
Research and Applications

Automated detection of substance use information from electronic health records for a pediatric population

Yizhao Ni,^{1,2} Alycia Bachtel,¹ Katie Nause,³ and Sarah Beal^{2,3}

¹Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA, ²Department of Pediatrics, College of Medicine, University of Cincinnati, Cincinnati, Ohio, USA, and ³Division of Psychology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

Corresponding Author: Yizhao Ni, PhD, Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA; Yizhao.Ni@cchmc.org

Received 26 March 2021; Revised 6 May 2021; Editorial Decision 24 May 2021; Accepted 26 May 2021

ABSTRACT

Objective: Substance use screening in adolescence is unstandardized and often documented in clinical notes, rather than in structured electronic health records (EHRs). The objective of this study was to integrate logic rules with state-of-the-art natural language processing (NLP) and machine learning technologies to detect substance use information from both structured and unstructured EHR data.

Materials and Methods: Pediatric patients (10-20 years of age) with any encounter between July 1, 2012, and October 31, 2017, were included ($n = 3890$ patients; 19 478 encounters). EHR data were extracted at each encounter, manually reviewed for substance use (alcohol, tobacco, marijuana, opiate, any use), and coded as lifetime use, current use, or family use. Logic rules mapped structured EHR indicators to screening results. A knowledge-based NLP system and a deep learning model detected substance use information from unstructured clinical narratives. System performance was evaluated using positive predictive value, sensitivity, negative predictive value, specificity, and area under the receiver-operating characteristic curve (AUC).

Results: The dataset included 17 235 structured indicators and 27 141 clinical narratives. Manual review of clinical narratives captured 94.0% of positive screening results, while structured EHR data captured 22.0%. Logic rules detected screening results from structured data with 1.0 and 0.99 for sensitivity and specificity, respectively. The knowledge-based system detected substance use information from clinical narratives with 0.86, 0.79, and 0.88 for AUC, sensitivity, and specificity, respectively. The deep learning model further improved detection capacity, achieving 0.88, 0.81, and 0.85 for AUC, sensitivity, and specificity, respectively. Finally, integrating predictions from structured and unstructured data achieved high detection capacity across all cases (0.96, 0.85, and 0.87 for AUC, sensitivity, and specificity, respectively).

Conclusions: It is feasible to detect substance use screening and results among pediatric patients using logic rules, NLP, and machine learning technologies.

Key words: automated substance use detection, electronic health records, pediatric population, natural language processing, deep learning

INTRODUCTION

Background and significance

Substance use and related morbidity and mortality contribute to 25% of all deaths in the United States and an annual cost of \$740 billion.^{1,2} Research examining substance use (eg, prevalence, initiation age) using nationally representative survey methodology has consistently demonstrated that initiation occurs during adolescence and use increases into early adulthood.^{3,4} Therefore, adolescence is an important stage for prevention. Pediatric healthcare providers play a critical role in identifying substance use initiation, monitoring substance use over time, and providing referrals to treatment when necessary.^{5,6} The Centers for Medicare and Medicaid Services recognize the need for targeted screening, prevention, and intervention programs in clinical settings and has incentivized implementing strategies for doing so.⁷ However, implementation in healthcare systems is fragmented. Studies of pediatric healthcare systems have suggested that substance use screening with adolescents is occurring, but documentation in patients' electronic health records (EHRs) is often unstandardized and exists in unstructured clinical notes, rather than in structured data fields.^{8–10} The complexity and intensity of retrieving substance use screening information from unstructured data has limited healthcare systems in supporting providers as they monitor substance use over time.^{8,11,12} Some studies have proposed incorporating standardized screening protocols to streamline documentation,^{6,13} which, however, does not entirely address the challenges of modifying pediatrician behavior,¹⁴ documentation,⁸ or outcomes.¹⁵ With a lack of widespread adoption of standardized substance use screening and documentation in pediatric settings,¹⁶ there is a critical need to develop an efficient and cost-effective approach to transform substance use information from EHRs to a reusable format that can be accessed and reviewed across providers and tracked over time.¹⁷

Decision support capabilities built on modern artificial intelligence (technologies such as natural language processing (NLP) and machine learning have promised great benefits to different aspects of healthcare delivery. NLP is a hybrid technology that utilizes linguistic knowledge (eg, substance-related lexicon) in computerized algorithms to identify relevant information (eg, substance use) from human language input, while machine learning employs algorithmic models to analyze the extracted information and make data-driven predictions (eg, results of substance use screening). The technologies, when combined, have shown to significantly improve predictive performance in detecting clinical conditions such as signs and symptoms, diseases, and adverse drug events from unstructured clinical narratives.^{18–22} Nevertheless, few studies have used NLP and machine learning for identifying substance use information from clinical notes. Wang et al²³ developed a knowledge-based NLP system that used a predefined lexicon to detect substance use (alcohol, drug, nicotine use) information from unstructured notes. Despite its encouraging performance, the study focused on analyzing social history sections from physical and history notes that did not represent the complexity and variation of language used in clinical data. Hylan et al²⁴ enriched a knowledge-based NLP system with regular expressions to identify indicators of problematic opioid use from clinical notes. Specifically, the authors looked for phrases indicative of a diagnosis or assessment of prescription opioid overuse, misuse, abuse, or addiction. The study used NLP-generated indicators only to facilitate manual validation of problematic opioid use; therefore, the detection performance of the system was not reported. Dligach et al²⁵ developed a convolutional neural network–based deep learning model to analyze clinical notes and extract information relevant

to specific billing codes (eg, International Classification of Diseases codes related to alcohol use disorders), which was then fed into machine learning classifiers for predicting substance use disorder. Limited by the data available, the study only applied the model to predict alcohol and opioid use disorder, and the work did not consider temporality (past or current use).

It is worth noting that all the studies focused on analyzing clinical notes in adult settings. Although terms for describing substance use behaviors (eg, alcohol use, smoker) are similar, findings from these studies may not generalize well to pediatric settings. First, rates of substance use disorder in adolescents are generally lower than in adults,⁴ resulting in a difference in frequency with which diagnostic and billing codes can be reliably used. Consequently, the utility of structured and unstructured data in identifying substance use information could be different in pediatric and adult settings. Second, screening for substance use is encouraged but often not required in pediatric settings, and hence data may be sparse compared with adult settings. The paucity of positive indicators in clinical notes might hinder the development of complex models such as deep learning algorithms. Finally, family history of substance use is more likely to be reviewed, discussed, and documented (as an indicator of child safety) in pediatric settings than in adult settings. Family substance use is associated with increased risk for adolescent substance use. As such, this meaningful indicator should be captured and shared with pediatricians in a way that would be different from adult settings. Importantly, the mixed mentions of substance use among patients and their families make automated detection more challenging, as additional disambiguation is needed to distinguish which subject is the substance user. For these reasons, additional study is required to fill the gap in knowledge in pediatric settings, in which the opportunity to prevent substance use is greatest.

