# RESEARCH

# **Open Access**



Integrating structured and unstructured data for predicting emergency severity: an association and predictive study using transformer-based natural language processing models

Xingyu Zhang<sup>1\*</sup>, Yanshan Wang<sup>2</sup>, Yun Jiang<sup>3</sup>, Charissa B. Pacella<sup>4</sup> and Wenbin Zhang<sup>5\*</sup>

# Abstract

Background Efficient triage in emergency departments (EDs) is critical for timely and appropriate care. Traditional triage systems primarily rely on structured data, but the increasing availability of unstructured data, such as clinical notes, presents an opportunity to enhance predictive models for assessing emergency severity and to explore associations between patient characteristics and severity outcomes. This study aimed to evaluate the effectiveness of combining structured and unstructured data to predict emergency severity more accurately.

Methods Data from the 2021 National Hospital Ambulatory Medical Care Survey (NHAMCS) for adult ED patients were used. Emergency severity was categorized into urgent (scores 1–3) and non-urgent (scores 4–5) based on the Emergency Severity Index. Unstructured data, including chief complaints and reasons for visit, were processed using a Bidirectional Encoder Representations from Transformers (BERT) model. Structured data included patient demographics and clinical information. Four machine learning models—Logistic Regression, Random Forest, Gradient Boosting, and Extreme Gradient Boosting—were applied to three data configurations: structured data only, unstructured data only, and combined data. A mean probability model was also created by averaging the predicted probabilities from the structured and unstructured models.

Results The study included 8,716 adult patients, of whom 74.6% were classified as urgent. Association analysis revealed significant predictors of emergency severity, including older age (OR = 2.13 for patients 65 +), higher heart rate (OR = 1.56 for heart rates > 90 bpm), and specific chronic conditions such as chronic kidney disease (OR = 2.28) and coronary artery disease (OR = 2.55). Gradient Boosting with combined data demonstrated the highest performance, achieving an area under the curve (AUC) of 0.789, an accuracy of 0.726, and a precision of 0.892. The mean probability model also showed improvements over structured-only models.

Conclusions Combining structured and unstructured data improved the prediction of emergency severity in ED patients, highlighting the potential for enhanced triage systems. Integrating text data into predictive models can

\*Correspondence: Xingyu Zhang xiz261@pitt.edu Wenbin Zhang wenbinzhang2008@gmail.com Full list of author information is available at the end of the article



© The Author(s) 2024. Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

provide more accurate and nuanced severity assessments, improving resource allocation and patient outcomes. Further research should focus on real-time application and validation in diverse clinical settings.

**Keywords** Emergency department, Predictive modeling, Association study, Unstructured data, Natural language processing, Clinical decision support

# Introduction

Emergency departments (EDs) are critical points of care that manage a diverse array of medical conditions with varying degrees of severity [1, 2]. Efficient triage and resource allocation are vital to ensuring timely and appropriate care for patients [3]. Traditionally, triage systems have relied on structured data such as patient demographics, vital signs, and medical history to assess the urgency of cases [4, 5]. Modern ED triage systems typically involve a nurse performing patient assessment and using an algorithm to determine triage acuity. The nurse incorporates subjective information obtained from the patient with structured data from the EHR into their triage decision. While triage systems are effective in predicting resource use and likelihood hospitalization [6], current systems rely heavily on the training and experience of the nurse to gather necessary data and correctly apply the algorithm. The wealth of unstructured data now available in EHRs presents an opportunity to refine and augment severity predictions by incorporating this rich source of information [7-9].

Natural Language Processing (NLP) techniques have shown promise in extracting valuable insights from unstructured text data. Specifically, Transformer-based models like Bidirectional Encoder Representations from Transformers (BERT) have revolutionized the field of NLP by enabling deep bidirectional understanding of text [10, 11]. BERT is uniquely suited for processing complex medical narratives because it captures the context of words in relation to all other words in a sentence. This ability sets BERT apart from traditional NLP techniques, such as Bag of Words or Term Frequency-Inverse Document Frequency (TF-IDF), which are limited by their reliance on isolated word frequencies and their inability to account for word dependencies and nuanced meanings in clinical language [12]. The use of BERT allows the models to understand the deeper semantics of clinical text, which is crucial in an ED setting where word context can dramatically alter the interpretation of patient symptoms [13]. Moreover, BERT is pre-trained on vast datasets and fine-tuned on specific tasks, making it effective even when applied to smaller datasets. This advantage of BERT has made it a key tool for integrating unstructured and structured data to enhance predictive tasks in healthcare [14, 15].

In recent years, NLP models have been applied to various clinical tasks, including analyzing clinical notes, predicting patient outcomes, and assisting in care prioritization [7–9, 16]. Appling NLP for ED triage has also grown significantly. For instance, Stewart et al. [17] reviewed various applications of NLP at ED triage, highlighting the potential of these techniques to enhance triage accuracy and efficiency. The advancements underscore the promising contribution of NLP to transforming triage practices in EDs. Despite these advances, the application of novel approaches to integrate structured and unstructured clinical data for predicting emergency severity in ED settings remains underexplored [17].

This study aimed to fill this gap by developing and comparing models that predict the emergency severity score, a critical triage indicator, using both structured and unstructured data from the National Hospital Ambulatory Medical Care Survey—Emergency Department (NHAMCS-ED) for the year 2021 [18]. We hypothesized that a combined approach integrating structured data with unstructured text data processed through a Transformer-based model would outperform models utilizing structured data alone or unstructured data alone. By leveraging the comprehensive NHAMCS-ED dataset, our goal was to harness the full spectrum of available information, enhancing the predictive power and clinical utility of our models.

# Method

## Data source

The data for this study was obtained from the NHAMCS-ED dataset for the year 2021. Only adult patients (age > 18 years) were included in the study, and patients with missing emergency severity scores were excluded. This dataset provides comprehensive information on patient visits to emergency departments across the United States, including demographic details, reasons for visit, and diagnostic and treatment data.

