dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240002

Analysis of Critical Incident Reports Using Natural Language Processing

Kerstin DENECKE[a,1] and Helmut PAULA[b]
[a] *Bern University of Applied Sciences, Bern, Switzerland*
[b] *Stiftung Patientensicherheit Schweiz, Zürich, Switzerland*

Abstract. A Critical Incident Reporting System (CIRS) collects anecdotal reports from employees, which serve as a vital source of information about incidents that could potentially harm patients. Objectives: To demonstrate how natural language processing (NLP) methods can help in retrieving valuable information from such incident data. Methods: We analyzed frequently occurring terms and sentiments as well as topics in data from the Swiss National CIRRNET database from 2006 to 2023 using NLP and BERTopic modelling. Results: We grouped the topics into 10 major themes out of which 6 are related to medication. Overall, they reflect the global trends in adverse events in healthcare (surgical errors, venous thromboembolism, falls). Additionally, we identified errors related to blood testing, COVID-19, handling patients with diabetes and pediatrics. 40-50% of the messages are written in a neutral tone, 30-40% in a negative tone. Conclusion: The analysis of CIRS messages using text analysis tools helped in getting insights into common sources of critical incidents in Swiss healthcare institutions. In future work, we want to study more closely the relations, for example between sentiment and topics.

Keywords. Text mining, Critical incident reporting system, Text analysis, Natural language processing

1. Introduction

The safety and well-being of patients is of paramount importance to all stakeholders in healthcare. As complex organizations, hospitals face several challenges in maintaining and improving patient safety. A key tool in this ongoing effort are Critical Incident Reporting Systems (CIRS), which serve as a vital source of information about incidents that could potentially harm patients [1], [2]. Analyzing CIRS messages can provide actionable insights into systemic weaknesses and risk factors, helping hospital administrators to implement effective preventive measures [3]. By identifying trends and predicting potential future incidents, hospitals can proactively address areas of concern before they result in adverse events.

Although the clear value of CIRS messages the wealth of this qualitative data is often under-utilized due to its unstructured nature [4]. Traditional methods of manual analysis are time consuming, potentially biased and often unable to cope with the scale and complexity of the data. The analysis of these reports using Natural Language Processing (NLP) methods offers a promising avenue for supporting this process [3], [5] since they automatically process CIRS messages, identify patterns and extract

[1] Corresponding Author: Kerstin Denecke, Bern University of Applied Sciences, Institute Patient-centered Digital Health, Quellgasse 21, Bern, Switzerland, E-Mail: kerstin.denecke@bfh.

meaningful insights [5]. Such automated analysis can lead to a more comprehensive understanding of the nature and frequency of incidents, as well as their underlying causes and in this way improving patient safety and healthcare quality. So far, research on NLP applied to CIRS messages is still rare, in particular in German-speaking countries. Tetzlaff et al. analyzed the German database "cirsmedical.de" [6] using NLP. Wang et al. applied a Support Vector Machine-based approach that considers also semantics for classifying incident types and compared it to the quality of a classifier based on a bag of words feature set [7]. In another work, they tested a convolutional neural network with word embedding to identify incidents by type and severity [8].

This research aims to use NLP to analyze CIRS reports of the Swiss national incident reporting system, providing an approach to understanding and mitigating risk in the hospital environment. Furthermore, it seeks to explore and demonstrate the value of NLP in the analysis of CIRS messages. In an era where data-driven decision-making is paramount, this research therefore aligns with the broader movement towards harnessing big data for improved outcomes in healthcare.

2. Methods and Dataset

Our dataset comprised 10'063 messages from CIRRNET starting from 21/05/2006 to 23/09/2023. CIRRNET, the Critical Incident Reporting & Reacting NETwork (https://patientensicherheit.ch/cirrnet/), is a supra-regional institution that performs a central networking function for local error reporting systems in Switzerland. All affiliated facilities can feed their local CIRS reports anonymously into the CIRRNET database and thus contribute to the comprehensive survey of patient risks. Currently, 73 healthcare institutions with their 128 locations are participating in CIRRNET on a voluntary basis. The organization Patient Safety Switzerland has been operating CIRRNET since 2006. CIRRNET differs from most CIRS networks in that all local CIRS reports are used to identify problem areas of supra-regional relevance, recommendations for improvement are developed together with experts and published in the form of Quick Alerts® by Patient Safety Switzerland. In average, 559 messages are included per year with a minimum of 84 messages in 2006 and a maximum of 1'196 messages in 2012. Between 2011 and 2014, more than 1'000 messages are available per year. Only intensive care units and anesthesia contributed their reports after the initial launch of CIRRNET. In 2011, CIRRNET was opened for all medical specialties, explaining this peak.

