

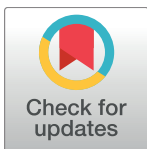
RESEARCH ARTICLE

# Predicting hospital admission at emergency department triage using machine learning

Woo Suk Hong<sup>1</sup>, Adrian Daniel Haimovich<sup>1</sup>, R. Andrew Taylor<sup>2\*</sup>

**1** Yale School of Medicine, New Haven, Connecticut, United States of America, **2** Department of Emergency Medicine, Yale School of Medicine, New Haven, Connecticut, United States of America

\* [richard.taylor@yale.edu](mailto:richard.taylor@yale.edu)



## Abstract

### Objective

To predict hospital admission at the time of ED triage using patient history in addition to information collected at triage.

### Methods

This retrospective study included all adult ED visits between March 2014 and July 2017 from one academic and two community emergency rooms that resulted in either admission or discharge. A total of 972 variables were extracted per patient visit. Samples were randomly partitioned into training (80%), validation (10%), and test (10%) sets. We trained a series of nine binary classifiers using logistic regression (LR), gradient boosting (XGBoost), and deep neural networks (DNN) on three dataset types: one using only triage information, one using only patient history, and one using the full set of variables. Next, we tested the potential benefit of additional training samples by training models on increasing fractions of our data. Lastly, variables of importance were identified using information gain as a metric to create a low-dimensional model.

### Results

A total of 560,486 patient visits were included in the study, with an overall admission risk of 29.7%. Models trained on triage information yielded a test AUC of 0.87 for LR (95% CI 0.86–0.87), 0.87 for XGBoost (95% CI 0.87–0.88) and 0.87 for DNN (95% CI 0.87–0.88). Models trained on patient history yielded an AUC of 0.86 for LR (95% CI 0.86–0.87), 0.87 for XGBoost (95% CI 0.87–0.87) and 0.87 for DNN (95% CI 0.87–0.88). Models trained on the full set of variables yielded an AUC of 0.91 for LR (95% CI 0.91–0.91), 0.92 for XGBoost (95% CI 0.92–0.93) and 0.92 for DNN (95% CI 0.92–0.92). All algorithms reached maximum performance at 50% of the training set or less. A low-dimensional XGBoost model built on ESI level, outpatient medication counts, demographics, and hospital usage statistics yielded an AUC of 0.91 (95% CI 0.91–0.91).

## OPEN ACCESS

**Citation:** Hong WS, Haimovich AD, Taylor RA (2018) Predicting hospital admission at emergency department triage using machine learning. PLoS ONE 13(7): e0201016. <https://doi.org/10.1371/journal.pone.0201016>

**Editor:** Qunfeng Dong, University of North Texas, UNITED STATES

**Received:** February 12, 2018

**Accepted:** July 6, 2018

**Published:** July 20, 2018

**Copyright:** © 2018 Hong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The raw data used in this study was derived from electronic health records of patient visits to the Yale New Haven Health system and is not publicly available due to the ubiquitous presence of protected health information (PHI). A de-identified, processed dataset of all patient visits included in the models, as well as scripts used for processing and analysis, is available on the Github repository (<https://github.com/yaleemmlc/admissionprediction>) (10.5281/zenodo.1308993). All other data is available within the paper and its Supporting Information files.

**Funding:** WH is supported by the James G. Hirsch Endowed Medical Student Research Fellowship at Yale University School of Medicine. AH is supported by National Institutes of Health grants 1F30CA196191 and T32GM007205. RT received no specific funding for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Conclusion

Machine learning can robustly predict hospital admission using triage information and patient history. The addition of historical information improves predictive performance significantly compared to using triage information alone, highlighting the need to incorporate these variables into prediction models.

## Introduction

While most emergency department (ED) visits end in discharge, EDs represent the largest source of hospital admissions [1]. Upon arrival to the ED, patients are first sorted by acuity in order to prioritize individuals requiring urgent medical intervention. This sorting process, called "triage", is typically performed by a member of the nursing staff based on the patient's demographics, chief complaint, and vital signs. Subsequently, the patient is seen by a medical provider who creates the initial care plan and ultimately recommends a disposition, which this study limits to hospital admission or discharge.

Prediction models in medicine seek to improve patient care and increase logistical efficiency [2,3]. For example, prediction models for sepsis or acute coronary syndrome are designed to alert providers of potentially life-threatening conditions, while models for hospital utilization or patient-flow enable resource optimization on a systems level [4–8]. Early identification of ED patients who are likely to require admission may enable better optimization of hospital resources through improved understanding of ED patient mixtures [9]. It is increasingly understood that ED crowding is correlated with poorer patient outcomes [10]. Notification of administrators and inpatient teams regarding potential admissions may help alleviate this problem [11]. From the perspective of patient care in the ED setting, a patient's likelihood of admission may serve as a proxy for acuity, which is used in a number of downstream decisions such as bed placement and the need for emergency intervention [12–14].

