Original Paper

A Machine Learning Approach with Human-AI Collaboration for Automated Classification of Patient Safety Event Reports: Algorithm Development and Validation Study

Hongbo Chen[1], MSc; Eldan Cohen[1], PhD; Dulaney Wilson[2], PhD; Myrtede Alfred[1], PhD

[1]Department of Mechanical & Industrial Engineering, Faculty of Applied Science & Engineering, University of Toronto, Toronto, ON, Canada

[2]Department of Public Health Sciences, College of Medicine, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Myrtede Alfred, PhD
Department of Mechanical & Industrial Engineering
Faculty of Applied Science & Engineering
University of Toronto
27 King's College Cir
Toronto, ON, M5S 1A1
Canada
Phone: 1 4372154739
Email: myrtede.alfred@utoronto.ca

Abstract

Background: Adverse events refer to incidents with potential or actual harm to patients in hospitals. These events are typically documented through patient safety event (PSE) reports, which consist of detailed narratives providing contextual information on the occurrences. Accurate classification of PSE reports is crucial for patient safety monitoring. However, this process faces challenges due to inconsistencies in classifications and the sheer volume of reports. Recent advancements in text representation, particularly contextual text representation derived from transformer-based language models, offer a promising solution for more precise PSE report classification. Integrating the machine learning (ML) classifier necessitates a balance between human expertise and artificial intelligence (AI). Central to this integration is the concept of explainability, which is crucial for building trust and ensuring effective human-AI collaboration.

Objective: This study aims to investigate the efficacy of ML classifiers trained using contextual text representation in automatically classifying PSE reports. Furthermore, the study presents an interface that integrates the ML classifier with the explainability technique to facilitate human-AI collaboration for PSE report classification.

Methods: This study used a data set of 861 PSE reports from a large academic hospital’s maternity units in the Southeastern United States. Various ML classifiers were trained with both static and contextual text representations of PSE reports. The trained ML classifiers were evaluated with multiclass classification metrics and the confusion matrix. The local interpretable model-agnostic explanations (LIME) technique was used to provide the rationale for the ML classifier’s predictions. An interface that integrates the ML classifier with the LIME technique was designed for incident reporting systems.

Results: The top-performing classifier using contextual representation was able to obtain an accuracy of 75.4% (95/126) compared to an accuracy of 66.7% (84/126) by the top-performing classifier trained using static text representation. A PSE reporting interface has been designed to facilitate human-AI collaboration in PSE report classification. In this design, the ML classifier recommends the top 2 most probable event types, along with the explanations for the prediction, enabling PSE reporters and patient safety analysts to choose the most suitable one. The LIME technique showed that the classifier occasionally relies on arbitrary words for classification, emphasizing the necessity of human oversight.

Conclusions: This study demonstrates that training ML classifiers with contextual text representations can significantly enhance the accuracy of PSE report classification. The interface designed in this study lays the foundation for human-AI collaboration in the classification of PSE reports. The insights gained from this research enhance the decision-making process in PSE report classification, enabling hospitals to more efficiently identify potential risks and hazards and enabling patient safety analysts to take timely actions to prevent patient harm.

(JMIR Hum Factors 2024;11:e53378) doi: 10.2196/53378

KEYWORDS

accident; accidents; black box; classification; classifier; collaboration; design; document; documentation; documents; explainability; explainable; human-AI collaboration; human-AI; human-computer; human-machine; incident reporting; interface design; interface; interpretable; LIME; machine learning; patient safety; predict; prediction; predictions; predictive; report; reporting; safety; text; texts; textual; artificial intelligence

Introduction

Since the publication of the seminal report on patient safety—*To Err Is Human* [1], the importance of preventing adverse events in health care has been widely recognized. Adverse events refer to unintended or unexpected incidents that occur during hospital care that cause harm to a patient [2]. Common adverse events include complications, falls, and medication errors. These events can lead to prolonged hospital stays, permanent harm to patients, life-saving interventions, or even contributing to patient deaths [2,3]. Unfortunately, adverse events remain one of the top 10 leading causes of death and disability worldwide, resulting in 251,454 deaths annually in the United States alone [4]. In Organization for Economic Cooperation and Development (OECD) countries, 15% of total hospital activity is the direct result of adverse events [5]. The global cost of adverse events has been estimated at 42 billion USD annually [6].

Patient safety event (PSE) reporting systems, also called incident reporting systems, have been widely adopted in hospitals across the world as part of their efforts to mitigate adverse events and improve patient safety [7,8]. Multiple nations, including Canada, Japan, England, and Norway, have made it mandatory for hospitals to establish and maintain a PSE reporting system, either with individual health care systems or through centralized national incident reporting platforms [9]. The primary purpose of the PSE reporting system is to provide health care organizations with a centralized system for tracking and analyzing PSEs, thereby facilitating continuous learning and maintaining a record of PSEs for risk assessment and prevention [7,10]. PSE reporting systems are tools that allow frontline health care personnel to voluntarily report adverse events, near-misses, and unsafe conditions [11]. Each PSE report includes structured data, such as event types, patient harm level, date, and location of the event, as well as unstructured data, including a free-text section that contains the factual description of the event and the patient's outcome [12]. Following submission, PSE reports are reviewed by relevant hospital staff, such as risk managers, patient safety analysts, nurse managers, physicians, and biomedical engineers, to identify areas for patient safety and quality improvement within the hospital [13].

Accurately classifying PSE reports into their appropriate event type is crucial to ensure that these reports are directed to the relevant patient safety analyst, support organizational learning, identify patterns and trends in adverse events, and ultimately prioritize measures to reduce adverse events [14,15]. An event type refers to a specific class of events that share common characteristics [16]. Examples of event types include falls, medication-related issues, and diagnosis errors [17,18]. PSE reporting systems may have upwards of 20 categories of events. The formulation of these classification taxonomies generally involves systematically grouping PSE reports based on common characteristics [19]. The descriptions of event types are not always readily accessible to PSE reporters and patient safety analysts [15]. Previous studies have found that the classification of PSE reports is inconsistent depending on the reporter's profession, interpretation of the adverse event, and understanding of the PSE classification taxonomy [15,20]. Furthermore, 25% of PSE reports are labeled with vague or nonspecific categories such as "miscellaneous" and "other" and require time-consuming retrospective analysis for reclassification [21]. These problems are further exacerbated by the growing volume of PSEs reported [18,22]. For instance, hospitals in the state of New South Wales in Australia reported close to 195,000 PSEs in 2020 [23], while there were approximately 2.3 million PSEs reported to the National Reporting and Learning System in England from April 2021 to March 2022 [24].

In light of these challenges, it is imperative to find an efficient solution to ensure the reliable classification of PSE reports. Recent studies have used static text representations and supervised machine learning (ML) techniques to automate the PSE report classification [17,25,26]. However, static text representations ignore the ordering of the words and do not account for the differences in word meaning across different contexts. These limitations may result in suboptimal classification performance. With the emergence of deep learning, contextual text representation produced from transformer-based deep learning models has achieved state-of-the-art performance on a wide range of natural language processing tasks, including text classification [27]. The contextual representation of each word is based on its surrounding context within the text, allowing for a more accurate understanding of its usage across different contexts and facilitating knowledge transfer across languages [28]. Therefore, using contextual text representation in training ML classifiers presents a promising opportunity for achieving a more precise classification of PSE reports.

