Canadian Journal of Emergency Medicine
https://doi.org/10.1007/s43678-024-00807-z

ORIGINAL RESEARCH

Machine learning outperforms the Canadian Triage and Acuity Scale (CTAS) in predicting need for early critical care

Lars Grant[1,2,3] · **Magueye Diagne**[1,3] · **Rafael Aroutiunian**[1,2] · **Devin Hopkins**[1,2] · **Tian Bai**[4] · **Flemming Kondrup**[5] · **Gregory Clark**[1,6]

Received: 24 June 2024 / Accepted: 6 October 2024

Abstract

Study objective This study investigates the potential to improve emergency department (ED) triage using machine learning models by comparing their predictive performance with the Canadian Triage Acuity Scale (CTAS) in identifying the need for critical care within 12 h of ED arrival.

Methods Three machine learning models (LASSO regression, gradient-boosted trees, and a deep learning model with embeddings) were developed using retrospective data from 670,841 ED visits to the Jewish General Hospital from June 2012 to Jan 2021. The model outcome was the need for critical care within the first 12 h of ED arrival. Metrics, including the areas under the receiver-operator characteristic curve (ROC) and precision-recall curve (PRC) were used for performance evaluation. Shapley additive explanation scores were used to compare predictor importance.

Results The three machine learning models (deep learning, gradient-boosted trees and LASSO regression) had areas under the ROC of 0.926 ± 0.003, 0.912 ± 0.003 and 0.892 ± 0.004 respectively, and areas under the PRC of 0.27 ± 0.01, 0.24 ± 0.01 and 0.23 ± 0.01 respectively. In comparison, the CTAS score had an area under the ROC of 0.804 ± 0.006 and under the PRC of 0.11 ± 0.01. The predictors of most importance were similar between the models.

Conclusions Machine learning models outperformed CTAS in identifying, at the point of ED triage, patients likely to need early critical care. If validated in future studies, machine learning models such as the ones developed here may be considered for incorporation in future revisions of the CTAS triage algorithm, potentially improving discrimination and reliability.

Keywords Emergency department triage · Machine learning · Artificial Intelligence · Emergency department operations

Résumé

✉ Lars Grant
lars.grant@mcgill.ca

[1] Department of Emergency Medicine, McGill University, Montreal, QC, Canada

[2] Emergency Department, Jewish General Hospital, Montreal, QC, Canada

[3] Lady Davis Research Institute at the Jewish General Hospital, Montreal, QC, Canada

[4] Department of Mathematics and Statistics, McGill University, Montreal, QC, Canada

[5] Quantitative Life Sciences Program, Faculty of Science, McGill University, Montreal, QC, Canada

[6] Emergency Department, Royal Victoria Hospital, Montreal, QC, Canada

Objectif de l'étude Cette étude vise à déterminer les possibilités d'amélioration du triage des services d'urgence (DE) au moyen de modèles d'apprentissage automatique en comparant leur rendement prédictif avec l'échelle canadienne d'acuité du triage (ETSC) Déterminer le besoin de soins intensifs dans les 12 heures suivant l'arrivée du DS

Méthodes Trois modèles d'apprentissage automatique (régression LASSO, arbres à gradient amplifié et modèle d'apprentissage profond avec intégration) ont été développés en utilisant des données rétrospectives de 670841 visites au ED de juin 2012 à janvier 2021. Le modèle a révélé un besoin de soins intensifs dans les 12 premières heures après l'arrivée des urgences. Les mesures, y compris les zones sous la courbe caractéristique du récepteur-opérateur (ROC) et la courbe de précision-rappel (PRC), ont été utilisées pour l'évaluation du rendement. Des scores d'explication additionnelle de Shapley ont été utilisés pour comparer l'importance du prédicteur.

Résultats Les trois modèles d'apprentissage automatique (apprentissage profond, arbres à gradient et régression LASSO) avaient des aires sous le ROC de 0,926 0,003, 0,912 0,003 et 0,892 0,004 respectivement, et des aires sous le PRC de 0,27 0,01, 0,24 0,01 et 0,23 0,01 respectivement. En comparaison, le score CTAS avait une aire sous le ROC de 0,804 0,006 et sous le PRC de 0,11 0,01. Les prédicteurs les plus importants étaient similaires entre les modèles.

Conclusions Les modèles d'apprentissage automatique ont surpassé l'ACST dans l'identification, au moment du triage des patients en urgence, de ceux qui pourraient avoir besoin de soins critiques précoces. Si les études futures sont validées, des modèles d'apprentissage automatique comme ceux développés ici pourraient être envisagés pour une intégration dans les révisions futures de l'algorithme de triage CTAS, ce qui pourrait améliorer la discrimination et la fiabilité.

