In or Out? Prediction of Inpatient Admission from Emergency Department Visits in a Veteran Population

Mirza S. Khan

Introduction

In the US, the number of emergency department (ED) visits has increased by approximately 50% over the past two decades. The number of ED visits in 2014 was estimated to be 138 million.[1] ED overcrowding is a common problem and is only worsening with the increase in ED visits in recent years.[2] ED overcrowding is associated with delays in care, increased morbidity and mortality, ambulance diversion, and poor hospital satisfaction.[1,3–5] Reducing ED overcrowding may improve healthcare outcomes, operational workflows, and patient and provider satisfaction.

One method to reduce ED overcrowding is to trigger the inpatient admission process as early as possible, which may be achieved by predicting each patient's likelihood of hospital admission early in the ED course. The initial interaction for most patients presenting to the ED will be triage. Triage is an initial assessment of patients to help prioritize patient care in the emergency department. This triage process often involves collecting the reason for presentation to the ED, vital signs and patient demographics. It is only after the patient is seen and evaluated by the healthcare provider that a decision is made about the patient's disposition. This process can be time consuming when factoring in obtaining laboratory studies or imaging and waiting for results to guide the admission decision. Nevertheless, current processes for bed requests and preparation for inpatient admission are only begun after admission is certain. Predicting admissions and demand for hospital beds earlier in the ED process may help improve ED overcrowding by reducing ED boarding and improve clinical outcomes.[6,7]

Ideally, the admission process can be triggered at the time of triage; however, current triage systems are rife with issues. There is great intra- and inter-provider variability in recommendation for hospital admission as clinical gestalt is a significant contributor to the decision-making process.[8] Brillman et al. also found significant variability in admission decisions within and between physicians, nurses and computer-guided triage.[2] Furthermore, a systematic review of current ED triage scales consist of variables that have limited or insufficient evidence to predict clinical outcomes.[9]

Although current systems have their limitations, predictive algorithms may be useful to help identify patients who are likely to require hospital admission during triage or early in the ED visit. Prior work has attempted to leverage ED visit data to predict admission,[1,10,11] but each has been studied in different patient populations. Peck et al. have developed prediction algorithms in the veteran population, but these were limited to regression and Naïve Bayes classification methods. In addition to logistic regression, I used additional methods to predict admission from ED visits in a veteran population that may yield improved predictive performance.

Methods

This study used retrospective data from one Veterans Affairs ED facility located in Nashville, TN. All patients who had an ED visit from January to February 2018 were included. Patients were predominantly male. Data was accessed from the Veterans Affairs Coroporate Data Warehouse using Microsoft SQL Server Management Studio 17 (Microsoft Corp., Redmond, WA). I used logistic regression and two tree-based models, random forest and extreme gradient boosting (XGBoost). Two subsets of data were used in this study: all patients who presented to the ED during the study period ($n = 3,764$) and those patients presenting to the ED who received at least one common lab ($n = 2,102$). For the complete patient sample ($n = 3,764$), I compared the performance of triage information compared to triage and order data on the ability to predict admission. When examining the subset of patients who had at least 1 lab, I fit each of the three models on four data

sources: 1. triage information only, 2. triage information and ED orders, 3. triage information and lab results, and 4. the full set of variables, i.e. triage, ED orders and lab results. Each of the data sources, i.e. triage, orders, and laboratory results, represent different time points in the ED course, i.e. at the time of triage, when orders are placed, and when laboratory results are reported. The target variable was 'admission' or 'no admission.'

Triage Data

Triage data pertains to information that may be available at the time of ED triage. For this analysis, the variables included past medical history, age, vital signs, marital status and if the day of encounter occurred on a weekend. For each patient, ICD-10 codes from the problem list were mapped to the 30 clinical categories of the Elixhauser comorbidity index. These values were structured as binary variables, i.e. 1 for the presence of an ICD-10 code in each Elixhauser category, 0 if the patient had no ICD-10 codes for a corresponding category. This method helped to reduce the problem list space to the 30 clinical categories, each being a binary variable. Marital status was also included given prior literature showing that non-married status was associated with increased mortality.[12] I removed unknown marital status to maintain full-rank and limit collinearity for linear models. Of note, only two individuals had an unknown marital status in this sample. I included only the first set of vital signs obtained to mimic the triage process. Vital signs included temperature, pulse, respiratory Rate, O_2 saturation, systolic blood pressure (SBP), diastolic blood pressure (DBP) and supplemental O_2 requirement. Furthermore, I included if the encounter occurred on a weekend, which was encoded as a binary variable of 1 if weekend, 0 if weekday. This was included as holidays and weekends may reflect limited resources and may influence admission rates.

Laboratory Results

Non-numeric lab results, such as comments were excluded. This primarily related to urine lab specimens. I attempted to normalize these results, but there was variation in how results were encoded, which made this difficult to automate. Labs with missing LOINC codes were also excluded. I used LOINC codes given variation in naming that may exist within and between VA sites to ensure that this method could be employed more generally. I excluded those labs where LOINC codes had greater than 60% missing values. I imputed the median lab value for each of the remaining missing values in the resulting dataset.

