








ORIGINAL RESEARCH

General Medicine

Automated analysis of unstructured clinical assessments improves emergency department triage performance: A retrospective deep learning analysis

Dana R. Sax MD, MPH¹  | E. Margaret Warton MPH²  | Oleg Sofrygin PhD³ |
Dustin G. Mark MD¹  | Dustin W. Ballard MD, MBE⁴  | Mamata V. Kene MD, MPH⁴  |
David R. Vinson MD⁵  | Mary E. Reed DrPH² 

¹Department of Emergency Medicine, Kaiser East Bay and Kaiser Permanente Northern, California Division of Research, Oakland, California, USA

²Kaiser Permanente Northern California Division of Research, Oakland, California, USA

³Uber, San Francisco, California, USA

⁴Department of Emergency Medicine, Kaiser San Rafael and Kaiser Permanente Northern California Division of Research, Oakland, California, USA

⁵Department of Emergency Medicine, Roseville, and Kaiser Permanente Northern California Division of Research, Oakland, California, USA

Correspondence

Dana R. Sax, MD, MPH, Department of Emergency Medicine, Kaiser Permanente East Bay and Kaiser Permanente Northern California Division of Research, Oakland, CA 94611, USA.

Email: dana.r.sax@kp.org

Funding information

Kaiser Permanente Lokahi Risk Reduction Funding Initiative, Large Project

Abstract

Objectives: Efficient and accurate emergency department (ED) triage is critical to prioritize the sickest patients and manage department flow. We explored the use of electronic health record data and advanced predictive analytics to improve triage performance.

Methods: Using a data set of over 5 million ED encounters of patients 18 years and older across 21 EDs from 2016 to 2020, we derived triage models using deep learning to predict 2 outcomes: hospitalization (primary outcome) and fast-track eligibility (exploratory outcome), defined as ED discharge with <2 resource types used (eg, laboratory or imaging studies) and no critical events (eg, resuscitative medications use or intensive care unit [ICU] admission). We report area under the receiver operator characteristic curve (AUC) and 95% confidence intervals (CI) for models using (1) triage variables alone (demographics and vital signs), (2) triage nurse clinical assessment alone (unstructured notes), and (3) triage variables plus clinical assessment for each prediction target.

Results: We found 12.7% of patients were hospitalized ($n = 673,659$) and 37.0% were fast-track eligible ($n = 1,966,615$). The AUC was lowest for models using triage variables alone: AUC 0.77 (95% CI 0.77–0.78) and 0.70 (95% CI 0.70–0.71) for hospitalization and fast-track eligibility, respectively, and highest for models incorporating clinical assessment with triage variables for both hospitalization and fast-track eligibility: AUC 0.87 (95% CI 0.87–0.87) for both prediction targets.

Conclusion: Our findings highlight the potential to use advanced predictive analytics to accurately predict key ED triage outcomes. Predictive accuracy was optimized when clinical assessments were added to models using simple structured variables alone.

Supervising Editor: Adam Landman, MD, MS.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *JACEP Open* published by Wiley Periodicals LLC on behalf of American College of Emergency Physicians.

1 | INTRODUCTION

1.1 | Background

Increases in emergency department (ED) visits and intensity of care (eg, number of tests, imaging, and medications ordered) have contributed to crowding, which has negatively affected the quality of care. ED triage, or the sorting of patients on arrival based on predicted acuity and resource needs, is essential to ensure the sickest patients get the time-sensitive care they need while prioritizing the remaining patients to optimize operational flow. Accurate triage in times of overcrowding is critical to ensure safe and timely care.

Prior studies have found that the most commonly used triage system in the United States, the 5-level Emergency Severity Index (ESI), has limited accuracy and reliability.¹⁻⁶ The ESI was developed in 1999³ and is used in over 70% of EDs across the United States.⁷ Several studies have found notable disparities in triage accuracy by patient characteristics when the ESI was used,^{8,9} potentially related to the subjective nature of the ESI algorithm's branch points.