The integration of ML models into PSE reporting systems has important implications for human–artificial intelligence (AI) collaboration, given the roles of the incident reporter (front end) and patient safety analyst (backend). Various approaches for using ML classifiers can be developed, including at different levels of automation; however, unifying the strengths of both human expertise and AI offers the most promising route for effective implementation [29-31]. A crucial determinant for successfully implementing the human-AI collaboration approach is decision transparency [32,33], which is often referred to as explainability. Explainability is the concept that an ML model's prediction can be explained in a way that human operators can comprehend and reconstruct the model's reasoning [33]. Incorporating explainability techniques in human-AI collaboration is paramount as it facilitates a deeper understanding of the factors influencing the predictions, thereby fostering trust and understanding between human experts and AI systems. Therefore, embedding explainability into the human-AI collaboration holds significant potential for enhancing PSE report classification.

The main aim of this study is to examine the efficacy of contextual text representation in improving the accuracy of PSE report classification. To accomplish this, we trained, evaluated, and compared various ML classifiers with both static and contextual text representations. Additionally, we developed an interface to illustrate the integration of the ML classifier in an event reporting system to support human-AI collaboration for PSE report classification. Moreover, we enhanced the explainability of the ML classifiers by using an explainable AI technique. Furthermore, we have investigated the ML classifier's performance under 2 conditions, differentiated by whether the explanation is valid for the predicted event type. Based on this analysis, we offer recommendations for optimizing human-AI collaboration in the context of PSE report classification.

Methods

Data Collection

The data set for this study was obtained from a large academic hospital located in the Southeastern United States. A total of 861 PSE reports from the labor and delivery and mother-baby units were extracted from the PSE reporting system from January 1, 2019, to December 31, 2020. Each PSE report was assigned to a single event type from a set of 25 classes, such as complication of the surgery, fall, medication-related, and supply issues. The ML classifiers were trained exclusively on PSE reports from the 7 most frequently occurring event types. This selection was intended to create a more balanced training data set to reduce sampling bias and the risk of overfitting. The selected PSE reports used for training ML classifiers constitute approximately 72.8% (627/861) of the extracted reports (Table 1).

Table 1. Prevalence of patient safety event reports by event type in this study.

Event type	Extracted reports (n=861), n (%)
Care coordination or communication	186 (21.6)
Laboratory test	122 (14.2)
Medication related	89 (10.3)
Omission or errors in assessment, diagnosis, and monitoring	67 (7.8)
Maternal	58 (6.7)
Equipment or devices	56 (6.5)
Supplies	49 (5.7)
Total	627 (72.8)

Data Preprocessing

The free-text section of PSE reports was preprocessed before feeding into ML classifiers as input features. The preprocessing procedures include text normalization, feature extraction, data splitting, and data augmentation (Multimedia Appendix 1 [28,34-39]).

Classifier Training

A range of ML classifiers, including multinomial logistic regression (MLR), support vector machine (SVM), extreme gradient boosting, light gradient boosting, random forest (RF), *k*-nearest neighbor (KNN), and multilayer perceptron, were used for the classification of PSE reports. While SVM is a binary classifier, it is also capable of performing multiclass classification using the one-versus-one strategy. This involves treating the multiclass classification problem as a series of binary classification problems, creating $n \times (n - 1) / 2$ binary classifiers for each pair of classes, where n represents the total number of classes, and the final classification is based on the majority vote of all binary classifiers. Extreme gradient boosting, light gradient boosting, and RF are tree-based ensemble algorithms that are commonly used in text classification tasks [17,40]. The KNN classifier predicts the class of a data point based on the majority class among its nearest neighbors in the training data set. Multilayer perceptron is a feedforward neural network consisting of multiple layers of interconnected neutrons and trained using backpropagation.

To optimize the performance of ML classifiers, we used the 5-fold cross-validation grid search technique to identify the best combination of hyperparameters. During this process, a range of values of important hyperparameters (ie, regularization strength) is assessed with 5-fold cross-validation. For each combination of hyperparameters, the training set is randomly split into 5 distinct folds, and then the ML classifier is trained and evaluated 5 times, picking a different fold for evaluation every time and training on the remaining 4 folds. The optimized combination of hyperparameters is determined based on the average performance of the classifier on the F_1-score across the 5-fold cross-validation runs.

Classifier Evaluation

We evaluated the performance of the trained classifiers on the testing set with standard classification metrics, including accuracy, precision, recall, F_1-score, and area under the receiver operating characteristic curve. We also evaluated classifiers on top-2 accuracy, which measures the proportion of predictions where the correct event type is among the top 2 highest probability event types predicted by the classifier. The definitions and mathematical formulas of the evaluation metrics are shown in Multimedia Appendix 2. Each of these metrics provides a distinct perspective on the performance of the classifier, and collectively, they offer a comprehensive understanding of how well the classifier is functioning. Since we framed PSE report classification as a multiclass text classification problem, the precision, recall, F_1-score, and area under the receiver operating characteristic curve are computed for each class and combined using a weighted average where the weights correspond to the number of data points in each class.

Development and Assessment of Explainability

As the contextual text representation is generated from transformer-based neural network, which has a black box nature, we used the local interpretable model-agnostic explanations (LIME) technique to analyze the top-performing ML classifier trained with the contextual text representation. LIME is a post hoc, local perturbation technique that provides the explanation for a single prediction. LIME generates perturbed data by randomly removing words from a text document and trains a locally explainable model with perturbed data to simulate the original classifier's prediction [41]. By measuring how the classifier's prediction changes under these perturbations, LIME reflects the contributions of each word to the prediction. The importance of each word can then be assessed for a single prediction, revealing whether the ML classifier has learned to use relevant words for classifying PSE reports. We used LIME to generate explanations for the top-performing classifier's prediction, specifically by highlighting the words that the classifier deems influential for the prediction. We presented 3 distinct cases: one where the classifier effectively leveraged relevant words for accurate prediction, another where it failed to do so, and a final case that illustrated the explanation for a misclassification. In addition, we analyzed the top 5 most prevalent words identified by LIME for each event type.

A total of 2 human factors graduate students were recruited to assess the quality of the LIME explanations. For each PSE report in the test data set, the reviewers were asked to determine independently if any of the highlighted words were relevant to the predicted event type. Based on these evaluations, the reports were then categorized into 2 distinct groups: those in which the highlighted terms were deemed relevant to the predicted event types and those where they were deemed irrelevant. Discrepancies were resolved through discussions. The interrater reliability index (Cohen κ) was calculated to quantify the level of agreement between the reviewers. The ML classifier's accuracy and F_1-score were evaluated for these 2 groups of PSE reports. A subsequent comparison will explore the influence of explanation quality on prediction reliability.

Interface Development

In the typical workflow of PSE report classification, reporters need to provide a narrative description of the event as well as key attributes such as the event type, level of harm, date, and location of the event. Subsequent to this initial classification, the patient safety analyst will review the submitted report and decide if it needs to be recategorized to better reflect the nature of the event [17,42]. To support efficient and reliable categorization, the classifier will need to provide reporters with real-time support during the reporting process. We developed a PSE reporting interface to illustrate the integration of the ML classifier and the LIME explainability technique. In the design, the ML classifier provides multiple high-probability event types along with explanations for its prediction and allows the user to select the most appropriate event type. The interface was developed in Figma [43] and designed using guidance from previous research on incident reporting systems, including question type, mandatory and optional questions, and taxonomy for event type and harm level [44,45].