Mots-clés Triage des services d'urgence · Apprentissage automatique · Intelligence artificielle · Opérations des services d'urgence

Clinician Capsule

What is known about the topic?

The current standard of care in ED triage is the use of a 5-point triage scale based on expert opinion such as CTAS—which is used uniformly across Canada.

What did this study ask?

Can machine learning models trained on historical ED data outperform CTAS in their ability to identify patients at ED triage who are likely to need early critical care?

What did this study find?

Three machine learning models all outperformed CTAS alone in their ability to predict the need for critical care within 12 h of ED triage.

Why does this study matter to clinicians?

Machine Learning models such as the one presented here may be considered for incorporation into CTAS triage protocols and may improve the reliability and discrimination of ED triage.

Introduction

Emergency Department (ED) triage is the first point of clinical contact with patients entering an ED and is intended to identify those most urgently in need of care [1, 2]. It attempts to prioritize high-risk visits and minimize the effects of delayed treatment among those who need it urgently [3]. It is especially important in the context of ED crowding—a longstanding problem [4–6] associated with delays in treatment, decreased quality of care [7–12] and increased mortality [13]. The current standard of care in ED triage employs validated 5-point triage scales [2] such as the Canadian Triage Acuity Scale (CTAS) [14–16] which is used across Canada and in several other countries [2, 17]. CTAS and other widely used 5-point triage scales demonstrate significant inter-rater variability and suboptimal accuracy and predictive ability [2, 17, 18]. The CTAS algorithm uses expert opinion to encode presenting complaints, vital signs, pain and some elements of past medical history or physical exam to generate a 5-level score associated with a recommended minimum time to physician contact (Appendix A).

Machine learning algorithms trained to predict clinical outcomes based on historical data have demonstrated superior predictive ability to some established 5-point triage scales [19–26]. Such tools can encode hard ED outcome data that could potentially improve current approaches to triage by decreasing inter-rater variability, increasing accuracy in the prediction of clinical outcomes and decreasing door-to-admission decision times [27]. Machine learning has recently been used with CTAS data to investigate the potential of assisting triage in a virtual care context [28, 29], but no reported machine learning triage tools use CTAS data to predict clinical outcomes.

This report uses historical data collected using CTAS to develop 3 machine learning models to predict, at the time of ED triage, critical illness within 12 h of ED triage. The machine learning models' ability to predict the need for early critical care was compared to the predictive performance of CTAS alone.

Methods

Study design and participants

This was a retrospective cohort study including all consecutive adult (≥ 18 years old) patient visits to the Jewish General Hospital ED between June 2012 and Jan 2021. The Jewish General Hospital is a tertiary care center receiving approximately 95,000 visits per year. Since May 30, 2012, our center has been using the MedUrge electronic triage tool which implements the 2008 version of CTAS. We used all available data up to 2021—reserving data after 2021 for validation of the derived models—therefore no sample size calculation was done a priori. The local research ethics board approved the study and waived the requirement for patient consent because it was retrospective and used no patient identifying data. We have followed the TRIPOD reporting guidelines [30].

Measurements

During ED triage at our hospital, a nurse enters triage data into MedUrge and assigns a triage score. The outcome data were extracted from MedUrge and other electronic health records. We excluded repeat, cancelled or incomplete triages (missing a triage score) and visits where patients left without being seen or against medical advice or were transferred to an outside hospital.

Predictors

In MedUrge, presenting symptoms, past medical history and physical exam findings relevant to CTAS are chosen from a large, but finite list of possible entries and were included as predictors along with demographic data, vital signs, a patient-reported pain level. All predictors were available at the time of triage (see Table 1).

Outcome

A composite, critical illness outcome was defined to have occurred if there was an ICU consultation placed in the ED, admission to a critical care bed (ICU, CCU or other) within 12 h after arrival or death within 24 h after arrival. We included ICU consultation as a surrogate for ICU admission as, due to boarding times, patients sometimes receive critical care interventions in our ED for long enough that they can ultimately be admitted to a lower acuity ward. We used a 12-h outcome window because we believe the clinical course of patients with a short-term need for critical care is more likely to be affected by a delay to care of minutes to hours caused by mis-triage. In our ED, 90% of patients are assessed by a physician within 4 h and 90% of ICU admissions take place within 34 h of arrival. Across Canada, 90% of ED patients see a physician within 5 h and 90% of admissions take place within 49 h of arrival [31, 32].