Recent studies have shown that the use of advanced predictive analytics can improve the accuracy of triage predictions⁸⁻²¹ and that implementation of such machine learning (ML)-based models is feasible.^{10,11} Early ML-based triage models have used structured variables from the electronic health record (EHR) to predict triage outcomes with a high degree of accuracy.¹⁰⁻¹⁹ Other models incorporated unstructured variables to predict triage outcomes: natural language processing to the patient's chief complaint,^{20,21} text mining methods to process data from early ED patient records,²² and textual information about currently used medications and medical/laboratorial exam descriptions.²³ As far as we know, none have incorporated triage nurse free-text clinical assessments into these predictions. It is unknown how electronic triage models that combine a documented brief clinical assessment with structured variables might affect predictive accuracy.

1.2 | Importance

An important goal of effective triage is to accurately identify the sickest patients to limit the underrecognition of significant diseases and associated delays in care. As such, most ML-based prediction models have used hospital or intensive care unit (ICU) admission as their prediction targets.^{10,11,14-17,20,23} Some have included other critical care outcomes^{10,12} or hospital transfer.²⁴ Early triage models using ML methods demonstrated superior discrimination compared to simple rule-based triage tools in predicting hospital or ICU admission.^{10,11}

A secondary goal of triage is to sort patients based on their predicted resource needs to help decide when and where different patients should be seen. Although hospitalization likelihood has been shown to correlate with ED resource use,²⁵ there is an opportunity to better discriminate among the 80%-90% of ED patients who are not admitted or transferred.²⁶ Triage systems would ideally safely identify true low-risk patients who can be treated in urgent care or fast-track type settings to help reduce main ED crowding. Interventions that

The Bottom Line

Novel triage models have the potential to help with emergency department flow. This study developed and validated deep learning models to predict hospital admission and fast-track eligible patients, finding the highest predictive accuracy when nurses' free-text clinical assessment was combined with traditional, structured triage variables.

efficiently separate out low-acuity patients early have been shown to decrease ED length of stay without negatively affecting care quality.²⁷

1.3 | Goals of this investigation

In this article, we present findings from the development of triage models that use deep learning to predict 2 important outcomes among adult ED patients. Our primary outcome is hospitalization, and our exploratory outcome is fast-track eligibility. We define "fast-track eligible" as 0-1 resource types used (including laboratory tests, imaging, non-oral medications, ECGs, specialty consultations, and procedures) and who are discharged home with no critical events. We also explore whether model incorporation of triage nurses' clinical assessments via unstructured notes might improve performance.

2 | METHODS

2.1 | Study design

We performed a retrospective cohort study of all ED encounters by adult patients during the study period. We describe patient characteristics and primary outcomes among the study population and use ML methods to predict 2 key outcomes: hospitalization and fast-track eligibility.

2.2 | Setting

This study was conducted in Kaiser Permanente Northern California (KPNC), an integrated health care delivery system with 21 hospital-based EDs that provides comprehensive medical care for more than 4 million members who are representative of the ethnic and socioeconomic diversity of the surrounding population.²⁸ Each year, patients make over 1.3 million visits to KPNC EDs. Annual ED volumes range from 28,000 to 128,000 across sites (mean of 60,611 encounters), and there is significant variation in specialty care services available. Although there are local nuances in implementation, all EDs in our system use ED nurses to triage all patients, and all EDs designate at least 3 types of care spaces: fast-track (for low-acuity patients), main ED (for

patients needing urgent and emergent evaluation and treatment), and resuscitation rooms (for critically ill patients). All EDs in our study use the ESI, the triage system used in over 70% of EDs across the United States.³ KPNC uses the Epic Systems EHR, Verona, WI.

2.3 | Selection of participants

We developed a study dataset of all ED encounters to any of the 21 EDs within KPNC between 2016 and 2020 by patients 18 years and older and extracted study data from the EHR. Encounters were excluded if the patient left against medical advice or before ED physician or advanced practice provider evaluation.

2.4 | Measurements

We describe patient sociodemographic and clinical characteristics. We collected patient demographic information, including age, gender, primary language, and race or ethnicity from EHR databases and neighborhood socioeconomic status at the census block group level using 2010 US Census data. We collected prior ED, inpatient, and ICU use from the EHR. We ascertained information on coexisting illnesses based on diagnoses using *International Classification of Diseases, Tenth Revision* (ICD-10) codes. For each patient, we obtained an internally derived and validated comorbidity risk score (Comorbidity Point Score, version 2 [COPS2]).²⁹ We assessed the ESI assignment of each patient, the number and types of resources used, the occurrence of critical outcomes, and the final ED disposition (coded as admission or discharge).