Ethical Considerations

The study was approved by the Medical University of South Carolina Hospital's institutional review board (Pro00105892). Following data extraction, PSE reports were anonymized in accordance with privacy regulation guidelines.

Results

Performance Comparison

We evaluated the trained ML classifier's classification performance on both static and contextual text representations (Multimedia Appendix 3). The performance of the top-performing ML classifier trained with static and contextual text representations is shown in Table 2. Our results showed that for static text representation, the MLR classifier trained with term frequency–inverse document frequency (TF-IDF) achieved the best performance, with an F_1-score of 0.631 and an accuracy of 66.7% (84/126). On the other hand, for contextual text representation, the SVM classifier trained with RoBERTa-base outperformed others, with an F_1-score of 0.753 and an accuracy of 75.4% (95/126). The SVM classifier trained with RoBERTa-base showed a 19.3% elative improvement in F_1-score and a 13% (11/85) relative improvement in accuracy compared to the MLR classifier trained with TF-IDF for contextual text representation. In addition, we compared the accuracy (95/126, 75.4%) and top 2 accuracy (107/126, 84.9%) of the SVM classifier trained with RoBERTa-base and observed that 9.5% (12/126) of PSE reports' true event type was predicted as the second highest probability event type by the classifier, which represents 39% (12/31) of misclassified PSE reports.

Table 2. Performance of top-performing ML classifiers trained with static and contextual text representations.

Metric	Top-performing ML[a] model trained with the static text representation			Top-performing ML model trained with the contextual text representation		
	Performance	ML classifier	Text Representation	Performance	ML classifier	Text Representation
Accuracy (%)	66.67	MLR[b]	TF-IDF[c]	75.40	SVM[d]	RoBERTa-base
Top 2 accuracy (%)	85.71	MLR	TF-IDF	88.10	MLP[e]	xlm-RoBERTa-base
Precision	0.707	KNN[f]	TF-IDF	0.757	SVM	RoBERTa-base
Recall	0.667	MLR	TF-IDF	0.754	SVM	RoBERTa-base
F_1-score	0.631	MLR	TF-IDF	0.753	SVM	RoBERTa-base

[a] ML: machine learning.

[b] MLR: multinomial logistic regression.

[c] TF-IDF: term frequency–inverse document frequency.

[d] SVM: support vector machine.

[e] MLP: multilayer perceptron.

[f] KNN: *k*-nearest neighbor.

Performance on Classifying Individual Event Types

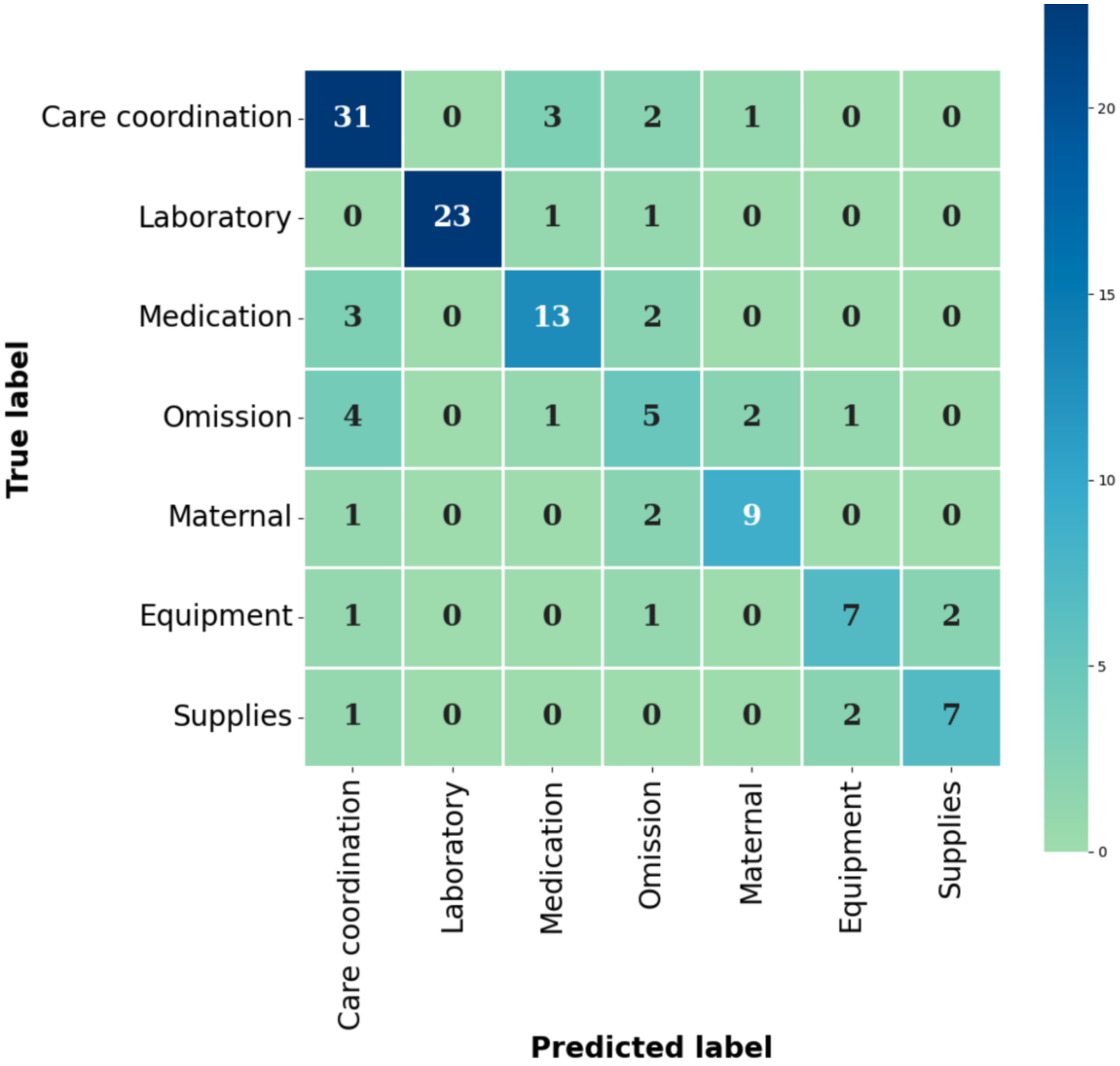
We analyzed the performance of the SVM classifier trained with RoBERTa-base on individual event types (Table 3). The F_1-score measure for different event types ranged from 0.958 (laboratory test) to 0.400 (omission or errors in assessment, diagnosis, and monitoring).

Table 3. Performance of support vector machine+RoBERTa-base on the individual event type.

