



Machine learning models to detect opioid misuse in emergency department patients at triage

Chirag Chhablani^{a,b}, Usman Shahid^{b,c}, Natalie Parde^{b,c}, Sami Muslmani^d, Huiyi Hu^b, Dillon Thorpe^b, Majid Afshar^e, Niranjan Karnik^{b,f,g}, Neeraj Chhabra^{a,b,g,*}

^a Department of Emergency Medicine, University of Illinois Chicago, Chicago, IL, United States of America

^b ALHealth4All Center, University of Illinois Chicago College of Medicine, Chicago, IL, United States of America

^c Department of Computer Science, University of Illinois Chicago, Chicago, IL, United States of America

^d University of Illinois Chicago College of Medicine, Chicago, IL, United States of America

^e Department of Medicine, School of Medicine and Public Health, University of Wisconsin Madison, Madison, WI, United States of America

^f Department of Psychiatry, University of Illinois Chicago, Chicago, IL, United States of America

^g Institute for Research on Addictions, University of Illinois Chicago, Chicago, IL, United States of America

ARTICLE INFO

Article history:

Received 20 November 2025

Received in revised form 23 January 2026

Accepted 25 February 2026

Keywords:

Emergency medicine

Opioid use disorder

Opioid misuse

Machine learning

ABSTRACT

Objective: Emergency department (ED) encounters represent valuable opportunities to initiate evidence-based treatments for patients with opioid misuse, but few receive such care. Universal manual screening has been proposed to improve patient identification but is uncommon due to its time and resource-intensive nature. We sought to determine the feasibility of identifying patients with opioid misuse at the time of ED triage using machine learning (ML).

Methods: We conducted a retrospective cohort study of 1123 ED encounters (September 2020 – March 2023) at a tertiary hospital. Encounters were enriched for opioid misuse, manually annotated, and chronologically split for training, validation, and testing. Candidate triage-time features included patient demographics, Emergency Severity Index, arrival time of day, chief complaint, comorbidities, and chronic medications. Model performance was evaluated using F1 score, area under the precision–recall curve (AUPRC), accuracy, recall, and AUROC. Post-hoc explainability analyses included SHapley Additive exPlanations (SHAP) and feature importance.

Results: All models performed comparably to opioid-related diagnosis codes placed at any time during the encounter. Random Forest (F1 = 0.75 [95%CI 0.70–0.83], AUPRC = 0.88 [0.81–0.93], accuracy = 0.79 [0.70–0.83]) and Gradient Boosting (F1 = 0.77 [0.71–0.82], AUPRC = 0.89 [0.85–0.93], accuracy = 0.81 [0.720.84]) had among the highest F1 score and AUPRC but confidence intervals overlapped with other methods. Explainability analyses highlighted prior drug-use diagnosis codes, triage acuity, and age as top predictors.

Conclusion: ML classifiers leveraging routinely collected triage codes data offer a feasible and scalable alternative to manual screening in flagging opioid misuse before physician evaluation, potentially enabling early harm-reduction interventions. Prospective multi-site validation, calibration, and bias assessments are warranted.

© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Despite recent progress, the opioid crisis remains a public health emergency in the United States with overdose the leading cause of death for multiple age groups [1,2]. The emergency department (ED) is a common first contact point between patients with opioid misuse and the broader healthcare system. Opioid-related ED encounters

have increased with approximately 2 million opioid-related ED encounters between 2022 and 2024 [3]. Patients with opioid misuse disproportionately access emergency medical services, making these encounters valuable opportunities for the initiation of treatment and harm reduction measures to prevent future fatal overdose [4,5]. Accurate identification of opioid misuse during ED encounters is essential for treatment, care coordination, and resource allocation.

Despite the existence of evidence-based therapies, only a minority of ED patients with opioid misuse are offered treatment due to a variety of factors including varied patient presentations, receptivity to treatment, stigma, time constraints of care teams, and difficulties in patient identification and prioritization [6–8]. While many patients with opioid-

* Corresponding author at: Department of Emergency Medicine, University of Illinois Chicago, Neeraj Chhabra, 808 S Wood, 4th Floor, Chicago, IL 60612, United States of America.

E-mail address: neeraj1@uic.edu (N. Chhabra).

<https://doi.org/10.1016/j.ajem.2026.02.037>

0735-6757/© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

related complaints are relatively easy for clinicians to identify, such as those with clear opioid toxidrome or withdrawal, others are more difficult or present with complaints unrelated to drug use. Similarly, stigma may have a chilling effect on patients disclosing substance use. Universal manual screening for ED patients has been advocated to improve detection of patient with opioid misuse but has not been adopted on a broad scale as results from experiments have been mixed [9–11]. Universal manual screening of patients entering the ED is a resource-intensive process requiring staff training and clinical workflow development which may conflict with other time-sensitive tasks and manual screening tools for other disease processes which have already been implemented. Automated identification of patients via machine learning (ML) may offer a scalable solution that does not take staff and resources away from other critical tasks [12]. Language model-based approaches have shown promising results in identifying opioid misuse from clinical notes [13–16]. These natural language processing-based algorithms for identifying opioid misuse require note processing and detection following clinical note completion which may only occur after a patient has physically left the ED. This limits the ability of ED staff to address dangerous opioid use while the patient is present in the ED. There remains a need for models designed for use early during an ED encounter allowing for interventions and care coordination in real-time.