We used the following triage variables in the models: age, sex, triage vital signs (blood pressure, heart rate, oxygen saturation, respiratory rate, and temperature), and triage nurse clinical notes. Although triage nurses must complete certain fields in the EHR, including vital signs and a chief complaint (structured data), they have the option to add a brief note about the patient's presentation in a free text field. These clinical assessments are typed in manually without use of standard templates or copied blocks of text. The triage nurse clinical note is the brief note about the patient's presentation in a free text field; up to 4 chief complaints can be added and there is 1 free text field for each entered chief complaint.

2.5 | Outcomes

We developed separate models for the 2 prediction targets. Our primary outcome was hospitalization, and our exploratory outcome was fast-track eligibility. Hospitalization was broadly defined to include multiple levels of further inpatient care, including admission to observation units, medical or surgical units, telemetry units, step-down units, ICUs, operating room or catheterization labs, labor and delivery units, psychiatric acute care units, and direct transfers to another acute care hospital. Fast-track eligible patients were defined as those who were

treated and discharged home with limited resource use. Specifically, this was defined as 0–1 resource type used and no hospitalization or critical events during the ED stay. Consistent with the ESI, resource types were defined as laboratory tests, imaging, non-oral medications, ECGs, specialty consultations, and procedures.⁷

Our study team of 4 emergency physicians and 1 dual emergency/critical care physician developed a hierarchical list of critical, time-sensitive events to define under triage for lower ESI scores (ESI II–V).¹⁰ These critical outcomes were similar, although more expansive, than critical events included in other ML-based triage models.^{10,12} Our list included advanced respiratory support; resuscitation and life-stabilizing medications; early blood transfusions; admissions to ICU, cardiac catheterization procedure suite, or operating room; or transfers to outside facilities. A full list of these critical events can be found in the [Appendix, Table S1](#). All critical events were captured in the EHR and were drawn from various sources, including procedure codes, blood product transfusion data, laboratory analysis, and radiology studies. To ensure we were identifying truly low-risk, fast-track eligible patients, the occurrence of any critical event excluded patients from meeting the fast-track eligibility outcome.

2.6 | Data analyses

We tested 3 sets of predictors for each outcome (a total of 6 models). We were interested to explore whether models that used a few universally available triage variables could accurately predict clinically meaningful outcomes. This goal informed the simplest model we tested, which included only electronically available triage data: age, sex, and triage vital signs (blood pressure, heart rate, oxygen saturation, respiratory rate, and temperature). The second model used only the triage nurses' clinical assessment, which nurses enter as free text in a triage data entry field. The third model combined the inputs of the first and second models (structured triage variables and unstructured triage nurse notes). All variables, including triage nurse notes, were collected from the EHR.

Before training the deep neural network on plain text triage notes, we had to process the notes into a suitable machine-readable digital format. First, the plain text from 6 triage comment fields was concatenated into a single plain text note. These included up to 6 chief complaints and up to 6 free text notes accompanying each chief complaint. Next, we changed all text (free text and chief complaints) to lower-case and removed any punctuation, special characters, and extra spaces. Next, we created a training set "vocabulary," whereby we mapped each unique word (across all reports) into a unique integer, $1, \dots, n$, where n represents the total number of unique words that appear across all triage reports ($n = 129,500$). The vocabulary allowed us to map each triage report from a sequence of words into a corresponding sequence of integers, thus making the plain text suitable for machine input. The resulting sequence of integers is referred to as a "tokenized" report, wherein each integer (representing each word from the original report) is referred to as a "token." Finally, we standardized all reports by making them the same length. Specifically, we truncated

any report that was longer than 40 words (approximately 1.3% of triage reports) and padded all reports shorter than 40 words with a special token "0" that was reserved just for padding (not used for any other words in the triage report).

We used median imputation for continuous variables if there were any missing and added a flag to indicate observations with imputed values.⁴ We included these flags in the models if they were significant. For categorical variables with >0.5% missing, we created a separate "missing" category.

We divided the study cohort into an 80% training set and a 20% testing set. We employed deep learning for all model training. Specifically, we employed a multilayer perceptron (MLP) neural network to develop predictions from structured predictors (electronic triage variables, chief complaint, age, and sex) and long short-term memory (LSTM) neural network for processing unstructured free-text predictors (triage nurse notes).³⁰

For the third model approach (structured and unstructured predictors), we concatenated the last layer outputs of the LSTM and MLP neural networks and passed them through another neural network layer. In all 3 approaches, we used the sigmoid activation function to output the probability predictions. The sigmoid activation function ensures that the output of the neural network is a number between 0 and 1, which can be interpreted as the probability of a positive class label.