## Study outcome

The emergency severity score used in this study is based on the Emergency Severity Index (ESI), a widely used triage tool in EDs. The ESI assigns scores from 1 to 5 to prioritize patients based on the urgency of their condition. A score of 1 represents the most critical cases requiring immediate life-saving interventions, while a score of 5 indicates non-urgent cases that can safely wait for care. These scores are typically assigned by a triage nurse during the initial evaluation, using a combination of objective measures (such as vital signs) and clinical judgment [19, 20]. ESI scores directly inform decisions about the urgency of treatment. For example, patients with a score of 1 need immediate intervention to prevent death, while those with a score of 2 are high-risk and must be seen quickly to avoid deterioration. Patients with a score of 3, though stable, still require timely care but can wait longer than those with more urgent scores. Meanwhile, scores of 4 and 5 represent minor conditions that can safely wait for extended periods [21, 22]. While the ESI is an ordinal scale, the differences in urgency between consecutive scores are not evenly spaced. The difference in urgency between a score of 1 and 2 is much greater than between scores 3 and 4. For this study, we grouped ESI scores into two categories: urgent (scores 1-3) and nonurgent (scores 4-5). This binary categorization reflects common clinical practice, where the main concern is whether a patient requires urgent intervention. Although this approach reduces some granularity, it aligns with the critical decision-making process in EDs, prioritizing the need for urgent care [23].

#### Structured predictors

The structured data extracted from the dataset included a variety of variables related to patient demographics, visit characteristics, and clinical information. Specifically, the structured data encompassed patient demographics such as age, sex, and race/ethnicity. Visit characteristics included arrival time, mode of arrival, day of the week, and whether the patient arrived by ambulance. Clinical information comprised vital signs (temperature, heart rate, diastolic blood pressure, systolic blood pressure, pulse oximetry, respiratory rate), pain level, and medical history (conditions such as Alzheimer's disease/dementia, asthma, cancer, cerebrovascular disease, chronic kidney disease, chronic obstructive pulmonary disease, congestive heart failure, coronary artery disease, depression, diabetes mellitus types I and II, end-stage renal disease, pulmonary embolism, HIV infection/AIDS, hyperlipidemia, hypertension, obesity, obstructive sleep apnea, osteoporosis, and substance abuse or dependence). Additional factors considered were the type of residence (private residence, nursing home, homeless, or other), insurance type, whether the visit was a follow-up or within the last 72 h, and the nature of any injury or trauma, overdose/poisoning, or adverse effect of medical/surgical treatment. Missing values in the structured data were handled using median imputation, and the data were standardized using StandardScaler [24].

## Unstructured data and BERT model

Unstructured data consisted of the chief complaints and reasons for the injury presented at the ED visits. To ensure the quality and consistency of the input data, a structured text cleaning process was applied. This involved converting all text to lowercase for uniformity, removing punctuation and numbers, and filtering out common stopwords (e.g., "and," "the") that do not contribute meaningfully to clinical interpretation. These steps ensured that the text data retained only relevant clinical information.

These cleaned text fields were tokenized using the BERT tokenizer from the HuggingFace library [25], preparing the text data for input into a BERT-based model. The BERT model represents a significant advancement in natural language processing by enabling deep bidirectional understanding of text [26, 27]. Unlike traditional models that read text either left-to-right or right-to-left, BERT processes text in both directions simultaneously, allowing it to understand the context of a word based on all surrounding words. This bidirectional approach enables BERT to capture the nuanced meanings of words and phrases in their specific contexts. BERT's architecture is based on transformers, a type of deep learning model that relies on self-attention mechanisms to weigh the importance of different words in a sentence. This allows BERT to excel at tasks that require understanding the relationships between words and the overall meaning of sentences. Pre-trained on a vast corpus of text data, including books and Wikipedia articles, BERT can be fine-tuned on specific tasks such as classification, question answering, and named entity recognition.

To prepare the unstructured text data for analysis, we used the BERT tokenizer. This process converts the clinical text into a structured format that BERT can interpret, ensuring that important contextual information is preserved. The tokenizer breaks down sentences into smaller units, allowing BERT to understand the relationships between words in a given sentence. Following tokenization, the text was passed through the BERT model to generate numerical embeddings-dense vectors that represent the semantic meaning of the text. These embeddings capture the context and meaning of the text, allowing the model to utilize the full depth of clinical narratives. The embeddings were then combined with the structured data, integrating both textual and numerical information to enhance the predictive capability of the model.

## Predictive model development

For this study, four different machine learning models were applied: Logistic Regression (LR), Random Forest

(RFM) [28], Gradient Boosting (GB) [29], and Extreme Gradient Boosting (XGB) [30]. We implemented four machine learning models using Python's scikit-learn and xgboost libraries to evaluate the predictive performance of structured data in classifying emergency severity. For each model, key parameters were configured, while all other parameters were set to their default values. The LogisticRegression function from sklearn.linear model was set with a maximum iteration limit of 1000 (max\_ iter = 1000) to ensure convergence. The RandomForest-Classifier function from sklearn.ensemble was employed with 500 estimators (n\_estimators = 500), balancing accuracy and computational efficiency. The GradientBoostingClassifier, also from sklearn.ensemble, was applied with default settings, allowing the model to iteratively adjust for errors made by prior trees, thereby focusing subsequent trees on misclassified instances to improve predictive precision. For XGBoost, we utilized XGBClassifier from the xgboost library, configuring it with logloss as the evaluation metric to prioritize probability calibration and classification accuracy. Each of these models was trained and evaluated using four distinct approaches: structured data, unstructured data, combined data, and a mean probability model.

The first approach used only structured data, including patient demographics, clinical information, and visit characteristics from the NHAMCS-ED dataset. Logistic Regression, Random Forest, Gradient Boosting, and XGBoost models were trained on this data. The second approach focused solely on unstructured data, processed using the BERT model to generate feature vectors. These vectors were used as input for the same machine learning models. The third strategy combined both structured and unstructured data, merging quantitative information with BERT-extracted features to provide a comprehensive input for the models. The final method employed a mean probability model, which averaged the predicted probabilities from the structured and unstructured models. This technique combined the strengths of both data types without retraining. All approaches were evaluated using fivefold cross-validation.

