Preliminary Processing and Analysis of an Adverse Event Dataset for Detecting Sepsis-Related Events

Melissa Y. Yan*
Norwegian University of Science and Technology
Trondheim, Norway
melissa.yan@ntnu.no

Lise Husby Høvik*
St. Olavs hospital, Trondheim University Hospital
Trondheim, Norway
lise.hovik@ntnu.no

André Pedersen
Norwegian University of Science and Technology
SINTEF
Trondheim, Norway
andre.pedersen@sintef.no

Lise Tuset Gustad
Norwegian University of Science and Technology
Trondheim, Norway
lise.t.gustad@ntnu.no

Øystein Nytrø
Norwegian University of Science and Technology
Trondheim, Norway
nytroe@ntnu.no

***Abstract*—Adverse event (AE) reports contain notes detailing procedural and guideline deviations, and unwanted incidents that can bring harm to patients. Available datasets mainly focus on vigilance or post-market surveillance of adverse drug reactions or medical device failures. The lack of clinical-related AE datasets makes it challenging to study healthcare-related AEs. AEs affect 10% of hospitalized patients, and almost half are preventable. Having an AE dataset can assist in identifying possible patient safety interventions and performing quality surveillance to lower AE rates. The free-text notes can provide insight into the cause of incidents and lead to better patient care. The objective of this study is to introduce a Norwegian AE dataset and present preliminary processing and analysis for sepsis-related events, specifically peripheral intravenous catheter-related bloodstream infections. Therefore, the methods focus on performing a domain analysis to prepare and better understand the data through screening, generating synthetic free-text notes, and annotating notes.**

***Index Terms*—Adverse events, Healthcare knowledge representation, Natural language processing, Quality improvement, Sepsis**

I. Introduction

Sepsis is the most common cause of death among hospitalized patients [1] and contributes to 30% to 50% of hospitalized deaths [2]. Caused by a dysregulated host response to an infection, sepsis can lead to multi-organ failure and death [3]. Bloodstream infections (BSIs) occur when bacteria enter the bloodstream [4]. A particularly lethal bacterium that commonly causes BSIs is *Staphylococcus aureus* (*S. aureus*) [5], a gram-positive bacteria frequently found on the skin. A range from 7.6% to 35% of *S. aureus* BSIs are due to peripheral intravenous catheters (PIVCs) [6]. PIVCs are inserted in a peripheral vein to administer IV fluids, medications, and blood transfusions. They are the most frequently used device in hospitals [7]; over one billion PIVCs are estimated to be inserted annually worldwide [8] and up to 80% of patients admitted to a hospital will receive at least one PIVC [9].

Improper management of PIVCs can lead to increased patient mortality risk via BSIs [6]. Four gateways to BSIs are described with PIVCs; migration of microbes down the catheter tract, via the catheter hub, by contaminated infusate, or by an existing infection where bacteria circulating the bloodstream can attach to the catheter [10]. Although frequently used, PIVCs are often not documented in clinical records [7]. Additionally, sepsis is poorly documented in departments outside the intensive care unit [11]. This makes retrospective and real-time systematic quality surveillance for PIVC difficult. However, failure related to PIVCs are more frequently reported in AE reports. Hence, the main motivation of this project was to use an AE dataset to facilitate systematic monitoring and quality of care improvements related to PIVCs for reducing sepsis and BSI cases.

II. Background

A. About Adverse Events

An estimated 1 in 10 hospitalized patients worldwide are affected by an adverse event (AE), and nearly 50% are preventable [12]. Commonly reported AEs include surgical or medication procedural errors, hospital-acquired infections, pressure ulcers, and falls [13]. AE incidents can be recorded in electronic health records (EHRs) or separate reporting systems. Current methods for detecting AEs include manual chart review and screening using ICD codes, keyword search, and natural language processing (NLP) [14]. Using the Global Trigger tool [15], various studies track health care quality indicators to identify triggers and measure AE rates by manually reviewing medical records [13]. In addition to the manual approach, other studies focus on developing automated trigger tools, such as extracting EHR data using NLP [16] and monitoring nursing notes for infection signs [17].