We used SAS, Version 9.4, Cary, NC, to build the data set, logistic regression to generate descriptive statistics, and PyTorch and Python to build and train our deep learning model, with graphical processing units-based neural network training and testing. We reserved 10% of the training data for hyperparameter tuning (number of MLP layers, hidden size of LSTM). All final area under the receiver operator characteristic curve (AUC) results were reported on the random 20% testing set. We report test data AUC and 95% confidence intervals (CI) for each of the 6 models, and sensitivity, specificity, positive and negative predictive variables for the 3 models at 3 different thresholds that predict our primary outcome of hospitalization.

3 | RESULTS

The study cohort consisted of 5,315,176 encounters, of which 4,252,141 were included in the training cohort and 1,063,035 in the testing cohort. Table 1 displays patient characteristics.

The mean age was 52 years, 56% were female, 44% were non-Hispanic White, 21% were Hispanic, 15% were Black, 11% were Asian, and 9% were other or multirace. Approximately 81% of encounters were made by patients with KPNC health plan membership. Table S2 in the Appendix displays the frequency of resource use in our overall study population and by ESI. Almost two thirds of ED encounters used at least 2 types of resources. In general, ESI assignment correlated with resource needs, with >90% of ESI I and II encounters and <10% of ESI IV and V encounters using at least 2 resource types.

There were 673,659 patients who met our hospitalization outcome (12.7%), and 1,966,615 who met our fast-track eligibility outcome

(37.0%). We found 3,262,047 encounters (61.4%) were assigned a midlevel triage category (ESI III). Of these ESI III encounters, we estimated that 913,373 (28%) were fast-track eligible.

There was substantial variation in the content, format, and style of triage nurse clinical notes. In about 15% of encounters, no notes were entered. See Table 2 for examples of triage notes. We found triage vital signs were missing in 0.6% of encounters, and outcome rates among encounters with missing triage vital signs were statistically significantly different than outcome rates without missing vital signs, so a missing flag was used.

Table 3 shows model performance for both prediction targets (hospitalization and fast-track eligibility) comparing the 3 different models. Predictive accuracy was high for both prediction targets in all models. The highest AUC was achieved for both hospitalization and fast-track eligibility (AUC of 0.87, 95% CI 0.87–0.87) when we included structured and unstructured triage data. For our primary outcome of hospitalization, the sensitivity, specificity, positive, and negative predictive values vary depending on the chosen threshold. Table 4 shows these values at different hospital admission risk strata for each model and shows that specificity and positive predictive values substantially increase in model 3 (compared to models 1 and 2) at each hospital admission risk stratum.

4 | LIMITATIONS

Our findings may be less generalizable to health systems with less comprehensive EHRs or different thresholds for considering hospitalization or fast-track eligibility. Fast-track eligibility is an exploratory prediction target and there may be substantial clinician and institutional variation and agreement on which patients are fast-track appropriate. The content and quality of nurse documentation were variable in our study cohort, and greater variation in other health systems may limit generalizability to those settings. It is possible that documentation of triage nurse clinical assessments varies by patient characteristics, and this could contribute to disparities in triage predictions. For example, patients who speak languages with limited interpreter services may have limited or missing assessments. About 20% of encounters in our study cohort were by patients who did not have KPNC health plan membership and generally have less detailed EHR history. Although we plan to explore using greater EHR data to inform triage predictions in future models, this preliminary work using only universally available triage data suggests triage predictions can be accurate even when more comprehensive data are not available.

For this exploratory analysis, we did not analyze the words or phrases in the nursing triage notes that had the most significant impact on triage predictions. Although variable importance lists are possible with neural networks, prior work has suggested the results can be misleading.³¹ Also, given fast-track eligibility is an exploratory prediction target that may have less acceptance and greater institutional variability compared to hospitalization, we did not report additional measures of model performance for fast-track eligibility models in this study.

TABLE 1 Patient characteristics of the study cohort, including 5,315,176 ED encounters from 2016 to 2020 across 21 EDs.