## **Evaluation metrics**

The evaluation of all models involved calculating ROC AUC, accuracy, F1 score, precision, recall, sensitivity, and specificity. The models' predictive probabilities and true labels were recorded, and ROC curves were plotted to visualize the performance of each model. The cutoff points for classification were determined by finding the thresholds closest to the top-left corner of the ROC curve [31, 32]. The ROC AUC quantifies the model's ability to differentiate between these two categories, with higher values (closer to 1.0) indicating better

discrimination across various threshold values. Accuracy reflects the proportion of correct classifications (both urgent and non-urgent) out of the total predictions; however, its utility may be limited when class distribution is imbalanced. Precision measures the proportion of true positives (correctly classified urgent cases) among all instances predicted as urgent, making it particularly useful when minimizing false positives is important. Sensitivity (Recall), on the other hand, evaluates the model's ability to correctly identify all urgent cases, which is crucial in emergency department settings where missing urgent cases could have serious consequences. Specificity assesses the model's ability to correctly classify non-urgent cases, thereby avoiding over-triage, where non-urgent patients are incorrectly labeled as urgent. Finally, the F1 score, which is the harmonic mean of precision and recall, offers a balanced evaluation of the model's handling of both false positives and false negatives, especially valuable in scenarios with uneven class distributions. ROC curves were plotted for each model to compare their performance. Additionally, visualizations such as forest plots of odds ratios and word clouds of unstructured variables were generated to illustrate the significance and frequency of different variables in the dataset.

## Results

Among the 8,716 patients included in the study, 25.4% were categorized as non-urgent or semi-urgent, while 74.6% were classified as urgent, emergent, or immediate. Table 1 and Supplement Table 1 present the baseline characteristics of U.S. patients presenting to the ED, stratified by emergency severity score. Significant differences were observed between the two groups in terms of gender, with a higher proportion of females in the urgent category (55.1%) compared to the non-urgent group (52.0%, p = 0.0096). Age also varied significantly, with older patients more likely to be in the urgent category (p < 0.0001). Specifically, 27.5% of patients aged 65 and above were in the urgent group, compared to 15.7% in the non-urgent group. Race/ethnicity did not show significant differences between groups (p = 0.0603). However, differences were noted in residence type (p < 0.0001), with a higher percentage of urgent patients residing in nursing homes (2.8% vs. 1.0%) and a greater proportion of non-urgent patients living in private residences (95.8% vs. 94.2%). Insurance type also showed significant differences (p < 0.0001), with a higher percentage of urgent patients covered by Medicare (29.7% vs. 19.0%) and a higher percentage of non-urgent patients being uninsured (10.4% vs. 8.0%). Arrival by ambulance was significantly more common in the urgent group (25.4% vs. 8.2%, p < 0.0001). Follow-up visits were slightly more frequent Table 1 Baseline characteristics of U.S. patients presenting to the ED, stratified by Emergency Severity Score, NHAMCS 2021

	Non-urgent or Semi-urgent	Urgent, Emergent or Immediate	<i>p</i> value
	2211(25.4)	6505 (74.6)	
Gender			0.0096
Female	1149(52.0)	3587(55.1)	
Male	1062(48.0)	2918(44.9)	
Age, y			< 0.0001
18–39	1071(48.4)	2264(34.8)	
40–65	792(35.8)	2455(37.7)	
>=65	348(15.7)	1786(27.5)	
Race/ethnicity			0.0603
White	1315(59.5)	3933(60.5)	
Black	537(24.3)	1496(23.0)	
Hispanic	293(13.3)	811(12.5)	
Other	66(3.0)	265(4.1)	
Residence type	()		< 0.0001
Private residence	2067(95.8)	5944(94.2)	
Nursing home	22(1,0)	175(2.8)	
Homeless	48(2,2)	118(1.9)	
Other	21(1.0)	74(1.2)	
	21(1.0)	/+(1.2)	< 0.0001
Brivata insurance	509(20.6)	1755(20.0)	< 0.0001
Modicare	396(29.0)	1733(29.9)	
Medicale	202(19.0) 710(25.5)	1740(29.7)	
	/ 18(35.5)	1/31(29.4)	
Other	211(10.4)	408(8.0)	
Other	110(5.4)	177(3.0)	0.0005
Day of week		(000) (75.0)	0.0085
Weekdays	1600(/2.4)	4891(75.2)	
Weekend	611(27.6)	1614(24.8)	
Arrival time			0.06/8
Morning	660(30.1)	1857(28.7)	
Afternoon	646(29.5)	1987(30.7)	
Evening	382(17.4)	1016(15.7)	
Night	504(23.0)	1604(24.8)	
Arrive by ambulance	179(8.2)	1615(25.4)	< 0.0001
Follow-up visit	187(8.9)	444(7.4)	0.0243
Seen within last 72 h	78(3.7)	243(4.0)	0.5704
Pain level			0.2265
No pain	644(37.1)	1933(39.4)	
Mild	569(32.8)	1556(31.7)	
Severe	524(30.2)	1416(28.9)	
Temperature			0.0035
36 °C−38 °C	2082(96.1)	5895(94.5)	
< = 36 °C	59(2.7)	199(3.2)	
>38 °C	26(1.2)	144(2.3)	
Heart Rate, times/min			< 0.0001
61–90	1329(60.1)	3504(53.9)	
<=60	131(5.9)	422(6.5)	
>90	751(34.0)	2579(39.6)	
DBP mm Hg			0.0002
- 60-80	978(44.2)	2762(42.5)	