Data analysis

We developed 3 machine learning models and a reference model based on the CTAS triage score to predict the outcome. The included visits were randomly split into training (used to teach the model how to predict the outcome—68%), tuning (used to tune key model characteristics—17%) and test sets (used exclusively for model evaluation after development complete—15%). The reference model used the CTAS triage score as a single predictor in a logistic regression. The LASSO regression model used 10 predictors selected by a machine learning process. The gradient boosting model used 298 predictors selected based on the frequency of occurrence and correlation with the outcome. The deep learning model used all available predictors using embeddings to handle text-based data such as presenting complaints and past medical history. The online supplement contains details of the machine learning model training (Appendix B) and missing data handling (Appendix C).

All four models (CTAS alone, LASSO, tree-based and deep learning) were evaluated in the same test set—which was not used during model development. We assessed the areas under the receiver-operator characteristic curve (ROC) and the precision-recall curve (PRC). In rare outcomes, such as ours, the positive predictive value is of great interest because it reflects the number of false positives produced by predicting critical illness at triage and the Precision-Recall curve, which plots the positive predictive value against the sensitivity, can be a useful and more discriminating tool than the ROC to assess performance. Confidence intervals were generated using 5000 bootstrapped samples from the test set. We used Shapley additive explanations [33] to compare the importance of predictors in the three machine learning models.

The area under the ROC may be interpreted as the probability that a model will correctly rank a random visit resulting in the outcome over a random visit not resulting in the outcome. We considered a difference in AUC of more than 0.05—corresponding to a 5% probability of mis-ranking such a pair of visits—to be clinically significant. The area under PRC can be interpreted as the average positive predictive value over all possible threshold choices. We considered a difference of more than 0.05 in the area under PRC,

Table 1 A list of predictors available at the time of triage and included as possible predictors of a patient's clinical outcome in all 3 machine learning models

Provenance - Where was patient before coming to the emergency department? Eg Home vs. rehab

Mode of arrival - eg ambulance vs. police vs. walking

Destination after triage - eg room with stretcher vs. waiting room

Flu symptoms score (positive or negative)

Hot/Cold with respect to COVID (since 2020)

Sex

Respiratory Rate

Systolic Blood Pressure

Diastolic Blood Pressure

Heart Rate

Temperature

Oxygen saturation

Capillary Blood Glucose Measurement

Glasgow Coma Scale

Triage Score

Age

PRISMA-75 Score

Litres of **oxygen by nasal prongs**

Fraction of inspired **oxygen by mask** (eg Venturi mask)

Referred to the ED by an outside physician (yes or no)

Chief Complaint (from a categorical list)

Presenting symptoms (each chosen from a categorical list, includes some findings that would be based on physical exam such as "appears septic" or "severe respiratory distress")

Items of the **Past Medical History** (each chosen from a categorical list)

Allergies

Triage alerts such as risk factors to be carrier of drug-resistant organism or social situation corresponding to a 0.05 difference in average positive predictive value, to be clinically significant.

Data analysis used python 3.9.13 (python.org).

Results

Characteristics of study subjects

The total number of visits included was 670,841. Overall, 9440 (1.4%) of the visits experienced the study outcome (4605 ICU consulted in ED, 1042 deaths and 4802 admitted to critical care within 12 h). The characteristics of the training, tuning and test cohorts are shown in Table 2. There were 2038 visits with a mid-range triage score of 3 that experienced the critical care outcome (0.7% of the CTAS 3 patients). Missingness in our data is described in Appendix C.

Decision curves for the deep learning and CTAS alone models (Appendix E) suggest there would be a benefit to using the deep learning model at any true ideal risk threshold for intervention but that the benefit would be particularly strong if the ideal risk threshold is less than 0.2.

Model Performance

ROC and PRC for CTAS and the 3 machine learning models are shown in Figs. 1 and 2, respectively. Table 3 shows the performance characteristics of the models. The deep learning model demonstrated statistically significantly improved discrimination in both the sensitivity–specificity and sensitivity-positive predictive value planes (area under ROC 0.926 ± 0.003 and area under PRC of 0.27 ± 0.01). The gradient boosted trees (area under ROC of 0.912 ± 0.003, area under PRC of 0.24 ± 0.01) and LASSO regression (area under ROC of 0.892 ± 0.004, area under PRC of 0.23 ± 0.01) models were inferior to the deep learning model, but all 3 machine learning models significantly outperformed CTAS alone (area under ROC of 0.804 ± 0.006 and area under PRC of 0.11 ± 0.01) in predicting the critical illness outcome. Calibration curves (Appendix D) for all models had slopes close to 1 with intercepts near 0.

Many of the most important predictors (Fig. 3) were similar between the models. In the deep learning and gradient-boosted trees models, the combined importance of the remaining predictors outweighed the importance of the most important predictors.