Patient characteristic		N in sample (%)
Age, years	18–29	1,015,662 (19.1)
	30–39	815,214 (15.3)
	40–49	707,641 (13.3)
	50–59	785,815 (14.8)
	60–69	733,716 (13.8)
	70–79	622,107 (11.7)
	80 and older	635,021 (12.0)
Gender	Female	2,962,867 (55.7)
Race or ethnicity	Asian	590,576 (11.1)
	Black	800,984 (15.1)
	Hispanic	1,137,459 (21.4)
	Non-Hispanic White	2,336,058 (44.0)
	Other/multirace	450,099 (8.5)
English is primary language	Yes	4,868,004 (91.6)
Patient arrived by ambulance	Yes	933,633 (17.6)
KPNC health plan membership	Yes	4,284,032 (80.6)
ESI level	I	33,491 (0.6)
	II	929,555 (18.1)
	III	3,262,047 (61.4)
	IV	104,806 (19.7)
	V	43,277 (0.8)
COPS2 comorbidity scores*	Low (<20)	2,655,487 (50.0)
	Medium (20–<65)	1,976,329 (37.2)
	High (>=65)	683,360 (12.9)
Recent health care use	Any hospitalization	217,835 (4.1)
	Any intensive care	41,512 (0.8)
	Any ED encounter	734,264 (13.8)

Notes: Other/multirace included American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, or multiple races or ethnicities, unknown, or missing race or ethnicity. Percentages for age, race, and COPS2 score represent column percentages within each respective type of characteristic (age, race, or COPS2). Other percentages represent the absolute percentage of encounters with these variables within each row and column. Please note that only age and gender (in addition to triage vital signs, chief complaint, and triage nurse free text) were included in the models.

Abbreviations: COPS2, Comorbidity Point Score, a longitudinal comorbidity score based on 12 months of patient data¹; ED, emergency department; ESI, Emergency Severity Index; KPNC, Kaiser Permanente Northern California.

5 | DISCUSSION

We present data exploring use of novel targets and incorporation of triage nurses' clinical assessments to make triage predictions using deep learning. Using a large sample of over 5 million ED encounters across 21 EDs, we found that simple models using readily available electronic triage data (age, sex, and triage vital signs) can accurately identify 2 important groups of ED patients: those likely needing hospitalization and those who can be treated in fast-track spaces. As far as we know, our study is the first to explore incorporation of unstructured triage clinician notes in triage prediction models. Nursing assessment

was highly predictive of both hospitalization and fast-track eligibility, highlighting the critical value of clinician gestalt in predicting acuity and resource needs. Our findings suggest that prediction models may be ideally used to assist and support triage nurses in determining triage assignments.

Ideally, a triage system can accurately identify both the sickest patients who need to be evaluated first, while simultaneously separating out true low-acuity patients who may be treated in a less emergent setting. A large multicenter study estimated the ESI to be accurate in <50% of high acuity cases, with no difference in overall accuracy according to nurse experience.² Using novel measures to define

TABLE 2 Examples of triage nurses' free text clinical assessments.

Chief complaint (structured data)	Triage nurse note (unstructured data)
Asthma	Onset yesterday. Pt neb/inhaler not effective. Dry cough noted. Afebrile
Chest pain	Pt c/o chest pain, pt sts pain started 2 days ago. Pt denies trauma, heavy lifting. Pt sts he does not recall what he was doing when chest pain started. Pt sts he laid down in bed but no help. Pt sts slight sob, no acute distress noted at this time.
Leg pain	Fell off the couch at 11:00 a.m., leg was straight. Took Tylenol after that, iced it. Can't walk.
Hip pain	Fall off scooter. 7/10
Accident	Pt fell out of a window of second story apartment building. Pt is acting appropriately in triage. No nausea/vomiting.
Referral	Here for transfusion. Chemo patient.
Lower abdominal pain	Lower abd. pain/cramping started an hour ago, with nausea. Pt currently on period at this time. States pain similar when on previous periods. Denies urinary symptoms.

Note: In approximately 15% of encounters, there was no unstructured triage nurse note entered. In these encounters, the chief complaint (structured text) was added automatically in the triage nurse note field.

TABLE 3 Area under the receiver operator characteristic curve (AUC) for models predicting hospitalization and fast-track eligibility.