## Table 1 (continued)

	Non-urgent or Semi-urgent	Urgent, Emergent or Immediate	p value
<60	125(5.7)	543(8.3)	
>80	1108(50.1)	3200(49.2)	
SBP mm Hg			0.1919
80–120	486(22.0)	1438(22.1)	
<80	31(1.4)	130(2.0)	
>120	1694(76.6)	4937(75.9)	
Pulse oximetry (percent)			< 0.0001
0–94	136(6.3)	731(11.5)	
95 +	2037(93.7)	5648(88.5)	
Respiratory rate per minute			< 0.0001
12–20	2066(93.4)	5651(86.9)	
<12	48(2.2)	154(2.4)	
>20	97(4.4)	700(10.8)	
Injury/trauma, overdose/poisoning or adverse effect of medi- cal/ surgical treatment			< 0.0001
Yes, injury/trauma	814(38.0)	1305(20.8)	
Yes, overdose/poisoning	24(1.1)	77(1.2)	
Yes, adverse effect of medical/surgical treatment	54(2.5)	200(3.2)	
No	1210(56.5)	4535(72.3)	
Questionable injury status	38(1.8)	155(2.5)	
Medical History			
Alzheimer's disease/Dementia	18(0.8)	132(2.0)	0.0001
Asthma	236(10.7)	761(11.7)	0.1909
Cancer	55(2.5)	408(6.3)	< 0.0001
Cerebrovascular disease/History of stroke (CVA)	42(1.9)	343(5.3)	< 0.0001
Chronic kidney disease (CKD)	59(2.7)	382(5.9)	< 0.0001
Chronic obstructive pulmonary disease (COPD)	106(4.8)	538(8.3)	< 0.0001
Congestive heart failure (CHF)	60(2.7)	417(6.4)	< 0.0001
Coronary artery disease (CAD)	96(4.3)	670(10.3)	< 0.0001
Depression	298(13.5)	1027(15.8)	0.009
Diabetes mellitus (DM)—Type I	13(0.6)	46(0.7)	0.5549
Diabetes mellitus (DM)—Type II	139(6.3)	747(11.5)	< 0.0001
End-stage renal disease (ESRD)	12(0.5)	113(1.7)	< 0.0001
Pulmonary embolism (PE), deep vein thrombosis (DVT), or venous thromboembolism (VTE)	30(1.4)	162(2.5)	0.0017
HIV infection/AIDS	16(0.7)	54(0.8)	0.628
Hyperlipidemia	197(8.9)	1069(16.4)	< 0.0001
Hypertension	566(25.6)	2466(37.9)	< 0.0001
Obesity $(BMI > = 30)$	164(7.4)	604(9.3)	0.0074
Obstructive sleep apnea (OSA)	60(2.7)	249(3.8)	0.0144
Osteoporosis	18(0.8)	108(1.7)	0.004
Substance abuse or dependence	178(8.1)	665(10.2)	0.0028

The variables "Payment type," "Seen within last 72 h," and "Episode of care" have missing proportions between 5 and 10%. The variables "Patient residence," "Arrival by ambulance," "Is visit related to an injury/trauma, overdose/poisoning, or adverse effect of medical/surgical treatment?", "Temperature," and "Pulse oximetry" have missing proportions of less than 5%

in the non-urgent group (8.9% vs. 7.4%, p = 0.0243). Pain levels, temperature, heart rate, diastolic blood pressure, systolic blood pressure, pulse oximetry, and respiratory rate all showed significant differences between the two

groups. In terms of medical history, conditions such as cancer, cerebrovascular disease, chronic kidney disease, chronic obstructive pulmonary disease, congestive heart failure, coronary artery disease, diabetes mellitus type II, end-stage renal disease, pulmonary embolism, hyperlipidemia, hypertension, obesity, obstructive sleep apnea, osteoporosis, and substance abuse were more common in the urgent group.

Figure 1a and Fig. 1b display forest plots of odds ratios with 95% confidence intervals for the various structured variables used in the study. These figures illustrate the significant predictors of emergency severity, highlighting the relative importance of different factors. In Fig. 1a, demographic and visit characteristics are detailed. Female patients had higher odds of being classified as urgent (OR=1.15, 95% CI: 1.05-1.26). Age was a significant predictor, with patients aged 40-65 having higher odds of urgency (OR=1.32, 95% CI: 1.21-1.44) compared to those aged 18-39. Patients aged 65 and above had even higher odds (OR=2.13, 95% CI: 1.88-2.41). Arrival by ambulance markedly increased the odds of being urgent (OR=3.65, 95% CI: 3.05-4.36). Medicare coverage was associated with higher odds of urgency (OR = 1.79, 95%CI: 1.58-2.02), while being uninsured was associated with lower odds (OR = 0.75, 95% CI: 0.61-0.91). Patients from nursing homes had higher odds of being classified as urgent (OR=2.80, 95% CI: 1.73-4.54). Figure 1b focuses on clinical information and medical history. Heart rate was a significant predictor, with patients having heart rates over 90 bpm showing higher odds of being urgent (OR=1.56, 95% CI: 1.42-1.72). Blood pressure was also significant; diastolic blood pressure less than 60 mm Hg was associated with higher odds of urgency (OR=1.53, 95% CI: 1.20–1.95), and DBP greater than 80 mm Hg showed increased odds (OR=1.29, 95% CI: 1.17-1.42). Systolic blood pressure greater than 120 mm Hg was associated with higher urgency (OR = 1.10, 95%CI: 1.00–1.22). Several medical conditions significantly increased the odds of being classified as urgent, including cancer (OR = 2.54, 95% CI: 1.91-3.37), chronic kidney disease (OR=2.28, 95% CI: 1.71-3.03), chronic obstructive pulmonary disease (OR=1.78, 95% CI: 1.46-2.17), congestive heart failure (OR = 2.45, 95% CI: 1.88–3.18), coronary artery disease (OR=2.55, 95% CI: 2.03-3.20), end-stage renal disease (OR=3.37, 95% CI: 1.88-6.05), diabetes mellitus type II (OR=1.90, 95% CI: 1.57-2.30), hyperlipidemia (OR=1.63, 95% CI: 1.40-1.90), hypertension (OR=1.73, 95% CI: 1.55-1.93), and obesity (OR = 1.25, 95% CI: 1.08–1.44).