* Authors contributed equally to this work.

978-1-6654-0126-5/21/$31.00 ©2021 IEEE

TABLE I. ADVERSE EVENT DATABASES (DB) WORLDWIDE

Type	Purpose	Database
Drug	Vigilance	Canada Vigilance Adverse Reaction Online DB; EU Drug Regulating Authorities Pharmacovigilance (EudraVigilance); German ABDA[a] DB; Japanese Adverse Drug Event Report (JADER) DB; Korean Adverse Event Reporting System (KAERS); UK MHRA[b] Interactive Drug Analysis Profiles (iDAPs); World Health Organization (WHO) VigiBase via VigiAccess
Drug	Post-market surveillance	US FDA[c] Adverse Event Reporting System (FAERS) DB; US Vaccine Adverse Event Reporting System (VAERS)
Device	Post-market surveillance	German Medical Devices Information and DB System (DMIDS); US Manufacturer and User Facility Device Experience (MAUDE)
Drug & Device	Post-market surveillance	Australian DB of Adverse Event Notifications (DAEN)
All	Near-miss or AE	Japan Council for Quality Health Care (JQ) project

[a] ABDA: Federal Union of German Associations of Pharmacists (Bundesvereinigung Deutscher Apothekerverbände)
[b] MHRA: Medicines and Healthcare products Regulatory Agency
[c] FDA: Food and Drug Administration

B. Purpose of Adverse Events

The primary purpose of documenting AEs is to promote patient safety. Along with collaborators worldwide, the World Health Organization has been promoting methods which contribute to effectively learning from AEs [18]. As AE documentation shifts from legal consequences with personal responsibility towards a learning perspective, it is becoming possible to better understand the causes resulting in AEs, and thus identify possible interventions to improve patient safety within hospitals [12].

C. Available Datasets and Databases

Available AE datasets and databases worldwide mainly focus on vigilance or post-market surveillance of drugs or devices (see Table I). As the focus is on either adverse drug events and side effects or device failures, it is unlikely that many of the databases will capture clinical healthcare-related AEs. To the best of our knowledge, the only publicly available AE database containing clinical healthcare-related AEs is the Japan Council for Quality Health Care's Project to Collect Medical Near-Miss/Adverse Event Information [19].

However, the Japanese healthcare system varies from Norwegian healthcare in terms of treatment, health system organization, and strategies to ensure the quality of care [20]. For instance, Norway has general practitioners who act as gatekeepers to specialist treatment; this is relatively new in Japan. In addition, most Norwegian hospitals are government-owned, whereas only 15% of Japanese hospitals are government-owned. Furthermore, Norwegian hospitals are obligated to participate and measure quality and safety improvement, and there is a national program for tracking health care indicators of survival and infection rates. In contrast, the Japanese government promotes hospitals to report quality indicators on their websites, and only advanced treatment Japanese hospitals are required to report AEs.

D. Objective

This paper aims to present a Norwegian AE dataset and preliminary results for characterizing a dataset on detecting sepsis-related events, to demonstrate further research potential using a dataset currently undergoing preparation for release. Various clinical events were found by inspecting the AE dataset with the initial motivation of identifying PIVC-related BSIs. This included events related to sepsis and phlebitis, which is inflammation of a vein near the skin's surface and can be an indicator of infection. Furthermore, falls and device failures were deemed relevant to interpret the AE dataset.

III. NORWEGIAN ADVERSE EVENTS

In Norway, a retrospective review of EHRs estimated that one-third of all hospital deaths were due to AEs [21]. Further exploration into two Norwegian hospitals identified that 11.2% of AEs were life-shortening. From these, 82.4% of the incidences were related to healthcare-associated infection. In general, comparing statistics based on AE-events from registries, EHR reviews, and automated methods is challenging.