Numerous prior studies have sought to predict hospital admission at the time of ED triage. Most models only include information collected at triage such as demographics, vital signs, chief complaint, nursing notes, and early diagnostics [11,14–19], while some models include additional features such as hospital usage statistics and past medical history [9,12,20,21]. A few models built on triage information have been formalized into clinical decision rules such as the Sydney Triage to Admission Risk Tool and the Glasgow Admission Prediction Score [22–25]. Notably, a progressive modeling approach that uses information available at later time-points, such as lab tests ordered, medications given, and diagnoses entered by the ED provider during the patient's current visit, has been able to achieve high predictive power and indicates the utility of these features [20,21]. We hypothesized that extracting such features from a patient's previous ED visits would lead to a robust model for predicting admission at the time of triage. Prior models that incorporate past medical history utilize simplified chronic disease categories such as heart disease or diabetes [9,12] while leaving out rich historical information accessible from the electronic health record (EHR) such as outpatient medications and historical labs and vitals, all of which are routinely reviewed by providers when evaluating a patient. As a recent work showed that using all elements of the electronic health record can robustly predict in-patient outcomes [26], a prediction model for admission built on comprehensive elements of patient history may improve on prior models.

Furthermore, many prior studies have been limited by technical factors, where continuous variables are often reduced to categorical variables through binning or to binary variables

encoding presence or missing-ness of data due to the challenges of imputation [9,15,16,19–21]. Logistic regression and Naive Bayes are commonly used [9,11,16,18–22], with few studies using more complex algorithms like random-forests, artificial neural networks, and support vector machines [12,15,17]. While gradient boosting and deep neural networks have been shown to be powerful tools for predictive modeling, neither has been applied to the task of predicting admission at ED triage to date.

Expanding on prior work [9,12,20], we build a series of binary classifiers on 560,486 patient visits, with 972 variables extracted per visit from the EHR, including previous healthcare usage statistics, past medical history, historical labs and vitals, prior imaging counts, and outpatient medications, as well as fine demographic details such as insurance and employment status. We use gradient boosting and deep neural networks, two of the best performing algorithms in classification tasks, to model the nonlinear relationships among these variables. Moreover, we test whether we have achieved maximum performance for our feature set by measuring performance across models trained on increasing fractions of our data. Lastly, we identify variables of importance using information gain as our metric and present a low-dimensional model amenable to implementation as clinical decision support.

## Materials and methods

### Study setting

Retrospective data was obtained from three EDs covering the period of March 2013 to July 2017 to ensure a 1-year of historical timeframe from the study start period of March 2014. The represented EDs include a level I trauma center with an annual census of approximately 85,000 patients, a community hospital-based department with an annual census of approximately 75,000 patients, and a suburban, free-standing department with an annual census of approximately 30,000 patients. All three EDs are part of a single hospital system utilizing the Epic EHR (Verona, WI) and the Emergency Severity Index (ESI) for triage. The study included all visits for adult patients with a clear, recorded disposition of either admission or discharge. Individuals with any other disposition, such as transfer, AMA, and eloped, were excluded. This study was approved, and the informed consent process waived, by the Yale Human Investigation Committee (IRB 2000021295).

### Data collection and processing

For each patient visit, we collected a total of 972 variables, divided into major categories shown in Table 1. The full list of variables is provided in S1 Table. All data elements were obtained from the enterprise data warehouse, using SQL queries to extract relevant raw-data in comma-separated value format. All subsequent processing was done in R. The link to a repository containing the de-identified dataset and R scripts are available in S1 Text. Below, we summarize the processing steps for each category.

**Response variable.** The primary response variable was the patient's disposition, encoded in a binary variable (1 = admission, 0 = discharge).

**Demographics.** Demographic information, either collected at triage or available from EHR at the time of patient encounter, included age, gender, primary language, ethnicity, employment status, insurance status, marital status, and religion. The primary language variable was recoded into a binary split (e.g., English vs. non-English), while the top twelve levels comprising >95% of all visits were retained for the religion variable and all other levels binned to one 'Other' category. All unique levels were retained for other demographic variables.

**Triage evaluation.** Triage evaluation included variables routinely collected at triage, such as the name of presenting hospital, arrival time (month, day, 4-hr bin), arrival method, triage

**Table 1. Variables included in models.**

Category	Number of Variables	Only Triage	Only History	Full
Response variable (Disposition)	1	X	X	X
Demographics	9	X	X	X
Triage evaluation	13	X		X
Chief complaint	200	X		X
Hospital usage statistic	4		X	X
Past medical history	281		X	X
Outpatient medications	48		X	X
Historical vitals	28		X	X
Historical labs	379		X	X
Imaging/EKG counts	9		X	X
Total	972	223	759	972

Only Triage—model using only triage information. Only History—model using only patient history. Full—model using the full set of variables. Note that demographic information is included in all three models.

<https://doi.org/10.1371/journal.pone.0201016.t001>

vital signs, and ESI level assigned by the triage nurse. Triage vital signs included systolic and diastolic blood pressure, pulse, respiratory rate, oxygen saturation, presence of oxygen device, and temperature. Values beyond physiologic limits were replaced with missing values.

**Chief complaint.** Given the high number of unique values (> 1000) for chief complaint, the top 200 most frequent values, which comprised >90% of all visits, were retained as unique categories and all other values binned into 'Other'.

**Hospital usage statistic.** The number of ED visits within one year, the number of admissions within one year, the disposition of the patient's previous ED visit, and the number of procedures and surgeries listed in the patient's record at the time of encounter were taken as metrics for prior hospital usage.

**Past medical history.** ICD-9 codes for past medical history (PMH) were mapped onto 281 clinically meaningful categories using the Agency for Healthcare Research and Quality (AHRQ) Clinical Classification Software (CCS), such that each CCS category became a binary variable with the value 1 if the patient's PMH contained one or more ICD-9 code belonging in that category and 0 otherwise.