Figure 2 presents the frequency and word cloud of the words in the unstructured variables, providing a visual representation of the most common terms found in the chief complaints and reasons for the injury presented at the ED visits. Table 2 and Fig. 3 summarize the performance metrics for the different models. The results demonstrate that integrating structured and unstructured data leads to improved model performance across all classifiers. Logistic Regression showed significant improvements when combining both data types, achieving an AUC of 0.784, an accuracy of 0.717, and a high precision of 0.894. Random Forest and Gradient Boosting models similarly benefited from the combination, with Random Forest achieving an AUC of 0.766 and Gradient Boosting reaching 0.789. In particular, Gradient Boosting demonstrated strong predictive capabilities with a precision of 0.892 and an F1 score of 0.797. Extreme Gradient Boosting, although slightly weaker with structured data alone, showed notable gains when unstructured data was included, with a combined AUC of 0.779 and a precision of 0.886.

## Discussion

Our study demonstrated that combining structured and unstructured data significantly improved the prediction of emergency severity in an ED setting. By integrating clinical narratives with traditional patient demographics, vital signs, and medical history, we were able to capture a more comprehensive representation of the patient's condition. The results showed that models incorporating both data types outperformed those relying solely on structured or unstructured data. This finding highlights the potential of leveraging advanced NLP techniques, such as BERT, in conjunction with structured clinical data to enhance decision-making in emergency care. While the BERT model effectively captured the contextual nuances in clinical notes, the combined approach proved most robust, supporting the idea that integrating diverse data sources can yield more accurate and actionable predictions in complex medical environments like the ED.

## Association analysis and clinical implications

While the association analysis identified several statistically significant predictors of emergency severity, it is

(See figure on next page.)

**Fig. 1** a Forest Plot of Odds Ratios with 95% CI (Log Scale). b Forest Plot of Odds Ratios with 95% CI (Log Scale). The odds ratios (ORs) presented in Figure. 1 were derived from a logistic regression model where all variables were mutually adjusted. This means that the ORs account for the influence of all other variables included in the model. For example, the OR for age reflects the effect of age on emergency severity while controlling for other factors such as gender, vital signs, and medical history. This mutual adjustment allows for a more accurate estimation of the individual contribution of each variable to the prediction of emergency severity, minimizing potential confounding effects



Fig. 1 (See legend on previous page.)





Fig. 2 Frequency and the word cloud of the word in the unstructured variables

Cutoff	AUC	Accuracy	Precision	Sensitivity	Specificity	F1 Score
0.744	0.704	0.635	0.847	0.623	0.668	0.718
0.783	0.760	0.696	0.882	0.685	0.730	0.771
0.779	0.784	0.717	0.894	0.705	0.753	0.788
0.735	0.787	0.732	0.888	0.733	0.728	0.803
0.726	0.701	0.652	0.845	0.654	0.647	0.737
0.715	0.749	0.696	0.868	0.698	0.687	0.774
0.720	0.766	0.712	0.875	0.716	0.698	0.787
0.718	0.779	0.723	0.882	0.726	0.716	0.797
0.737	0.711	0.651	0.845	0.652	0.649	0.736
0.747	0.763	0.714	0.876	0.719	0.700	0.790
0.759	0.789	0.726	0.892	0.719	0.745	0.797
0.740	0.789	0.729	0.889	0.727	0.734	0.800
0.803	0.678	0.623	0.836	0.615	0.645	0.709
0.897	0.734	0.682	0.866	0.679	0.691	0.761
0.893	0.779	0.714	0.886	0.708	0.733	0.787
0.782	0.759	0.699	0.870	0.702	0.691	0.777
	Cutoff 0.744 0.783 0.779 0.735 0.726 0.715 0.720 0.718 0.737 0.747 0.759 0.740 0.803 0.897 0.893 0.782	Cutoff         AUC           0.744         0.704           0.783         0.760           0.779         0.784           0.735         0.787           0.726         0.701           0.715         0.749           0.726         0.701           0.715         0.749           0.720         0.766           0.718         0.779           0.737         0.711           0.747         0.763           0.759         0.789           0.740         0.789           0.803         0.678           0.897         0.734           0.893         0.779           0.782         0.759	Cutoff         AUC         Accuracy           0.744         0.704         0.635           0.783         0.760         0.696           0.779         0.784         0.717           0.735         0.787         0.732           0.726         0.701         0.652           0.715         0.749         0.696           0.720         0.766         0.712           0.715         0.749         0.696           0.720         0.766         0.712           0.718         0.779         0.723           0.737         0.711         0.651           0.749         0.6951         0.749           0.737         0.789         0.726           0.740         0.789         0.726           0.740         0.789         0.729           0.803         0.678         0.623           0.897         0.734         0.682           0.893         0.779         0.714           0.789         0.714         0.699	Cutoff         AUC         Accuracy         Precision           0.744         0.704         0.635         0.847           0.783         0.760         0.696         0.882           0.779         0.784         0.717         0.894           0.735         0.787         0.732         0.882           0.735         0.787         0.732         0.884           0.726         0.701         0.652         0.845           0.715         0.749         0.696         0.868           0.720         0.766         0.712         0.875           0.718         0.779         0.723         0.882           0.720         0.766         0.712         0.875           0.718         0.779         0.723         0.882           0.737         0.711         0.651         0.845           0.747         0.763         0.714         0.876           0.749         0.789         0.726         0.892           0.740         0.789         0.729         0.889           0.803         0.678         0.623         0.836           0.897         0.734         0.682         0.866           0.893	Cutoff         AUC         Accuracy         Precision         Sensitivity           0.744         0.704         0.635         0.847         0.623           0.783         0.760         0.696         0.882         0.685           0.779         0.784         0.717         0.8944         0.705           0.735         0.787         0.732         0.888         0.733           0.726         0.701         0.652         0.845         0.654           0.715         0.749         0.696         0.868         0.698           0.720         0.766         0.712         0.875         0.716           0.718         0.799         0.723         0.882         0.726           0.718         0.799         0.723         0.875         0.716           0.718         0.799         0.723         0.882         0.726           0.737         0.711         0.651         0.845         0.652           0.747         0.763         0.714         0.876         0.719           0.759         0.789         0.726         0.892         0.719           0.740         0.789         0.729         0.886         0.615	Cutoff         AUC         Accuracy         Precision         Sensitivity         Specificity           0.744         0.704         0.635         0.847         0.623         0.668           0.783         0.760         0.696         0.882         0.685         0.730           0.779         0.784         0.717         0.894         0.705         0.753           0.735         0.787         0.732         0.888         0.733         0.728           0.726         0.701         0.652         0.845         0.654         0.647           0.715         0.749         0.696         0.868         0.698         0.687           0.715         0.749         0.696         0.868         0.698         0.687           0.720         0.766         0.712         0.875         0.716         0.698           0.718         0.779         0.723         0.882         0.726         0.649           0.747         0.763         0.714         0.876         0.719         0.700           0.749         0.726         0.892         0.719         0.745           0.740         0.789         0.726         0.892         0.719         0.745