The objective of this study was to examine the ability of machine learning classifiers in detecting opioid misuse using data already available or collected at the time of ED patient triage. We hypothesize that ML classifiers can accurately detect patients with opioid misuse, suggesting a possible role for automated screening at patient arrival in the ED immediately following the triage process.

2. Methods

2.1. Study design and cohort development

This is a retrospective cohort study of ED patients at a single tertiary care center from September 2020 to March 2023. A cohort of 1200 ED encounters and corresponding data were extracted from the clinical data warehouse of the primary study institution. The sample included patients aged 18 years or older at the time of encounter. The sample was enriched for opioid misuse using previously described methods involving sampling ED encounters with an opioid-related International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) diagnosis code or a positive urine opiate screen without coexisting opioid prescription or administration and matching 1:1 with encounters lacking these criteria [13,17,18]. Enrichment was pursued to create a balanced dataset for efficient model training as the primary study institution has a baseline proportion of opioid misuse among ED patients of 3–6%. Matching was performed on disposition status rather than age and sex to minimize bias from expanded clinical documentation for admitted patients, which may have affected the manual labeling process even though model features were limited to triage-only data. The cohort was temporally split for model training, validation, and testing. Urine results and ICD-10-CM codes were used solely to enrich the sample and final labeling was done by manual annotation for opioid misuse. Due to the low sensitivity of ICD-10-CM codes for detecting opioid misuse in the ED setting, we used expert manual annotation to assign ground truth labels for model training and testing [17]. All included encounters, regardless of urine toxicology testing status or results, underwent expert manual adjudication to determine the presence or absence of opioid misuse based on comprehensive review of the clinical record. Opioid misuse is defined as use of opioids in any way that was not directed by a doctor and represents a heterogeneous pattern of drug use that is difficult to categorize solely from structured variables in the EHR [19]. We therefore followed a blinded and structured manual annotation schema using a definition for opioid misuse consistent with those from the National Institute of Drug Abuse (NIDA) and the National Survey on Drug Use and Health (NSDUH)

[20,21]. Annotators underwent training and evaluation of interrater reliability with an expert prior to independent annotation and following completion of every 100 cases. The study was exempted as non-human subjects research by the primary study institution, was not pre-registered, and conforms, where appropriate, to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis, Artificial Intelligence (TRIPOD + AI) guidelines (Supplementary material) [22].

2.2. Data preprocessing

The dataset used in this study included a set of features available at the time of ED triage or entered during the triage process (Table 1). These features included both patient and encounter-specific variables. Patient-specific variables included demographics, past medical history, and medication lists. Encounter-specific variables included triage acuity, arrival time of day, and chief complaint. Triage acuity was measured by the Emergency Severity Index (ESI) which is used by the majority of EDs in the United States. The ESI is measured as a 5-point scale categorizing patients based on severity of their conditions from level 1 (most urgent) to level 5 (least urgent) [23,24].

All variables except for chief complaint represent structured data fields. Free-text chief complaints were preprocessed to enhance clinical interpretability and generalizability. We developed an expert-derived lexicon of keywords to map each complaint to one of several broad clinical categories: respiratory, cardiac, gastrointestinal, mental health, neurological, musculoskeletal, dermatological, drug use, pain/injury, genitourinary, ophthalmological, infectious disease, and other (see supplementary material). This keyword rule-based categorization allowed us to transform diverse textual complaints into a simple structured format that could be used as easily interpretable model inputs, while also offering a low-complexity approach requiring minimal computing resources that could be readily adapted to other settings. For medical history, we evaluated ICD-10-CM codes recorded prior to the ED arrival time, including those categorized as “Medical History” or as part of the “Problem List” with a designation of active. To reduce the dimensionality of medical history codes, we grouped codes into clinical categories using a standardized classification defined by the Elixhauser Comorbidity Index (ECI). The ECI maps comorbid diagnosis codes to 38 pre-existing condition categories and represents comorbidities that co-exist at the time of hospital encounter [25–27]. Each category was converted into a binary feature, resulting in a structured representation of the patient’s diagnostic history leading up to the ED visit. Preprocessing categories for chief complaint and medical history are in the supplementary materials.

Each patient’s medication list was cleaned by removing dosage, frequency, and quantity. To reduce dimensionality, resulting medications

Table 1
Candidate features for model inclusion.

| Feature Name | Description (Value Format) |
|----------------------|---|
| Age | Patient’s age in years (numerical) |
| ED Acuity | 5-point scale indicating acuity of patient presentation (Emergency Severity Index) |
| Arrival category | Time of arrival at emergency department (categorical); categorized into: 0 = Day (6 AM–1:59 PM) 1 = Evening (2 PM–9:59 PM) 2 = Night (10 PM–5:59 AM) |
| Chief Complaint | Patient’s presenting complaint (categorical), categorized into: Respiratory, Cardiac, Mental Health, Neurological, Gastrointestinal, Musculoskeletal, Genitourinary, Dermatological, Overdose, Pain/Injury, Ophthalmological, Infectious Disease, Other |
| Past Medical History | Binary vector indicating presence of 38 diagnosis categories assigned prior to ED arrival (Problem list or medical history) |
| Medication History | Binary vector indicating presence of medication categories (categorical vector with values 0 or 1) |

were grouped into therapeutic classes defined by the United States Pharmacopeial Convention (USP) Drug Classification 2025 updated schema [28]. Temporal patterns in ED arrivals were also encoded to capture potential circadian effects. The ED arrival time was discretized into three categories: day (6 AM to 1:59 PM), evening (2 PM to 9:59 PM), and night (10 PM to 5:59 AM). There was no missingness noted in variables apart from the ED acuity field, which contained 2 missing values (0.18% of the dataset). Given the minimal degree of missingness, we used replacement with median value. Results were robust to other methods of imputation (Supplement).