Order Data

The mean difference in time from first vitals to ED order placement was 2.41 hours. I limited orders to those that were placed 1 hour before and upto 3 hours after the initial set of vitals were obtained in the ED. Limiting the time window for orders ensures that these orders do not reflect decisions after the admission decision has already occurred. For each order, I collected the count of these orders that were obtained within the specified time window. Counts were used rather than one-hot encoding to avoid potential loss of information. For example, serial labs or imaging may be more informative of the severity of illness, and thus likelihood of admission, rather than treating each order as a binary variable.

Analysis

For all patients presenting to the emergency department, triage information and ED orders are available. If no orders were placed, this is represented as 0 for each order count. Thus, I was able to study the predictive performance of triage information alone to triage information and ED order data using each of the three models. For the subset of patients who have at least one laboratory result, I arranged the available data into four datasets: 1. triage information only, 2. triage and ED order data, 3. triage and laboratory results, and 4. all variables. For each of the four datasets, I fit three models: logistic regression, random forest, and XGBoost.

I used stratified sampling such that seventy-five percent of each of the four datasets was used for training each model and the remaining twenty-five percent was kept as a hold-out set to assess model performance and generalizability. Metrics for performance on the hold-out set included Accuracy, No-information Rate (NIR), Precision, Recall and F1-score.

Prior to fitting each logistic regression model, I performed several pre-processing steps to maintain linearity given the underlying assumptions with this parametric method. Pre-processing included Box Cox transformation of variables with non-normal distributions and centering and scaling of variables to ensure that the predictors have a mean of zero and standard deviation of one. Additionally, collinear variables were excluded to ensure a full rank matrix and variables with zero or near-zero variance were also excluded as they are likely uninformative. The pre-processed data for each of the four datasets was only used for fitting the logistic regression model. Given that random forest and XGBoost are tree-based models and do not hold assumptions about the distribution of the underlying data, I did not use the pre-processed data when fitting these two models.

Each model was fit using 10-fold cross validation. I used the `ranger` package in R for fitting a random forest model due to improved computation performance and parallelizability compared to the `randomForest` package.[13] When possible, I used parallel processing for model fitting using the `doParalel` package to improve efficiency of model fitting. Optimal parameters for the random forest and XGBTree models were identified by tuning hyperparameters using grid search. The optimal model was selected based on the parameters that yielded the highest area under the receiver operating characteristic curve (AUC). The DeLong confidence interval for AUC was computed using 2,000 stratified bootstrap replicates using the `pROC` package. I used AUC because it is not affected by class imbalance, e.g. when compared to accuracy. To ensure reproducibility of results, I used a fixed seed during model fitting.

To examine the variables which were most important to each model, I calculated variable importance using permutation importance for each random forest model and variable importance for each XGBTree model. I also used the Boruta algorithm on the entirety of each of the four datasets to identify important features.[14]

Data analysis and modeling was performed in R (R Foundation for Statistical Computing, Vienna, Austria) and RStudio (RStudio, Inc., Boston, MA) using the `caret`, `ranger`, and `xgboost` packages.

Results

During January to February, 2018, there were 3,764 patients who presented to the Nashville VAMC ED. Of these patients, 948 (25.2% of the total) were admitted to the hospital for further care. The mean age of patients who were admitted was 66.9 years compared to 58.5 years for those who were not admitted. 21.8% of these 948 admissions occurred on weekends. Triage information was available for each of these 3,764 patients.

Models trained on the entire sample included triage information only and triage information and ED orders. Results from each of the models is shown in Table 1. Models trained on triage information resulted in an AUC of 0.705 with a 95% Confidence Interval (CI) of 0.682-0.725 for logistic regression, 0.739 (95% CI: 0.717-0.760) for random forest, and 0.738 (95% CI: 0.717-0.759) for XGBoost. For models trained on triage and ED order information, the AUC for logistic regression was 0.907 (95% CI :0.895-0.919], for random forest was 0.909 (95% CI: 0.898-0.921), and for XGBoost was 0.919 (95% CI: 0.908-0.929). Receiver operating characteristic curves (ROC) for each model trained using samples for each of the four datasets is shown in Figure 1. The distribution of AUC values, sensitivity and specificity for each model from cross-validation is shown in Supplemental Figure 1.