A. Health Care Policy

Many countries have their own AE-related legislation to monitor the safety of drugs and medical devices. In Norway, the Regulations on Medicinal Products maintain drug safety and marketing[1], whereas the Medical Equipment Act regulates medical equipment safety and post-market surveillance[2]. Additionally, under the Norwegian Specialized Health Services Act of 1999, all health and care services are obligated to notify the Norwegian Board of Health Supervision of unexpected incidents related to patient injury and death[3]. Furthermore, according to the Regulations on Management and Quality Improvement in the Health and Care Service, those services are required to manage quality improvement and patient safety systematically by reviewing deviations (i.e., AE), evaluating implemented preventative measures, and rectifying activities[4].

B. Norwegian Adverse Event Dataset Description

There are 18 555 AE reports from the electronic incident reporting registry system at St. Olavs hospital, Trondheim University Hospital in Trondheim, Norway between September 30, 2015 and December 31, 2019. Intentionally written for a specific purpose and directed at someone specific, these reports are not written routinely by a clinician and differ in quality, purpose, and structure from EHR clinical notes. These reports describe various events in addition to AEs, such as procedural and guideline deviations, near-miss events that could have

[1] https://lovdata.no/dokument/LTI/forskrift/2009-12-18-1839
[2] https://lovdata.no/dokument/NL/lov/2020-05-07-37
[3] https://lovdata.no/dokument/NL/lov/1999-07-02-61
[4] https://lovdata.no/dokument/SF/forskrift/2016-10-28-1250

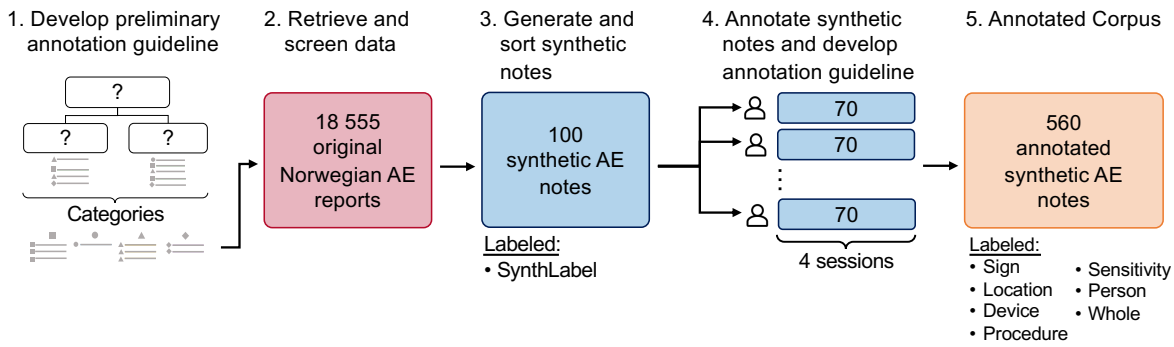

Fig. 1. Steps for preprocessing adverse event (AE) data. **1)** The preliminary guideline was developed from clinical questions of interest, which are used to create categories. **2)** The categories are used to screen the original 18 555 Norwegian AE reports. **3)** A total of 100 unique synthetic AE notes were generated based on notes from the original reports, and each note was given a **SynthLabel** label indicating if it contained infection, faulty device malfunctioning, or fall information. **4)** Eight annotators each annotated 70 notes over four sessions using a guideline that was revised after each session. **5)** This resulted in 560 annotated synthetic AE notes labeled with seven categories (i.e., **Sign**, **Location**, **Device**, **Procedure**, **Sensitivity**, **Person**, and **Whole**) used to capture and represent documented healthcare knowledge.

harmed patients, misunderstandings, resource needs, and patients with poor behavior who pose a risk to others. Each report has: an identifier, title, registration date, changed date, report to and from units, booleans for security-related or patient-related event, event type and severity, clinical division, and an unstructured free-text note. Lastly, it also contains a status indicating if the incident is open or closed; a closed status indicates specific solutions for patient safety problems have been developed and implemented. Use of AE notes for the purpose of this study was approved by the Norwegian Regional Committees for Medical and Health Research Ethics (REK), approval no 26814.