2.3. Model development and explainability

The model training/testing cohort was developed using the sample of 1200 ED encounters, of which 77 were excluded due to enhanced privacy protections (62) or lack of any clinical documentation in the clinical encounter to allow for manual label assignment (15), leaving 1123 for analysis. Encounters were sorted chronologically by date and divided into training, validation, and testing sets at a 70:15:15 ratio with the test set representing the most recent encounters to provide temporal validation and diminish the effects of selection bias due to COVID-related changes in healthcare utilization. We evaluated a diverse set of statistical and machine learning classifiers, including logistic regression with elastic net regularization, support vector machines (SVM), decision trees, K-nearest neighbors (KNN), naive bayes, bagging, random forest, gradient boosting (with different estimator settings), and AdaBoost. We also evaluated a feedforward neural network (FFNN). For our FFNN, we employed a fully connected feedforward neural network comprising three hidden layers with 128, 64, and 32 neurons respectively, each followed by ReLU activations and dropout for regularization. The final output layer produced a single logit for binary classification. Final model parameters are included in the supplementary materials. Models were evaluated along multiple metrics with the primary being Area Under the Precision Recall Curve (AUPRC), F1 score, precision (positive predictive value), recall (sensitivity), accuracy, and area under the receiver operator curve (AUROC). We used model classification threshold probability = 0.5 across all the classifiers. Confidence intervals were estimated using bootstrap resampling with replacement ($n = 1000$ iterations). The 95% confidence intervals were defined as the 2.5th and 97.5th percentiles of the bootstrap distribution. Calibration curves were created for each model and included in the supplementary material. To provide context for the performance of our triage-only data models, we also calculated the performance of opioid-related ICD-10-CM diagnosis codes that were placed at any time during the encounter [29].

We performed explainability analyses on the best performing classifier, selected by F1 score. Using SHAP (SHapley Additive exPlanations), we visualized the influence of individual features on model predictions in a beeswarm plot, with each point indicating one encounter. SHAP allows for both global and individual-level interpretability, showing the direction and magnitude of effect normalized to a 0–1 scale. We also examined global feature importance using random forest's built-in Gini metric from scikit-learn, which ranks features by their contribution to overall classification [30]. While Gini scores reflect general model behaviour, SHAP reveals how specific features drive individual predictions. Analyses were performed using Python (version 3.11). Given the use of protected health information, data cannot be shared but code will be shared with reasonable request to the study team.

3. Results

There were 1123 ED encounters included for analysis with the most recent 169 encounters excluded from model training as the test set. Demographic information of the cohort is shown in Table 2. The opioid misuse cohort contained a greater proportion of male, White, Black, non-Hispanic, and Medicaid insured patients than the cohort without

Table 2
Demographic features.

| Feature | Opioid misuse ($n = 570$) | No opioid misuse ($n = 568$) |
|--------------------------------------|--------------------------------|-----------------------------------|
| Age (mean, SD) | 51.3 (14.0) | 46.5 (19.0) |
| Sex - Male | 379 (66.5%) | 207 (36.4%) |
| Sex - Female | 191 (33.5%) | 361 (63.6%) |
| Race - American Indian/Alaska Native | 1 (0.2%) | 3 (0.5%) |
| Race - Asian | 2 (0.4%) | 21 (3.5%) |
| Race - Black/African American | 347 (60.9%) | 308 (54.2%) |
| Race - White | 122 (21.4%) | 63 (11.1%) |
| Race - Hispanic | 88 (15.4%) | 169 (29.8%) |
| Race - Patient Declined | 10 (1.8%) | 4 (0.7%) |
| Ethnicity - Non-Hispanic or Latino | 465 (81.6%) | 388 (68.3%) |
| Ethnicity - Hispanic or Latino | 91 (16.0%) | 167 (29.4%) |
| Ethnicity - Patient Declined | 4 (0.7%) | 7 (1.2%) |
| Ethnicity - Unknown | 10 (1.8%) | 6 (1.1%) |
| Insurance Type - Private Insurance | 14 (2.5%) | 119 (21.0%) |
| Insurance Type - Medicaid | 395 (69.3%) | 281 (49.5%) |
| Insurance Type - Medicare | 126 (22.1%) | 119 (21.0%) |
| Insurance Type - Other | 10 (1.8%) | 8 (1.4%) |
| Insurance Type - No Information | 25 (4.4%) | 41 (7.2%) |

opioid misuse. Most of the triage-data only models performed comparably to the performance of ICD-10-CM codes placed at any time during the encounter (Table 3). Among machine learning models, ensemble methods such as Random Forest (Accuracy = 0.79, F1 Score = 0.75, AUROC = 0.87, AUPRC = 0.88) and Gradient Boosting (Accuracy = 0.81, F1 Score = 0.77, AUROC = 0.88, AUPRC = 0.89) achieved among the best overall performance of the classic ML methods but confidence intervals of most performance metrics overlapped with other model types including those for simpler models such as logistic regression.