For the subset of patients with laboratory data, models trained on triage information resulted in an AUC of 0.673 (95% CI: 0.644-0.698) for logistic regression, 0.700 (95% CI: 0.672-0.724) for random forest, and 0.692 (95% CI: 0.665-0.718) for XGBoost. Triage information and ED order data for this subset yielded an AUC of 0.838 (95% CI: 0.818-0.856) for logistic regression, 0.856 (95% CI: 0.836-0.872) for random forest, and 0.866 (95% CI: 0.849-0.883) for XGBoost. Those models trained on triage information and lab results yielded an AUC of 0.723 (95% CI: 0.698-0.749) for logistic regression, 0.855 (95% CI: 0.836-0.872) for random forest,

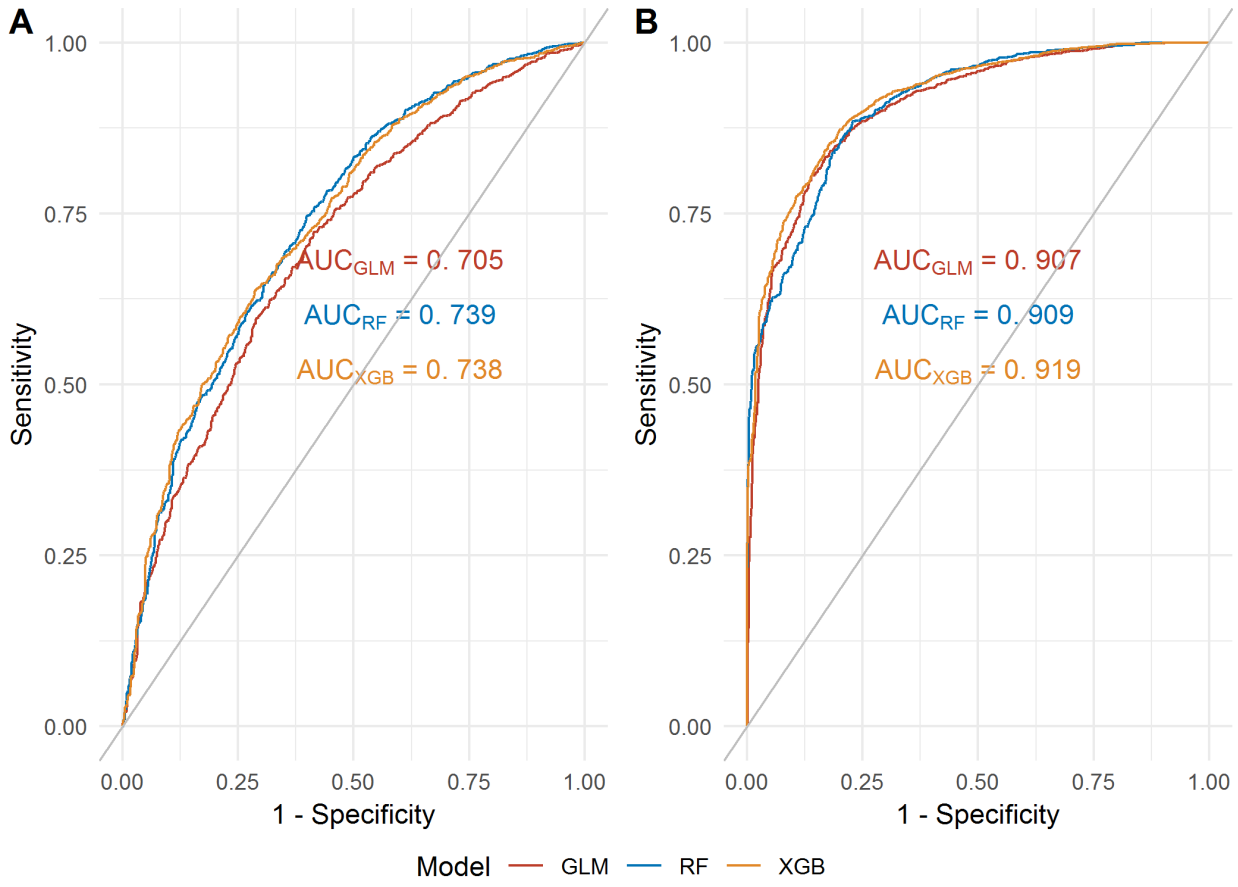

Figure 1: **ROC curves using all ED patient data**: (A) triage information only, (B) triage information and ED provider orders.