crucial to differentiate between statistical significance and clinical relevance. Chronic conditions such as coronary artery disease, chronic kidney disease, and chronic obstructive pulmonary disease were significant predictors in our model. However, the practical relevance of these findings for real-time decision-making in ED settings should be critically examined. For instance, while the presence of chronic conditions may inform longterm risk stratification, their immediate impact on triage decisions may be limited unless the condition is actively



**Fig. 3** Receiver Operating Characteristic (ROC) curves for the four models evaluated in the study. Each model include the structured data model, which uses only structured data such as patient demographics, visit characteristics, vital signs, and medical history; the unstructured data model, a BERT-based natural language processing (NLP) model that uses only unstructured data, including chief complaints and reasons for injury; the combined input model, a machine learning classification model that integrates both structured data and BERT-extracted features from the unstructured data; and the mean probability model, which averages the predicted probabilities from the structured data model and the unstructured data model

contributing to the acute presentation. Thus, although these conditions were associated with increased urgency, further research is needed to explore their practical role in ED triage processes.

In addition, older age and higher heart rate emerged as significant predictors, aligning with clinical expectations that elderly patients and those with abnormal vital signs require urgent attention. However, it is important to interpret these findings with caution, particularly in the context of retrospective analysis. While our model can identify factors associated with higher acuity, it does not substitute clinical judgment, which remains critical in real-time decision-making. The inclusion of unstructured data, particularly chief complaints, offers a way to incorporate nuanced patient information that is often missing in structured data, thus improving the predictive accuracy of the models.

A notable finding was the association between insurance status and emergency severity. Patients covered by Medicare had higher odds of being classified as urgent, while uninsured patients were less likely to be classified as urgent. This result raises important questions regarding access to care and its influence on triage outcomes. One potential explanation is that uninsured patients may delay seeking care due to financial concerns, leading to underrepresentation in our dataset or potentially presenting with less acute conditions. Alternatively, these findings may reflect broader disparities in healthcare access and utilization, where insurance status influences not only access to primary care but also ED triage decisions [33–35]. The association between Medicare coverage and higher urgency might reflect the higher baseline health risks of the elderly population, who are more likely to suffer from multiple comorbidities. Further exploration of how insurance status interacts with other social determinants of health, such as socioeconomic status and healthcare access, is warranted. Future studies should aim to validate these findings and examine whether controlling for other factors, such as pre-existing health conditions, changes the relationship between insurance status and triage classification. Moreover, this finding highlights the need for ED policies that address potential biases in triage based on insurance status and other social determinants of health.

### Model performance and clinical application

Our results demonstrated that integrating structured and unstructured data improves the performance of predictive models, particularly in complex cases where traditional triage systems may fall short. The Gradient Boosting and Extreme Gradient Boosting models achieved the highest performance, with AUCs of 0.789 and 0.779, respectively, when both data types were combined. The strong performance of these models underscores the value of using machine learning techniques that can account for non-linear interactions and complex relationships between variables, which are often present in clinical data.

Our findings can be compared to the findings of Brouns et al. (2019) [36] and Veldhuis et al. (2022) [37]. Brouns et al. evaluated the Manchester Triage System in older emergency department patients, reporting an AUC of 0.74 for predicting hospital admissions, a result similar to the AUCs achieved by our combined data models. However, their study noted that MTS had a lower AUC of 0.71 for predicting in-hospital mortality, highlighting the limitations of relying solely on structured triage systems in medically complex populations. Our results demonstrate that combining structured and unstructured data can address some of these limitations by improving predictive accuracy, particularly in more complex cases. Similarly, Veldhuis et al. compared clinical judgment to early warning scores and found that clinical judgment outperformed risk stratification models, with AUCs between 0.70 and 0.89, especially for ICU admissions and severe adverse events [37]. While our models performed similarly, this emphasizes the need to integrate machine learning with clinical judgment. Our models, combining structured and unstructured data, outperformed singlesource models, aligning with Veldhuis et al.'s suggestion that clinical tools combined with automated systems yield the best results.

In comparison with traditional triage systems, such as the Manchester Triage System [36], our models show promise in enhancing predictive accuracy by leveraging a broader range of patient data, particularly unstructured clinical narratives. However, it is essential to emphasize that clinical judgment remains a critical component of ED decision-making. Predictive models, while valuable, should complement—not replace—the expertise of healthcare providers, who are best equipped to make nuanced decisions in real-time clinical settings.

### Limitations and future directions

There are several limitations to this study. First, the study is retrospective and relies on the accuracy and completeness of the NHAMCS-ED dataset. Any missing or inaccurately recorded data could impact the model's performance. Although the proportion of missing data was relatively low (<10%), the method of imputation (median) might not capture the true underlying values in all cases, and different imputation techniques could lead to slightly different results. Second, the study focuses on data from a single year (2021), which may limit the generalizability of the findings to other years or different hospital settings. Emergency presentations can vary over time due to factors such as seasonal changes, pandemics, or other public health events. Future studies should validate these findings with data from multiple years and diverse clinical environments to ensure the robustness and applicability of the models across varying contexts. Third, while BERT proved highly effective in processing unstructured clinical text, it is computationally intensive compared to simpler models such as TF-IDF or logistic regression. The complexity and resource demands of BERT may limit its use in real-time ED settings, particularly in resource-constrained environments. For realtime applications, it may be beneficial to explore lighter models like DistilBERT [38] or other simplified NLP approaches that balance computational efficiency with performance. Additionally, another important limitation involves the potential bias inherent in machine learning models [39]. Bias can emerge from the data used to train the model, particularly if the dataset reflects existing disparities in healthcare access, treatment, or outcomes. For instance, the underrepresentation of uninsured patients in the dataset may skew the model's ability to predict outcomes for this group, potentially reinforcing inequities in healthcare delivery. Furthermore, models trained on past data may perpetuate historical biases in clinical decision-making, such as differences in treatment recommendations based on race, gender, or insurance status. Addressing this issue will require careful evaluation of the model's performance across diverse patient populations and the implementation of fairness-enhancing techniques, such as bias mitigation algorithms, to ensure that the model does not exacerbate existing healthcare disparities.