IV. Materials and Methods

The original 18 555 Norwegian AE reports dataset was used to create a synthetic dataset annotated for PIVC-related BSI events (see Fig. 1). The synthetic notes were annotated to capture data, information, and knowledge in the text at different levels; word- or phrase-level indicates an annotation that spans a word or phrase, whereas note-level indicates an annotation representing the span of the whole text. This resulted in 100 synthetic notes with **SynthLabel** note-level labels (i.e., infection-related, faulty device malfunction-related, and fall-related incidents) and 560 annotated synthetic notes.

A. Preliminary Annotation Guideline Development

The preliminary annotation guideline was developed based on the proposed clinical question: "Is there a connection between BSIs and PIVCs at the hospital?" The clinical question was simplified to:

- How can sepsis or BSIs be identified when the symptoms are similar to other diseases?
- How can poorly documented PIVCs be identified?

Those questions were then modified based on the clinical perspectives of the nurses; for example, some catheters are documented distinctly (for data extraction), whereas others can be distinguished based on anatomical insertion site (for information extraction) or procedures (for knowledge extraction). This resulted in the following domain-specific *questions of interest*:

- What are the different signs of infections, specifically for BSIs, sepsis, or infected PIVCs?
- What are the signs for different types of catheters?
- Where are the anatomical insertion sites of catheters?
- What events can be related to catheter use?

Domain-specific questions were answered by nurses who provided a list of keywords, phrases, sentences, and examples from the clinic. As shown in Fig. 2(a), these answers were sorted into four categories (technically known as entities in annotation or classes in ontologies) for word- or phrase-level labels: **Sign**, **Location**, **Device**, and **Procedure**. Next, a total of 700 randomly selected notes, from the original AE dataset, were manually screened to ensure that the four categories related to catheters and BSIs could be found and occurred frequently enough for downstream analysis (see Fig. 2(c)). After screening, three additional categories (i.e., **Sensitivity**, **Person**, and **Whole**) were included to ensure that sensitive data was correctly anonymized, actions related to an individual could be determined, and a note-level label was available. This resulted in seven categories:

1) **Sign**: infection signs
2) **Location**: anatomical insertion sites
3) **Device**: signs of catheter types
4) **Procedure**: procedures, interventions, or activities related to catheters
5) **Sensitivity**: protected health information
6) **Person**: individuals (i.e., patient, clinician, or relative)
7) **Whole**: note-level label indicating whether the note contains infection, BSI, sepsis, faulty device malfunctioning, catheter, PIVC, or sensitive information.

Each category can form a hierarchy with more specific subcategories (e.g., the **Device** category contains a general subcategory "catheter" that has a more specific "PIVC" subcategory). In addition, six relationships (see Fig. 2(b)) were added to link categories 1-4 together to ensure that information was not lost for downstream analysis (e.g., infection sign at a specific