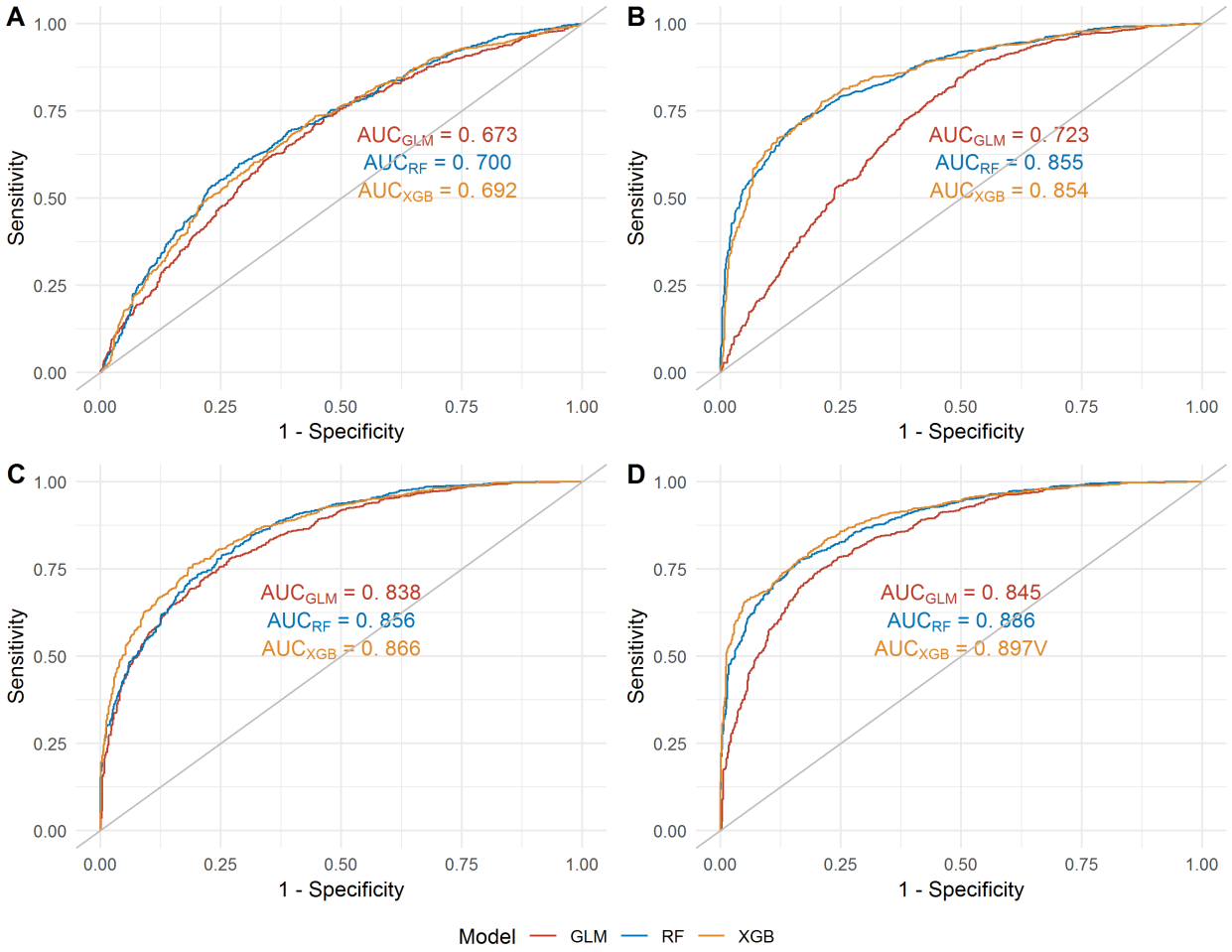

Figure 2: **ROC curves using data from ED patients with laboratory results**: (A) triage information only, (B) triage information and laboratory results, (C) triage information and ED provider orders, (D) triage information, laboratory results and ED provider orders.

and 0.854 (95% CI: 0.834-0.871) for XGBoost. Lastly, models trained using the full set of variables resulted in an AUC of 0.845 (95% CI: 0.826-0.864) for logistic regression, 0.886 (95% CI: 0.870-0.902) for random forest, and 0.897 (95% CI: 0.880-0.911) for XGBoost. These results are reported in Table 2. ROC curves of each model for each of the four datasets is shown in Figure 2. The distribution of AUC values for each model from cross-validation is shown in Supplemental Figure 2.

Tables 1 and 2 includes information from model fitting with 10-fold cross validation and performance metrics when the model was used to predict admission outcomes on the hold-out set. The triage information alone was a poor predictor of hospital admission. For both sets of analyses, i.e. all patients and patients with labs, the accuracy for each of the three models was not significantly greater than the no-information rate NIR with the exception of random forest in the subset of patients with labs (Accuracy 0.611; 95% CI 0.568-0.653; NIR 0.563; $p = 0.015$). Inclusion of ED order data with triage information improved predictive performance for both sets of analyses - all patients and those with labs. Similarly, laboratory results and the full set of variables improved improved predictive accuracy that was statistically significant when compared to the NIR. Each of the three models yielded comparable AUC values for each set of analysis with the exception of logistic regression for the triage and laboratory results dataset, which was performed less well than random forest and XGBoost. Nevertheless, accuracy of this logistic regression model on the hold-out set was no different

than for the other two models for said dataset.

Using the combined triage, lab results and orders dataset, feature importance using Boruta included several laboratory results. The ten most important LOINC codes correspond with Urea Nitrogen (LOINC code 3094-0), Glomerular Filtration Rate (33914-3), Red Blood Cells (RBC; 789-8), Albumin (1751-7), RBC Distribution Width (788-0), Prothrombin Time (5902-2), Neutrophils (751-8), Leukocytes (6690-2), Creatinine (2160-0), and Hemoglobin (718-7). Among the triage information, age was identified as the most important feature.