**Outpatient medications.** Outpatient medications listed in the EHR as active at the time of patient encounter were binned into 48 therapeutic subgroups (e.g. cardiovascular, analgesics) used internally by the Epic EHR system, with each corresponding variable representing the number of medications in that subgroup.

**Historical vitals.** A time-frame of one year from the date of patient encounter was used to calculate historical information, which included vital signs, labs and imaging previously ordered from any of the three EDs. Historical vital signs were represented by the minimum, maximum, median, and the last recorded value of systolic blood pressure, diastolic blood pressure, pulse, respiratory rate, oxygen saturation, presence of oxygen device, and temperature. Values beyond physiologic limits were replaced with missing values.

**Historical labs.** Given the diversity of labs ordered within the ED, the 150 most frequent labs comprising 94% of all orders were extracted then divided into labs with numeric values and those with categorical values. The cutoff of 150 was chosen to include labs ordered commonly enough to be significant in the management of most patients (e.g., Troponin T, BNP, CK, D-Dimer), even if they were not as frequent as routine labs like CBC, BMP, and urinalysis. The minimum, maximum, median, and the last recorded value of each numeric lab were

included as features. Categorical labs, which included urinalysis and culture results, were recoded into binary variables with 1 for any positive value (e.g. positive, trace, +, large) and 0 otherwise. Any growth in blood culture was labeled positive as were urine cultures with > 49,000 colonies/mL. The number of tests, the number of positives, and the last recorded value of each categorical lab were included as features.

**Imaging and EKG counts.** The number of orders were counted for each of the following categories: electrocardiogram (EKG), chest x-ray, other x-ray, echocardiogram, other ultrasound, head CT, other CT, MRI, and all other imaging.

## Model fitting and evaluation

A series of nine binary classifiers were trained using logistic regression (LR), gradient boosting (XGBoost), and deep neural networks (DNN) on three dataset types: one using only triage information, one using only patient history, and one using the full set of variables (Table 1). All analyses were done in R using the caret, xgboost, and keras packages [27–29].

A randomly chosen test set of 56,000 (10%) samples was held out, then the remaining 504,486 (90%) samples split randomly five times to create five independent validation sets of 56,000 (10%) and training sets of 448,486 (80%). Hyperparameters for each model were optimized by maximizing the average validation AUC across the five validation sets. The optimized set of hyperparameters was then used to train the model on all 504,486 (90%) samples excluding the test set. Finally, the test AUC was calculated on the held-out test set, with 95% confidence intervals constructed using the DeLong method implemented in the pROC package [30,31]. Youden's index was used to find the optimal cutoff point on the ROC curve to calculate the sensitivity, specificity, positive predictive value, and negative predictive value for each model [32,33]. Details of the tuning process are provided in S1 Text.

Categorical variables were converted into binary variables prior to training using one-hot encoding [34]. The median for each variable post normalization was used to impute the input matrix for LR and DNN. The sparsity of our dataset prevented taking a more sophisticated imputation approach such as k-nearest neighbors or random forests [35,36]. An alternative to imputation is to transform all continuous variables into categorical variables, binning NA into a separate category, then performing one-hot encoding [37]. However, this approach loses all ordinal information and thus was not taken. Imputation was not performed for XGBoost, since the algorithm learns a default direction for each split in the case that the variable needed for the split is missing [28].

## Testing the benefit of additional training samples

One key question in predictive modeling is whether additional training samples will improve performance or whether a model has reached its maximum performance given the inherent noise in its features [38]. To test the potential benefit of additional training samples, we trained full-variable models using each of the three algorithms on randomly selected fractions of the training set (1%, 10%, 30%, 50%, 80%, 100%), then calculated their AUCs on the held-out test set in order to quantify the incremental gain in performance.

## Variables of importance

Information gain is a metric that quantifies the improvement in accuracy of a tree-based algorithm from a split based on a given variable [39]. We calculated the mean information gain for each variable based on 100 training iterations of the full XGBoost model. We then trained a low-dimensional XGBoost model using a subset of variables with high information gain to test whether such a model could predict hospital admission as robustly as the full model.

## Results

### Characteristics of study samples

A total of 560,486 ED visits were available for analysis after filtering for exclusion criteria, with 13% of the samples excluded due to disposition other than admission or discharge. The visits represented 202,953 unique patients, with a median visit count of 1 and a mean visit count of 2.76 per patient during the study duration. The overall hospital admission risk was 29.7% and decreased by triage level: ESI-1 85.6%, ESI-2 55.0%, ESI-3 29.1%, ESI-4 2.2%, and ESI-5 0.4% (S1 Fig). Characteristics of the study samples are presented in Table 2.

### Model performance

Models trained on triage information yielded a test AUC of 0.87 for LR (95% CI 0.86–0.87), 0.87 for XGBoost (95% CI 0.87–0.88) and 0.87 for DNN (95% CI 0.87–0.88). Models trained on patient history yielded an AUC of 0.86 for LR (95% CI 0.86–0.87), 0.87 for XGBoost (95% CI 0.87–0.87) and 0.87 for DNN (95% CI 0.87–0.88). Models trained on the full set of variables yielded an AUC of 0.91 for LR (95% CI 0.91–0.91), 0.92 for XGBoost (95% CI 0.92–0.93) and 0.92 for DNN (95% CI 0.92–0.92). The addition of historical information improved predictive performance significantly compared to using triage information alone (Fig 1). Notably, we were able to achieve an AUC of over 0.86 by using patient history alone, which excludes triage level. XGBoost and DNN outperformed LR on the full dataset, while there was no significant difference in performance between XGBoost and DNN across all three dataset types. The sensitivity, specificity, positive-predictive-value (PPV), and negative-predictive-value (NPV) of each model are shown in Table 3. The optimized hyperparameters for each model, as well as its the training and validation AUCs, are provided in S1 Text.