Event type	Precision	Recall	F_1-score
Care coordination or communication	0.721	0.838	0.775
Laboratory test	1.000	0.920	0.958
Medication related	0.765	0.722	0.743
Omission or errors in assessment, diagnosis, and monitoring	0.417	0.385	0.400
Maternal	0.750	0.750	0.750
Equipment or devices	0.700	0.636	0.667
Supplies	0.778	0.700	0.737

Figure 1 shows the confusion matrix for the SVM classifier trained with RoBERTa-base evaluated on the test set. A confusion matrix is a table that visualizes the performance of a classifier. The main diagonal value is the number of PSE reports that have been classified as true event types, whereas off-diagonal values are the number of PSE reports that have been wrongly classified. While the classifier was able to classify the majority of event types of PSE reports correctly, there is a consistent misclassification of the omission or errors in assessment, diagnosis, or monitoring PSE report as the care coordination or communication (coordination) event type.

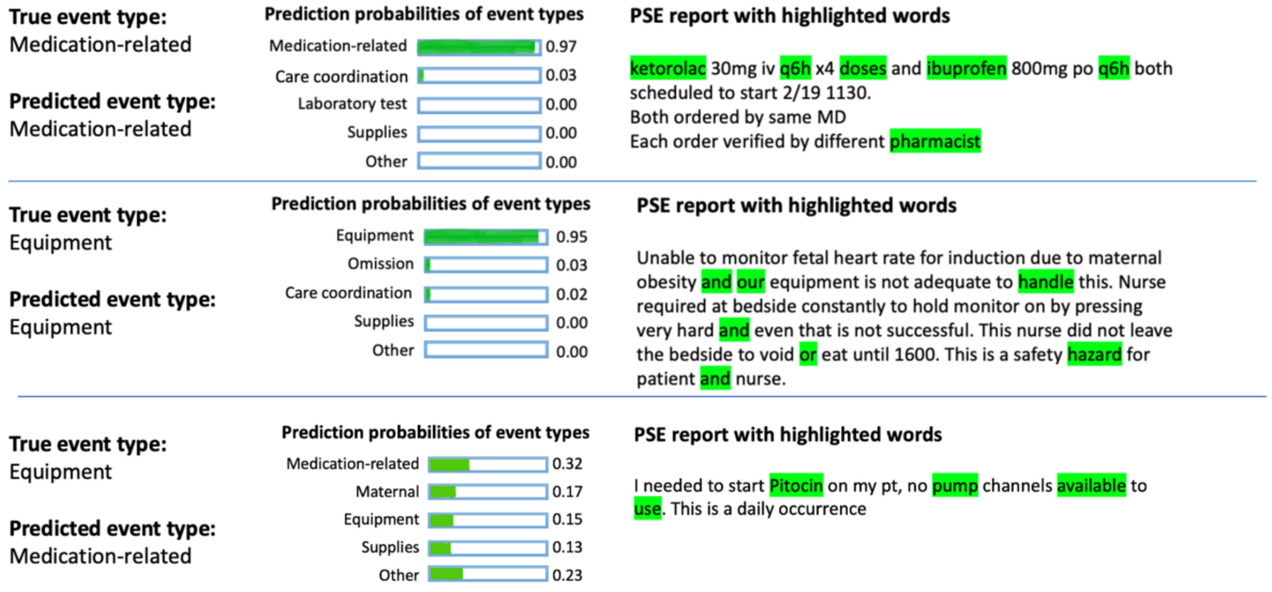
Figure 1. Confusion matrix for the testing set evaluation with a support vector machine classifier trained with RoBERTa-base.



LIME-Based Explainability Analysis

We used LIME to evaluate whether the SVM classifier trained with RoBERTa-base has leveraged informative words for classification. Figure 2 presents 3 examples of explanations for the classifier's predictions. At the top of Figure 2, LIME identified "ketorolac," "ibuprofen," and "doses" from the PSE report as important words for classifying the report into the medication-related event type, which is reasonable given the report's association with incorrect medication doses. Conversely, in the middle of Figure 2, LIME highlighted "our," "handle," and "or" from the text as important words for classifying the report into the equipment or device event type. Although the predicted event type was correct, the classifier relied on irrelevant words for the classification. At the bottom of Figure 2, a case of misclassification is shown. LIME highlighted "pitocin," "pump," "available," and "use" as influential words for classifying the PSE report into medication-related event type when it belongs to the equipment class. In addition, for each event type, we extract the 5 most prevalent words that were deemed important for the classifier's prediction across the whole data set (Table 4). This inclusion of stop words (ie, "was," "not," and "till") among influential terms, as shown in Table 4, demonstrated that the classifier does not always rely on relevant words for making classifications.

Figure 2. Local interpretable model-agnostic explanations of support vector machine classifiers trained with RoBERTa-base. MD: medical doctor; PSE: patient safety event; pt: patient.



ketorolac 30mg iv q6h x4 doses and ibuprofen 800mg po q6h both scheduled to start 2/19 1130.
Both ordered by same MD
Each order verified by different pharmacist

Unable to monitor fetal heart rate for induction due to maternal obesity and our equipment is not adequate to handle this. Nurse required at bedside constantly to hold monitor on by pressing very hard and even that is not successful. This nurse did not leave the bedside to void or eat until 1600. This is a safety hazard for patient and nurse.

I needed to start Pitocin on my pt, no pump channels available to use. This is a daily occurrence

Table 4. The 5 most prevalent and important words for each event type were derived from the support vector machine classifier trained with RoBERTa-base.

Event type	Prevalent influential words highlighted by local interpretable model-agnostic explanations
Care coordination or communication	requested, delayed, patient, not, follow
Laboratory test	specimen, lab, labels, collection, results
Medication related	patches, doses, orders, medication, pitocin
Omission or errors in assessment, diagnosis, and monitoring	warning, patient, was, till, late
Maternal	baby, hysterectomy, stable, pumping, hemorrhage
Equipment or devices	instruments, trays, notified, malfunctioning, faulty
Supplies	vendor, sterile, available, needed, OR

After reviewing the LIME explanations for each PSE report in the test data set, 73.8% (93/126) of the reports were categorized into a subset where at least 1 highlighted word was deemed relevant to the predicted event type. The remaining reports comprised a second subset where no highlighted words were relevant. The interrater reliability index measured by Cohen κ between the 2 reviewers was 0.83, indicating substantial agreement. Table 5 presents the performance of the top-performing ML classifier for both subsets. For the first subset, the classifier achieved an accuracy of 84% (78/93) and an F_1-score of 0.825. In contrast, the second subset showed a classifier accuracy of 52% (17/33) and an F_1-score of 0.549.

Table 5. Performance of a top-performing machine learning classifier on reports that have relevant words highlighted and reports with irrelevant words highlighted.