Discussion

Triage information consisting of a patient's age, vital signs, marital status, past medical history encoded using the Elixhauser comorbidity index, and if the encounter occurred on a weekend provided no significant predictive capability on ED disposition for each of the models studied. By contrast, incorporating data from either ED provider orders, laboratory results, or both yielded significant predictive performance when compared to the no-information rate (Tables 1 and 2). In other words, I found that acute and short-term predictors were the most important to each model, rather than historical information, such as past medical history.

For each of the 6 sets of models, logistic regression performed comparably to the tree-based methods used with the exception of models built using triage information and laboratory results (Figure 2B, Table 2). This may reflect some degree of non-linearity or interactions that were not accounted for in the logistic regression model. As such, this may be remedied by using spline functions, polynomial features or interaction terms. Furthermore, other linear approaches, such as lasso or ridge regression may have given better performance over logistic regression, especially when increasing the feature space. Neural networks could also be considered as these are better equipped to model higher-order interactions.

The target variable for this study was hospital admission or not. Thus, I was effectively modeling provider behavior and decision-making, rather than the optimal decision about hospitalization for each patient. This target is suitable for the proposed study, namely to identify patients with the highest likelihood of admission to mitigate ED overcrowding and improve hospital bed utilization and flow. Nevertheless, studying the optimal admission decision for each patient is an important research question to pose in future studies.

The pattern of missing laboratory data appeared consistent with informative missing, which can be classified as "missing not at random." Thus, there may be some selection bias in the laboratory studies obtained. For instance, the presence of a troponin test may be informative about the severity of illness and the reason for presentation. I used median imputation in an attempt to impute a within normal lab value for missing laboratory results. Imputation with normal laboratory results is a commonly used approach when laboratory results are missing as in this setting. The rationale is that providers are intentionally not ordering certain laboratory tests because they expect them to be normal or uninformative. Future studies should consider other approaches, such as multiple imputation, nearest neighbors or other regression methods, such as random forest. Waljee et al. found imputation error to be comparable using the aforementioned approaches with the lowest error using random forest imputation in patients with cirrhosis and inflammatory bowel disease.[15] Additionally, some implementations of tree-based methods are capable of handling missing data without the need for imputation, such as XGBoost.

Other limitations of this study include the use of order names when one-hot encoding orders. Given that naming conventions are not standardized within and across hospitals, I should consider using standardized alternatives such as RxNorm or National Drug Codes. Bias and variance in the estimates using 10-fold cross validation could have been improved by using repeat cross-validation, but this was not done due to prolonged compute time. Overfitting may also be an issue. An approach to mitigate this effect would be to allow for some tolerance when selecting to optimal model that would favor selection of a more parsimonious model. Another limitation is the lack of the current triage or admission model for reference. Ideally, I would compare each model to the triaging system currently employed within VA EDs or other widely used methods. For

example, Emergency Severity Index is the most commonly used triaging system in the US, but relies heavily on clinical gestalt and presents significant inter-rater variability.[1]

Future studies should incorporate additional patient and provider-level information, including medication history, number of current medications or if a patient is taking a select class of medications, e.g. chemotherapy, immunosuppressants, cardiac medications, or antibiotics. Additionally, medication severity such as dose, e.g. morphine-equivalent dose for opioids, may also be more informative to the models. Including frailty, mobility, loneliness, proximity to hospital, and social determinants of health may also be useful. Furthermore, radiology results and results of other diagnostic tests should also be included in future studies.

I found that 21.8% of admissions through the ED occur on weekends, which is similar to what is shown in the literature.[16] Similar to prior studies, among the most important demographic features was a patient's age.[11]

Conclusion

ED overcrowding is a significant problem in healthcare systems that is worsening with rising rates of ED utilization. This poses a strain on hospital resources and is associated with worse healthcare outcomes and patient satisfaction. Predicting the likelihood of hospitalization early in ED encounters may help improve patient flow and allocate hospital resources in a more timely manner. I have identified that use of triage information, ED order data and laboratory results using logistic regression or tree-based methods may be useful in identifying patients with high likelihood of hospital admission. This may be useful to mitigate ED overcrowding and improve hospital workflows and patient outcomes.