We analyzed the CIRS messages automatically using various NLP methods. To simplify reporting, we grouped the data into five batches: years 2006-2009, years 2010-2014, years 2015-2019, years 2020-2021, years 2022-2023. This grouping was motivated by the distribution of messages per year. Beyond, we wanted to consider the two years of COVID-10 pandemic separately, hypothesizing that at the time of the pandemic reporting of incidents differs from other years. We were also interested in the incidents reporting during the pandemic.

To analyze the content of the messages, we determined the frequent words and generated word clouds using the Python library WordCloud. Further, we identified the main topics of the messages using BERTopic analysis [9]. BERTopic uses a transformer-based method to process documents. It involves embedding documents, reducing their dimensional complexity, clustering them, creating bag-of-words models and using c-TF-IDF to create topic representations. Each incident report was also classified as positive, negative or neutral using the Python library TextBlobDE to study the sentiment of the incident messages. By all these efforts, we wanted to identify common patterns and trends in the data, i.e. to uncover frequent types of incidents, common causes, and areas within the hospital where incidents occur most often.

Table 1. Top 20 Commonly used terms in the messages per period (presented in descending order of frequency)

2006-2009 (n=1'146)	2010-2014 (n=4'703)	2015-2019 (n=3'061)	2020-2021 (n=526)	2022-2023 (n=627)
OP	verordnet	verordnet	verordnet	Verordnet
Seite	Verordnung	Verabreicht	Station	Pflege
Bein	bemerkt	Uhr	Zimmer	OP
Operateur	Uhr	Pflege	Notfall	Verabreicht
Anästhesist	mg	Bemerkt	Kind	Erhalten
Immer	OP	Verordnung	Covid	Notfall
Ab	Tag	Medikament	Uhr	Zimmer
Intubation	Infusion	Erhalten	Pflege	Covid
Gabe	verabreicht	Arzt	Verabreicht	Verordnung
Wegen	Arzt	OP	Tag	Bemerkt
Infusion	erhalten	Informiert	Informiert	Informiert
Saal	Pflege	Labor	Verordnung	Abteilung
Sei	Station	Gegeben	Arzt	Uhr
Kammer	tbl	Abteilung	Abteilung	Station
Postoperativ	Medikamente	Bekommen	Infusion	Sei
Arzt	Erhielt	Medikamente	Bemerkt	Medikamente
Darauf	Worden	Statt	Isoliert	IPS
Einleitung	Fehler	Tag	Gemacht	Verordnungen
Falsche	Statt	Station	Wegen	Worden
informiert	gegeben	falschen	Zeit	falsch

3. Results

The messages had an average number of words per message of 52.5 words (minimum of 1 word, maximum of 645 words). 97 texts were in French, 3 in Italian and 6 in English. All other texts were in German. The top 20 commonly used terms (Table 1) reflected that a frequent term is related to the prescription of medications. The word clouds (Figure 1) together with the top 20 words demonstrated a topic shift over the five considered periods: In the first period (2006-2009), many reported incidents seemed to concern the surgery. In the second (2010-2014) and third period (2015-2019), the frequent terms addressed the medication and their prescription. In 2020-2021, incidents concerned COVID-19 and isolation. In the period 2022-2023, the main topic was again related to medication and prescription, but also incident reports related to nursing, surgery and emergency become more frequent.

The BERTopic analysis provided more in-depth insights into the topics addressed in the messages (Table 2). Five topics (Topic 3,4,6,7,8,10) were related to medication in general or even to specific medication handling problems. For example, topic 4 deals with insulin injection in the context of diabetes handling. Topic 6 concerned postoperative thromboembolism prophylaxis. Topic 1 concerned laboratory procedures related to blood analysis. In particular, exchange of names and patients when labelling the blood tests seemed to be an issue. Topic 2 dealt with children and emergencies. Topic 5 pointed to various problems of patient safety related to fall and mobility of patients. Keywords let us assume that the messages concerned patients falling out of their beds or on their way to the toilet. Topic 7 concerned incidents related to anesthesia. Topic 9 was related to COVID-19, testing and isolation.