OBJECTIVES

This study represents the first step toward developing an accurate and scalable informatics-based solution to minimize existing technology and workflow barriers to support substance use screening. By integrating logic rules with NLP and machine learning techniques, we developed an automated substance use detection system (ASUDS) that analyzed both structured EHR data and unstructured clinical notes to identify substance use information in pediatric settings. We hypothesized that (1) automated algorithms exploiting clinical notes would capture more substance use information than structured EHR data, based on the lack of consistency with which substance use screening is documented; and (2) with using state-of-the-art informatics technologies, the ASUDS could detect substance use information for individual patients with high sensitivity and specificity, given the evidence in literature for adult settings.^{23,25} Creation of such a system would allow current provider-led preferences and practices in substance use documentation to continue while simultaneously improving access to documented information. As such, our research has the potential to aid in long-term efforts to target prevention and intervention in adolescence, where substance use behaviors often emerge.²⁶

MATERIALS AND METHODS

Participants

Pediatric patients (10–20 years of age) with any encounter at the freestanding pediatric children's hospital where the study occurred between July 1, 2012, and October 31, 2017, may have been eligible

Table 1. The substance use–related indicators collected from the electronic health records

Indicators	Description	Example	Format
Diagnoses	A patient's substance use–related diagnoses documented in an encounter.	Alcohol abuse; alcohol intoxication; tobacco dependence; cannabis dependence, unspecified.	S: Descriptive
Laboratory results	Substance use–related laboratory results documented in an encounter.	Opioid mass spec: opioid inter— not detected. Alcohol screen: ethanol level—104 mg/dL. Nicotine mass spec: cotinine—10 ng/mL.	S: Quantitative and descriptive
Social history flowsheet	A patient's social histories (alcohol, tobacco, drug use (documented in an encounter).	Tobacco use: never. Drug/alcohol use: current (within past month). If current use, what? Alcohol, marijuana, cocaine, other.	S: Descriptive
Substance and chemical abuse assessment	A structured questionnaire performed in an encounter to evaluate a patient's substance use behaviors.	Have you ever used any of the following substances or chemicals? Yes, alcohol, marijuana, pills. How used? Oral/ingest. Last used? Over a month ago.	S: Descriptive
Clinical notes	Clinical notes with potential mentions of substance use.	Brother has history of heroin addiction, and another adult brother has a history of opioid pill addiction.	U: 27 141 unique narratives

S: structured electronic health record data field; U: unstructured electronic health record data field.

for inclusion. Data were collected as part of a larger study to understand the emergence of substance use among adolescents in foster care.²⁷ Therefore, participants were either in foster care (eligible: $n = 2787$) or demographically matched to foster youths (eligible: $n = 2787$) based on sex, birthdate within 6 months, race, ethnicity, and public insurance status. The dataset for this study included 3890 participants (58.3% foster youths; 41.7% matched youths) who had at least 1 substance use screening during the study period. There were 121 656 patient visits, of which 19 478 (16.0%) encounters contained screening documentation. Approval for this study was given by the institutional review board where the study occurred (Institutional Review Board: 2017-4747) and the county child welfare agency who holds custody of foster youths. A written waiver of consent was authorized.

Substance use–related indicators from EHRs

All substance use–related indicators from structured EHR data, including patient social histories, encounter diagnoses, substance and chemical abuse assessments, and laboratory results, were extracted along with all unstructured clinical notes created during encounters. A comprehensive list of substance-related key phrases was applied to identify paragraphs containing potential substance use mentions from the notes. The process generated 130 998 free-text paragraphs from 26 931 notes. Removing duplicated paragraphs resulted in a set of 27 141 (of 130 998) unique paragraphs. [Table 1](#) summarizes the substance use–related indicators collected by the study. The complete list of structured indicators is presented in the [Supplementary Appendix](#) file. The list of substance-related key phrases is presented in [Supplementary Appendix A](#).

Gold-standard substance use screening results

We applied 2 distinct processes to extract substance use information for all 19 478 encounters containing screening documentation. First, substance use information was manually annotated from all 27 141

narrative paragraphs using a double annotation schema.²⁸ Two data analysts manually reviewed each narrative to identify 5 categories of substance use: alcohol, tobacco, marijuana, opiate, and any use (ie, all categories in [Supplementary Appendix Table A.1](#)). If a positive mention was detected, the analysts further classified the behavior into 3 assertions—lifetime (past or current use), current (use within the past 12 months), or family use (referred to substance use by a family member)—to be consistent with other surveillance studies of adolescent substance use.^{29,30} Differences between the annotators' decisions were resolved by adjudication and the interannotator agreement was assessed using Cohen's kappa.³¹ For structured data, manual review was performed on all collected records to derive substance use categories and assertions. Finally, the results from structured and unstructured data were merged for each encounter and served as a gold-standard set to evaluate automated approaches.

Automated substance use detection

[Figure 1](#) diagrams an overview of ASUDS. Encounters without structured indicators or potential mentions of substance use in clinical notes were classified as encounters without screening (process 1 in [Figure 1](#)). For remaining encounters, a logic-based rule matcher (LRM) was developed to classify structured indicators into screening results (process 2). An NLP- and machine learning–based substance information screener (SIS) was developed to detect substance use categories and assertions from unstructured narratives (process 3). The results were merged as a final prediction of whether substance use screening occurred and the associated results for each encounter (process 4).

Logic-based rule matcher

The LRM utilized logic rules to map structured indicators to screening results. For descriptive indicators (eg, social histories), regular expressions were developed to determine substance use categories and assertions. For example, given the description “Tobacco use:

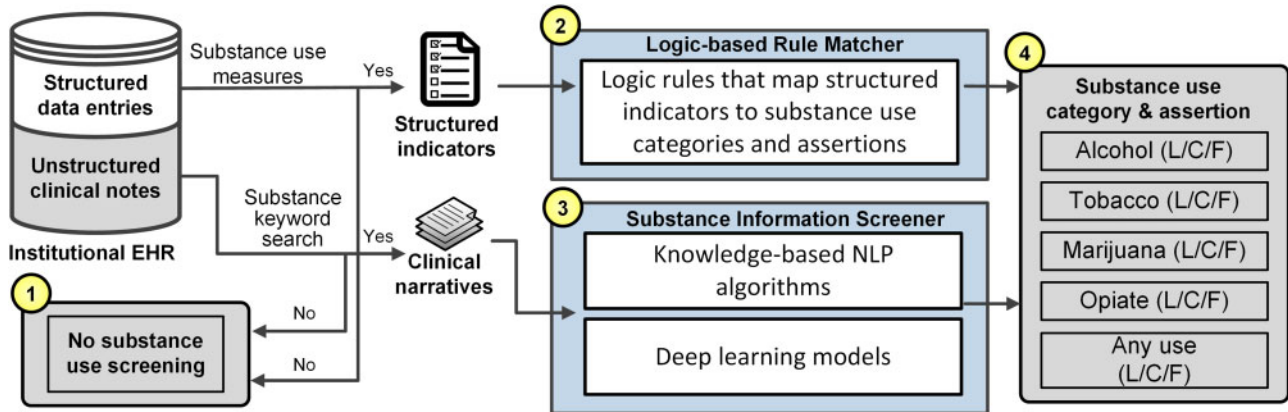


Figure 1. An overview of the automated substance use screening system. C: current; EHR: electronic health record; F: family use; L: lifetime; NLP: natural language processing.