Future research should explore the integration of additional data sources, such as imaging and laboratory results, to further enhance predictive models [40]. These data sources could provide additional valuable information that can improve the accuracy of severity predictions. Prospective studies are also needed to validate the performance of these models in real-time clinical settings. Implementing these models in actual ED workflows and assessing their impact on clinical outcomes and operational efficiency will provide crucial insights [41]. Exploring the use of other advanced NLP models and techniques could yield further improvements in handling unstructured data. For example, models that incorporate contextual embeddings or use transfer learning from larger clinical datasets could enhance the performance of text-based predictions. Additionally, investigating the specific contributions of different types of unstructured data to the model's performance can provide valuable insights for improving triage protocols and decisionmaking processes in EDs.

## Conclusion

The integration of structured and unstructured data shows promise in enhancing the prediction of emergency severity in ED settings. By leveraging advanced NLP techniques and comprehensive data sources, healthcare providers may improve the accuracy of severity predictions, potentially leading to more informed resource allocation and better patient outcomes. However, while these findings are encouraging, further validation in diverse clinical environments is necessary before definitive claims can be made about the model's broader applicability. Prospective studies involving multi-year datasets and real-world implementation will be essential to confirm the impact on clinical decision-making. This study underscores the potential of combining diverse data types to support predictive modeling, but caution is required when interpreting the results, given the need for additional validation. The findings contribute to the growing body of work on data integration in healthcare and suggest that this approach has the potential to support clinical decisionmaking in emergency settings, pending further research.

### Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s12911-024-02793-9.

Supplementary Material 1.

#### Acknowledgements

None.

#### Authors' contributions

X.Z. and W.Z. conceived and designed the study. X.Z., Y.W., and Y.J. drafted the manuscript. C.B.P. provided clinical insights and contributed to the interpretation of the results. All authors reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

#### Funding

This research was supported by internal funding allocated to Dr. Xingyu Zhang from the Department of Communication Science and Disorders at the University of Pittsburgh.

#### Data availability

The NHAMCS-ED dataset can be accessed through the website of the US Centers for Disease Control and Prevention (CDC) (https://www.cdc.gov/nchs/ ahcd/index.htm). The detailed explanation of the survey data for each year and the code book can be found here: https://ftp.cdc.gov/pub/Health\_Stati stics/NCHS/dataset\_documentation/nhamcs/. The SAS dataset for each year can be found here: https://ftp.cdc.gov/pub/Health\_Statistics/NCHS/Datasets/ NHAMCS/. All the data were in a SAS format. To get the unstructured data, one needs to run the SAS format files under the following link before import the data to the analysis software. https://ftp.cdc.gov/pub/Health\_Statistics/NCHS/ dataset\_documentation/nhamcs/sas/

https://ftp.cdc.gov/pub/Health\_Statistics/NCHS/dataset\_documentation/nhamcs/.

The SAS dataset for each year can be found here:

https://ftp.cdc.gov/pub/Health\_Statistics/NCHS/Datasets/NHAMCS/.

All the data were in a SAS format. To get the unstructured data, one needs to run the SAS format files under the following link before import the data to the analysis software.

https://ftp.cdc.gov/pub/Health\_Statistics/NCHS/dataset\_documentation/nhamcs/sas/.

#### Declarations

#### Ethics approval and consent to participate

Not applicable, as this study utilized publicly available, anonymized data from the NHAMCS-ED dataset.

Not applicable, as the research was conducted using publicly available, anonymized data from the NHAMCS-ED dataset.

#### **Consent for publication**

Not applicable, as this study utilized publicly available, anonymized data.

#### **Competing interests**

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Communication Science and Disorders, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, USA. <sup>2</sup>Department of Health Information Management, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, USA. <sup>3</sup>School of Nursing, University of Michigan, Ann Arbor, MI, USA. <sup>4</sup>Department of Emergency Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. <sup>5</sup>Knight Foundation School of Computing & Information Sciences, Florida International University, Miami, USA.

### Received: 18 August 2024 Accepted: 28 November 2024 Published online: 05 December 2024

#### References

- 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.
- Mostafa R, El-Atawi K. Strategies to measure and improve emergency department performance: a review. Cureus. 2024;16(1):e52879.