Table 2 Characteristics of the included visits

Characteristic	Training Set $N = 456{,}171$ (%)	Tuning Set $N = 114{,}043$ (%)	Test Set $N = 100{,}627$ (%)
Age	52.3 ± 21.4	52.5 ± 21.4	52.5 ± 21.5
Gender (male)	44.0	43.7	43.7
Arrival by ambulance	18.5	18.6	18.7
CTAS 1	0.72	0.73	0.79
CTAS 2	22.6	22.4	22.4
CTAS 3	46.5	46.6	46.7
CTAS 4	24.3	24.4	24.2
CTAS 5	5.97	5.85	5.89
Chief complaints			
Abdominal pain	9.6	9.7	9.6
Chest pain with cardiac features	5.4	5.4	5.4
Shortness of breath	4.5	4.5	4.6
Lower extremity pain	3.9	4.0	4.0
Minor complaints NOS	3.7	3.6	3.7
Back pain	3.4	3.5	3.4
Headache	3.0	2.9	2.8
General weakness	2.9	2.9	3.0
Lower extremity injury	2.9	2.9	2.8
Localized swelling and redness	2.7	2.8	2.7
Outcome			
Critical illness outcome	2	2.1	2.1

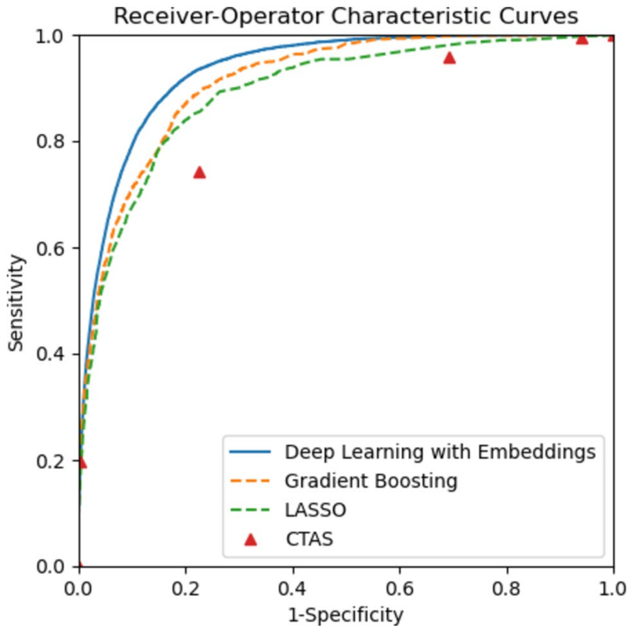


Fig. 1 Receiver Operator Characteristic Curves for the reference model (CTAS) and the 3 machine learning models. Note the curve for the CTAS alone model is displayed as 5 single points rather than a continuous curve. This is because there are effectively only 5 possible threshold choices in this model, corresponding to the 5 CTAS triage scores. *CTAS* Canadian Triage and Acuity Scale

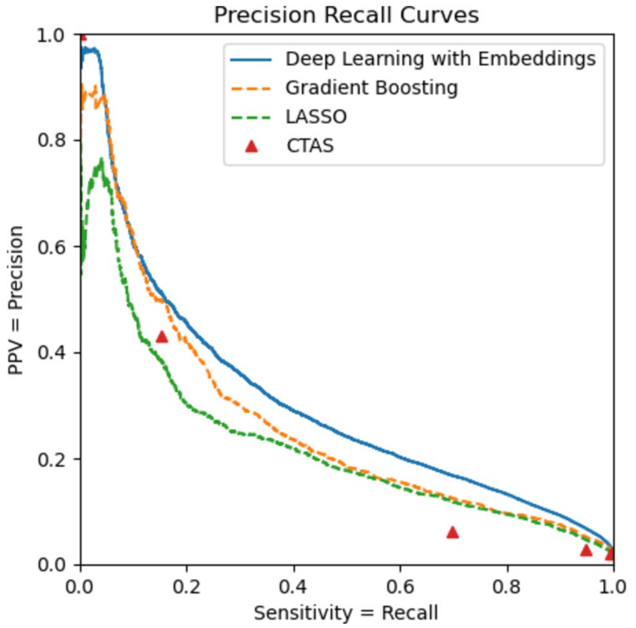


Fig. 2 Precision-Recall Curves for the reference model (CTAS) and the 3 machine learning models. Note the curve for the CTAS alone model is displayed as 5 single points rather than a continuous curve. This is because there are effectively only 5 possible threshold choices in this model, corresponding to the 5 CTAS triage scores. *CTAS* Canadian Triage and Acuity Scale

Discussion

Interpretation of findings

All 3 machine learning models demonstrated superior discrimination (by area under ROC or PRC) to a reference model using the CTAS score alone as a predictor. The difference in area under the ROC was greater than 0.088 (well over our threshold of 0.05 for clinical significance) for all 3 machine learning models—corresponding to a 9% difference in probability of mis-ranking a random visit resulting in the outcome over a random visit not resulting in the outcome. The difference in area under PRC—was greater than 0.11 for all 3 machine learning models (well over our threshold of 0.05 for clinical significance) corresponding to an 11% difference in average positive predictive value. The deep learning model outperformed the other models on all measures. Decision Curve Analysis suggests an increased net clinical benefit with the use of machine learning models as compared to CTAS alone.