Outcome	Model/variables used	AUC (95% confidence interval)
Hospitalization	1. Triage nurse clinical notes alone	0.80 (0.80–0.81)
	2. Structured triage data: age, sex, triage vital signs	0.77 (0.77–0.78)
	3. Triage nurse clinical notes plus structured triage data	0.87 (0.87–0.87)
Fast-track eligible (ED discharge home with <2 resource types and no critical events)	4. Triage nurse clinical notes alone	0.84 (0.83–0.84)
	5. Structured triage data: age, sex, triage vital signs	0.70 (0.70–0.71)
	6. Triage nurse clinical notes plus structured triage data	0.87 (0.87–0.87)

Abbreviation: ED, emergency department.

TABLE 4 Test characteristics of each model among various hospital admission risk strata: Sensitivity, specificity, positive, predictive value, and negative predictive values.

Hospital admission threshold	Model 1: triage nurse clinical notes alone	Model 2: structured triage data: age, sex, triage vital signs	Model 3: triage nurse clinical notes plus structured triage data
10%	Sens: 94.99%	Sens: 99.64%	Sens: 92.94%
	Spec: 35.88%	Spec: 7.39%	Spec: 57.48%
	PPV: 15.22%	PPV: 11.50%	PPV: 20.94%
	NPV: 98.33%	NPV: 99.41%	NPV: 98.53%
30%	Sens: 32.37%	Sens: 25.50%	Sens: 32.19%
	Spec: 95.44%	Spec: 96.39%	Spec: 97.66%
	PPV: 46.23%	PPV: 46.12%	PPV: 62.52%
	NPV: 92.09%	NPV: 91.44%	NPV: 92.24%
50%	Sens: 9.38%	Sens: 10.47%	Sens: 9.46%
	Spec: 99.31%	Spec: 99.27%	Spec: 99.66%
	PPV: 62.12%	PPV: 63.59%	PPV: 76.88%
	NPV: 90.04%	NPV: 90.15%	NPV: 90.08%

Abbreviations: NPV, negative predictive value; PPV, positive predictive value; sens, sensitivity; spec, specificity.

under- and overtriage for each ESI level, we recently reported that mistriage with ESI occurs in over 30% of encounters, with overtriage much more common than under triage.⁴

Early ML-based triage prediction models demonstrated superior ability to identify high-acuity patients compared to ESI.^{10,25} The AUC in our hospitalization models using only universally available triage variables and deep learning is comparable or slightly lower than those reported in earlier models^{10,11} that used a significantly greater number of features and random forest or gradient boosted methods. Adding triage nurse clinical assessments to the prediction model led to a similar or higher AUC for predicting hospitalization compared to results reported in earlier models.^{10,11}

Although we were not able to directly compare model results in our study with predictive accuracy of the ESI because of distinct prediction targets (the ESI predicts 5 levels of patient acuity and resource needs, whereas our models predict hospital admission and fast-track eligibility), our findings suggest ML-models offer an opportunity for improvement. In prior work, we found the sensitivity and specificity of ESI to predict low-acuity, low resource-needs patients (correctly assigning an ESI IV or V among patients who used <2 resources and had no critical interventions) were 50.0% and 96.8%, respectively, whereas the sensitivity and specificity of ESI to predict a critically ill patient (defined as correctly assigning an ESI I or II among patients with a critical care intervention) were 65.9% and 83.4%, respectively.⁴ In this study, we found that deep learning methods could achieve high sensitivity at lower hospital admission risk thresholds (>90% in all models at 10% admission threshold) and high specificity at higher hospital risk thresholds (>95% in each model at 30% and 50% risk thresholds). There were notable improvements in specificity and positive predictive value at each risk threshold in model 3, again highlighting the added value of triage nurse clinical notes in predicting hospital admission. Our findings of sensitivity and specificity are comparable to ML-based hospitalization prediction models reported in earlier studies.^{13,17}

A significant drawback of the ESI is its limited ability to discriminate among the many midacuity patients. Similar to other studies,¹⁰ we found the majority of patients in our study were assigned a midacuity (ESI III) level. We estimate that nearly one-third of these patients were low-acuity and low resource-need patients who were discharged home from the ED without need for critical interventions. A model predicting fast-track eligibility could be used to identify these low-acuity patients accurately and efficiently at triage to reduce main ED crowding and improve patient flow.