Model explainability analyses were performed on the random forest model (Figs. 1 and 2). As the gradient boosting model had the highest AUPRC and FFNN had similar F1 score, we included the SHAP analyses for those models in the supplementary material. The presence of a prior drug use diagnosis code was top predictor across models. ED acuity and patient age were also among the top predictive features. Prior diagnosis codes for uncomplicated and complicated diabetes were negative predictors for opioid misuse while chronic pulmonary disease, alcohol abuse, and depression were positive predictive features. Non-daytime ED arrival times were also predictive of opioid misuse.

4. Discussion

These results demonstrate the ability of statistical and machine learning models to identify opioid misuse among ED patients using only data available within the EHR following patient triage. The sensitivity of models rivaled those of ICD-10-CM codes placed at any point during the encounter. Among the models tested, tree-based approaches like gradient boosting and random forest, achieved the best performance on the held-out test set but confidence intervals overlapped with other approaches. Feature importance evaluations revealed that pre-existing substance use-related diagnosis codes were the top predictors for opioid misuse. Given the limited sample of this pilot study, these results should be considered preliminary and confirmed in larger-scale evaluations on true prevalence for screening.

Real-world utility of such models in emergency medicine requires further exploration. Because models described here rely on data collected at the time of triage, model-informed interventions resulting could theoretically be deployed at anytime following the triage process. Potential interventions include best practice alerts to ED clinicians, social workers, peer recovery specialists, or addiction medicine colleagues to encourage manual evaluations for OUD, initiation of MOUD, or provision of take-home naloxone when indicated. Further evaluation is needed to determine the optimal timing, modality, and governance of such interventions during an ED course. Although some models had

Table 3
Model performance using triage-only data.

| Model | Accuracy | Precision | Recall | F-1 Score | AUROC | AUPRC |
|---------------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| ICD-10-CM Baseline* | 0.87 | 0.96 | 0.73 | 0.83 | – | – |
| Support Vector Machine | 0.75 (0.70–0.77) | 0.74 (0.65–0.81) | 0.72 (0.68–0.78) | 0.73 (0.67–0.77) | – | – |
| Logistic Regression | 0.76 (0.71–0.77) | 0.73 (0.66–0.77) | 0.71 (0.70–0.77) | 0.74 (0.70–0.77) | 0.84 (0.81–0.86) | 0.84 (0.81–0.86) |
| With ElasticNet | 0.79 (0.70–0.83) | 0.80 (0.68–0.87) | 0.71 (0.67–0.74) | 0.75 (0.70–0.83) | 0.87 (0.82–0.92) | 0.88 (0.81–0.93) |
| Random Forest | 0.76 (0.70–0.80) | 0.75 (0.71–0.80) | 0.71 (0.60–0.77) | 0.73 (0.67–0.78) | 0.80 (0.75–0.84) | 0.81 (0.76–0.84) |
| Bagging | 0.81 (0.72–0.84) | 0.84 (0.76–0.87) | 0.71 (0.61–0.76) | 0.77 (0.71–0.82) | 0.88 (0.83–0.93) | 0.89 (0.85–0.93) |
| Gradient Boosting | 0.69 (0.61–0.76) | 0.64 (0.53–0.73) | 0.75 (0.67–0.80) | 0.69 (0.61–0.72) | 0.75 (0.72–0.80) | 0.68 (0.62–0.73) |
| K-Nearest Neighbors | 0.58 (0.50–0.69) | 0.52 (0.45–0.59) | 0.99 (0.85–1) | 0.68 (0.59–0.72) | 0.79 (0.75–0.84) | 0.75 (0.72–0.81) |
| Naive Bayes | 0.78 (0.72–0.84) | 0.72 (0.67–0.78) | 0.75 (0.71–0.80) | 0.75 (0.70–0.78) | 0.85 (0.80–0.89) | 0.86 (0.81–0.90) |
| AdaBoost | 0.74 (0.70–0.79) | 0.70 (0.66–0.74) | 0.75 (0.73–0.77) | 0.73 (0.70–0.76) | 0.74 (0.71–0.77) | 0.64 (0.61–0.69) |
| Decision Tree | 0.77 (0.72–0.81) | 0.77 (0.67–0.79) | 0.70 (0.61–0.79) | 0.73 (0.63–0.76) | 0.83 (0.80–0.86) | 0.80 (0.78–0.82) |
| Linear Discriminant Analysis | 0.51 (0.46–0.62) | 0.48 (0.43–0.55) | 0.97 (0.87–1.0) | 0.64 (0.56–0.68) | 0.60 (0.55–0.67) | 0.52 (0.48–0.61) |
| Quadratic Discriminant Analysis | 0.79 (0.75–0.84) | 0.73 (0.67–0.77) | 0.71 (0.66–0.73) | 0.72 (0.68–0.75) | 0.82 (0.78–0.86) | 0.83 (0.78–0.87) |
| Extra Trees | 0.76 (0.71–0.80) | 0.78 (0.72–0.82) | 0.67 (0.61–0.72) | 0.72 (0.69–0.77) | 0.85 (0.81–0.89) | 0.81 (0.78–0.84) |
| XGBoost | 0.78 (0.71–0.83) | 0.73 (0.71–0.82) | 0.75 (0.73–0.86) | 0.74 (0.72–0.84) | 0.79 (0.70–82) | 0.76 (0.70–0.82) |
| Feed Forward Neural Network | | | | | | |

* Documented by the end of hospital encounter, not specifically at time of triage.