Comparison to previous studies

Our deep learning model had an area under the ROC of 0.926 which was higher than most similar previously reported machine learning triage models of which we are aware [19, 20, 22]. Differences may be related to a combination of using CTAS vs. other triage scales, differences in outcome definition, lower heterogeneity because of single site data or the use of embeddings to process patients' symptoms, which may allow the deep learning model to develop a deeper understanding of the clinical significance of patients' symptoms.

Strengths and limitations

The large sample size, accumulated over 9 years in a clinical environment, makes it a good representation of triage practice at the study site, but the model weights developed here may not generalize to other centers. For incorporation into a revised version of CTAS, the model would need to be retrained on a nationally representative data set. An institution-specific machine learning model may maximize predictive performance but make comparison or use across hospitals difficult. Also, we included the first year of the COVID-19 pandemic which significantly altered the pattern of ED critical care. A time-based validation of the machine learning tools developed here would evaluate the impact such a change in case-mix over time would have on performance. The MedUrge CTAS triage system used at our center collects reasons for visit, past medical history and modifier

Table 3 Performance Characteristics of the reference model and the 3 machine learning models in the test set

	Deep Learning	Gradient Boosted Trees	LASSO regression	CTAS alone
ROC-AUC For a random classifier = 0.5	0.926 95% CI (0.923, 0.929)	0.912 95% CI (0.909, 0.915)	0.892 95% CI (0.888, 0.896)	0.804 95% CI (0.798, 0.810)
PR-AUC For a random classifier = 0.02	0.27 95% CI (0.26, 0.28)	0.24 95% CI (0.23, 0.25)	0.23 95% CI (0.22, 0.24)	0.11 95% CI (0.10, 0.12)

data in a categorical format which allowed us to generate and train the custom embeddings in the deep learning model. This approach may allow the model to increase the depth of understanding of the patient's presentation and would be straightforward to implement across Quebec because most sites employ a similar electronic implementation of CTAS, but these systems are not currently used elsewhere in Canada to our knowledge.

Not all relevant clinical data can be easily entered into an electronic triage tool. For example, a triage nurse's subjective impression that a patient is more ill than their complaints suggest may improve the quality of triage, but be difficult to quantify reliably [34]. The critical illness outcome we chose may not reflect all needs for urgent care. For example, the need for urgent pain control cannot easily be incorporated into a machine learning model because pain at triage is poorly correlated with actual clinical outcomes [35], so any such model trained on clinical outcome data will deprioritize pain. Furthermore, some visits (e.g. anaphylaxis or severe agitation) require rapid intervention but rarely require critical care. Machine learning triage tools must be part of a larger triage system that involves human judgment and assessment of variables that cannot easily be handled electronically or are essential, but poorly captured by machine learning.