Supplemental Material

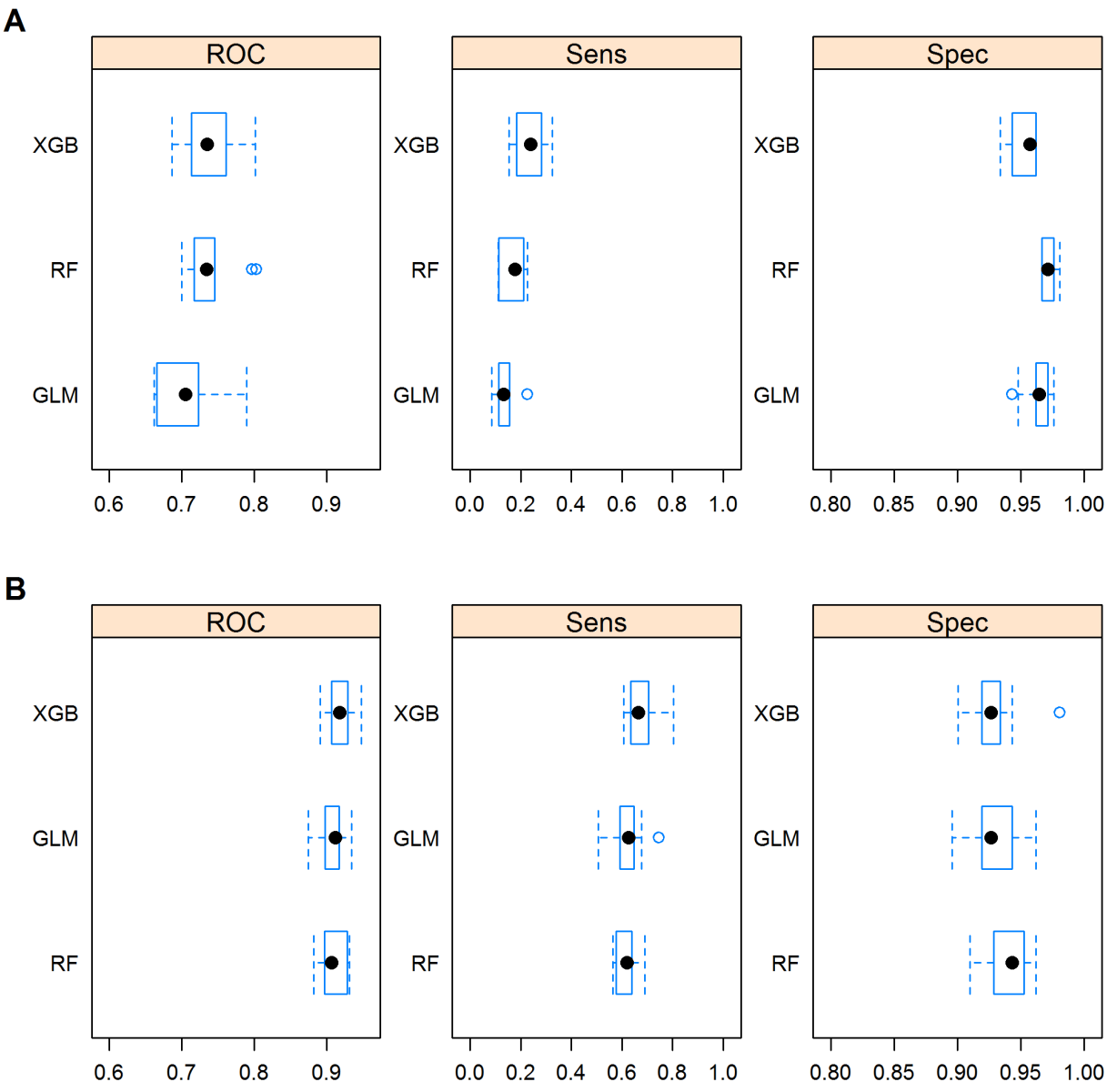

Figure 3: **Supplemental Figure 1 - all ED patients**: Distribution of AUC, Sensitivity and Specificity based upon (A) triage information only and (B) triage information and ED orders.

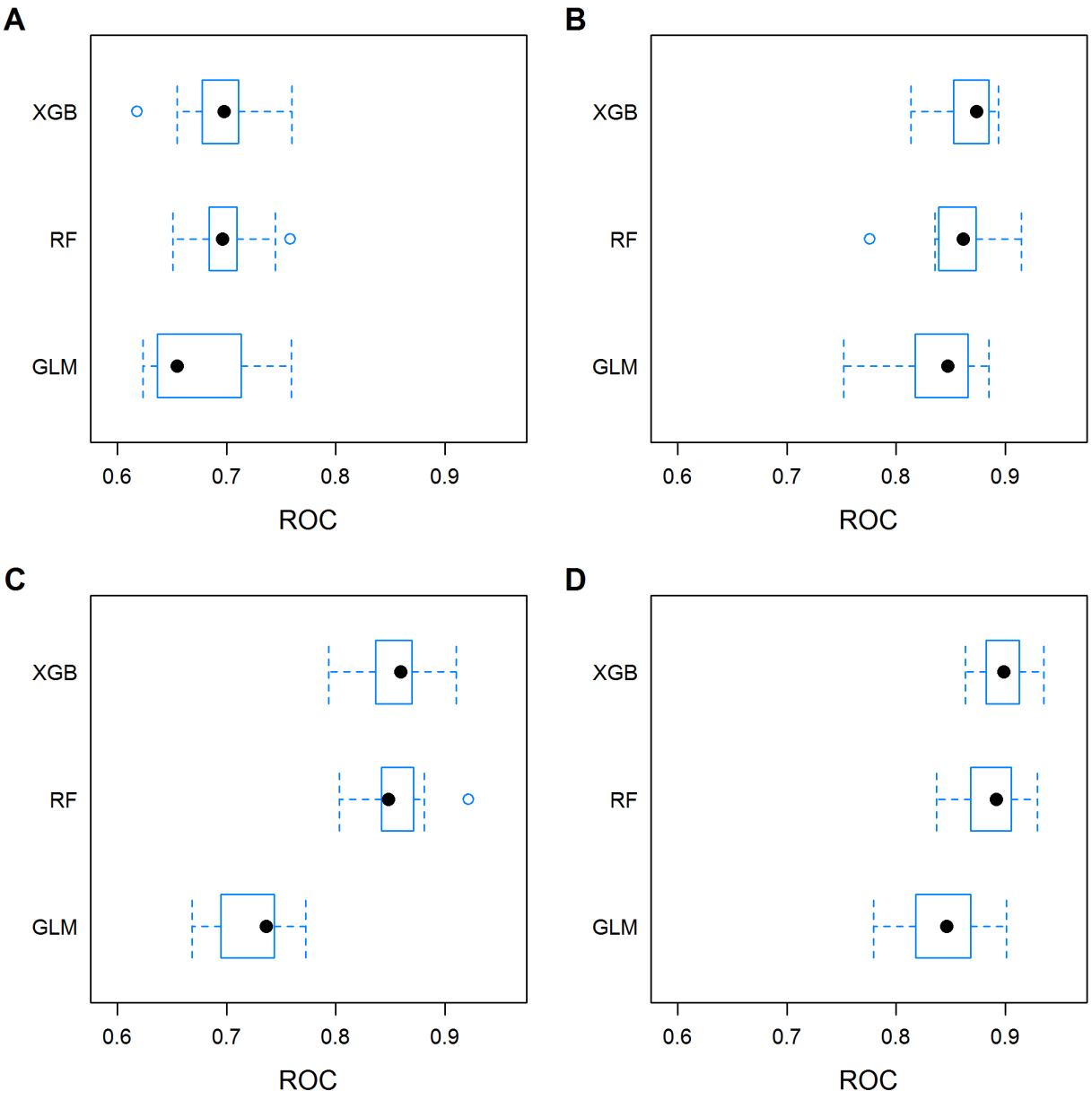

Figure 4: **Supplemental Figure 2 - ED patients with laboratory results**: Distribution of AUC for (A) triage information only, (B) triage information and ED orders, (C) triage information and laboratory results, and (D) triage information, laboratory results and ED provider orders.

Table 1: All ED Patients - Model Performance with Cross-Validation and Hold-out Set

	Cross-Validation			Hold-out Set					
Model	AUC (95% CI)	Sensitivity	Specificity	Accuracy (95% CI)	NIR	p-value[a]	Precision	Recall	F1-score
Triage									
GLM	0.705 (0.682, 0.725)	0.136	0.963	0.766 (0.738, 0.793)	0.748	0.107	0.681	0.135	0.225
Random Forest	0.739 (0.717, 0.760)	0.170	0.972	0.768 (0.740, 0.795)	0.748	0.081	0.694	0.143	0.238
XGBTree	0.738 (0.717, 0.759)	0.238	0.953	0.765 (0.737, 0.792)	0.748	0.122	0.600	0.203	0.303
Triage & Orders									
GLM	0.907 (0.895, 0.919)	0.619	0.930	0.859 (0.835, 0.880)	0.748	$<$ 2.2e-16	0.765	0.633	0.693
Random Forest	0.909 (0.898, 0.921)	0.619	0.941	0.873 (0.850, 0.893)	0.748	$<$ 2.2e-16	0.806	0.650	0.720
XGBTree	0.919 (0.908, 0.929)	0.675	0.929	0.864 (0.840, 0.885)	0.748	$<$ 2.2e-16	0.763	0.667	0.712

Abbreviations: AUC, Area under the ROC curve; NIR, No Information Rate

[a] One-sided significance test: Pr[Accuracy $>$ NIR]

Table 2: Subset of Patients with Labs - Model Performance with Cross-Validation and Hold-out Set

	Cross-Validation			Hold-out Set					
Model	AUC (95% CI)	Sensitivity	Specificity	Accuracy (95% CI)	NIR	p-value[a]	Precision	Recall	F1-score
Triage									
GLM	0.673 (0.644, 0.698)	0.487	0.765	0.599 (0.560, 0.642)	0.563	0.051	0.553	0.437	0.488
Random Forest	0.700 (0.672, 0.724)	0.529	0.738	0.611 (0.568, 0.653)	0.563	0.015	0.558	0.528	0.543
XGBTree	0.692 (0.665, 0.718)	0.509	0.753	0.599 (0.556, 0.641)	0.563	0.051	0.544	0.511	0.527
Triage & Orders									
GLM	0.838 (0.818, 0.856)	0.675	0.807	0.758 (0.719, 0.794)	0.563	$<$ 2e-16	0.745	0.677	0.709
Random Forest	0.856 (0.836, 0.872)	0.683	0.847	0.760 (0.721, 0.796)	0.563	$<$ 2e-16	0.759	0.659	0.706
XGBTree	0.866 (0.849, 0.883)	0.728	0.821	0.771 (0.733, 0.806)	0.563	$<$ 2e-16	0.740	0.734	0.737
Triage & Labs									
GLM	0.723 (0.698, 0.749)	0.532	0.807	0.679 (0.6375, 0.719)	0.563	3.23e-08	0.663	0.542	0.596
Random Forest	0.855 (0.836, 0.872)	0.752	0.790	0.750 (0.711, 0.787)	0.563	$<$ 2e-16	0.715	0.712	0.713
XGBTree	0.854 (0.834, 0.871)	0.751	0.808	0.756 (0.717, 0.792)	0.563	$<$ 2e-16	0.721	0.721	0.721
Triage, Orders, & Labs									
GLM	0.845 (0.826, 0.864)	0.706	0.816	0.771 (0.733, 0.806)	0.563	$<$ 2e-16	0.756	0.703	0.728
Random Forest	0.886 (0.870, 0.902)	0.755	0.822	0.786 (0.749, 0.821)	0.563	$<$ 2e-16	0.762	0.742	0.752
XGBTree	0.897 (0.880, 0.911)	0.781	0.832	0.790 (0.753, 0.824)	0.563	$<$ 2e-16	0.767	0.747	0.757

Abbreviations: AUC, Area under the ROC curve; NIR, No Information Rate

[a] One-sided significance test: Pr[Accuracy $>$ NIR]

References

1. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. *Critical Care*. 2019;23(1):64. doi:10.1186/s13054-019-2351-7

2. Brillman JC, Doezema D, Tandberg D, et al. Triage: Limitations in Predicting Need for Emergent Care and Hospital Admission. *Annals of Emergency Medicine*. 1996;27(4):493-500. doi:10.1016/S0196-0644(96)70240-8

3. Sprivulis PC, Silva J-AD, Jacobs IG, Jelinek GA, Frazer ARL. The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments. *Medical Journal of Australia*. 2006;184(5):208-212. doi:10.5694/j.1326-5377.2006.tb00203.x

4. Boyle J, Jessup M, Crilly J, et al. Predicting emergency department admissions. *Emergency Medicine Journal*. 2012;29(5):358-365. doi:10.1136/emj.2010.103531

5. Pines JM, Iyer S, Disbot M, Hollander JE, Shofer FS, Datner EM. The Effect of Emergency Department Crowding on Patient Satisfaction for Admitted Patients. *Academic Emergency Medicine*. 2008;15(9):825-831. doi:10.1111/j.1553-2712.2008.00200.x

6. Peck JS, Benneyan JC, Nightingale DJ, Gaehde SA. Predicting Emergency Department Inpatient Admissions to Improve Same-day Patient Flow. *Academic Emergency Medicine*. 2012;19(9):E1045-E1054. doi:10.1111/j.1553-2712.2012.01435.x

7. Sun Y, Heng BH, Tay SY, Seow E. Predicting Hospital Admissions at Emergency Department Triage Using Routine Administrative Data. *Academic Emergency Medicine*. 2011;18(8):844-850. doi:10.1111/j.1553-2712.2011.01125.x

8. Wuerz R, Fernandes CM, Alarcon J. Inconsistency of Emergency Department Triage. *Annals of Emergency Medicine*. 1998;32(4):431-435. doi:10.1016/S0196-0644(98)70171-4

9. Farrohknia N, Castrén M, Ehrenberg A, et al. Emergency Department Triage Scales and Their Components: A Systematic Review of the Scientific Evidence. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*. 2011;19(1):42. doi:10.1186/1757-7241-19-42

10. Dexheimer JW, Leegon J, Aronsky D. Predicting hospital admission at triage in an emergency department. *AMIA Annual Symposium proceedings AMIA Symposium*. October 2007:937.

11. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLOS ONE*. 2018;13(7):e0201016. doi:10.1371/journal.pone.0201016

12. Johnson NJ, Backlund E, Sorlie PD, Loveless CA. Marital Status and Mortality: The National Longitudinal Mortality Study. *Annals of Epidemiology*. 2000;10(4):224-238. doi:10.1016/S1047-2797(99)00052-6

13. Wright MN, Ziegler A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*. 2017;77(1). doi:10.18637/jss.v077.i01

14. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. *Journal of Statistical Software*. 2010;036(i11).

15. Waljee AK, Mukherjee A, Singal AG, et al. Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*. 2013;3(8):e002847. doi:10.1136/bmjopen-2013-002847

16. Ryan K, Levit K, Davis PH. Characteristics of Weekday and Weekend Hospital Admissions: Statistical Brief #87. In: *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006.