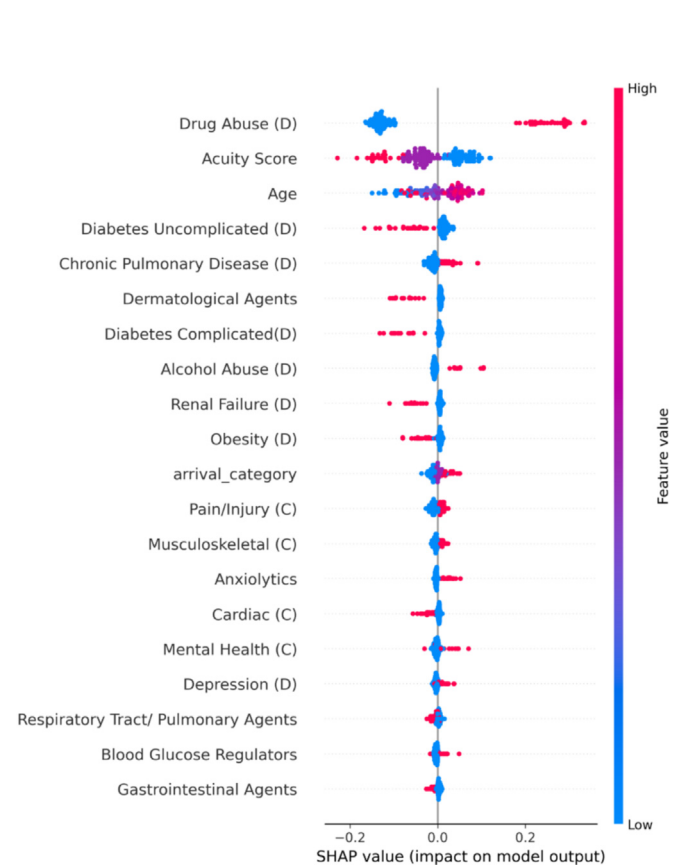


Fig. 1. SHAP beeswarm plot for Random Forest Classifier.
**Each point in the plot represents a single patient record. Points are color-coded according to the feature value, with red denoting opioid misuse and blue denoting no opioid misuse. D = diagnosis category; C = chief complaint.

sensitivities rivaling diagnosis codes places at anytime during the encounter, none achieved perfect performance, as indicated by accuracy scores and precision below 0.85. Consequently, model deployment raised important ethical and legal considerations related to false detection and omission. False detection of opioid misuse may contribute to stigma and mistrust. Conversely, false omission may result in delayed care or missed treatment opportunities. It should be emphasized that such models are meant as decision-support and not relied on for diagnostic determinations. Any results require clinician verification and models must be deployed with safeguards emphasizing transparency and auditability.

Top performing modeling included the decision tree-based approaches (random forest and gradient boosting models). These methods leverage the strength of multiple base learners, reducing variance and bias, and are particularly well-suited for structured datasets with heterogeneous feature importance. Simpler models like logistic regression and SVM yielded similar performance to the more computationally complex methods, including FFNN, indicating that relatively simpler approaches may have utility in this domain. The similarities in performance among models may be indicative of the relatively small sample size or a ceiling effect due to the limited nature of data available at the time of triage. Larger scale evaluations on cohorts with a true prevalence of opioid misuse will help identify top-performing models. While, in general, simpler models are preferred to more complex ones, model selection should also be guided by interpretability, calibration, and ease of clinical deployment.

Explainability is critical for clinical machine learning models in healthcare to ensure transparency and trust. In our classifier, explainability analyses revealed that pre-existing diagnosis codes for opioid misuse were the top predictor, supporting the model's face validity. Codes for alcohol use and depression were also predictive, reflecting known associations of opioid misuse with co-occurring substance use and mental health comorbidities [31–33]. Chronic pulmonary disease also emerged as a key feature, possibly reflecting known links with insufflated opioids common in the location of the study. [34] These findings highlight the value of both diagnostic history and comorbidities in