Clinical implications

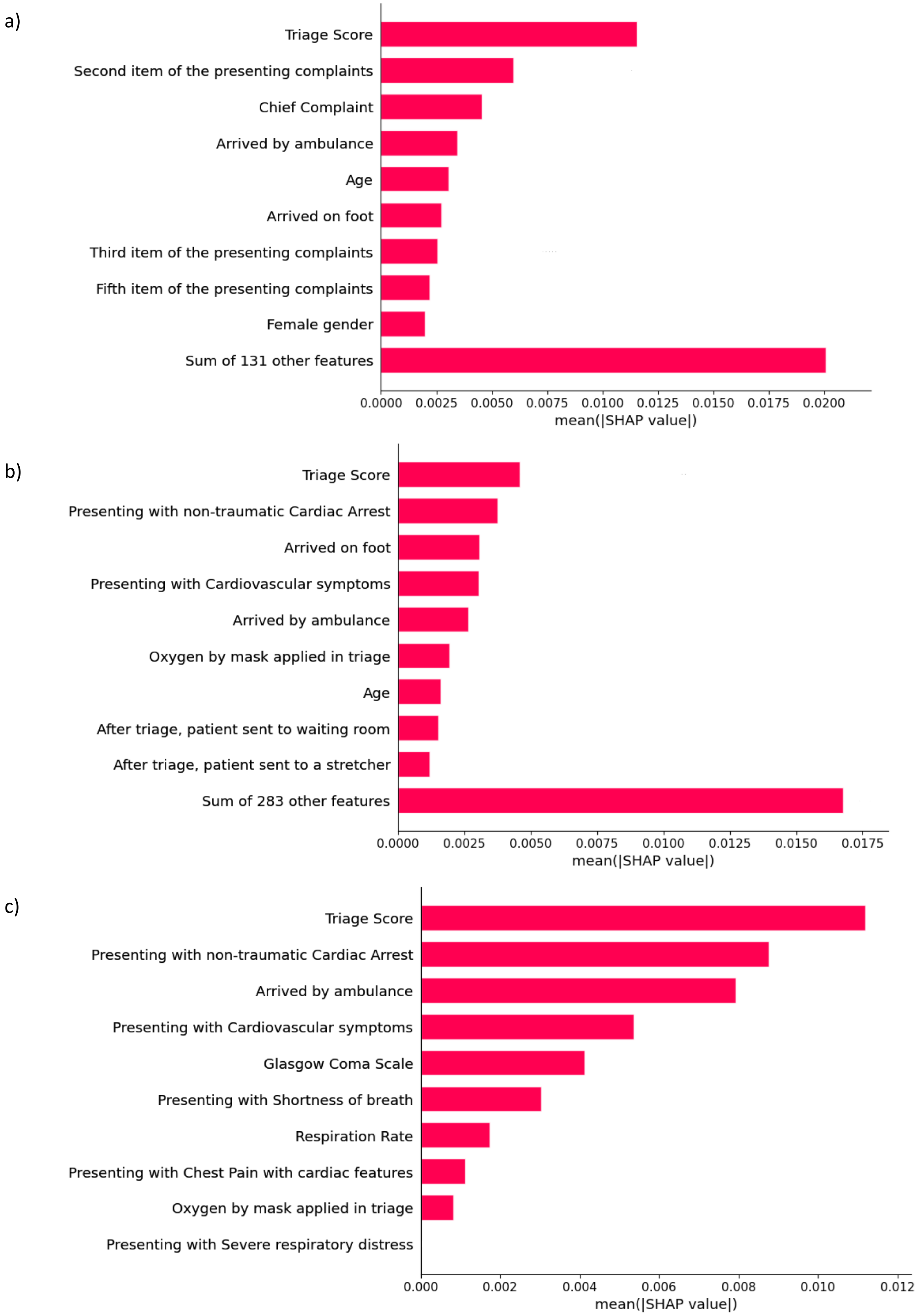
The machine learning models ranked the CTAS score as the top predictor of critical illness alongside several others (e.g. age, arrival by ambulance, requirement for a stretcher) that are not currently part of the CTAS triage algorithm. The discrimination of CTAS for critical illness might be improved by modifying its algorithm to include some of these variables. On the other hand, the predictor importance plots (Fig. 3) of the deep learning and gradient-boosted trees models, show that the combined importance of the remaining hundreds of predictors is greater than the importance of the most important predictors. Combining so many predictors in a human-calculable system may not be feasible and incorporating a validated machine learning model into future CTAS revisions may be a simpler way to improve the discrimination and reliability of ED triage. Such a machine learning tool could be added to the CTAS algorithm as a "modifier" in a manner analogous to the way the "hemodynamic stability" modifier is currently used. The predicted probability of early critical illness would set a minimum triage score for each visit which could be superseded by other factors considered in the CTAS algorithm or by nursing judgement. The chief benefit of this approach would be to improve the sensitivity of CTAS in detecting certain patients who might otherwise be deprioritized. Other approaches might include displaying to the triage nurse at the completion of triage a flag highlighting cases at risk of critical illness, the actual machine learning estimated probability of critical illness or a machine learning recommended triage score based on the predicted probability of critical illness. The second option would identify low-risk visits as well as high-risk visits, which might be useful among visits triaged as CTAS 3 (approximately 46% of visits [35]). The third option would allow the triage nurse to choose the CTAS or machine learning triage score or a score they felt more appropriate based on their own judgment.

Research implications

This work suggests that machine learning approaches may offer clinical benefits at the point of CTAS ED triage. We are currently planning a time-based validation of the models developed here. The development of a national-level ED triage dataset would be a great step forward in the development of machine learning triage tools for use with CTAS. Such a data set, which should contain the predictors in Table 1 as well as uniform, high-quality outcome data, would be beneficial to ED research generally and for public health monitoring of ED use. Further research is required to determine the optimal approach to incorporating a machine learning triage tool into CTAS (including approaches to explainability).