We view our study findings as important preliminary work that has potential for significant clinical impact. Next steps in this work will include studying the incremental benefit of including additional predictor variables in models (eg, patient comorbidity, health care use, and pharmacy data), comparison of different modeling methods, and prospective validation. Further, the feasibility of incorporating triage nurse notes in real-time needs to be explored. Multiple recent studies have applied deep learning methods with LSTM neural networks for prediction of clinical targets. These studies suggest use of clinician notes for real-time triage predictions is potentially feasible.³²

In addition, the optimal way to include multiple prediction targets (ie, hospitalization and fast-track eligibility) to assist with triage determinations at the point of care would need to be considered. Work by Levin et al provides a framework for combining separate models with different prediction targets to assign triage levels defined by likelihood of each outcome.¹⁰ Significant stakeholder engagement to develop and refine the triage clinician user interface would be needed. The potential impacts of implementing novel triage models on patient outcomes, equity, and resource use would need to be investigated during prospective evaluation.

In this study, we explored the use of deep learning to improve triage predictions using novel prediction targets and incorporating unstructured nurse triage notes. We found hospitalization and fast-track eligibility can be accurately predicted using universally available triage variables and that incorporation of nursing assessments significantly improved model discrimination.

AUTHOR CONTRIBUTIONS

Dana R. Sax, Dustin G. Mark, Dustin W. Ballard, Mamata V. Kene, David R. Vinson, and Mary E. Reed conceived of the study and obtained research funding. Dana R. Sax and Mary E. Reed supervised the conduct of the study and data collection. E. Margaret Warton and Oleg Sofrygin managed the data and provided statistical advice and analyzed the data. Dana R. Sax and Mary E. Reed led data quality control. Dana R. Sax drafted the manuscript, and all authors contributed substantially to its revision. Dana R. Sax takes responsibility for the paper as a whole.

CONFLICT OF INTEREST STATEMENT

All authors report no conflicts of interest.

ORCID


Dana R. Sax MD, MPH  <https://orcid.org/0000-0002-9061-7746>

E. Margaret Warton MPH  <https://orcid.org/0000-0001-6613-3947>

Dustin G. Mark MD  <https://orcid.org/0000-0002-5001-7228>

Dustin W. Ballard MD, MBE  <https://orcid.org/0000-0002-5801-8794>

Mamata V. Kene MD, MPH  <https://orcid.org/0000-0001-7493-8438>

David R. Vinson MD  <https://orcid.org/0000-0001-6559-1858>

Mary E. Reed DrPH  <https://orcid.org/0000-0002-8203-1016>

REFERENCES

1. Martin A, Davidson CL, Panik A, Buckenmyer C, Delpais P, Ortiz M. An examination of ESI triage scoring accuracy in relationship to ED nursing attitudes and experience. *J Emerg Nurs*. 2014;40(5):461-468.
2. Mistry B, Stewart De Ramirez S, Kelen G, et al. Accuracy and reliability of emergency department triage using the emergency severity index: an international multicenter assessment. *Ann Emerg Med*. 2018;71(5):581-587.e3.
3. McHugh M, Tanabe P, McClelland M, Khare RK. More patients are triaged using the emergency severity index than any other triage acuity system in the United States. *Acad Emerg Med*. 2012;19(1):106-109.
4. Sax DR, Warton EM, Mark DG, et al. Evaluation of the emergency severity index in US emergency departments for the rate of mistriage. *JAMA Network Open*. 2023;6(3):e233404.