Metric	PSE[a] reports with relevant words highlighted	PSE reports with irrelevant words highlighted
Number of PSE reports, n	93	33
Percentage of test data set (%)	73.81	26.19
Accuracy (%)	83.87	51.51
F_1-score	0.825	0.549

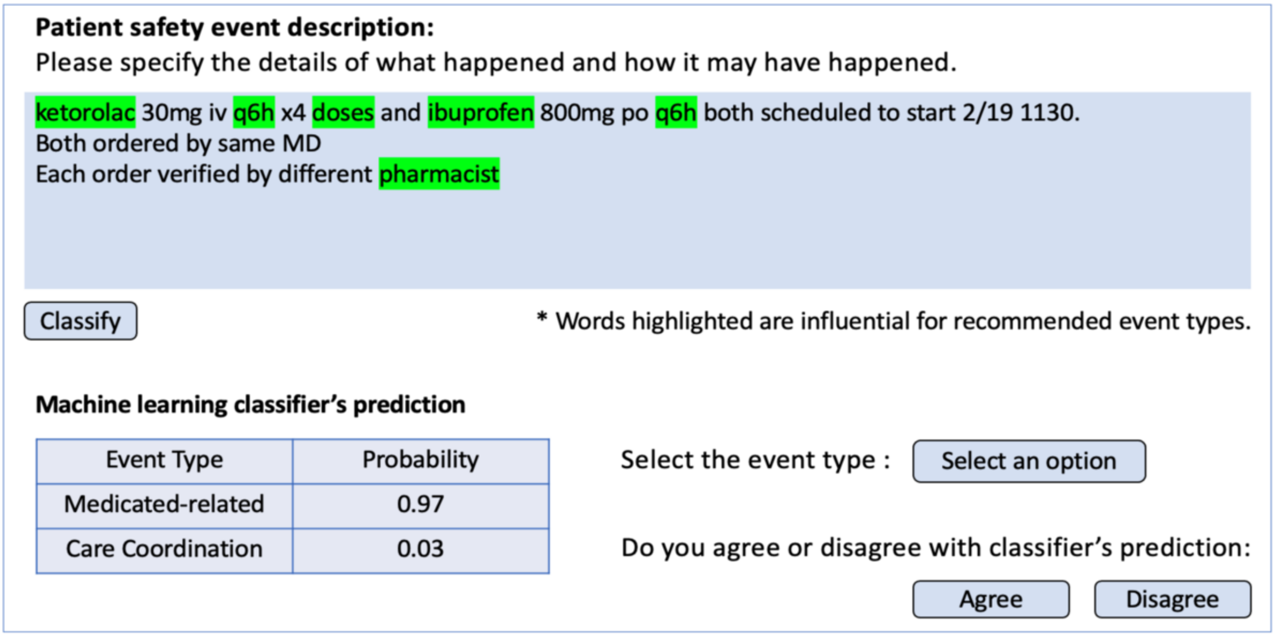
[a]PSE: patient safety event.

PSE Reporting System Interface

We designed an event reporting interface that integrates both the ML classifier and the LIME explainability technique. Figure 3 shows the event classification screen, where reporters enter a narrative description of the event after providing the details of the event, including date, time, unit, and information about the patient and reporter. Before describing the event in narrative form, reporters also choose among factors that contributed to the incident and the level of harm experienced by the patient.

Once the reporter enters their narrative and selects the "classify" button, the system activates the ML classifier. Subsequently, the interface displays the top 2 most probable event types, along with their associated probability distributions, in the lower left section. Simultaneously, the LIME technique will identify influential words that significantly contributed to the predicted event type, highlighting these words in green in the upper section of the dashboard. Based on the predicted event types and words highlighted for their influence on the prediction, the reporter may select the most suitable event type from a drop-down menu located in the lower-right section of the dashboard. Following this selection, reporters are queried on whether they agree with the classifier's prediction, and the collected data can be used to guide subsequent refinement of the ML classifier.

Figure 3. Interface visualization of a patient safety event report classifier coupled with the local interpretable model-agnostic explanations technique. MD: medical doctor.



Patient safety event description:

Please specify the details of what happened and how it may have happened.

ketorolac 30mg iv q6h x4 doses and ibuprofen 800mg po q6h both scheduled to start 2/19 1130.
Both ordered by same MD
Each order verified by different pharmacist

* Words highlighted are influential for recommended event types.

Machine learning classifier's prediction

Event Type	Probability
Medicated-related	0.97
Care Coordination	0.03

Select the event type : Select an option

Do you agree or disagree with classifier's prediction:

Discussion

Overview

PSE event reporting systems are commonly used in health systems and hospitals across the world [46]. Data collected in PSE reporting systems drive quality improvement and patient safety efforts and supports regulatory reporting requirements for hospitals. The erroneous classification of PSE reports can impede the learning capabilities of the PSE reporting system, leading to suboptimal performance in detecting and preventing potential patient safety hazards [20]. It can also result in a substantial time cost for reclassifying PSE reports and compromise the integrity of a PSE database when analysts are investigating trends in events to develop effective solutions [17]. Previous studies have trained ML classifiers with static text representations for automatic PSE classification [12,17,25,26]. This study aimed to investigate whether using contextual text representations can further improve the accuracy of classifying PSE reports. We trained and evaluated a range of ML classifiers using both static and contextual text representations. To the best of our knowledge, this is the first time that contextual text representation has been used for training ML classifiers for PSE report classification. We analyzed the confusion matrix of the top-performing classifier to identify prevalent misclassified event types. Furthermore, aiming for more accurate and reliable PSE report classification, we incorporated an explainability technique to support human-AI collaboration and designed an interface to illustrate the possible integration of the ML classifier in PSE reporting systems.

Principal Findings

In this study, we extensively investigated the potential of using contextual representation for improving PSE report classification. The leading classifier trained with the static text representation (MLR trained with TF-IDF) was able to achieve an accuracy of 66.7% (84/126). This accuracy considerably exceeds the baseline accuracy of 29.4% (37/126), which involves classifying all PSE reports into the majority event type. However, using contextual text representation proved more efficacious. The SVM trained with contextual text representation (RoBERTa-base) was able to achieve an accuracy of 75.4% (95/126), reflecting a 13% (11/84) relative improvement in accuracy compared to the best-performing classifier trained with static text representation. While the achieved accuracy of 75.4% may not appear outstanding in isolation, it represents a significant advance compared with static text representations and exceeds the baseline, given the limited size of the data set. The improvement in classifier performance can be attributed to the use of contextual text representations, which can capture not only the meaning of individual words but also the complex and subtle ways in which words interact with each other in a specific context. Therefore, contextual text representation overcomes some limitations of static text representation, which relies primarily on word frequency and co-occurrence to represent text. Moreover, contextual text representation does not require explicit text normalization while also avoiding issues associated with high-dimensionality and sparsity commonly found in static text representations. Hence, when training ML classifiers for PSE reporting systems, contextual text representation should be prioritized over static text representation to ensure the highest level of accuracy in classifying PSE reports.

As part of our investigation, we evaluated the performance of the top-performing classifier trained with contextual text representation on individual event types. While the classifier demonstrated impressive performance in accurately classifying laboratory test PSE reports (F_1-score=0.958), it struggled with classifying omissions or errors in assessment, diagnosis, and monitoring PSE reports, resulting in an unsatisfactory F_1-score of 0.400. To investigate this discrepancy, we analyzed the confusion matrix for the classifier and discovered that omissions or errors in assessment, diagnosis, and monitoring PSE reports were frequently misclassified as the coordination event type. This misclassification can be attributed to the multiclass nature of PSE reports. For example, a failure to document the removal of a patient's epidural catheter (omission or errors in assessment, diagnosis, and monitoring) could lead to a medication ordered by a physician (such as Lovenox) being withheld by the pharmacy due to a complication risk (coordination). On the other hand, the laboratory test is a more distinct event type in comparison to the other event types, and the classifier was able to correctly classify the majority of these reports. The observation obtained from the confusion matrix implies that PSE reports can potentially have more than 1 event type. This finding is consistent with previous studies [25,26]. The finding also underscores the need for further refinement in the development of the PSE taxonomy to create more distinctive event types. Another potential solution for addressing the multiclass nature of PSEs is to enable multiple event-type assignments [47]. Alternatively, the ML classifier can provide several probable event types, allowing the user to select the most appropriate one. We evaluated the top 2 accuracy of the top-performing ML classifier trained with contextual text representation and observed that 39% (12/31) of misclassified PSE reports' true event type was predicted as the second-highest probability event type by the classifier. The finding suggests that there is a greater chance for the ML classifier to provide the correct event type when considering multiple options. As event reporting systems usually encompass over 20 event types, which can be difficult to memorize or access [17], narrowing down the PSE report's potential event types to a smaller range also reduces the cognitive workload for PSE reporters during the classification process [48] and enhances the efficiency of reclassifying PSE reports for patient safety analysts.