Conclusion

Machine learning models demonstrated improved discrimination, average positive predictive value and net benefit compared to the CTAS alone in predicting the need for early critical care. Incorporation of such machine learning tools


a)
Triage Score
Second item of the presenting complaints
Chief Complaint
Arrived by ambulance
Age
Arrived on foot
Third item of the presenting complaints
Fifth item of the presenting complaints
Female gender
Sum of 131 other features
0.0000
0.0025
0.0050
0.0075
0.0100
0.0125
0.0150
0.0175
0.0200
mean(|SHAP value|)
b)
Triage Score
Presenting with non-traumatic Cardiac Arrest
Arrived on foot
Presenting with Cardiovascular symptoms
Arrived by ambulance
Oxygen by mask applied in triage
Age
After triage, patient sent to waiting room
After triage, patient sent to a stretcher
Sum of 283 other features
0.0000
0.0025
0.0050
0.0075
0.0100
0.0125
0.0150
0.0175
mean(|SHAP value|)
c)
Triage Score
Presenting with non-traumatic Cardiac Arrest
Arrived by ambulance
Presenting with Cardiovascular symptoms
Glasgow Coma Scale
Presenting with Shortness of breath
Respiration Rate
Presenting with Chest Pain with cardiac features
Oxygen by mask applied in triage
Presenting with Severe respiratory distress
0.000
0.002
0.004
0.006
0.008
0.010
0.012
mean(|SHAP value|)

◂**Fig. 3** Predictor importance in the machine learning models. **a** Deep learning model, **b** Gradient boosted tress model and **c** LASSO regression model. In **a** and **b**, the Shapley additive explanation scores of the 9 most important predictors are shown, alongside the sum of the scores of all the remaining predictors. Even though the importance of the remaining predictors is individually small, their sum outweighs the importance of any of the most important predictors for these models. In **c** the Shapley additive explanation scores of the 10 selected predictors are shown. In **a** "item from the presenting complaints" refers to the list of presenting symptoms and modifiers entered by the triage nurse. The "first" item in this list always describes the general organ system involved in the visit (e.g. cardiovascular), the "second" item is the primary presenting symptom and the "third" and other items represent secondary items entered by the triage nurse for a given visit, which may include other symptoms or CTAS modifiers. The deep learning model relies heavily on the data about the reasons for visit and modifiers entered by the nurse but does not rely heavily on the general organ category of the presenting complaint

into ED triage protocols may enhance the performance of CTAS triage by improving the reliability and discrimination of triage.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43678-024-00807-z.

Author contributions LG conceived the study and obtained funding. LG, RA, DH and MD designed the study. MD extracted and cleaned the data from the electronic sources. LG, RA and DH audited the data for quality. MD, and LG analyzed the data. LG drafted the manuscript and RA, TB, FK and GC contributed essential ideas to its revision. LG supervised the project and took responsibility for the paper as a whole.

Funding We would like to thank the New Frontiers in Research Fund for their support of this work (NFRFE-2019–01541). None of the authors has a conflict of interest to declare. Specifically, LG, MD, RA, DH, TB, FK and GC have no conflicts of interest to declare.

Data availability The data that support the findings of this study are not openly available because they contain Protected Health Information (PHI). Data are located in controlled access data storage at the Jewish General Hospital.