5. Ghafarypour-Jahrom M, Taghizadeh M, Heidari K, Derakhshanfar H. Validity and reliability of the emergency severity index and Australasian triage system in pediatric emergency care of Mofid Children's Hospital in Iran. *Bull Emerg Trauma*. 2018;6(4):329-333.
6. Hinson JS, Martinez DA, Schmitz PSK, et al. Accuracy of emergency department triage using the emergency severity index and independent predictors of under-triage and over-triage in Brazil: a retrospective cohort analysis. *Int J Emerg Med*. 2018;11(1):3.
7. Agency for Healthcare Research and Quality. Emergency Severity Index (ESI): a triage tool for emergency department. May 2020. Accessed March 16, 2023. Available from: <https://www.ahrq.gov/patient-safety/settings/emergency-dept/esi.html>
8. Dennis JA. Racial/Ethnic Disparities in Triage Scores Among Pediatric Emergency Department Fever Patients. *Pediatr Emerg Care*. Published online first 3/2/2020.
9. Zook HG, Kharbanda AB, Flood A, Harmon B, Puumala SE, Payne NR. Racial differences in pediatric emergency department triage scores. *J Emerg Med*. 2016;50(5):720-727.
10. Levin S, Toerper M, Hamrock E, 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*. 2018;71(5):565-574.e2.
11. Fenn A, Davis C, Buckland DM, et al. Development and validation of machine learning models to predict admission from emergency department to inpatient and intensive care units. *Ann Emerg Med*. 2021;78(2):290-302.
12. Goto T, Camargo CA, Faridi MK, Freishtat RJ, Hasegawa K. Machine learning-based prediction of clinical outcomes for children during emergency department triage. *JAMA Netw Open*. 2019;2(1):e186937. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6484561/>. [Internet]. 2019 Jan 11 [cited 2021 Jan 15].
13. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care*. 2019;23(1):64.
14. Araz OM, Olson D, Ramirez Nafarrate A. Predictive analytics for hospital admissions from the emergency department using triage information. *Int J Prod Econ*. 2019;208:199-207.
15. Barak-Corren Y, Agarwal I, Michelson KA, et al. Prediction of patient disposition: comparison of computer and human approaches and a proposed synthesis. *J Am Med Inform Assoc*. 2021;28(8):1736-1745. ocab076.
16. Handy N, Thompson DA, Li J, Chuirazzi DM, Venkat A. Evaluation of a hospital admission prediction model adding coded chief complaint data using neural network methodology. *Eur J Emerg Med*. 2015;22(2):87-91.
17. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLoS One*. 2018;13(7):e0201016.
18. 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.
19. Sun Y, Heng BH, Tay SY, Seow E. Predicting hospital admissions at emergency department triage using routine administrative data. *Acad Emerg Med*. 2011;18(8):844-850.
20. Fernandes M, Mendes R, Vieira SM, et al. Predicting intensive care unit admission among patients presenting to the emergency department using machine learning and natural language processing. *PLoS One*. 2020;15(3):e0229331.
21. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Med Inform Decis Mak*. 2020;20(1):280.
22. Lucini FR, Fogliatto FS, da Silveira GJC, et al. Text mining approach to predict hospital admissions using early medical records from the emergency department. *Int J Med Inform*. 2017;100:1-8.
23. Roquette BP, Nagano H, Marujo EC, Maiorano AC. Prediction of admission in pediatric emergency department with deep neural networks and triage textual data. *Neural Netw*. 2020;126:170-177.
24. Sterling NW, Patzer RE, Di M, Schragger JD. Prediction of emergency department patient disposition based on natural language processing of triage notes. *Int J Med Inform*. 2019;129:184-188.
25. Dugas AF, Kirsch TD, Toerper M, et al. An electronic emergency triage system to improve patient distribution by critical outcomes. *J Emerg Med*. 2016;50(6):910-918.
26. Lin MP, Baker O, Richardson LD, Schuur JD. Trends in emergency department visits and admission rates among US Acute Care Hospitals. *JAMA Intern Med*. 2018;178(12):1708-1710.
27. Wiler JL, Ozkaynak M, Bookman K, et al. Implementation of a front-end split-flow model to promote performance in an Urban Academic Emergency Department. *Jt Comm J Qual Patient Saf*. 2016;42(6):271-280.
28. Gordon N, Lin T. The Kaiser Permanente Northern California adult member health survey. *Perm J*. 2016;20(4):215-225.
29. Escobar GJ, Ragins A, Scheirer P, Liu V, Robles J, Kipnis P. Nonelective rehospitalizations and postdischarge mortality: predictive models suitable for use in real time. *Med Care*. 2015;53(11):916-923.
30. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735-1780.
31. Hooker G, Mentch L, Zhou S. Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. *Stat Comput*. 2021;31(6):82.
32. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2018;19(6):1236-1246.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sax DR, Warton EM, Sofrygin O, et al. Automated analysis of unstructured clinical assessments improves emergency department triage performance: A retrospective deep learning analysis. *JACEP Open*. 2023;4:e13003. <https://doi.org/10.1002/emp2.13003>

AUTHOR BIOGRAPHY



Dana Sax, MD, MPH, is an adjunct investigator with the Kaiser Permanente Division of Research and an emergency medicine physician with The Permanente Medical Group (TPMG) in Oakland, California.