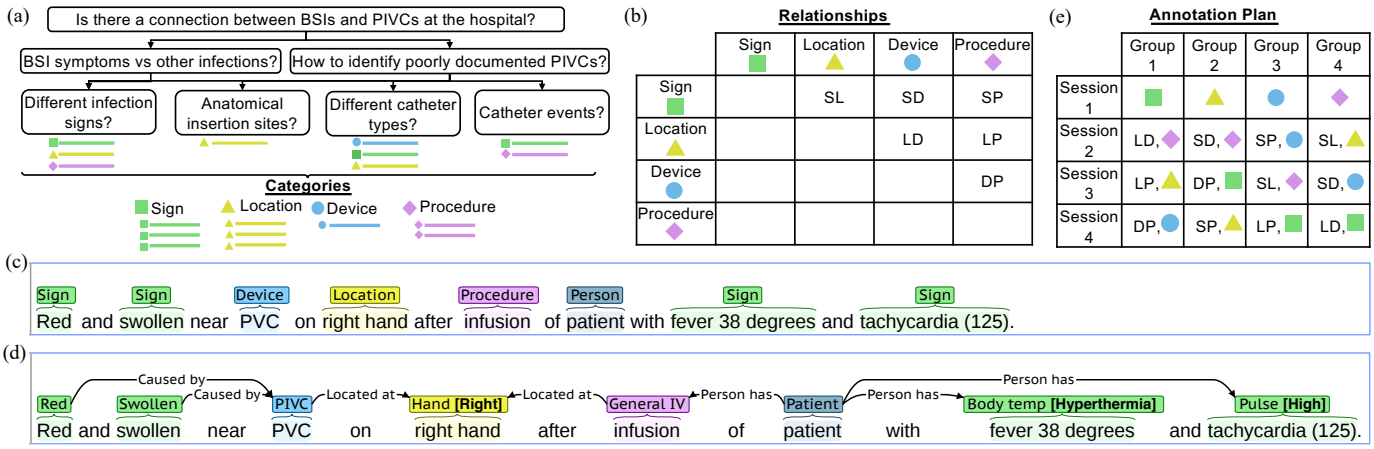

Fig. 2. Annotation guideline development and annotation. (a) The clinical question of interest was simplified into domain-specific *questions of interest* which were answered by clinicians and sorted into four different main categories (i.e., **Sign**, **Location**, **Device**, and **Procedure**). (b) To capture knowledge about peripheral intravenous catheters (PIVCs) and bloodstream infections for downstream analysis, relationships linking categories to each other were included in the guideline. There are six relationships: sign-location, sign-device, sign-procedure, location-device, location-procedure, and device-procedure (i.e., SL, SD, SP, LD, LP, and DP). (c) Randomly selected adverse event notes were manually screened to ensure that the four main categories were detectable. (d) During annotation, annotators used the Brat rapid annotation tool to label notes in more detail using subcategories and attributes. For instance, "right hand" which was previously labeled as **Location** in (c) in now labeled using **Location**'s subcategory "Hand" and given the attribute [Right]. Relationships are also used to link one label to another (e.g., "Red" is linked to "PIVC" using the sign-device relationship "Caused by"). (e) Categories and relationships were sorted to create an annotation plan with four sessions and four groups of two annotators each.

location). Fig. 2(d) provides an example of how relationships link categories together and how detailed information can be provided by using subcategories and attributes. Using categories, relationships, and screening results as examples and counterexamples, a preliminary annotation guideline was created. The preliminary annotation guideline describes how to annotate each category and relationship to remove annotator confusion and disagreements.

B. Synthetic Adverse Event Dataset Generation

The 100 synthetic notes were generated and validated by a nurse, and thereafter divided into 10 sets with 10 notes each for the four main categories and six possible relationships. The 10 sets were sorted into four groups such that the sets for the four main categories were annotated by each group once and the sets for the six relationships were annotated at least twice by a different group. This was done to assess guideline revision improvements among different annotators using the same set of notes. The combination of these sets resulted in an annotation plan with four annotation sessions and four groups each with two annotators (see Fig. 2(e)). Thus, each annotator would annotate 10 notes in the first session and 20 notes in the remaining three sessions for a total of 70 notes.

C. Annotation Guideline Development and Annotation

Synthetic notes were annotated in four annotation sessions. In each session, two annotators annotated notes using the annotation guideline and Brat rapid annotation tool (BRAT) [22]. Annotations were evaluated by group using the inter-annotator agreement (IAA) F_1-score and assessed for whether clinical question information was captured. Then, ambiguities and annotator comments were discussed with nurses and incorporated into annotation guidelines revisions, and the process was repeated (guidelines for each session are available online[5]).

D. Annotated Dataset

Annotation by eight annotators produced a dataset of 560 notes stored in the BRAT standoff format. Each note has a note-level label (i.e., **Whole**). All AE notes can have overlapping note-level topics. Additionally, each note can have word- or phrase-level labels for the remaining six categories (i.e., **Sign**, **Location**, **Device**, **Procedure**, **Sensitivity**, and **Person**). Each word- or phrase-level label can have additional attribute information and can be linked to other labels to form relationships.