Figure 1. Word clouds generated from the incident reports from the five indicated time periods

Table 2. Topics as determined by BERTopic for the complete database

Manual labeling of topic	Topic-relevant terms as identified by BERTopic
Topic 1: Blood testing and laboratory	[labor, blutentnahme, falschen, falsche, namen, röhrchen, blut, abgenommen, be, beschriftet]
Topic 2: Pediatric emergency	[kind, station, notfall, op, kommt, uhr, ca, arzt, kam, uns]
Topic 3: Medication prescription	[medikamente, medikament, tbl, medis, mg, haldol, medi, verordnung, eingenommen, morgen]
Topic 4: Handling of diabetes patients	[insulin, bz, novorapid, gespritzt, blutzucker, glucose, diabetes, einheiten, pen, nachspritzschema]
Topic 5: Fall and mobility	[boden, bett, bettgitter, wollte, rollstuhl, aufstehen, dabei, trage, wc, gestürzt]
Topic 6: Thromboembolism prophylaxis	[clexane, 40mg, sc, verordnung, injiziert, 20mg, clex, postop, gespritzt, 40]
Topic 7: Anesthesia procedures	[intubation, adrenalin, einleitung, mg, gabe, fentanyl, tubus, propofol, minuten, problemlose]
Topic 8: Heparin treatment	[heparin, heparinperfusor, perfusor, 10000, gestartet, ie, eingestellt, lief, heparininfusion, fehler]
Topic 9: Covid 19 treatment	[covid, isoliert, abstrich, getestet, positiv, verdacht, mrsa, verlegt, test, covid19]
Topic 10: Treatment with Calium chlorid	[kcl, infusion, 24h, 40mmol, eingestellt, 500ml, 20mmol, 21mlh, ringerfundin, mlh]

Regarding sentiment (Figure 2), we can recognize that most messages were formulated in a neutral manner. They became slightly more positive since 2019. In 2008, the percentage of positive messages was the smallest with 16.4% positively and 51.1% of neutrally formulated reports.

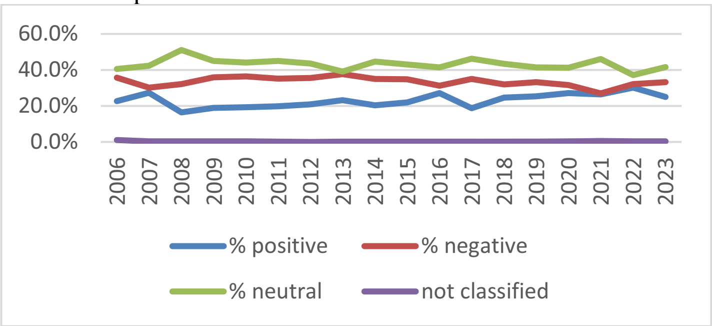

Figure 2. Sentiment of the CIRS messages

4. Discussion

Our results reflect that the most common incidents are related to medication. This corresponds to the global trends that medication-related harm affects 1 in 30 patients in healthcare settings [10], [11]. The increase in terms related to medication prescription since 2010 can be associated with the introduction of electronic prescription systems in Swiss hospitals. On a global level, surgical errors and health care-associated infections also belong to the most frequent sources of patient harm [11], [12]. While surgery occurs as frequent term also in our data, we could not find information related to infections. A reason is that CIRS reports refer to events where the relation between cause and event can be recognized. Hospital-acquired infections occur not immediately and can often not be related to a specific event. Patient falls and incidents related to venous thromboembolism in turn are known to be frequent adverse events in hospitals [13], [14] which corresponds to our results. The COVID-19 pandemic was clearly reflected in the incident reports. Interestingly, we identified the topic related to errors in handling patients with diabetes. Insulin injection schemata are complex and not integrated in electronic health records which may lead to errors. Further, multiple devices are used for diabetes management, in particular patient-owned medical devices such as insulin pens providing a source of errors.

The study results demonstrate the usefulness of applying NLP methods to CIRS reports, and show that it is beneficial to use multiple methods to identify all trends and topics and get a more complete picture. Tetzlaff et al. conducted a similar analysis using NLP applied to the German database “cirsmedical.de” [6]. They focused on their analysis rather on the reporting differences among clinical disciplines. In contrast to the work by Wang et al. we were not interested in classifying the messages into predefined categories, but to explore the dataset by the chosen methods [7], [8].

Our study has some limitations: The CIRRNET database is not a primary reporting system, but contains only reports that have been triaged and forwarded in the affiliated organizations. The dataset contained some messages in languages others than German. However, we only applied sentiment analysis methods specialized for German texts. The sentiment analysis method was not assessed regarding accuracy on our data set. The chosen method was not specifically developed for analyzing sentiments in CIRS messages and might have led to wrong classifications.