We used LIME to showcase 3 predictions' explanations and demonstrated cases where the ML classifier used informative words for classifying the PSE report and where it used irrelevant words for classification. These results highlight the importance of not solely relying on the ML classifier's prediction and underscore the need for explainability and transparency in using the ML classifier for PSE report classification. Additionally, we showed the top 5 most prevalent words the ML classifier deemed important in the PSE reports for each event type. These words are indicative of the prevalent themes and issues within specific event types. Understanding the context and relationships between these prevalent informative words and specific event types can potentially provide valuable insights into the factors contributing to different types of PSEs. Furthermore, we have evaluated the top-performing ML classifier's performance on 2 subsets of PSE reports, differentiated by whether the highlighted word by LIME is relevant to the predicted event type. Our findings reveal that the majority of PSE reports (93/126) have at least 1 relevant word highlighted, with the classifier achieving an accuracy of 84% (78/93) on these reports. Conversely, accuracy drops to 52% (17/33) when irrelevant words are highlighted. Such a disparity in performance emphasizes the necessity for additional scrutiny from reporters and patient safety analysts, particularly when dealing with PSE reports that have irrelevant words highlighted.

While previous research has focused on the development of ML classifiers, none of these previous works have investigated the potential integration of the classifier within the PSE reporting system in a manner that aligns with the workflow of the front-end reporter. We designed an interface to demonstrate the feasibility of a collaborative human-AI approach for event categorization. The interface provides the PSE reporter with multiple probable event types and associated explanations for the ML classifier's prediction. This approach aligns with the principles of level 2 automation, where ML classifiers aids human decision-making rather than fully automating it [49]. This collaboration optimally combines human expertise with ML capabilities, potentially reducing cognitive workload and memorization of the taxonomy while also reducing the risks associated with overreliance on automation. Numerous studies have shown that the human-AI collaboration approach can improve the decision-making process [50-52], indicating its potential for enhancing PSE report classification. Furthermore, the interface also integrates the LIME explainability technique, which offers real-time insights into the rationale for the probable event types. Given the role of reporters and patient safety analysts in the incident reporting process, the use of explainability techniques can also increase trust in the recommendation provided by the ML classifier as it provides transparent and interpretable reasoning for the classification decisions [50,51]. Using LIME to highlight top informative words in real time for a PSE report can assist PSE reporters by emphasizing keywords in their narratives that are linked to the proposed classification. Highlighting informative words can also facilitate patient safety analysts working at the back end by providing insights into why a specific event type was chosen for classification. Such transparency not only clarifies current recommendations but also guides analysts in identifying influential terms for future report classifications. Previous research has illustrated the value of automation transparency in supporting appropriate levels of trust in the system, including decision support systems [32]. Additionally, regularly checking the explanations of the ML classifier's prediction enables continuous monitoring of the classifier's performance, identification of issues, and refinement [52]. As we have only designed the interface, additional research is needed to test the effectiveness of this approach in PSE report classification. Assessing the interface's impact on cognitive workload and decision-making accuracy is essential for ensuring its usability and adoption in the event reporting system. We plan to undertake a usability testing study with health care professionals in a subsequent study.

Comparison With Previous Work

Research into the use of ML classifiers for the automation of PSE report classification has been relatively scarce. Wang et al [26] used logistic regression and SVM with the binary count, term frequency, and TF-IDF text representation to classify ten types of PSE reports, reaching an F_1-score as high as 0.783. However, they used a considerably larger data set (n=2860). Fong et al [17] achieved an accuracy rate of 92.0% (284/309) when they examined the usage of an ML classifier for classifying miscellaneous PSE reports using SVM, RF, and logistic regression with TF-IDF [17]. They also used a much larger data set (n=70,051). Ong et al [12] investigated the feasibility of using an ML classifier to automatically classify 2 types of PSE reports, including inadequate clinical handover and incorrect patient identification. They used Bag of Words model for text representation and trained both SVM and naive Bayes on classifying PSE reports, reaching accuracy as high as 98% (364/372). However, they framed the problem as a binary classification problem, which inherently has a higher baseline accuracy compared to our investigation. In this study, we've performed an in-depth comparative analysis with the available PSE data set and compared the established methods of classifying PSE reports and our novel method of using contextual text representations for classification. Our findings reveal that our proposed method outperforms the traditional models in terms of accuracy (ie, 84/126, 66.7% vs 95/126, 75.4%) and F_1-score (ie, 0.631 vs 0.753). This underlines the significance of our approach and its potential to advance the field of using ML classifiers for PSE report classification.

Limitations

There are several limitations to this study. First, the PSE reports used to train the ML classifiers were obtained from the maternal care units of a single hospital in the United States; therefore, the classifier might not generalize well to other settings. Second, this research's scope was constrained by the limited amount of PSE report data, and only 7 prevalent classes were incorporated for training the ML classifiers. The restricted quantity of PSE reports might also result in an underestimation of the ML classifier's actual capabilities [12]. Third, the quality of the LIME explanations was assessed by 2 graduate students; thus, further investigation is needed for a more robust validation of explanation quality. Furthermore, we have not yet empirically tested the interface for potential decision-making biases it may introduce.

Future research should investigate the performance of ML classifiers trained with contextual text representations on a larger and more diverse data set. Additionally, while we plan to refine the interface and test whether it supports event classification, future research can continue to investigate the appropriate way of incorporating the ML classifier into the reporting and reviewing workflow of PSE report classification and examine various human-AI collaboration approaches. Future studies should explore the potential biases (ie, automation bias) that the interface may introduce into the analysts' decision-making process.

Conclusions

Improving the precision of PSE report classifications is a multifaceted task, involving both the refinement of the event type taxonomy and adequate training of hospital staff on the event reporting system. Despite these challenges, ML classifiers offer substantial potential to support accurate classification throughout the reporting and reviewing process. The findings of this study contribute to the advancement of ML classifiers for PSE report classification by demonstrating the superior performance of contextual text representation over static text representations in achieving more accurate classification outcomes. The integration of explainability techniques in ML classifiers fosters trust in their usage and provides valuable insights for informed decision-making and potential adjustments to the classifier. An event reporting interface that integrates an ML classifier with collaborative decision-making capabilities offers the potential to achieve an efficient and reliable PSE report classification process. These approaches can ultimately help hospitals identify risks and hazards promptly and take timely and informed actions to mitigate adverse events and reduce patient harm.

Acknowledgments

This research was funded by the Agency for Healthcare Research and Quality (1R03HS027680).