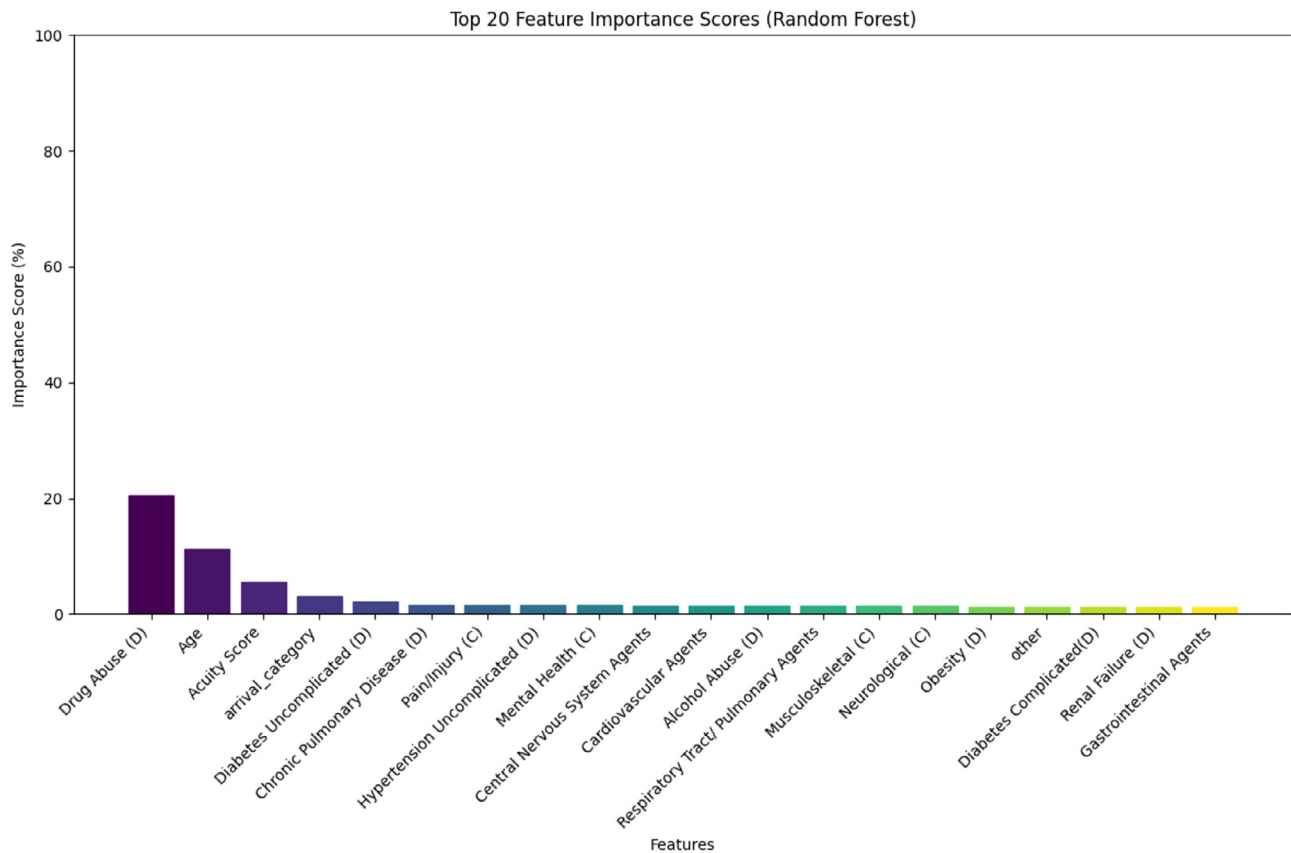


Fig. 2. Global Feature Importance for Random Forest Classifier. D = diagnosis category; C = chief complaint.

substance use prediction. ED acuity had a mixed effect: high ESI (low acuity encounters) generally aligned with no misuse, but lower ESI did not consistently predict misuse. Although unadjusted demographic differences were observed between those with and without opioid misuse, we excluded race, ethnicity, sex, and insurance from model inputs out of concern for fairness and the embedding of structural inequities into clinical decision support.

Machine learning has the potential to automate clinical processes and reduce cognitive burden. The volume of EHR data can be overwhelming for clinicians, contributing to poor patient care and burnout [35,36]. By leveraging routinely collected clinical data, machine learning models can surface relevant information in real time, streamlining workflows and reducing staff burden. Prior models using clinical text performed well but rely on data not available until later in the ED visit, limiting real-time utility [13]. Our triage-based model provides a real-world proof of concept showing how EHR data can support timely disease detection in EDs.

To enhance generalizability and reduce the risk of overfitting in the context of our limited sample size, we incorporated preprocessing steps such as keyword searching and categorical encoding. While such preprocessing was done to improve model robustness, it introduced additional complexity that may hinder implementation. This tradeoff underscores the challenge of balancing performance with real-world feasibility. Future work should evaluate streamlined approaches that preserve accuracy with less operational burden.

Our findings suggest that predictive modeling using triage data can support early screening for opioid misuse which can be used to prompt diagnostic workups, harm reduction, or MOUD initiation. Such models could standardize and scale opioid misuse assessment as a routine part of ED care, potentially reducing variability and selection bias in practice. Prior to clinical deployment, model calibration and further

evaluations of operational factors such as workup-to-detection ratios and numbers needed to evaluate are needed for safe integration into care. Future work should validate these models in larger cohorts and assess their impact on patient outcomes and clinical workflows.

All results should be interpreted within the context of their limitations. Data used in model training and testing were entered into the EHR for non-research purposes at a single institution and may not be universally representative. As ICD-10-CM codes were a portion of the criteria for cohort development, performance metrics for diagnosis codes were likely inflated. We acknowledge that reliance on urine toxicology screening to enrich the cohort may preferentially select encounters in which clinicians suspected substance use, potentially introducing selection bias. However, this approach was chosen to improve the efficiency of identifying sufficient positive cases which ultimately underwent manual review, rather than to define misuse itself. Given the limited sample size, evaluations for bias on demographic variables were uninformative. Models should undergo calibration and testing for bias and fairness in a larger cohort with true disease prevalence prior to testing in clinical trials.

We demonstrated the development and internal validation of high-performing machine learning models for opioid misuse screening. Future studies should evaluate the potential for triage-based machine learning models to improve care for patients with opioid misuse in the emergency department.

CRediT authorship contribution statement

Chirag Chhablani: Writing – original draft, Methodology, Formal analysis, Data curation. **Usman Shahid:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Natalie Parde:** Writing – review & editing, Supervision, Methodology, Formal

analysis. **Sami Muslmani:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Huiyi Hu:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Dillon Thorpe:** Writing – review & editing, Project administration, Investigation. **Majid Afshar:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Niranjan Karnik:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Neeraj Chhabra:** Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Financial support

Authors report support from the following grants from the National Institute on Drug Abuse (NIDA)/NIH: K23DA055061 (Chhabra), R61DA057629 (Karnik/Chhabra/Parde). This work was supported, in part, by the National Center for Advancing Translational Sciences (NCATS), through Grant UL1TR002003. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Neeraj Chhabra reports financial support was provided by National Institute on Drug Abuse. Niranjan Karnik reports financial support was provided by National Institute on Drug Abuse. Majid Afshar reports financial support was provided by National Institute on Drug Abuse. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ajem.2026.02.037>.

Data availability

The data used in this study contain protected health information (PHI) and cannot be shared publicly due to institutional and regulatory constraints. The analytic code used to develop models will be made available upon reasonable request to the corresponding author, subject to institutional approval.

References

- [1] Dowell D. Treatment for opioid use disorder: population estimates – United States, 2022. *MMWR Morb Mortal Wkly Rep.* 2024;73. doi:10.15585/mmwr.mm7325a1.
- [2] CDC. Statement from CDC's National Center for Injury Prevention and Control on provisional 2024 overdose death data. CDC Newsroom. <https://www.cdc.gov/media/releases/2025/2025-statement-from-cdcs-national-center-for-injury-prevention-and-control-on-provisional-2024.html>; 2025.
- [3] Das S. Trends in opioid-involved emergency department visits: Overall and by type of opioid, 2021–2024; 2026.
- [4] D'Onofrio G, McCormack RP, Hawk K. Emergency departments - a 24/7/365 option for combating the opioid crisis. *N Engl J Med.* 2018;379:2487–90. doi:10.1056/NEJMp1811988.
- [5] Weber W, Heins A, Jardine L, Stanford K, Duber H. Principles of screening for disease and health risk factors in the emergency department. *Ann Emerg Med.* 2023;81:584–91. doi:10.1016/j.annemergmed.2022.06.015.
- [6] Faude S, Delgado MK, Perrone J, McFadden R, Xiong RA, O'Donnell N, et al. Variability in opioid use disorder clinical presentations and treatment in the emergency department: a mixed-methods study. *Am J Emerg Med.* 2023;66:53–60. doi:10.1016/j.ajem.2023.01.009.
- [7] Chhabra N, Smith D, Dickinson G, Caglianone L, Taylor RA, D'Onofrio G, et al. Trends and disparities in initiation of buprenorphine in US emergency departments, 2013–2022. *JAMA Netw Open.* 2024;7:e2435603. doi:10.1001/jamanetworkopen.2024.35603.