There are many additional questions we want to study in future: How does the sentiment in reports correlate with topics of incidents? Which linguistic features (e.g., specific terminologies, phrasing) are commonly associated with different types of incidents? Answers to the latter question would help in identifying linguistic markers that could automate the categorization of reports or signal the criticality of incidents. Another question of interest is: Can NLP help in identifying under-reported or misclassified incidents? We also want to closer analyze the emotional tone of the reports, which could reflect the urgency or severity of incidents. In addition, the emotional and psychological impact of critical incidents on healthcare staff is often overlooked. Sentiment analysis of CIRS messages can offer insights into relations between emotional responses associated with different incidents, helping to address staff well-being and foster a supportive work environment. In particular, this form of analysis could provide important conclusions about the willingness of employees to participate in a more constructive or negative manner and thus about the safety culture in the various organizations. It could be interesting to study whether at certain times in a year the topics of incidents are changing. Such information could be useful to predict potential future incidents based on historical data, aiding in proactive risk management.

References

[1] J. Mack, 'CIRS als Werkzeug des Qualitäts-und Risikomanagement im deutschen Versorgungssystem.', *Qualitätsmanagement in der Gesundheitsversorgung*, pp. 171–88, 2020.

[2] W. Petschnig and E. Haslinger-Baumann, 'Critical Incident Reporting System (CIRS): a fundamental component of risk management in health care systems to enhance patient safety', *Saf Health*, vol. 3, no. 1, p. 9, Dec. 2017, doi: 10.1186/s40886-017-0060-y.

[3] I. J. B. Young, S. Luz, and N. Lone, 'A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis', *International Journal of Medical Informatics*, vol. 132, p. 103971, Dec. 2019, doi: 10.1016/j.ijmedinf.2019.103971.

[4] C. Neuhaus, M. Holzschuh, C. Lichtenstern, and M. St.Pierre, 'Erkenntnisse aus 10 Jahren CIRS-AINS: Eine Analyse von Nutzerverhalten und Ausblick auf neue Herausforderungen', *Anaesthesist*, vol. 69, no. 11, pp. 793–802, Nov. 2020, doi: 10.1007/s00101-020-00829-z.

[5] K. Denecke, 'Automatic Analysis of Critical Incident Reports: Requirements and Use Cases', *Stud Health Technol Inform*, vol. 223, pp. 85–92, 2016.

[6] L. Tetzlaff, A. S. Heinrich, R. Schadewitz, C. Thomeczek, and T. Schrader, '[The analysis of CIRSmedical.de using Natural Language Processing]', *Z Evid Fortbild Qual Gesundhwes*, vol. 169, pp. 1–11, Apr. 2022, doi: 10.1016/j.zefq.2021.12.002.

[7] Y. Wang, E. Coiera, and F. Magrabi, 'Can Unified Medical Language System–based semantic representation improve automated identification of patient safety incident reports by type and severity?', *Journal of the American Medical Informatics Association*, vol. 27, no. 10, pp. 1502–1509, Oct. 2020, doi: 10.1093/jamia/ocaa082.

[8] Y. Wang, E. Coiera, and F. Magrabi, 'Using convolutional neural networks to identify patient safety incident reports by type and severity', *Journal of the American Medical Informatics Association*, vol. 26, no. 12, pp. 1600–1608, Dec. 2019, doi: 10.1093/jamia/ocz146.

[9] M. Grootendorst, 'BERTopic: Neural topic modeling with a class-based TF-IDF procedure', *arXiv preprint arXiv:2203.05794*, 2022.

[10] A. Hodkinson *et al.*, 'Preventable medication harm across health care settings: a systematic review and meta-analysis', *BMC Med*, vol. 18, no. 1, p. 313, Dec. 2020, doi: 10.1186/s12916-020-01774-9.

[11] M. Panagioti *et al.*, 'Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis', *BMJ*, p. l4185, Jul. 2019, doi: 10.1136/bmj.l4185.

[12] S. Raoofi *et al.*, 'Global prevalence of nosocomial infection: A systematic review and meta-analysis', *PLoS One*, vol. 18, no. 1, p. e0274248, 2023, doi: 10.1371/journal.pone.0274248.

[13] J. H. LeLaurin and R. I. Shorr, 'Preventing Falls in Hospitalized Patients: State of the Science', *Clin Geriatr Med*, vol. 35, no. 2, pp. 273–283, May 2019, doi: 10.1016/j.cger.2019.01.007.

[14] G. E. Raskob *et al.*, 'Thrombosis: A Major Contributor to Global Disease Burden', *ATVB*, vol. 34, no. 11, pp. 2363–2371, Nov. 2014, doi: 10.1161/ATVBAHA.114.304488.