Authors' Contributions

HC was responsible for the conceptualization, data analysis, and drafting of the manuscript. EC contributed to the conceptualization, methodology design, and review and revision of the document. DW contributed to data acquisition and funding acquisition. MA contributed to data acquisition, conceptualization, funding acquisition, and the review and revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Data preprocessing procedures.
[DOCX File , 14 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Evaluation metrics for examining patient safety event machine learning classifiers.
[DOCX File , 15 KB-Multimedia Appendix 2]

Multimedia Appendix 3

The performance of machine learning classifiers in classifying patient safety event report event type.
[DOCX File , 20 KB-Multimedia Appendix 3]

References

1. Kohn LT, Corrigan JM, Donaldson MS, Committee on Quality of Health Care in America. To Err Is Human: Building a Safer Health System. Volume 627. Washington (DC). National Academies Press (US); 2000.
2. Nebeker JR, Barach P, Samore MH. Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. Ann Intern Med. May 18, 2004;140(10):795-801. [FREE Full text] [doi: 10.7326/0003-4819-140-10-200405180-00009] [Medline: 15148066]
3. Chopard D, Treder MS, Corcoran P, Ahmed N, Johnson C, Busse M, et al. Text mining of adverse events in clinical trials: deep learning approach. JMIR Med Inform. 2021;9(12):e28632. [FREE Full text] [doi: 10.2196/28632] [Medline: 34951601]
4. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ. 2016;353:i2139. [FREE Full text] [doi: 10.1136/bmj.i2139] [Medline: 27143499]
5. Slawomirski L, Auraaen A, Klazinga NS. The Economics of Patient Safety: Strengthening a Value-based Approach to Reducing Patient Harm at National Level. Paris. OECD Publishing; 2017.
6. Medication without harm. World Health Organization. 2023. URL: https://www.who.int/initiatives/medication-without-harm [accessed 2023-03-08]
7. Hewitt TA, Chreim S. Fix and forget or fix and report: a qualitative study of tensions at the front line of incident reporting. BMJ Qual Saf. 2015;24(5):303-310. [FREE Full text] [doi: 10.1136/bmjqs-2014-003279] [Medline: 25749025]
8. Ward JK, Armitage G. Can patients report patient safety incidents in a hospital setting? a systematic review. BMJ Qual Saf. 2012;21(8):685-699. [FREE Full text] [doi: 10.1136/bmjqs-2011-000213] [Medline: 22562875]
9. Li E, Clarke J, Ashrafian H, Darzi A, Neves AL. The impact of electronic health record interoperability on safety and quality of care in high-income countries: systematic review. J Med Internet Res. 2022;24(9):e38144. [FREE Full text] [doi: 10.2196/38144] [Medline: 36107486]
10. Vincent C. Incident reporting and patient safety. BMJ. 2007;334(7584):51. [FREE Full text] [doi: 10.1136/bmj.39071.441609.80] [Medline: 17218667]
11. Albolino S, Tartaglia R, Bellandi T, Amicosante AMV, Bianchini E, Biggeri A. Patient safety and incident reporting: survey of Italian healthcare workers. Qual Saf Health Care. 2010;19(Suppl 3):i8-i12. [FREE Full text] [doi: 10.1136/qshc.2009.036376] [Medline: 20959324]
12. Ong MS, Magrabi F, Coiera E. Automated categorisation of clinical incident reports using statistical text classification. Qual Saf Health Care. 2010;19(6):e55. [FREE Full text] [doi: 10.1136/qshc.2009.036657] [Medline: 20724392]
13. Herzer KR, Mirrer M, Xie Y, Steppan J, Li M, Jung C, et al. Patient safety reporting systems: sustained quality improvement using a multidisciplinary team and "good catch" awards. Jt Comm J Qual Patient Saf. 2012;38(8):339-347. [FREE Full text] [doi: 10.1016/s1553-7250(12)38044-6] [Medline: 22946251]
14. Fong A, Hettinger AZ, Ratwani RM. Exploring methods for identifying related patient safety events using structured and unstructured data. J Biomed Inform. 2015;58:89-95. [FREE Full text] [doi: 10.1016/j.jbi.2015.09.011] [Medline: 26432354]
15. Brubacher JR, Hunte GS, Hamilton L, Taylor A. Barriers to and incentives for safety event reporting in emergency departments. Healthc Q. 2011;14(3):57-65. [FREE Full text] [doi: 10.12927/hcq.2011.22491] [Medline: 21841378]
16. Conceptual framework for the international classification for patient safety version 1.1: final technical report January 2009. World Health Organization. 2010. URL: https://apps.who.int/iris/handle/10665/70882 [accessed 2023-03-08]
17. Fong A, Behzad S, Pruitt Z, Ratwani RM. A machine learning approach to reclassifying miscellaneous patient safety event reports. J Patient Saf. 2021;17(8):e829-e833. [FREE Full text] [doi: 10.1097/PTS.0000000000000731] [Medline: 32555052]
18. Ngo J, Lau D, Ploquin J, Receveur T, Stassen K, Del Castilho C. Improving incident reporting among physicians at south health campus hospital. BMJ Open Qual. 2022;11(4):e001945. [FREE Full text] [doi: 10.1136/bmjoq-2022-001945] [Medline: 36207052]
19. Schrager JD, Schuler K, Isakov AP, Wright DW, Yaffee AQ, Jacobson KL, et al. Development and usability testing of a web-based COVID-19 self-triage platform. West J Emerg Med. 2020;21(5):1054-1058. [FREE Full text] [doi: 10.5811/westjem.2020.7.48217] [Medline: 32970554]
20. Gong Y, Song HY, Wu X, Hua L. Identifying barriers and benefits of patient safety event reporting toward user-centered design. Saf Health. 2015;1(1):1-9. [FREE Full text] [doi: 10.1186/2056-5917-1-7]
21. Gong Y. Data consistency in a voluntary medical incident reporting system. J Med Syst. 2011;35(4):609-615. [FREE Full text] [doi: 10.1007/s10916-009-9398-y] [Medline: 20703528]