References

1. Farrohknia N, Castrén M, Ehrenberg A, et al. Emergency department triage scales and their components: a systematic review of the scientific evidence. Scand J Trauma Resuscit Emerg Med. 2011;19(1):42.
2. Hinson JS, Martinez DA, Cabral S, et al. Triage performance in emergency medicine: a systematic review. Ann Emerg Med. 2019;74(1):140–52.
3. Gilboy N, Travers D, Wuerz R. Re-evaluating triage in the new millennium: a comprehensive look at the need for standardization and quality. J Emerg Nurs. 1999;25(6):468–73.
4. Higginson I. Emergency department crowding. Emerg Med J. 2012;29(6):437–43.
5. Javidan AP, Hansen K, Higginson I, Jones P, Lang E, International Federation Emergency Department Crowding and Access Block Task Force. The international federation for emergency medicine report on emergency department crowding and access block: a brief summary. Emerg Med J 2021;38(3):245–6.
6. Affleck A, Parks P, Drummond A, Rowe BH, Ovens HJ. Emergency department overcrowding and access block. Can J Emerg Med. 2013;15(6):359–70.
7. Ortiz SS, Huang Y, Rowe BH, Zheng B, Rosychuk RJ. Emergency department crowding negatively influences outcomes for adults presenting for chronic obstructive pulmonary disease. Can J Emerg Med. 2023;25(5):411–20.
8. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: a systematic review of causes, consequences and solutions. PLoS ONE. 2018;13(8): e0203316.
9. Sun BC, Hsia RY, Weiss RE, et al. Effect of emergency department crowding on outcomes of admitted patients. Ann Emerg Med. 2013;61(6):605-611.e6.
10. Guttmann A, Schull MJ, Vermeulen MJ, Stukel TA. Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from Ontario, Canada. BMJ. 2011;342: d2983.
11. Miró O, Antonio MT, Jiménez S, et al. Decreased health care quality associated with emergency department overcrowding. Eur J Emerg Med. 1999;6(2):105–7.
12. Gaieski DF, Agarwal AK, Mikkelsen ME, et al. The impact of ED crowding on early interventions and mortality in patients with severe sepsis. Am J Emerg Med. 2017;35(7):953–60.
13. Richardson DB. Increase in patient mortality at 10 days associated with emergency department overcrowding. Med J Aust. 2006;184(5):213–6.
14. Beveridge R, John S, Clarke B, et al. Implementation guidelines for the Canadian emergency department triage and acuity scale (CTAS). Canadian Association of Emergency Medicine. 1998.
15. Murray M, Bullard M, Grafstein E, the CTAS. Revisions to the Canadian emergency department triage and acuity scale implementation guidelines. CJEM. 2004;6(06):421–7.
16. Bullard MJ, Musgrave E, Warren D, et al. Revisions to the Canadian emergency department triage and acuity scale (CTAS) guidelines 2016. CJEM. 2017;19(S2):S18-27.
17. Zachariasse JM, van der Hagen V, Seiger N, Mackway-Jones K, van Veen M, Moll HA. Performance of triage systems in emergency care: a systematic review and meta-analysis. BMJ Open. 2019;9(5): e026471.
18. Mirhaghi A, Ebrahimi M, Heydari A, Mazlom R. The reliability of the Canadian triage and acuity scale: meta-analysis. North Am J Med Sci. 2015;7(7):299.
19. 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.
20. 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.
21. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. PLoS ONE. 2018;13(7): e0201016.
22. 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.
23. Goto T, Camargo CA, Faridi MK, Yun BJ, Hasegawa K. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. Am J Emerg Med. 2018;36(9):1650–4.
24. Zhang X, Kim J, Patzer RE, Pitts SR, Patzer A, Schrager JD. Prediction of emergency department hospital admission based on natural language processing and neural networks. Methods Inf Med. 2017;56(5):377–89.

25. Miles J, Turner J, Jacques R, Williams J, Mason S. Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review. Diagnos Prognos Res. 2020;4(1):16.
26. Yu JY, Xie F, Nan L, et al. An external validation study of the Score for Emergency Risk Prediction (SERP), an interpretable machine learning-based triage score for the emergency department. Sci Rep. 2022;12(1):17466.
27. Levin S. HopScore: an Electronic Outcomes-based emergency triage system—final report [Internet]. Johns Hopkins University, Department of Emergency Medicine. 2018. Available from: https://digital.ahrq.gov/sites/default/files/docs/citation/r21hs023641-levin-final-report-2018.pdf
28. Hall JN, Galaev R, Gavrilov M, Mondoux S. Development of a machine learning-based acuity score prediction model for virtual care settings. BMC Med Inform Decis Mak. 2023;23(1):200.
29. Nasser L, McLeod SL, Hall JN. Evaluating the reliability of a remote acuity prediction tool in a Canadian Academic Emergency Department. Ann Emerg Med. 2024;83(4):373–9.
30. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350: g7594.
31. CIHI. Canadian Institute for Health Information. Emergency Department Wait Time for Physician Initial Assessment (90% Spent Less, in Hours). Available from: https://www.cihi.ca/en/indicators/emergency-department-wait-time-for-physician-initial-assessment-90-spent-less-in-hours. Accessed June 5, 2024. [Internet].
32. CIHI. Canadian Institute for Health Information. Total time spent in emergency department for admitted patients (90% Spent Less, in Hours). Available from: https://www.cihi.ca/en/indicators/total-time-spent-in-emergency-department-for-admitted-patients-90-spent-less-in-hours. Accessed June 5, 2024. [Internet].
33. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765–74.
34. Davis S, Ju C, Marchandise P, Diagne M, Grant L. The effect of human supervision on an electronic implementation of the Canadian triage acuity scale (CTAS). J Emerg Med. 2022;63(4):498–506.
35. Davis S, Ju C, Marchandise P, Diagne M, Grant L. Impact of pain assessment on Canadian triage and acuity scale prediction of patient outcomes. Ann Emerg Med. 2022;79(5):433–40.