- [8] Chhabra N, Smith D, Parde N, Hsing-Smith N, Bianco JM, Taylor RA, et al. Racial, ethnic, and sex disparities in buprenorphine treatment from emergency departments by discharge diagnosis. *Acad Emerg Med Off J Soc Acad Emerg Med.* 2025. doi:10.1111/acem.70035.
- [9] Moe J, Koh J, Ma JA, Pei LX, MacLean E, Keech J, et al. Screening for harmful substance use in emergency departments: a systematic review. *Int J Emerg Med.* 2024;17:52. doi:10.1186/s12245-024-00616-2.
- [10] Lowenstein M, Perrone J, McFadden R, Xiong RA, Meisel ZF, O'Donnell N, et al. Impact of universal screening and automated clinical decision support for the treatment of opioid use disorder in emergency departments: a difference-in-differences analysis. *Ann Emerg Med.* 2023;82:131–44. doi:10.1016/j.annemergmed.2023.03.033.
- [11] Chalmers CE, Mullinax S, Brennan J, Vilke GM, Oliveto AH, Wilson MP. Screening tools validated in the outpatient pain management setting poorly predict opioid misuse in the emergency department: a pilot study. *J Emerg Med.* 2019;56:601–10. doi:10.1016/j.jemermed.2019.03.018.
- [12] Chhabra N. Death by a thousand screens: a practical role for machine learning in emergency medicine. *Ann Emerg Med.* 2023;82:531–2. doi:10.1016/j.annemergmed.2023.04.030.
- [13] Shahid U, Parde N, Smith DL, Dickinson G, Bianco J, Thorpe D, et al. Development and evaluation of machine learning models for the detection of emergency department patients with opioid misuse from clinical notes. *MedRxiv Prepr Serv Health Sci.* 2024;2024.12.11.24318875. doi:10.1101/2024.12.11.24318875.
- [14] Singleton J, Li C, Akpunonu PD, Abner EL, Kucharska-Newton AM. Using natural language processing to identify opioid use disorder in electronic health record data. *Int J Med Inform.* 2023;170:104963. doi:10.1016/j.ijmedinf.2022.104963.
- [15] Afshar M, Sharma B, Dligach D, Oguss M, Brown R, Chhabra N, et al. Development and multimodal validation of a substance misuse algorithm for referral to treatment using artificial intelligence (SMART-AI): a retrospective deep learning study. *Lancet Digit Health.* 2022;4:e426–35. doi:10.1016/S2589-7500(22)00041-3.
- [16] Lopez I, Swaminathan A, Vedula K, Narayanan S, Nateghi Haredasht F, Ma SP, et al. Clinical entity augmented retrieval for clinical information extraction. *NPJ Digit Med.* 2025;8:45. doi:10.1038/s41746-024-01377-1.
- [17] Chhabra N, Smith D, Pachwiczewicz P, Lin Y, Bhalla S, Maloney CM, et al. Performance of international classification of Disease-10 codes in detecting emergency department patients with opioid misuse. *Addict Abingdon Engl.* 2024;119:766–71. doi:10.1111/add.16394.
- [18] Sharma B, Dligach D, Swope K, Salisbury-Afshar E, Karnik NS, Joyce C, et al. Publicly available machine learning models for identifying opioid misuse from the clinical notes of hospitalized patients. *BMC Med Inform Decis Mak.* 2020;20. doi:10.1186/s12911-020-1099-y.
- [19] 2022 NSDUH Annual National Report | CBHSQ Data. <https://www.samhsa.gov/data/report/2022-nsduh-annual-national-report>; 2026. (accessed April 2, 2025).
- [20] Han B, Compton WM, Blanco C, Crane E, Lee J, Jones CM. Prescription opioid use, misuse, and use disorders in U.S. adults: 2015 National Survey on drug use and health. *Ann Intern Med.* 2017;167:293–301. doi:10.7326/M17-0865.
- [21] Abuse NI on D. Summary of Misuse of Prescription Drugs. National Institute on Drug Abuse (NIDA); 2026. <https://nida.nih.gov/publications/research-reports/misuse-prescription-drugs/overview> (accessed April 30, 2025).
- [22] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. doi:10.1136/bmj-2023-078378.
- [23] Sax DR, Warton EM, Mark DG, Vinson DR, Kene MV, Ballard DW, et al. Evaluation of version 4 of the emergency severity index in US emergency departments for the rate of Misriage. *JAMA Netw Open.* 2023;6:e233404. doi:10.1001/jamanetworkopen.2023.3404.
- [24] Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. *Acad Emerg Med Off J Soc Acad Emerg Med.* 2000;7:236–42. doi:10.1111/j.1553-2712.2000.tb01066.x.
- [25] Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care.* 1998;36:8–27. doi:10.1097/00005650-199801000-00004.
- [26] Elixhauser Comorbidity Software Refined for ICD-10-CM. https://hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp; 2026. (accessed April 22, 2025).
- [27] Concept: Elixhauser Comorbidity Index. <http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?printer=Y&conceptID=1436>; 2026. (accessed April 22, 2025).
- [28] USP DC | USP. <https://www.usp.org/health-quality-safety/usp-drug-classification-system>; 2026. (accessed April 22, 2025).
- [29] Weiss AJ, McDermott KW, Heslin KC. Opioid-related hospital stays among women in the United States, 2016: Statistical brief #247. *Healthc. Cost util. Proj. HCUP stat. Briefs, Rockville (MD): Agency for Healthcare Research and Quality (US);* 2006.
- [30] scikit-learn: machine learning in Python – scikit-learn 1.7.0 documentation. <https://scikit-learn.org/stable/>; 2026. (accessed June 26, 2025).
- [31] Chhabra N, Smith DL, Maloney CM, Archer J, Sharma B, Thompson HM, et al. The identification of subphenotypes and associations with health outcomes in patients with opioid-related emergency department encounters using latent class analysis. *Int J Environ Res Public Health.* 2022;19:8882. doi:10.3390/ijerph19148882.
- [32] Montiel Ishino FA, Gilreath T, Williams F. Finding the hidden risk profiles of the United States opioid epidemic: using a person-centered approach on a national dataset of noninstitutionalized adults reporting opioid misuse. *Int J Environ Res Public Health.* 2020;17. doi:10.3390/ijerph17124321.
- [33] Rhee TG, Ross JS, Rosenheck RA, Grau LE, Fiellin DA, Becker WC. Accidental drug overdose deaths in Connecticut, 2012–2018: the rise of polysubstance detection? *Drug Alcohol Depend.* 2019;205:107671. doi:10.1016/j.drugalcdep.2019.107671.

- [34] Weeks MA, Clark EP, Mycyk MB. Characteristics of heroin-dependent patients seeking asthma care in the ED. *Am J Emerg Med.* 2016;34:895–8. doi:[10.1016/j.ajem.2016.02.005](https://doi.org/10.1016/j.ajem.2016.02.005).
- [35] Gal DB, Han B, Longhurst C, Scheinker D, Shin AY. Quantifying electronic health record data: a potential risk for cognitive overload. *Hosp Pediatr.* 2021;11:175–8. doi:[10.1542/hpeds.2020-002402](https://doi.org/10.1542/hpeds.2020-002402).
- [36] Asgari E, Kaur J, Nuredini G, Balloch J, Taylor AM, Sebire N, et al. Impact of electronic health record Use on cognitive load and burnout among clinicians: narrative review. *JMIR Med Inform.* 2024;12:e55499. doi:[10.2196/55499](https://doi.org/10.2196/55499).