### Testing the benefit of additional training samples

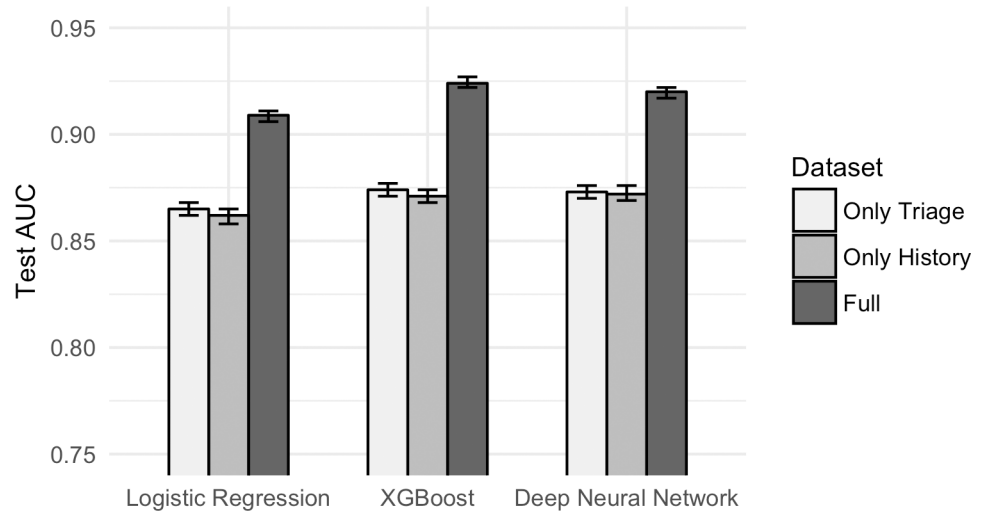
For LR, the 95% CI of the AUC of the model trained on 10% of the training set contained the AUC of the model trained on the entire training set. For XGBoost and DNN, the point at which this occurred was at 50% of the training set (Fig 2). All AUC values are provided in S1 Text.

**Table 2. Characteristics of study samples.**

	Admitted (n = 166,638)	Discharged (n = 393,848)
Age in mean years (95% CI)	61.6 (61.5–61.7)	44.9 (44.9–45.0)
Gender—Male (%)	77,093 (46.3%)	173,740 (44.1%)
Language—English (%)	154,831 (92.9%)	359,985 (91.4%)
Arrival mode—Ambulance (%)	89,955 (54.0%)	100,415 (25.5%)
Mean triage heart rate (95% CI)	88.9 (88.7–89.0)	84.6 (84.5–84.6)
Mean triage systolic blood pressure (95% CI)	134.7 (134.6–134.9)	132.9 (132.9–133.0)
Mean triage diastolic blood pressure (95% CI)	79.4 (79.3–79.5)	80.8 (80.8–80.9)
Mean triage respiratory rate (95% CI)	18.0 (18.0–18.0)	17.5 (17.5–17.5)
Mean triage oxygen saturation (95% CI)	96.6 (96.6–96.7)	97.5 (97.5–97.5)
Mean triage temperature (95% CI)	98.2 (98.2–98.2)	98.1 (98.1–98.1)
Median ESI Level	2	3

All comparisons were significant with  $p < 2.2e-16$

<https://doi.org/10.1371/journal.pone.0201016.t002>



**Fig 1. Test AUC by dataset type by algorithm.** Addition of historical information improves predictive performance significantly compared to using triage information alone. Patient history alone can predict admission to a reasonable degree.

<https://doi.org/10.1371/journal.pone.0201016.g001>

### Variables of importance

Variables of importance extracted from a hundred iterations of the full XGBoost model are shown in Fig 3 (numeric values provided in S2 Table). Variables representing ESI level, outpatient medication counts, demographics, and hospital usage statistics showed high information gain. A low-dimensional XGBoost model built on variables from these four categories, which include ESI level, age, gender, marital status, employment status, insurance status, race, ethnicity, primary language, religion, number of ED visits within 1 year, number of admissions within 1 year, disposition of the previous ED visit, total number of prior surgeries or procedures, and outpatient medication counts by therapeutic category, yielded a test AUC of 0.91 (95% CI 0.91–0.91).

### Discussion

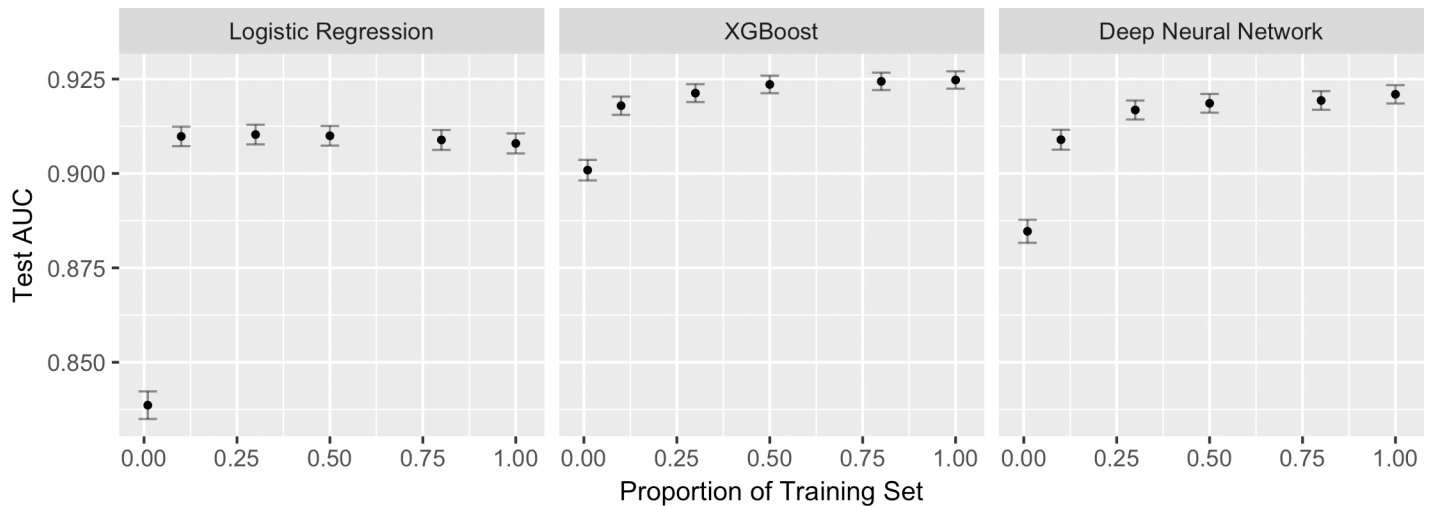
We describe a series of prediction models for hospital admission that leverage gradient boosting and deep neural networks on a dataset of 560,486 patient visits. Our study shows that

**Table 3. Summary of statistical measures for each model.**

Algorithm	Dataset	Test AUC (95% CI)	Sensitivity	Specificity	PPV	NPV
LR	Only Triage	0.865 (0.862–0.868)	0.68	0.85	0.65	0.87
LR	Only History	0.862 (0.858–0.865)	0.72	0.85	0.67	0.88
LR	Full	0.909 (0.906–0.911)	0.80	0.85	0.69	0.91
XGBoost	Only Triage	0.874 (0.871–0.877)	0.69	0.85	0.66	0.87
XGBoost	Only History	0.871 (0.868–0.874)	0.73	0.85	0.67	0.88
XGBoost	Full	0.924 (0.922–0.927)	0.83	0.85	0.70	0.92
DNN	Only Triage	0.873 (0.870–0.876)	0.70	0.85	0.66	0.87
DNN	Only History	0.872 (0.869–0.876)	0.74	0.85	0.67	0.89
DNN	Full	0.920 (0.917–0.922)	0.82	0.85	0.70	0.92
XGBoost	Top Variables	0.910 (0.908–0.913)	0.79	0.85	0.69	0.91

95% CI for all measures < ± 0.01. The cutoff threshold for each model was set to match a fixed specificity of 0.85 to facilitate comparison. The value of 0.85 was chosen by using Youden’s Index on the full XGBoost model. Models achieving a test AUC greater than 0.9 are shaded in gray.

<https://doi.org/10.1371/journal.pone.0201016.t003>



**Fig 2. Model performance on increasing fractions of the training set.** 95% CIs are shown in gray bars. All three algorithms reach maximum performance at 50% of the training set or less. LR reaches maximum performance earlier than XGBoost or DNN.

<https://doi.org/10.1371/journal.pone.0201016.g002>

machine learning can robustly predict hospital admission at emergency department (ED) triage and that the addition of patient history improves predictive performance significantly compared to using triage information alone, highlighting the need to incorporate these variables into predictive models.

In order to test whether additional training samples will improve performance, we train our models on increasing fractions of the training set and show that the AUC plateaus well below the full training set. This result suggests that we have likely maximized the discriminatory capability for our feature set. More studies will be required to develop features that may further improve performance. We expect that many of these features will be derived from free-text data in the electronic health record (EHR). Specifically, natural language processing of medical notes may provide an informative set of features that capture information absent in tabular data. Recent prediction models on outcomes ranging from sepsis to suicide have demonstrated success with these approaches [4,26,40,41].

The ranking of information gain extracted from the gradient boosting (XGBoost) model present a number of notable features. In particular, the importance of medication counts in the model may either reflect a proxy feature for medical complexity or indicate that polypharmacy itself is a risk factor [42,43]. Not surprisingly, the triage level encoded by the Emergency Severity Index (ESI) had high information gain [44,45], as did prior hospital usage statistics such as the number of admissions within the past year and the disposition of the previous ED visit. Variables correlated with age and markers of socioeconomic status such as insurance type were some of the other features identified by our model that have been previously linked to hospital admission [46–48]. We show that these features can be combined to create a low-dimensional model amenable to implementation in EHR systems as clinical decision support.

This study has a number of limitations. We chose to restrict patient history to information gathered from previous ED visits and anticipate that expanding the sources of historical data may improve model performance. Importing historical data from outpatient clinics or inpatient wards may present technical difficulties regarding EHR integration and differing standards of care. Furthermore, the data utilized in this study came from a hospital system that includes multiple emergency departments with a large catchment area. We anticipate difficulties extending this study to datasets from dense urban areas with multiple independent EDs,





**Fig 3. Variables from the full XGBoost model ordered by information gain.** Row names represent the variables in the design matrix post one-hot encoding (see [S1 Table](#) for name descriptions). Points represent the mean information gain from a hundred runs of XGBoost. Horizontal lines show bootstrapped 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0201016.g003>

given that patients may not consistently present to the same hospital system. Ongoing progress with inter-system information sharing presents one path forward, and this study highlights the importance of those efforts [49,50].

Throughout this study, we predict patient disposition by using the ED provider's prior decision as our true label. In doing so, we are unable to address the appropriateness of individual clinical decisions. This and similar studies would benefit from further research into a gold-standard metric for hospital admission. Studies have suggested that such a metric will remain elusive [51,52]. However, we expect that future work will align response variables of interest with patient-oriented outcomes to create a standardized metric for hospital admission. Future studies may try to adjust the response variable for discharged patients who returned to the ED the next day to get admitted and for admitted patients who in retrospect did not require admission.

Lastly, this study does not address the implementation and efficacy barriers present in clinical practice [53]. While we propose a low-dimensional model with the explicit intent of facilitating implementation into an EHR system, there is no uniform method by which clinical decision support tools are implemented. More systems-based research will be required to analyze methods of implementation and its effect on patient outcomes, with the ultimate goal of providing a standardized evaluation metric for prediction models.

## Supporting information

**S1 Table. Full list of variables.** Descriptions of all 972 variables included in the study. (XLSX)

**S2 Table. Variables ranked by information gain.** Mean information gain for top variables from 100 runs of XGBoost. (XLSX)

**S1 Fig. Patient census by ESI level.** (TIF)

**S1 Text. Model fitting protocol.** Details on the model fitting process, including the link to the de-identified dataset and R scripts. (DOCX)

## Author Contributions

**Conceptualization:** Woo Suk Hong, Adrian Daniel Haimovich, R. Andrew Taylor.

**Data curation:** Woo Suk Hong, R. Andrew Taylor.

**Formal analysis:** Woo Suk Hong.

**Investigation:** Woo Suk Hong.

**Methodology:** Woo Suk Hong, Adrian Daniel Haimovich, R. Andrew Taylor.

**Project administration:** R. Andrew Taylor.

**Resources:** R. Andrew Taylor.

**Software:** Woo Suk Hong.

**Supervision:** R. Andrew Taylor.

**Validation:** Woo Suk Hong.

**Visualization:** Woo Suk Hong.

**Writing – original draft:** Woo Suk Hong, Adrian Daniel Haimovich.

**Writing – review & editing:** Woo Suk Hong, Adrian Daniel Haimovich, R. Andrew Taylor.

## References

1. Moore B, Stocks C, Owens P. Trends in Emergency Department Visits, 2006–2014. HCUP Statistical Brief #227 [Internet]. Agency for Healthcare Research and Quality; 2017 Sep. Available: <https://www.hcup-us.ahrq.gov/reports/statbriefs/sb227-Emergency-Department-Visit-Trends.pdf>
2. Obermeyer Z, Emanuel EJ. Predicting the Future—Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med*. 2016; 375: 1216–1219. <https://doi.org/10.1056/NEJMp1606181> PMID: 27682033
3. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012; 98: 683–690. <https://doi.org/10.1136/heartjnl-2011-301246> PMID: 22397945
4. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLOS ONE*. 2017; 12: e0174708. <https://doi.org/10.1371/journal.pone.0174708> PMID: 28384212
5. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Med Inform*. 2016; 4. <https://doi.org/10.2196/medinform.5909> PMID: 27694098
6. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. *Acad Emerg Med*. 2016; 23: 269–78. <https://doi.org/10.1111/acem.12876> PMID: 26679719
7. Weng SF, Reys J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLOS ONE*. 2017; 12: e0174944. <https://doi.org/10.1371/journal.pone.0174944> PMID: 28376093
8. Haimovich JS, Venkatesh AK, Shojaee A, Coppi A, Warner F, Li S-X, et al. Discovery of temporal and disease association patterns in condition-specific hospital utilization rates. *PLOS ONE*. 2017; 12: e0172049. <https://doi.org/10.1371/journal.pone.0172049> PMID: 28355219
9. Sun Y, Heng BH, Tay SY, Seow E. Predicting hospital admissions at emergency department triage using routine administrative data. *Acad Emerg Med*. 2011; 18: 844–50. <https://doi.org/10.1111/j.1553-2712.2011.01125.x> PMID: 21843220
10. Bernstein SL, Aronsky D, Duseja R, Epstein S, Handel D, Hwang U, et al. The Effect of Emergency Department Crowding on Clinically Oriented Outcomes. *Acad Emerg Med*. 2009; 16: 1–10. <https://doi.org/10.1111/j.1553-2712.2008.00295.x> PMID: 19007346
11. Peck JS, Benneyan JC, Nightingale DJ, Gaehde SA. Predicting emergency department inpatient admissions to improve same-day patient flow. *Acad Emerg Med*. 2012; 19: E1045–54. <https://doi.org/10.1111/j.1553-2712.2012.01435.x> PMID: 22978731
12. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index. *Ann Emerg Med*. 2017; <https://doi.org/10.1016/j.annemergmed.2017.08.005> PMID: 28888332
13. Arya R, Wei G, McCoy JV, Crane J, Ohman-Strickland P, Eisenstein RM. Decreasing length of stay in the emergency department with a split emergency severity index 3 patient flow model. *Acad Emerg Med Off J Soc Acad Emerg Med*. 2013; 20: 1171–1179. <https://doi.org/10.1111/acem.12249> PMID: 24238321
14. Dugas AF, Kirsch TD, Toerper M, Korley F, Yenokyan G, France D, et al. An Electronic Emergency Triage System to Improve Patient Distribution by Critical Outcomes. *J Emerg Med*. 2016; 50: 910–918. <https://doi.org/10.1016/j.jemermed.2016.02.026> PMID: 27133736
15. Leegon J, Jones I, Lanaghan K, Aronsky D. Predicting Hospital Admission in a Pediatric Emergency Department using an Artificial Neural Network. *AMIA Annu Symp Proc*. 2006; 2006: 1004.
16. Leegon J, Jones I, Lanaghan K, Aronsky D. Predicting Hospital Admission for Emergency Department Patients using a Bayesian Network. *AMIA Annu Symp Proc*. 2005; 2005: 1022.
17. Lucini FR, S. Fogliatto F, C. da Silveira GJ, L. Neyeloff J, Anzanello MJ, de S. Kuchenbecker R, et al. Text mining approach to predict hospital admissions using early medical records from the emergency department. *Int J Med Inf*. 2017; 100: 1–8. <http://dx.doi.org/10.1016/j.ijmedinf.2017.01.001>

18. Savage DW, Weaver B, Wood D. P112: Predicting patient admission from the emergency department using triage administrative data. *Can J Emerg Med.* 2017; 19: S116–S116. <https://doi.org/10.1017/cem.2017.314>
19. Lucke JA, Gelder J de, Clarijs F, Heringhaus C, Craen AJM de, Fogteloo AJ, et al. Early prediction of hospital admission for emergency department patients: a comparison between patients younger or older than 70 years. *Emerg Med J.* 2017; emermed-2016-205846. <https://doi.org/10.1136/emered-2016-205846> PMID: 28814479
20. Barak-Corren Y, Israelit SH, Reis BY. Progressive prediction of hospitalisation in the emergency department: uncovering hidden patterns to improve patient flow. *Emerg Med J.* 2017; 34: 308–314. <https://doi.org/10.1136/emered-2014-203819> PMID: 28188202
21. Barak-Corren Y, Fine AM, Reis BY. Early Prediction Model of Patient Hospitalization From the Pediatric Emergency Department. *Pediatrics.* 2017; 139: e20162785. <https://doi.org/10.1542/peds.2016-2785> PMID: 28557729
22. Dinh MM, Russell SB, Bein KJ, Rogers K, Muscatello D, Paoloni R, et al. The Sydney Triage to Admission Risk Tool (START) to predict Emergency Department Disposition: A derivation and internal validation study using retrospective state-wide data from New South Wales, Australia. *BMC Emerg Med.* 2016; 16: 46. <https://doi.org/10.1186/s12873-016-0111-4> PMID: 27912757
23. Ebker-White AA, Bein KJ, Dinh MM. The Sydney Triage to Admission Risk Tool (START): A prospective validation study. *Emerg Med Australas EMA.* 2018; <https://doi.org/10.1111/1742-6723.12940> PMID: 29417732
24. Cameron A, Ireland AJ, McKay GA, Stark A, Lowe DJ. Predicting admission at triage: are nurses better than a simple objective score? *Emerg Med J.* 2017; 34: 2–7. <https://doi.org/10.1136/emered-2014-204455> PMID: 26864326
25. Cameron A, Jones D, Logan E, O’Keeffe CA, Mason SM, Lowe DJ. Comparison of Glasgow Admission Prediction Score and Amb Score in predicting need for inpatient care. *Emerg Med J.* 2018; 35: 247–251. <https://doi.org/10.1136/emered-2017-207246> PMID: 29444899
26. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Liu PJ, et al. Scalable and accurate deep learning for electronic health records. *Npj Digit Med.* 2018; 1. <https://doi.org/10.1038/s41746-018-0029-1>
27. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw.* 2008;028. Available: [https://econpapers.repec.org/article/jssjstsof/v\\_3a028\\_3ai05.htm](https://econpapers.repec.org/article/jssjstsof/v_3a028_3ai05.htm)
28. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *ArXiv160302754 Cs.* 2016; 785–794. <https://doi.org/10.1145/2939672.2939785>
29. Arnold TB. kerasR: R Interface to the Keras Deep Learning Library. In: *The Journal of Open Source Software [Internet].* 22 Jun 2017 [cited 2 Feb 2018]. <https://doi.org/10.21105/joss.00231>
30. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988; 44: 837–845. PMID: 3203132
31. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011; 12: 77. <https://doi.org/10.1186/1471-2105-12-77> PMID: 21414208
32. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950; 3: 32–35. PMID: 15405679
33. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and Optimal Cut-Point Estimated from Observations Affected by a Lower Limit of Detection. *Biom J Biom Z.* 2008; 50: 419–430. <https://doi.org/10.1002/bimj.200710415> PMID: 18435502
34. Kuhn M, Johnson K. Chapter 3. Data Pre-processing. *Applied Predictive Modeling.* Springer Science & Business Media; 2013. pp. 27–50.
35. Acuña E, Rodríguez C. The Treatment of Missing Values and its Effect on Classifier Accuracy. *Classification, Clustering, and Data Mining Applications.* Springer, Berlin, Heidelberg; 2004. pp. 639–647. [https://doi.org/10.1007/978-3-642-17103-1\\_60](https://doi.org/10.1007/978-3-642-17103-1_60)
36. Stekhoven DJ. missForest: Nonparametric missing value imputation using random forest. *Astrophys Source Code Libr.* 2015; ascl:1505.011.
37. Lin JH, Haug PJ. Exploiting missing clinical data in Bayesian network modeling for predicting medical problems. *J Biomed Inf.* 2008; 41: 1–14. <https://doi.org/10.1016/j.jbi.2007.06.001> PMID: 17625974
38. Junqué de Fortuny E, Martens D, Provost F. Predictive Modeling With Big Data: Is Bigger Really Better? *Big Data.* 2013; 1: 215–226. <https://doi.org/10.1089/big.2013.0037> PMID: 27447254
39. Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. *J Mach Learn Res.* 2003; 3: 1157–1182.