- Ahsan KB, Alam M, Morel DG, Karim M. Emergency department resource optimisation for improved performance: a review. J Indust Eng Int. 2019;15(Suppl 1):253–66.
- 4. Yancey CC, O'Rourke MC: Emergency department triage. 2020.
- Christ M, Grossmann F, Winter D, Bingisser R, Platz E. Modern triage in the emergency department. Dtsch Arztebl Int. 2010;107(50):892–8.
- 6. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. Acad Emerg Med. 2000;7(3):236–42.
- Chiu CC, Wu CM, Chien TN, Kao LJ, Li C, Chu CM. Integrating Structured and Unstructured EHR Data for Predicting Mortality by Machine Learning and Latent Dirichlet Allocation Method. Int J Environ Res Public Health. 2023;20(5):4340.
- Zhang X, Bellolio MF, Medrano-Gracia P, Werys K, Yang S, Mahajan P. Use of natural language processing to improve predictive models for imaging utilization in children presenting to the emergency department. BMC Med Inform Decis Mak. 2019;19(1):287.
- 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(05):377–89.
- Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Dig Med. 2021;4(1):86.
- Tang R, Yao H, Zhu Z, Sun X, Hu G, Li Y, Xie G: Embedding Electronic Health Records to Learn BERT-based Models for Diagnostic Decision Support. In: 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI): 9–12 Aug. 2021 2021; 2021: 311–319.
- Lu H, Ehwerhemuepha L, Rakovski C. A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. BMC Med Res Methodol. 2022;22(1):181.
- Turchin A, Masharsky S, Zitnik M. Comparison of BERT implementations for natural language processing of narrative medical documents. Inform Med Unlocked. 2023;36:101139.
- Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep. 2016;6(1):26094.
- Suresh H, Hunt N, Johnson AEW, Celi LA, Szolovits P, Ghassemi M: Clinical Intervention Prediction and Understanding using Deep Networks. ArXiv 2017, abs/1705.08498.
- Su D, Li Q, Zhang T, Veliz P, Chen Y, He K, Mahajan P, Zhang X. Prediction of acute appendicitis among patients with undifferentiated abdominal pain at emergency department. BMC Med Res Methodol. 2022;22(1):18.
- Stewart J, Lu J, Goudie A, Arendts G, Meka SA, Freeman S, Walker K, Sprivulis P, Sanfilippo F, Bennamoun M, et al. Applications of natural language processing at emergency department triage: A narrative review. PLoS ONE. 2023;18(12):e0279953.
- Cairns C, Kang K: National hospital ambulatory medical care survey: 2019 emergency department summary tables. 2022.
- Eitel DR, Travers DA, Rosenau AM, Gilboy N, Wuerz RC. The emergency severity index triage algorithm version 2 is reliable and valid. Acad Emerg Med. 2003;10(10):1070–80.
- Green NA, Durani Y, Brecher D, DePiero A, Loiselle J, Attia M. Emergency Severity Index version 4: a valid and reliable tool in pediatric emergency department triage. Pediatr Emerg Care. 2012;28(8):753–7.
- Tanabe P, Gimbel R, Yarnold PR, Adams JG: The Emergency Severity Index (version 3) 5-level triage system scores predict ED resource consumption. J Emerg Nurs 2004;30(1):22–29.
- 22. Hinson JS, Martinez DA, Schmitz PS, Toerper M, Radu D, Scheulen J, Stewart de Ramirez SA, Levin S: 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. International journal of emergency medicine 2018, 11:1-10.
- Alnasser S, Alharbi M, AAlibrahim A, Aal Ibrahim A, Kentab O, Alassaf W, Aljahany M. Analysis of Emergency Department Use by Non-Urgent Patients and Their Visit Characteristics at an Academic Center. Int J Gen Med. 2023;16:221–32.
- Zollanvari A: Supervised Learning in Practice: the First Application Using Scikit-Learn. In: Machine Learning with Python: Theory and Implementation. edn.: Springer; 2023: 111–131.
- Jain SM: Hugging face. In: Introduction to transformers for NLP: With the hugging face library and models to solve problems. edn.: Springer; 2022: 51–67.

- Deepa MD. Bidirectional encoder representations from transformers (BERT) language model for sentiment analysis task. Turkish J Comput Math Educ. 2021;12(7):1708–21.
- 27. Alaparthi S, Mishra M: Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey. arXiv preprint arXiv:200701127 2020.
- Parmar A, Katariya R, Patel V: A review on random forest: An ensemble classifier. In: International conference on intelligent data communication technologies and internet of things (ICICI) 2018: 2019: Springer; 2019: 758–763.
- 29. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot. 2013;7:21.
- 30. Chen T: Xgboost: extreme gradient boosting. R package version 04–2 2015, 1(4).
- Zhou J, Gandomi AH, Chen F, Holzinger A. Evaluating the quality of machine learning explanations: A survey on methods and metrics. Electronics. 2021;10(5):593.
- Naidu G, Zuva T, Sibanda EM: A review of evaluation metrics in machine learning algorithms. In: Computer Science On-line Conference: 2023: Springer; 2023: 15–25.
- Zhang X, Carabello M, Hill T, Bell SA, Stephenson R, Mahajan P. Trends of racial/ethnic differences in emergency department care outcomes among adults in the United States from 2005 to 2016. Front Med. 2020;7:300.
- Myran D, Hsu A, Kunkel E, Rhodes E, Imsirovic H, Tanuseputro P. Socioeconomic and geographic disparities in emergency department visits due to alcohol in Ontario: a retrospective population-level study from 2003 to 2017. Can J Psychiatry. 2022;67(7):534–43.
- Pierce A, Marquita Norman M, Rendon J, Rucker D, Velez L, Powers R: Health Disparities in the Emergency Department. Emerg Med Rep 2021;42(20).
- Brouns SH, Mignot-Evers L, Derkx F, Lambooij SL, Dieleman JP, Haak HR. Performance of the Manchester triage system in older emergency department patients: a retrospective cohort study. BMC Emerg Med. 2019;19:1–11.
- Veldhuis LI, Ridderikhof ML, Bergsma L, Van Etten-Jamaludin F, Nanayakkara PW, Hollmann M. Performance of early warning and risk stratification scores versus clinical judgement in the acute setting: a systematic review. Emerg Med J. 2022;39(12):918–23.
- Adoma AF, Henry N-M, Chen W: Comparative analyses of bert, roberta, distilbert, and xInet for text-based emotion recognition. In: 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP): 2020: IEEE; 2020: 117–121.
- Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM computing surveys (CSUR). 2021;54(6):1–35.
- Zhang X, Bellolio MF, Medrano-Gracia P, Werys K, Yang S, Mahajan P. Use of natural language processing to improve predictive models for imaging utilization in children presenting to the emergency department. BMC Med Inform Decis Mak. 2019;19:1–13.
- 41. Chan SL, Lee JW, Ong MEH, Siddiqui FJ, Graves N, Ho AFW, Liu N. Implementation of prediction models in the emergency department from an implementation science perspective—determinants, outcomes, and real-world impact: a scoping review. Ann Emerg Med. 2023;82(1):22–36.

# **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.