22. Koike D, Ito M, Horiguchi A, Yatsuya H, Ota A. Implementation strategies for the patient safety reporting system using consolidated framework for implementation research: a retrospective mixed-method analysis. BMC Health Serv Res. 2022;22(1):409. [FREE Full text] [doi: 10.1186/s12913-022-07822-9] [Medline: 35346182]
23. Biannual incident report. Clinical Excellence Commission. URL: https://www.cec.health.nsw.gov.au/Review-incidents/Biannual-Incident-Report [accessed 2023-05-16]
24. Organisation patient safety incident report up to March 2022. NHS England. 2022. URL: https://www.england.nhs.uk/publication/organisation-patient-safety-incident-report-up-to-march-2022/ [accessed 2023-03-08]
25. Evans HP, Anastasiou A, Edwards A, Hibbert P, Makeham M, Luz S, et al. Automated classification of primary care patient safety incident report content and severity using supervised Machine Learning (ML) approaches. Health Informatics J. 2020;26(4):3123-3139. [FREE Full text] [doi: 10.1177/1460458219833102] [Medline: 30843455]
26. Wang Y, Coiera E, Runciman W, Magrabi F. Using multiclass classification to automate the identification of patient safety incident reports by type and severity. BMC Med Inform Decis Mak. 2017;17(1):84. [FREE Full text] [doi: 10.1186/s12911-017-0483-8] [Medline: 28606174]
27. Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, Gao J. Deep learning—based text classification: a comprehensive review. ACM Comput Surv. 2021;54(3):1-40. [FREE Full text] [doi: 10.1145/3439726]
28. Liu Q, Kusner MJ, Blunsom P. A survey on contextual embeddings. ArXiv. Preprint posted online on April 13, 2020. [doi: 10.48550/arXiv.2003.07278]
29. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res. 2020;22(6):e15154. [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
30. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. J Med Internet Res. 2022;24(10):e40238. [FREE Full text] [doi: 10.2196/40238] [Medline: 36197712]
31. Sezgin E. Artificial intelligence in healthcare: complementing, not replacing, doctors and healthcare providers. Digit Health. 2023;9:20552076231186520. [FREE Full text] [doi: 10.1177/20552076231186520] [Medline: 37426593]
32. Hemmer P, Schemmer M, Riefle L, Rosellen N, Vössing M, Kühl N. Factors that influence the adoption of human-AI collaboration in clinical decision-making. ArXiv. Preprint posted online on April 19, 2022. [doi: 10.48550/arXiv.2204.09082]
33. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310. [FREE Full text] [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
34. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. Minneapolis, Minnesota. Association for Computational Linguistics; Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2-7, 2019, 2019;4171-4186; Minneapolis, Minnesota.
35. Pennington J, Socher R, Manning C. GloVe: global vectors for word representation. Doha, Qatar. Association for Computational Linguistics; Presented at: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); October 25-29, 2014, 2014;1532-1543; Doha, Qatar. [doi: 10.3115/v1/d14-1162]
36. Models. Hugging Face. 2022. URL: https://huggingface.co/models [accessed 2023-03-09]
37. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. ArXiv. Preprint posted online on July 26 2019. [doi: 10.5260/chara.21.2.8]
38. Kumar P, Bhatnagar R, Gaur K, Bhatnagar A. Classification of imbalanced data:review of methods and applications. IOP Conf Ser Mater Sci Eng. 2021;1099:012077. [FREE Full text] [doi: 10.1088/1757-899x/1099/1/012077]
39. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Int Res. 2002;16(1):321-357. [FREE Full text] [doi: 10.1613/jair.953]
40. Wang L, Zhang Y, Chignell M, Shan B, Sheehan KA, Razak F, et al. Boosting delirium identification accuracy with sentiment-based natural language processing: mixed methods study. JMIR Med Inform. 2022;10(12):e38161. [FREE Full text] [doi: 10.2196/38161] [Medline: 36538363]
41. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016, 2016;1135-1144; San Francisco, California. URL: https://dl.acm.org/doi/abs/10.1145/2939672.2939778 [doi: 10.1145/2939672.2939778]
42. Palojoki S, Saranto K, Reponen E, Skants N, Vakkuri A, Vuokko R. Classification of electronic health record-related patient safety incidents: development and validation study. JMIR Med Inform. 2021;9(8):e30470. [FREE Full text] [doi: 10.2196/30470] [Medline: 34245558]
43. The collaborative interface design tool. Figma. URL: https://www.figma.com/ [accessed 2023-09-24]
44. Delio J, Catalanotti JS, Marko K, Paul C, Taffel M, Ho G, et al. Integrating adverse event reporting into a free-text mobile application used in daily workflow increases adverse event reporting by physicians. Am J Med Qual. 2020;35(5):374-379. [FREE Full text] [doi: 10.1177/1062860619891995] [Medline: 31795736]
45. Klemp K, Zwart D, Hansen J, Hellebek T, Luettel D, Verstappen W, et al. A safety incident reporting system for primary care. A systematic literature review and consensus procedure by the LINNEAUS collaboration on patient safety in primary care. Eur J Gen Pract. 2015;21(sup1):39-44. [FREE Full text] [doi: 10.3109/13814788.2015.1043728] [Medline: 26339835]

46. Hasegawa T, Fujita S. Patient safety policies experiences, effects and priorities; lessons from OECD member states (version 2.0). Patient Safety Global Ministerial Summit 2018. 2018. URL: https://www.mhlw.go.jp/psgms2018/pdf/document/5_Document.pdf [accessed 2023-12-21]
47. Wang Y, Coiera E, Runciman W, Magrabi F. Automating the identification of patient safety incident reports using multi-label classification. Stud Health Technol Inform. 2017;245:609-613. [Medline: 29295168]
48. Buettner R. Cognitive workload of humans using artificial intelligence systems: towards objective measurement applying eye-tracking technology. In: KI 2013: Advances in Artificial Intelligence Berlin. Heidelberg. Springer Berlin Heidelberg; Presented at: 36th Annual German Conference on AI, Koblenz; September 16-20, 2013, 2013;37-48; Germany. [doi: 10.1007/978-3-642-40942-4_4]
49. Berretta S, Tausch A, Ontrup G, Gilles B, Peifer C, Kluge A. Defining human-AI teaming the human-centered way: a scoping review and network analysis. Front Artif Intell. 2023;6:1250725. [FREE Full text] [doi: 10.3389/frai.2023.1250725] [Medline: 37841234]
50. Lai Y, Kankanhalli A, Ong DC. Human-AI collaboration in healthcare: a review and research agenda. Presented at: 54th Hawaii International Conference on System Sciences (HICSS); January 5-8, 2021, 2021; Grand Wailea, Maui, Hawaii. URL: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/495a50c0-16e9-4c3e-a00b-34aeec9dff6b/content [doi: 10.24251/hicss.2021.046]
51. Reverberi C, Rigon T, Solari A, Hassan C, Cherubini P, GI Genius CADx Study Group; et al. Experimental evidence of effective human-AI collaboration in medical decision-making. Sci Rep. 2022;12(1):14952. [FREE Full text] [doi: 10.1038/s41598-022-18751-2] [Medline: 36056152]
52. Lee MH, Siewiorek DP, Smailagic A, Bernardino A, Bermúdez i Badia S. A human-AI collaborative approach for clinical decision making on rehabilitation assessment. New York, NY, USA. Association for Computing Machinery; Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8-13, 2021, 2021;1-14; Yokohama Japan. URL: https://dl.acm.org/doi/10.1145/3411764.3445472 [doi: 10.1145/3411764.3445472]

Abbreviations

AI: artificial intelligence
KNN: k-nearest neighbor
LIME: local interpretable model-agnostic explanations
ML: machine learning
MLR: multinomial logistic regression
OECD: Organization for Economic Cooperation and Development
PSE: patient safety event
RF: random forest
SVM: support vector machine
TF-IDF: term frequency–inverse document frequency

Edited by E Borycki, A Kushniruk; submitted 06.10.23; peer-reviewed by H Tibble, S Matathil; comments to author 19.11.23; revised version received 30.11.23; accepted 03.12.23; published 25.01.24

Please cite as:
Chen H, Cohen E, Wilson D, Alfred M
A Machine Learning Approach with Human-AI Collaboration for Automated Classification of Patient Safety Event Reports: Algorithm Development and Validation Study
JMIR Hum Factors 2024;11:e53378
URL: https://humanfactors.jmir.org/2024/1/e53378
doi: 10.2196/53378
PMID: