the high performance of these models with only 1 hour of data is encouraging for quick prediction of the need for invasive ICU procedures as early as possible.

Conclusion/Future Work

We used the MIMIC-III database and machine learning approaches, including lasso regression, gradient boosting, and support vector machines, to predict invasive ICU procedures. We considered different timepoints in the ICU stay, including 1 hour, 12 hours, and 24 hours, and found modest improvements in accuracy at later time points. Our SVM model (accuracy = 0.7729) had better prediction performance than our lasso regression model (0.7441) and our gradient boosting model (accuracy = 0.7595). We hypothesize that our result may be a consequence of the proclivity of SVMs to overfit; we would need to use additional data sources and cross-validation to confirm this result. Going forward, we would recommend using an SVM as the default choice due to its quick implementation and ease of interpretation. The main disadvantage of using SVM is the massive amount of computational power required, especially in order to complete cross-validation. A future extension of this project would be using additional computational resources to determine how cross-validation could further improve our results. Taken together, prediction of invasive ICU procedures may identify high-risk patients and prompt preventative interventions to reduce the massive morbidity and mortality associated with ICU admissions.

Contributions

Jiwoo, Keaton, and Alex contributed equally to the execution and synthesis of this analysis. Jiwoo was responsible for processing the raw data, engineering the outcome, and generating the final data. Keaton was responsible for implementing initial models on the reduced dataset. Alex was responsible for implementing such models to the full dataset, adding additional optimizations, and further extensions.

References/Bibliography

[1] Rauché, A. (n.d.). Invasive procedures in intensive care unit. Invasive procedures in intensive care unit – Rauché. http://www.rauche.net/izdanja/broj-5/invasive-procedures-in-intensive-care-unit/.

[2] Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L.-wei H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016, May 24). MIMIC-III, a freely accessible critical care database. Nature News. https://www.nature.com/articles/sdata201635.

[3] Zhao, Z., Chen, A., Hou, W., Graham, J. M., Li, H., Richman, P. S., Thode, H. C., Singer, A. J., & Duong, T. Q. (n.d.). Prediction model and risk scores of ICU admission and mortality in COVID-19. PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0236618.

[4] Johnson, A. E. W., & Mark, R. G. (2018, April 16). Real-time mortality prediction in the Intensive Care Unit. AMIA ... Annual Symposium proceedings. AMIA Symposium. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977709/.

[5] Sinuff T;Adhikari NK;Cook DJ;Schünemann HJ;Griffith LE;Rocker G;Walter SD; (n.d.). Mortality predictions in the intensive care unit: comparing physicians with scoring systems. Critical care medicine. https://pubmed.ncbi.nlm.nih.gov/16505667/.

[6] Rojas, J. C., Carey, K. A., Edelson, D. P., Venable, L. R., Howell, M. D., & Churpek, M. M. (2018, July). *Predicting Intensive Care Unit Readmission with Machine Learning Using Electronic Health Record Data*. Annals of the American Thoracic Society. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6207111/.

[7] Maximilian Hammer, S. D. G. (n.d.). *A Tool to Predict Readmission to the Intensive Care Unit in Surgical Critical Care Patients-The RISC Score - Maximilian Hammer, Stephanie D. Grabitz, Bijan Teja, Karuna Wongtangman, Marjorie Serrano, Sara Neves, Shahla Siddiqui, Xinling Xu, Matthias Eikermann, 2020*. SAGE Journals. https://journals.sagepub.com/doi/abs/10.1177/0885066620949164.

[8] Y;, X. Y. K. D. L. (n.d.). Predicting ICU readmission using grouped physiological and medication trends. Artificial intelligence in medicine. https://pubmed.ncbi.nlm.nih.gov/30213670/#:~:text=ICU%20readmission%20risk%20prediction%20may,built%20based%20on%20physiological%20variables.

[9] A;, H. A. J. L. C. (n.d.). Reduction of intensive care unit length of stay: the case of early mobilization. The health care manager. https://pubmed.ncbi.nlm.nih.gov/24776831/#:~:text=The%20average%20ICU%20length%20of,in%20a%20non%2DICU%20bed.