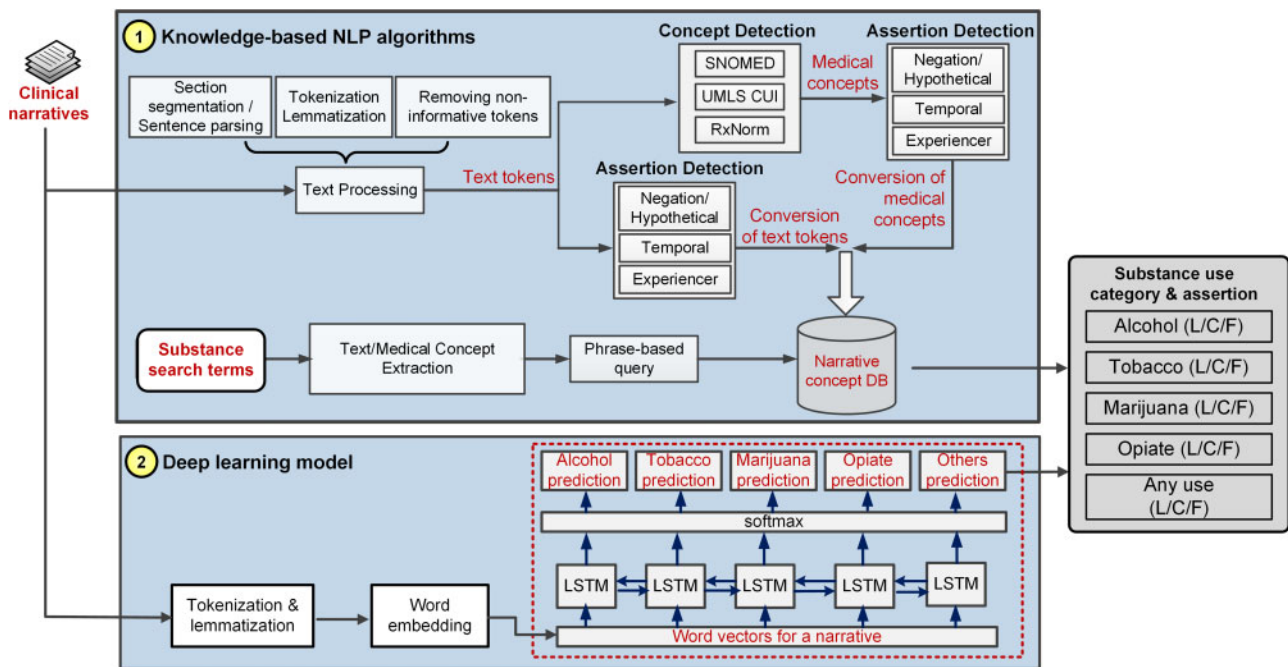


Figure 2. An overview of the substance information screener. C: current; CUI: concept unique identifier; F: family use; L: lifetime; LSTM: long-short term memory; RxNorm: normalized names for clinical drugs; SNOMED: Systematized Nomenclature of Medicine Clinical Terms; UMLS: Unified Medical Language System.

past” in a social history flowsheet, the LRM classified tobacco use with a lifetime assertion as positive and a current assertion as negative. For quantitative indicators (eg, laboratory results), any detected value was classified as positive for both lifetime use and current use. For instance, a detected value of ≥ 3.0 mg/dL for “ethanol level” in an “ethanol-only blood alcohol screen” indicated lifetime and current alcohol use. A descriptive overview of the LRM including the numbers of rules developed for each indicator is available in [Supplementary Appendix B](#). The logic rules were coded in Java.

Substance information screener

The SIS consisted of 2 modules ([Figure 2](#)). The first module was a knowledge-based NLP system developed in our earlier study to extract medical information from clinical narratives (module 1 in

[Figure 2](#)).^{18,20,32,33} Details of the system can be found in our earlier publications.²⁰ To summarize, the system first tokenized and lemmatized clinical narratives. Concept identification was then applied to detect substance use–related words/phrases (eg, alcohol, marijuana) using clinical terminologies including concept unique identifiers from the UMLS (Unified Medical Language System), SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms) codes, and normalized names for clinical drugs (RxNorm).^{34,35} Finally, assertion detection determined negation (absence of findings), temporality (historical findings), and experiencer (object of a finding) expressions from the context and converted the detected terms to the corresponding format. For example, the word *heroin* in “patient has history of heroin addiction” would be converted to a temporal format “HISTORY-heroin” at word level and “HISTORY-C0011892” at concept level. The processes represented each narra-

tive by a vector of substance use findings. A search query was then built for each substance use category and assertion (eg, HISTORY-opiate representing historical opiate use) based on the keyword list described in [Supplementary Table A.1](#) and matched with the narrative vectors to classify screening results. The knowledge-based NLP system was coded in house in Java.

Although knowledge-based systems with well-defined rules remain an effective approach to NLP, data-driven machine learning techniques such as deep learning have been adopted to improve information retrieval from free-text narratives.³⁶ To achieve the best detection capacity, we also implemented a deep learning model to analyze clinical narratives and classify screening results (module 2 in [Figure 2](#)). After tokenization and lemmatization, unique words from the clinical narratives were grouped into 100 clusters using the word embedding technique, in which each word was represented by a 100-dimensional numeric vector such that words with similar semantic and syntactic patterns were close to each other in this Euclidean space.³⁷ The prediction process then involved aggregating information from the word vector sequence of a narrative to make a binary decision (yes/no) for each substance use category and assertion. A deep learning model based on bidirectional long-short term memory (LSTM) networks (the dashed box in [Figure 2](#)) was implemented.³⁸ A descriptive illustration of the LSTM network and its training strategy is available in [Supplementary Appendix C](#). Different from the deep learning frameworks used in the literature,²⁵ this model aggregated information from a word sequence forward and backward to capture local and global semantic and syntactic connections between words. The bidirectional LSTM model was implemented with the TensorFlow library in Python.³⁹

Experiments

Experiment setup

A retrospective evaluation of substance use detection on structured and unstructured EHR data individually and in combination was performed. LRM predictions were compared against manual review of structured indicators at each encounter. SIS predictions from the knowledge-based NLP system and the deep learning model were compared against gold-standard annotations of clinical narratives. The deep learning model required a training set to tune its parameters and a nested 10-fold cross-validation was applied to avoid overfitting.⁴⁰ A descriptive illustration of the cross-validation process is available in [Supplementary Appendix C](#). The approach randomly split the dataset into 10 rotating subsets—9 for model training and hyperparameter tuning and 1 for testing at each run. At the end of cross-validation, predictions for each test set were combined to restore predictions for the entire dataset. Finally, all note predictions from the best model in SIS were aggregated at the encounter level using sum aggregation. The predictions were merged with those from LRM and compared against the gold-standard results at the encounter level as the final performance of ASUDS.

Evaluation metrics

System performance was assessed with 5 customary evaluation metrics: positive predictive value, sensitivity, negative predictive value, specificity, and area under the receiver-operating characteristic curve (AUC).^{41–43} We adopted AUC as the primary measure and reported the evaluation metrics with optimal cutoffs maximizing sensitivity and specificity in the receiver-operating characteristic analysis.⁴⁴ The significance of performance differences was assessed with 95% confidence intervals.⁴⁵ The evaluation metrics and 95%

confidence intervals were calculated with the pROC library in R.⁴⁶ The LRM generated determinate classification rather than probabilistic predictions; therefore, we did not report AUCs in its evaluation.

RESULTS

Descriptive statistics of the dataset

Among the 121 656 encounters in EHRs, 19 478 (16.0%) encounters had screening information and 11 063 (9.1%) encounters had documented substance use information. Whether a screening occurred at a given encounter differed by sex ($\chi^2_1 = 16.03$, $P < .001$), in which a higher proportion of females were screened (16.2% female vs 15.4% male). The screening rate also differed for participants who were non-Hispanic White compared with Black, indigenous, or person of color (BIPOC) ($\chi^2_1 = 8.78$, $P = .003$), in which a higher proportion of BIPOC participants were screened (16.1% BIPOC vs 15.5% White). For the 19 478 encounters with screening information, whether a screening outcome was positive did not differ by gender ($\chi^2_1 = 0.99$, $P = 0.319$) but did differ between participants who were non-Hispanic White and BIPOC ($\chi^2_1 = 4.31$, $P = 0.038$), in which a higher proportion of BIPOC participants were identified as having a lifetime or current substance use (33.3% BIPOC vs 31.9% White). While the differences were statistically significant in part due to the large sample size, proportional differences were small.

[Table 2](#) presents descriptive statistics of the substance use information documented in the EHRs. There were 17 235 substance use-related structured records, including 567 encounter diagnoses, 13 585 laboratory orders, 2095 social history flowsheets, and 988 substance and chemical abuse assessments. The average number of non-punctuation tokens (eg, words, numbers) in the clinical narratives was 39 (minimum and maximum tokens = 2 and 425 tokens, respectively). Manual annotation with clinical narratives captured 94.0% ($n = 37\,448$ of 39 849) of positive screening results across the substance use categories and assertions, while structured EHR data captured 22.0% ($n = 8753$ of 39 849). The overall IAAs (Cohen's kappa) between the data analysts and the consensus were 0.921 and 0.922, respectively, indicating good reliability of annotation. Sensitivities for capturing substance use information using structured indicators and clinical narratives are available in [Supplementary Appendix D](#).

Performance of the LRM in classifying structured indicators

[Figure 3](#) shows the performance of LRM in classifying the 17 235 structured indicators. By analyzing structured data, the logic rules achieved 100% sensitivity and over 99.8% specificity across all substance use categories and assertions.

Performance of the SIS in analyzing clinical narratives

[Figures 4](#) and [5](#) present the system performances in detecting substance use information from the 27 141 clinical narratives. The knowledge-based NLP system achieved over 0.858 AUCs on detecting specific categories including alcohol, marijuana, opiates, and tobacco, with sensitivity $\geq 78.5\%$ and specificity $\geq 87.6\%$. Its performances were lower on detecting any lifetime or current substance use. The deep learning model achieved over 0.959 AUCs on most substance use categories and assertions, with sensitivity $\geq 87.5\%$ and specificity $\geq 89.0\%$. Its performances were lower in detecting opiates use, with over 0.882 AUC, sensitivity $\geq 81.0\%$,

Table 2. Numbers of encounters with substance use information documented in structured indicators and clinical notes

Category	Structured indicators			Clinical notes			Total		
	Lifetime	Current	Family	Lifetime	Current	Family	Lifetime	Current	Family
Alcohol	434	311	0 ^a	1740	1315	3702	1817	1387	3702
Marijuana	1108	916	0 ^a	3406	2765	264	3596	2953	264
Opiates	67	61	0	123	99	164	171	145	164
Tobacco	1015	858	0	2729	2234	796	3094	2605	796
Any use	2143	1840	0	5881	5095	7135	6402	5618	7135
Total	4767	3986	0	13 879	11 508	12 061	15 080	12 708	12 061

^a16 (0.4%) subjects had indicators of fetal drug exposure (fetal alcohol syndrome, neonatal abstinence syndrome) in structured problem lists, indicating potential maternal drug use. These indications were patient/family reported and were not accompanied by any encounter diagnoses made by clinicians. For that reason, they were excluded from the analysis.

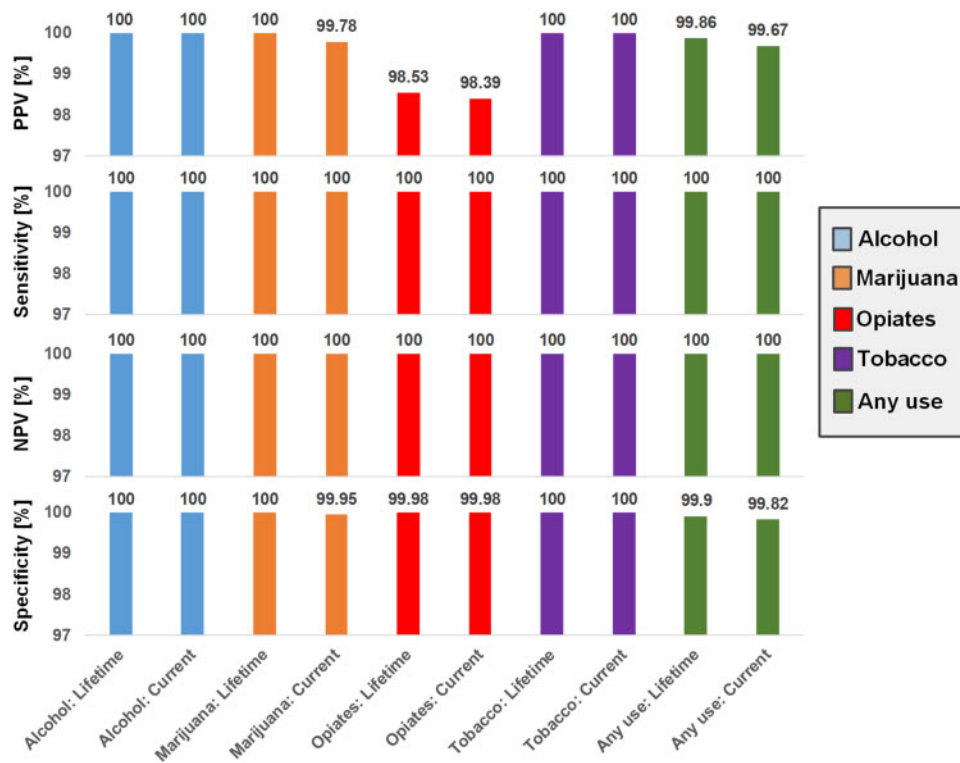


Figure 3. Performance of the logic-based rule matcher in classifying structured indicators. Note that the structured indicators did not contain assertion of family use. The logic-based rule matcher generated determinate classification rather than probabilistic predictions; therefore, we did not report area under the receiver-operating characteristic curve in the evaluation. NPV: negative predictive value; PPV: positive predictive value.

and specificity $\geq 85.0\%$. The deep learning model outperformed the knowledge-based system statistically significantly for all assertions except opiates ($P = 0.05$). Given its superior performances, the deep learning model was selected to implement the SIS.

Integrated performance of the ASUDS

Table 3 presents the performance of ASUDS, in which LRM and SIS predictions were integrated at the encounter level. The system achieved over 0.957 AUCs across all categories and assertions at a significance level of 0.05, with sensitivity $\geq 84.6\%$ and specificity $\geq 87.2\%$. SIS outperformed LRM statistically significantly on all

cases ($P = .05$) (see Supplementary Appendix E for all performance details).

Error analysis

The confusion matrices of LRM and SIS are presented in Supplementary Appendix F. The LRM made 13 false positives on 5 encounters, all owing to conflicting EHR records (eg, an encounter with 2 social history flowsheets, one indicating current tobacco use and the other indicating no substance use). Data analysts screened additional EHR information in these instances (eg, record overriding logs), which were not available to the LRM. The SIS made 26 696

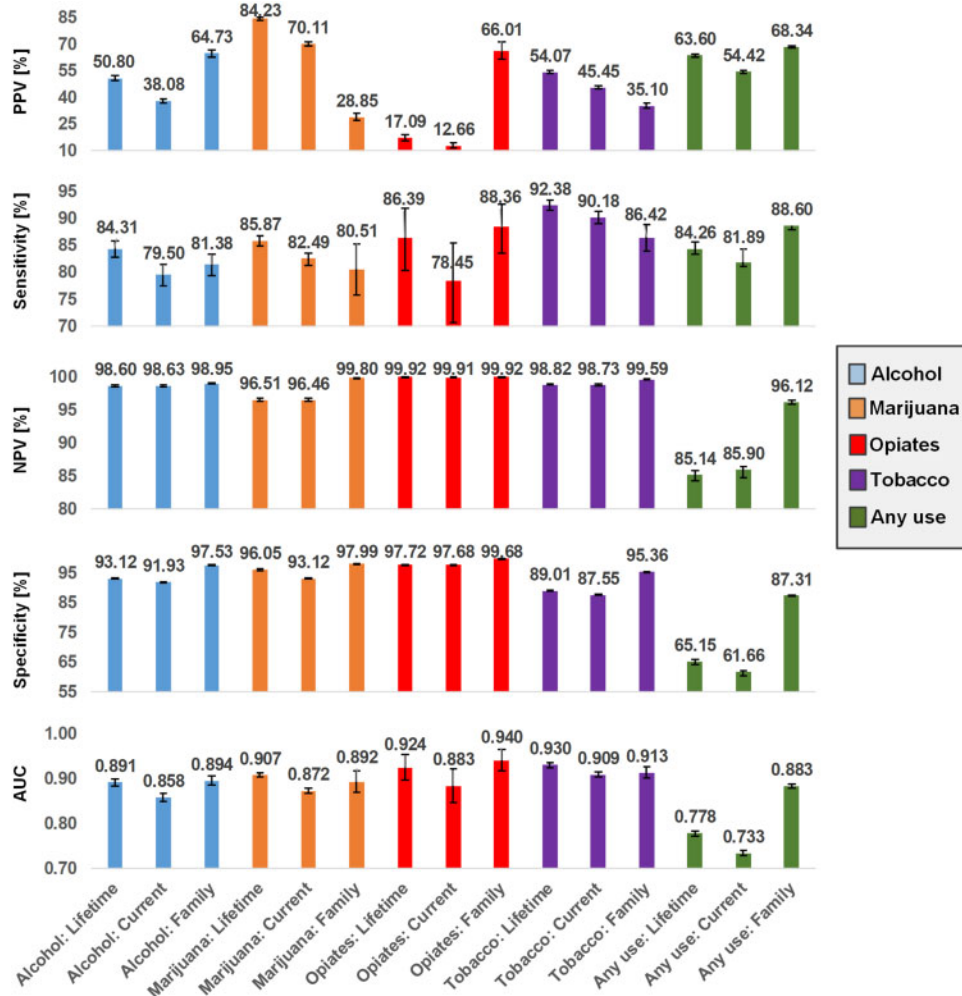


Figure 4. Performance of the knowledge-based natural language processing system in detecting substance use categories and assertions on individual clinical narratives. Error bars indicate 95% confidence intervals. AUC: area under the receiver-operating characteristic curve; NPV: negative predictive value; PPV: positive predictive value.

false positives and 3134 false negatives across the substance use categories and assertions. To identify its limitations, we stratified the data, sampled 10% (n = 2983) of the errors based on the categories and assertions, and performed an error analysis. The causes were grouped into 11 categories, which are presented in Table 4.

DISCUSSION

This study developed an ASUDS to identify substance use information using structured EHR data and unstructured clinical narratives from a pediatric population and setting. Detecting screening results from the structured data using LRM achieved close to perfect performance (100% sensitivities; ≥99.8% specificities). Although rules might need customization when applied to additional institutions, implementation effort is likely minimal given the limited number of rules required (see Supplementary Appendix B). However, our analysis suggested that structured EHR data only documented 22.0% of screening results, consistent with the literature.⁸⁻¹⁰ Despite its high performance, LRM alone was unable to capture substance use information comprehensively.

Using NLP and machine learning technologies, the SIS showed good capability in detecting substance use information from unstructured narratives. The knowledge-based NLP system achieved over 0.858 AUCs in detecting specific substance uses, but it performed lowest when detecting any use due to variation in keywords or phrases (Figure 4). By learning linguistic patterns from clinical narratives, the deep learning model further improved detection capacity, with over 0.882 AUCs across all substance use categories and assertions (Figure 5). Compared with the knowledge-based system, the deep learning model achieved significantly better performance in most cases. This illustrates the advantage of machine learning technologies over knowledge-driven rules by learning latent patterns from human language. However, deep learning was less advantageous when there was paucity of data (eg, opiates), revealing a major limitation of machine learning-based models.⁴⁷ The SIS achieved significantly better performance than the LRM (P = .05) (Supplementary Appendix E), confirming our hypothesis that automated algorithms exploiting clinical notes would capture more substance use information than structured EHR data. By integrating the predictions from SIS (unstructured clinical notes) and LRM (structured EHR data), the ASUDS achieved over 0.957 AUC across all

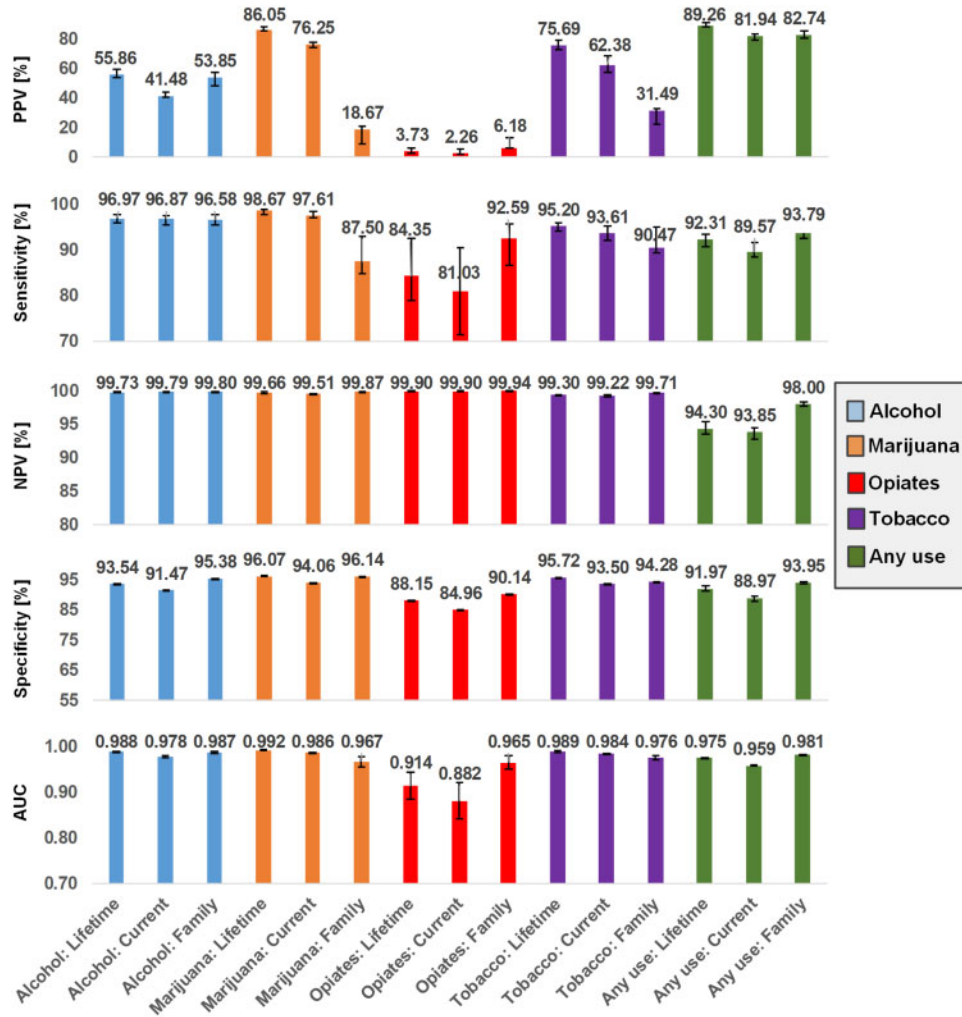


Figure 5. Performance of the deep learning model in detecting substance use categories and assertions on individual clinical narratives. Error bars indicate 95% confidence intervals. AUC: area under the receiver-operating characteristic curve; NPV: negative predictive value; PPV: positive predictive value.

substance use categories and assertions, with sensitivity $\geq 84.6\%$ and specificity $\geq 87.2\%$ at a significance level of .05.

The results and findings demonstrate a significant advance to the field of substance use research in pediatric settings. Performance using NLP and machine learning was consistent with the best results reported in the literature for alcohol and nicotine detection (91.8% and 96.6% for sensitivity, respectively),²³ and opioid and alcohol detection (0.951 and 0.730 for AUC, respectively).²⁵ Importantly, our results cover more comprehensive substance use categories and examine a pediatric population, in which rates of substance use disorder are generally lower than in adults.⁴ In addition, screening in pediatric settings often involve discussion of family substance use (Table 2), which have not been investigated in previous studies. Detecting family use and distinguishing it from subject use in this study demonstrates feasibility and fills a gap in knowledge that is unique to pediatrics. Finally, integration of evidence-based screening (eg, SBIRT [Screening, Brief Intervention and Referral to Treatment], CRAFFT)⁴⁸ requires providers to modify how they document screening results to be effective, which has been identified as a barrier to implementation in pediatric settings.¹⁶ The promising performance achieved by the ASUDS suggests potential for technology to support implementation of evidence-based screening without the

need to modify clinician behavior around documentation, addressing a critical first barrier to the delivery of substance use prevention and intervention.^{49,50} For instance, the computerized system could assemble a report of historic substance use information before an encounter, presenting both predictions regarding previous screening results and supportive evidence (eg, sentences mentioning substance use in notes, values of laboratory results) to facilitate clinician review, assist with determining change over time in substance use, and inform decisions about intervention during the encounter. Further, creation of such a system would improve secondary analysis of substance use information in EHRs, which could facilitate the dissemination of screening results to external institutions or to researchers interested in understanding substance use onset and risk.⁵¹

Error analysis

Despite high performance, false positives from LRM suggested that additional rules were needed to address conflicting records in the EHRs. Future studies of rule-based approaches could examine whether ordering records chronologically and using most recent record(s) reduces false positives. The error analysis on SIS also uncovered several areas of improvement. The majority of errors were

Table 3. Performance of the automated substance use detection system in detecting substance use categories and assertions on individual encounters

Category	Assertion	PPV (%) (95% CI)	SEN (%) (95% CI)	NPV (%) (95% CI)	SPEC (%) (95% CI)	AUC (95% CI)
Alcohol	Lifetime	70.78 (69.33-74.07)	91.19 (89.82-92.52)	99.07 (98.92-99.20)	96.13 (95.85-96.72)	0.979 (0.977-0.982)
	Current	41.63 (40.59-42.70)	97.84 (97.04-98.56)	99.82 (99.75-99.88)	89.48 (89.03-89.92)	0.969 (0.966-0.973)
	Family	89.58 (87.13-91.18)	98.22 (97.73-98.92)	99.57 (99.46-99.74)	97.32 (96.58-97.78)	0.991 (0.990-0.992)
Marijuana	Lifetime	89.66 (88.16-91.22)	98.89 (98.47-99.28)	99.74 (99.65-99.83)	97.42 (96.98-97.85)	0.993 (0.992-0.994)
	Current	83.01 (76.93-83.99)	97.46 (97.16-99.12)	99.53 (99.48-99.83)	96.44 (94.70-96.68)	0.990 (0.989-0.991)
	Family	19.36 (12.59-20.50)	91.29 (88.64-95.83)	99.87 (99.84-99.94)	94.77 (91.04-95.10)	0.972 (0.961-0.982)
Opiates	Lifetime	9.95 (6.34-31.06)	90.06 (84.63-93.74)	99.91 (99.85-99.95)	92.79 (87.82-98.30)	0.960 (0.941-0.980)
	Current	8.48 (6.22-20.10)	91.03 (84.83-95.86)	99.93 (99.88-99.97)	92.63 (89.53-97.36)	0.963 (0.944-0.981)
	Family	7.95 (7.61-21.84)	93.29 (86.59-96.34)	99.94 (99.88-99.97)	90.83 (90.48-97.32)	0.971 (0.957-0.985)
Tobacco	Lifetime	58.83 (57.55-60.81)	97.12 (96.06-97.83)	99.38 (99.16-99.53)	87.16 (86.42-88.28)	0.966 (0.964-0.969)
	Current	82.32 (80.25-83.81)	94.40 (93.55-95.36)	99.11 (98.98-99.26)	96.87 (96.40-97.18)	0.983 (0.980-0.985)
	Family	36.55 (25.41-41.06)	87.06 (85.05-92.09)	99.41 (99.33-99.62)	93.56 (88.58-94.69)	0.967 (0.962-0.973)
Any use	Lifetime	80.00 (79.21-80.75)	91.11 (90.44-91.80)	95.33 (94.99-95.67)	88.85 (88.29-89.37)	0.959 (0.957-0.962)
	Current	83.84 (83.00-84.71)	91.03 (90.25-91.78)	96.23 (95.92-96.53)	92.89 (92.45-93.33)	0.963 (0.960-0.966)
	Family	88.41 (87.80-90.89)	96.61 (94.77-97.02)	97.93 (96.90-98.18)	92.68 (92.24-94.49)	0.976 (0.975-0.978)

AUC: area under the receiver-operating characteristic curve; CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value; SEN: sensitivity; SPEC: specificity.

caused by omitting semantic information including experiencer (category 1; 24.28%), negation (category 2; 16.86%), temporality (category 6; 7.27%), and hypothetical statement (category 8; 4.59%). The errors suggested limitations of word-embedding features that were not sensitive to granularity of semantic meanings. For instance, the words *month* and *year* had similar vector values and the granularity of these timestamps was omitted. Incorporating meaningful linguistic features provided by knowledge-based NLP systems, including negation (for converting negated terms), temporal expressions (for capturing timestamps), and experiencer expressions (for identifying appropriate objects) might alleviate this problem and should be evaluated in future studies.^{18–20} In addition, over 10% of errors (category 3) were attributed to ambiguous content (eg, confusion of pronouns in object clauses) that required reasoning in the context. For example, “she discovered he began drinking alcohol” might imply mother found her son drinking (subject substance use) or daughter found her father drinking (family substance use). Enriching the feature set with syntax-based pronoun disambiguation should be explored in future studies to mitigate this issue.⁵² Another 9.65% of errors (category 4) was due to confusion between substance use categories, which often occurred when 2 substances had similar activities (eg, smoking tobacco vs marijuana). Likewise, the SIS confused between substance use and medication use because of similar content in a sentence (category 7; 5.94%). These issues could potentially be addressed by n-grams and syntactic n-grams that capture

contextual information and may contribute to better performance.^{53,54} Finally, a notable portion of false positive errors was triggered by substance description that implied potential use (category 5), which may be important to capture as an indicated need for screening or close monitoring and could be explored as an additional classification in future studies. Similarly, substance use secondary to another concern (category 10), including instances in which youths reported (eg, forcible use of substances, substance use as a method of suicide) are important to capture and relevant for clinical evaluation but represent a departure from the goals of this ASUDS.

Limitations

While this study makes an important contribution to advancing methods to extract substance use information from EHRs, there are limitations to be considered. First, system performance is based on one specific population (ie, youths 10 years of age and older, all enrolled in Medicaid, 50% in foster care) at a single institution. Additional evaluation is required to assess generalizability with diverse patient populations (eg, adult patients, private pediatric practices), institutions, and clinical settings. Effort to customize the LRM for other patient populations and settings may be minimal, and disseminating these findings is an important first step for advancing that work. In addition, NLP and machine learning techniques support retraining the SIS when new data become available. As such, if generalizability is not satisfactory, appropriate active learning

Table 4. Categorization and distribution of errors made by the substance information screener

ID	Category	Subcategory	Example	Rate (%)
1	Wrong experiencer (24.28%)	Misclassified substance use by family members as subject use	Patient had in utero exposure to drugs and alcohol.	15.59
		Misclassified substance use by the subject as family use	Mom found marijuana in patient's pocket tonight.	7.30
		Substance use by the third person (eg, friends, neighbors)	The patient was hanging out with kids that were doing drugs.	1.39
2	Missing negation (16.86%)	Omitting negation expression in the context	She denies marijuana and tobacco use, and she reports that she drinks beer occasionally.	16.86
3	Ambiguous content (11.45%)	Required reasoning based on the context to determine the user	She discovered he began drinking alcohol, using marijuana and smoking. (she: patient; he: patient's father)	8.63
		Failed in capturing subtle implications in the context	Patient is exposed to secondhand smoke outside the home. (no substance use for both subject and family)	1.43
		Two substance categories were presented in the same sentence	Drug and EtOH counseling is recommended. (could not infer if it was drug use or alcohol use)	1.39
4	Wrong category (9.65%)	The system captured the substance use behavior but categorized it into wrong category	Patient smoke 8 joints total.	9.65
5	Potential use (8.39%)	The context implied substance use such as drug possession and trafficking, drug screening, and instruction	He has been in temporary custody due to drug trafficking in home.	8.39
6	Missing temporality (7.27%)	Failed in capturing and reasoning dates in the context	She reported that she has not smoked marijuana since May 2015.	3.13
		Failed in reasoning vague temporal expression in the context	Patient has tested positive for marijuana in the past, per chart review.	4.14
7	No substance (5.94%)	The triggered terms were not related to substance use	She is needing increased pain control with opiates.	5.94
8	Hypothetical statement (4.59%)	Failed in detecting hypothetical statements or awareness	Mom reported to social worker that she suspects the client has been drinking alcohol.	4.59
9	Rare expressions (3.97%)	Errors caused by rare phrases such as concatenated words, rare semantic expressions or conflict findings	Drug abusesister. (The words abuse and sister were concatenated) Patient has said many things about drugs but is not clear about use.	3.97
10	Substance use secondary to another concern (1.83%)	Use of substance only in specific circumstance (forced use, suicidal attempts)	Suicidal description: patient overdosed today with alcohol, marijuana, and pills.	1.83
11	Unknown (5.77%)	Errors with unidentified reasons	N/A	5.77

N/A: No representative examples.

approaches could be implemented to retune the disseminated system automatically as new data become available.⁵⁵ Second, these analyses were restricted to retrospective data. Once reliability and generalizability are established, the ASUDS can be transferred to a production environment to adequately assess its usability and utility with prospective data in future studies. Finally, published studies have suggested that certain substance use screenings (eg, blood and urine test) are used more frequently in underserved adult populations such as in African Americans.^{56,57} To mitigate screening bias, the ASUDS did not only rely on structured indicators such as blood and urine testing to determine if screening had occurred or the screening results. It utilized a comprehensive analysis of clinical data, in which predictions were only made using descriptive language and structured indicators associated with substance use. Across all screening data, a similar effect was found in the current

study, in which BIPOC participants were slightly more likely to be screened and identified as using substances than would non-Hispanic White youths. Whether an ASUDS trained on the dataset would propagate screening bias or help reduce bias in provider-documented substance use screening warrants systematic investigation and will be a focus of our future work.

CONCLUSION

By integrating logic rules with NLP and machine learning, we developed an ASUDS to identify substance use information from both structured EHR data and unstructured clinical notes among pediatric patients. In a double-annotated, gold-standard-based evaluation of pediatric clinical data, the computerized system showed

good capacity for detecting substance use screening occurrences and results. The ASUDS achieved AUCs ≥ 0.957 across 15 substance categories and assertions, with sensitivity $\geq 84.6\%$ and specificity $\geq 87.2\%$ at a significance level of .05. In addition to demonstrating feasibility and rigor, this work confirmed the value of NLP and machine learning for detecting family substance use, a unique screening characteristic in pediatric settings. Validating the system's generalizability and utility across patient populations and clinical settings is a logical next step. Given its high performance in this stage of development, ASUDS holds great potential to facilitate research and healthcare delivery addressing substance use screening in adolescence and, ultimately, when combined with prevention and intervention, reduce risk of substance use disorders across the lifespan.

FUNDING

This work was supported by the National Institutes of Health (1K01DA041620, 2UL1TR001425-05A1, 1R01HD103630) and the Patient-Centered Outcomes Research Institute (PCORI/PCS-2018C1-11111). YN was also supported by internal funds from Cincinnati Children's Hospital Medical Center.

AUTHOR CONTRIBUTIONS

YN conceptualized the study, coordinated the data cleaning and annotation, developed the automated substance use detection system, analyzed the results, created the tables and figures, and wrote the manuscript. AB performed data cleaning and annotation, coordinated the result analysis, and contributed to the manuscript. KN performed data collection, data cleaning, and annotation; coordinated the result analysis; and contributed to the manuscript. SB conceptualized the study; provided specialist guidance on data collection, data cleaning, system development, and result analysis; and contributed to the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors thank Jay Ghalop for his support in providing the clinical dataset. Particular thanks go to Imani Crosby, Madeline Converse, Jenny Duma, Kim Nguyen, Kelsey Childress, and Elizabeth Hamik, who helped review the clinical data.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

DATA AVAILABILITY STATEMENT

The data underlying this article are from the National Institutes of Health-funded project to understand the emergence of substance use among adolescents in foster care (grant number: 1K01DA041620). The data cannot be shared publicly due to protected health information of individuals that participated in the study. All variables used in the study have been listed in the [Supplementary Appendix](#). Investigators who would like to use the data may make specific requests to collaborate via contacting the authors.

COMPETING INTERESTS STATEMENT

The authors have no competing interests to declare.

REFERENCES

1. McGinnis JM, Foege WH. Mortality and morbidity attributable to use of addictive substances in the United States. *Proc Assoc Am Physicians* 1999; 111 (2): 109–18.
2. National Institute on Drug Abuse. Cost of substance abuse. 2020. <https://www.drugabuse.gov/drug-topics/trends-statistics/costs-substance-abuse>. Accessed February, 19, 2021.
3. Johnston L, Miech R, O'Malley P, Bachman J, Schulenberg J, Patrick M. *Monitoring the Future National Survey Results on Drug Use, 1975-2018: Overview Key Findings on Adolescent Drug Use*. Ann Arbor, MI: Institute for Social Research, University of Michigan; 2019.
4. Substance Abuse and Mental Health Services Administration. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health (HHS Publication No. PEP19-5068, NSDUH Series H-54)*. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2019.
5. Trenz RC, Scherer M, Harrell P, Zur J, Sinha A, Latimer W. Early onset of drug and polysubstance use as predictors of injection drug use among adult drug users. *Addict Behav* 2012; 37 (4): 367–72.
6. Levy SJL, Williams JF; Committee on Substance use and Prevention. Substance use screening, brief intervention, and referral to treatment. *Pediatrics* 2016; 138 (1): e20161211.
7. Ghitza UE, Gore-Langton RE, Lindblad R, Shide D, Subramaniam G, Tai B. Common data elements for substance use disorders in electronic health records: The NIDA clinical trials network experience. *Addiction* 2013; 108 (1): 3–8.
8. Wu LT, Payne EH, Roseman K, et al. Clinical workflow and substance use screening, brief intervention, and referral to treatment data in the electronic health records: a national drug abuse treatment clinical trials network study. *EGEMS (Wash DC)* 2019; 7 (1): 35.
9. Levy S, Ziemnik RE, Harris SK, et al. Screening adolescents for alcohol use. *J Addict Med* 2017; 11 (6): 427–34.
10. Harris BR, Shaw BA, Sherman BR, Lawson HA. Screening, brief intervention, and referral to treatment for adolescents: attitudes, perceptions, and practice of New York school-based health center providers. *Subst Abuse* 2016; 37 (1): 161–7.
11. Jha AK. Meaningful use of electronic health records: the road ahead. *JAMA* 2010; 304 (15): 1709–10.
12. Kuhns LM, Carlino B, Greeley K, et al. A chart review of substance use screening and related documentation among adolescents in outpatient pediatric clinics: implications for practice. *Subst Abuse Treat Prev Policy* 2020; 15 (1): 36.
13. Tai B, Wu LT, Clark HW. Electronic health records: essential tools in integrating substance abuse treatment with primary care. *Subst Abuse Rehabil* 2012; 3: 1–8.
14. Sterling S, Kline-Simon AH, Satre DD, et al. Implementation of screening, brief intervention, and referral to treatment for adolescents in pediatric primary care: a cluster randomized trial. *JAMA Pediatr* 2015; 169 (11): e153145.
15. Sterling S, Kline-Simon AH, Weisner C, Jones A, Satre DD. Pediatrician and behavioral clinician-delivered screening, brief intervention and referral to treatment: substance use and depression outcomes. *J Adolesc Health* 2018; 62 (4): 390–6.
16. Palmer A, Karakus M, Mark T. Barriers faced by physicians in screening for substance use disorders among adolescents. *Psychiatr Serv* 2019; 70 (5): 409–12.
17. Mannelli P, Wu LT. Commentary on Winhusen et al. (2019): Substance use disorders, chronic diseases, and electronic health records—a paradigm for screening and intervention. *Addiction* 2019; 114 (8): 1471–2.
18. Li Q, Spooner SA, Kaiser M, et al. An end-to-end hybrid algorithm for automated medication discrepancy detection. *BMC Med Inform Decis Mak* 2015; 15 (1): 37.

19. Ni Y, Barzman D, Bachtel A, Griffey M, Osborn A, Sorter M. Finding warning markers: leveraging natural language processing and machine learning technologies to detect risk of school violence. *Int J Med Inform* 2020; 139: 104137.
20. Ni Y, Kennebeck S, Dexheimer JW, et al. Automated clinical trial eligibility prescreening: Increasing the efficiency of patient identification for clinical trials in the emergency department. *J Am Med Inform Assoc* 2015; 22 (1): 166–78.
21. Pestian JP, Sorter M, Connolly B, et al.; STM Research Group. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. *Suicide Life Threat Behav* 2017; 47 (1): 112–21.
22. Tang H, Solti I, Kirkendall E, et al. Leveraging food and drug administration adverse event reports for the automated monitoring of electronic health records in a pediatric hospital. *Biomed Inform Insights* 2017; 9: 1178222617713018.
23. Wang Y, Chen E, Pakhomov S, et al. Automated extraction of substance use information from clinical texts. *AMIA Annu Symp Proc* 2015; 2015: 2121–30.
24. Hylan TR, Von Korff M, Saunders K, et al. Automated prediction of risk for problem opioid use in a primary care setting. *J Pain* 2015; 16 (4): 380–7.
25. Dligach D, Afshar M, Miller T. Toward a clinical text encoder: pretraining for clinical natural language processing with applications to substance misuse. *J Am Med Inform Assoc* 2019; 26 (11): 1272–8.
26. Compton WM, Jones CM, Baldwin GT, Harding FM, Blanco C, Wargo EM. Targeting youth to prevent later substance use disorder: an underutilized response to the us opioid crisis. *Am J Public Health* 2019; 109 (S3): S185–S9.
27. Beal SJ. Using administrative and clinical data to detect drug use and HIV risk in foster care. NIH RePORT: RePORTER. 2017. <https://reporter.nih.gov/project-details/10086077>. Accessed March, 26, 2021.
28. Roberts A, Gaizauskas R, Hepple M, et al. Building a semantically annotated corpus of clinical texts. *J Biomed Inform* 2009; 42 (5): 950–66.
29. National Institute on Drug Abuse. Monitoring the Future 2020 Survey Results. 2020. <https://www.drugabuse.gov/drug-topics/related-topics/trends-statistics/infographics/monitoring-future-2020-survey-results>. Accessed April, 29, 2021.
30. National Institute on Drug Abuse. National Survey of Drug Use and Health. 2021. <https://www.drugabuse.gov/drug-topics/trends-statistics/national-drug-early-warning-system-ndews/national-survey-drug-use-health>. Accessed April, 29, 2021.
31. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012; 22 (3): 276–82.
32. Ni Y, Bermudez M, Kennebeck S, Liddy-Hicks S, Dexheimer J. A real-time automated patient screening system for clinical trials eligibility in an emergency department: design and evaluation. *JMIR Med Inform* 2019; 7 (3): e14185.
33. Ni Y, Wright J, Perentesis J, et al. Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients. *BMC Med Inform Decis Mak* 2015; 15 (1): 28.
34. De Silva TS, MacDonald D, Paterson G, Sikdar KC, Cochrane B. Systematized nomenclature of medicine clinical terms (SNOMED CT) to represent computed tomography procedures. *Comput Methods Programs Biomed* 2011; 101 (3): 324–9.
35. U.S. National Library of Medicine. Unified Medical Language System (UMLS). 2016. <https://www.nlm.nih.gov/research/umls/>. Accessed February, 19, 2021.
36. Stubbs A, Filannino M, Soysal E, Henry S, Uzun O. Cohort selection for clinical trials: N2c2 2018 shared task track 1. *J Am Med Inform Assoc* 2019; 26 (11): 1163–71.
37. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems*; 2013: 3111–9.
38. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw* 2005; 18 (5–6): 602–10.
39. Abadi M, Agarwal A, Barham P, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv, doi: <https://arxiv.org/abs/1603.04467>, 16 Mar 2017, preprint: not peer reviewed.
40. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 2006; 7: 91.
41. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. *BMJ* 1994; 308 (6943): 1552.
42. Altman DG, Bland JM. Diagnostic tests 2: predictive values. *BMJ* 1994; 309 (6947): 102.
43. Rice JA. *Mathematical Statistics and Data Analysis*. 3rd ed. Pacific Grove, CA: Duxbury Press; 2006.
44. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; 3 (1): 32–5.
45. McDonald JH. *Handbook of Biological Statistics*. 3rd ed. Baltimore, MD: Sparky House Publishing; 2014.
46. Robin X, Turck N, Hainard A, et al. Proc: An open-source package for r and s+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12: 77.
47. Jain AK, Duin PW, Jianchang M. Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intelligence* 2000; 22 (1): 4–37.
48. Mitchell SG, Gryczynski J, Schwartz RP, et al. Adolescent SBIRT implementation: generalist vs. Specialist models of service delivery in primary care. *J Subst Abuse Treat* 2020; 111: 67–72.
49. Ozechowski TJ, Becker SJ, Hogue A. SBIRT-a: Adapting SBIRT to maximize developmental fit for adolescents in primary care. *J Subst Abuse Treat* 2016; 62: 28–37.
50. Levy S, Ziemnik RE, Harris SK, et al. Screening adolescents for alcohol use: tracking practice trends of Massachusetts pediatricians. *J Addict Med* 2017; 11 (6): 427–34.
51. Sanchez-Roige S, Palmer AA. Electronic health records are the next frontier for the genetics of substance use disorders. *Trends Genet* 2019; 35 (5): 317–8.
52. Cardie C. Learning to disambiguate relative pronouns. In: *AAAI '92: Proceedings of the Tenth National Conference on Artificial Intelligence*; 1992: 38–43.
53. Manning CD, Schütze H. *Foundation of Statistical Natural Language Processing*. Cambridge, MA: MIT Press; 1999.
54. Sidorov G, Velasquez F, Stamatatos E, Gelbukh A, Chanona-Hernández L. Syntactic n-grams as machine learning features for natural language processing. *Expert Syst Appl* 2014; 41 (3): 853–60.
55. Tong S, Koller D. Support vector machine active learning with applications to text classification. *J Mach Learn Res* 2001; 2: 45–66.
56. Kunins HV, Bellin E, Chazotte C, Du E, Arnsten JH. The effect of race on provider decisions to test for illicit drug use in the peripartum setting. *J Womens Health* 2007; 16 (2): 245–55.
57. Roberts SC, Nuru-Jeter A. Universal screening for alcohol and drug use and racial disparities in child protective services reporting. *J Behav Health Serv Res* 2012; 39 (1): 3–16.