40. Poulin C, Shiner B, Thompson P, Vepstas L, Young-Xu Y, Goertzel B, et al. Predicting the Risk of Suicide by Analyzing the Text of Clinical Notes. *PLOS ONE*. 2014; 9: e85733. <https://doi.org/10.1371/journal.pone.0085733> PMID: 24489669
41. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc*. 2016; 23: 1007–1015. <https://doi.org/10.1093/jamia/ocv180> PMID: 26911811
42. Hohl CM, Dankoff J, Colacone A, Afilalo M. Polypharmacy, adverse drug-related events, and potential adverse drug interactions in elderly patients presenting to an emergency department. *Ann Emerg Med*. 2001; 38: 666–671. <https://doi.org/10.1067/mem.2001.119456> PMID: 11719747
43. Banerjee A, Mbamalu D, Ebrahimi S, Khan AA, Chan TF. The prevalence of polypharmacy in elderly attenders to an emergency department—a problem with a need for an effective solution. *Int J Emerg Med*. 2011; 4: 22. <https://doi.org/10.1186/1865-1380-4-22> PMID: 21635734
44. Tanabe P, Gimbel R, Yarnold PR, Kyriacou DN, Adams JG. Reliability and validity of scores on The Emergency Severity Index version 3. *Acad Emerg Med*. 2004; 11: 59–65. PMID: 14709429
45. Baumann MR, Strout TD. Triage of geriatric patients in the emergency department: validity and survival with the Emergency Severity Index. *Ann Emerg Med*. 2007; 49: 234–240. <https://doi.org/10.1016/j.annemergmed.2006.04.011> PMID: 17141145
46. Selassie AW, McCarthy ML, Pickelsimer EE. The Influence of Insurance, Race, and Gender on Emergency Department Disposition. *Acad Emerg Med*. 2003; 10: 1260–1270. [https://doi.org/10.1197/S1069-6563\(03\)00497-4](https://doi.org/10.1197/S1069-6563(03)00497-4) PMID: 14597503
47. Jiang HJ, Russo CA, Barrett ML. Nationwide Frequency and Costs of Potentially Preventable Hospitalizations, 2006. HCUP Statistical Brief #72 [Internet]. Agency for Healthcare Research and Quality; 2009 Apr. Available: <http://www.hcup-us.ahrq.gov/reports/statbriefs/sb72.pdf>
48. Kangovi S, Barg FK, Carter T, Long JA, Shannon R, Grande D. Understanding Why Patients Of Low Socioeconomic Status Prefer Hospitals Over Ambulatory Care. *Health Aff (Millwood)*. 2013; 32: 1196–1203. <https://doi.org/10.1377/hlthaff.2012.0825> PMID: 23836734
49. Shapiro JS, Kannry J, Lipton M, Goldberg E, Conocenti P, Stuard S, et al. Approaches to Patient Health Information Exchange and Their Impact on Emergency Medicine. *Ann Emerg Med*. 2006; 48: 426–432. <https://doi.org/10.1016/j.annemergmed.2006.03.032> PMID: 16997679
50. Shapiro JS, Johnson SA, Angiolillo J, Fleischman W, Onyile A, Kuperman G. Health Information Exchange Improves Identification Of Frequent Emergency Department Users. *Health Aff (Millwood)*. 2013; 32: 2193–2198. <https://doi.org/10.1377/hlthaff.2013.0167> PMID: 24301405
51. Sabbatini AK, Kocher KE, Basu A, Hsia RY. In-Hospital Outcomes and Costs Among Patients Hospitalized During a Return Visit to the Emergency Department. *JAMA*. 2016; 315: 663. <https://doi.org/10.1001/jama.2016.0649> PMID: 26881369
52. Adams JG. Ensuring the Quality of Quality Metrics for Emergency Care. *JAMA*. 2016; 315: 659. <https://doi.org/10.1001/jama.2015.19484> PMID: 26881367
53. Schenkel SM, Wears RL. Triage, Machine Learning, Algorithms, and Becoming the Borg. *Ann Emerg Med*. 2018;0. <https://doi.org/10.1016/j.annemergmed.2018.02.010> PMID: 29530655