E. Preprocessing

For purposes of this study, only note-level labels were used. The 560 notes were converted into a comma-separated value file. The file contains basic information for each note, such as annotation session number, annotator identifier, filename, and text. Annotations included all word- or phrase-level labels from the seven categories and any annotator provided attributes. **Whole** category labels were separated into individual columns to identify note-level labels indicating whether a note contains infection, BSI, sepsis, faulty device malfunction, catheter, PIVC, or sensitive information. In addition, **SynthLabel** note-level labels were also separated into individual columns for infection, faulty device malfunction, or fall. Two additional merged labels, "Merged Infections" and "Merged Device Fails", were formed by combining parts of note-level labels **Whole** and **SynthLabel**. "Merged Infections" was comprised of **SynthLabel** label infection-related

[5]https://folk.ntnu.no/melissay/ae-guidelines/

and **Whole** category labels infection-related, BSI-related, and sepsis. "Merged Device Fails" was comprised of **SynthLabel** label device malfunction-related and **Whole** category labels device malfunction.

As multiple annotators labeled the same note, a max-voting strategy was conducted to produce a ground truth. No ties occurred between annotators (e.g., four annotators assigned 1 and four other annotators assigned 0).

F. Experiment

To assess the usefulness of the annotated dataset, experiments were conducted on a selection of tasks using a machine learning pipeline. Firstly, two datasets were defined: the training set of 18 555 original Norwegian notes and the test set of 560 annotated synthetic notes.

Each note was preprocessed using the following pipeline: 1) The common, redundant phrase "Hele_Notater" and other stop words were removed. 2) Capitalization was converted to lowercase. 3) Redundant characters such as newlines and quotation marks were removed. 4) Rare words with less than three occurrences were discarded. 5) Only notes with more than n_{min} and less than n_{max} words were kept. 6) Notes having less than l_{min} characters were discarded.

The following topic analysis pipeline was used to perform classification: 1) A word count vectorization was applied, keeping the top N words. Only unigrams and bigrams were generated. 2) Latent Dirichlet Allocation (LDA) [23] was then applied using K number of topics, trained concurrently using 16 workers for M iterations. 3) The word vectorizer and the LDA model were then trained using the training set only. 4) As LDA is an unsupervised method, it does not produce classification labels directly. Hence, the topic with the highest overlap with the task's labels in the test set was assigned for each respective task. This enabled evaluation of the unsupervised pipeline without manually choosing which topic(s) corresponded to each respective task(s), which is infeasible for a large number of topics.

Manually tuning relevant hyperparameters such as the number of topics K and the number of iterations M for the LDA model is challenging. Thus, an automatic hyperparameter search utilizing Bayesian optimization was conducted for 1000 iterations. To initialize the Bayesian search, the first 20 iterations were a random search. The test set's macro-averaged F_1-score was used as the objective function.

Models were trained using an Intel Core Processor with 32 cores and 128 GB of RAM. Implementation was done in Python 3.6. The topic model and feature extractor were implemented using scikit-learn (v0.16.1) [24]. Bayesian hyperparameter optimization was conducted using scikit-optimize (v0.8.1) [25]. The source code used in this study is made openly available on GitHub[6].

V. Preliminary Results

The resultant F_1-scores varied considerably between tasks (see Table II). Overall, the modeling technique performed well on all tasks, but performed best on the Fall and the Catheters tasks. It performed poorer on the infection and the merged tasks.

TABLE II. Test set performance of a selection of tasks using the hyperparameters chosen by the Bayesian optimization.

Task	F_1-score	Hyperparameters				
		N[a]	K[b]	n_{min}[c]	n_{max}[d]	l_{min}[e]
Infection	0.791	9268	35	1	48	34
Fall	0.997	5405	13	1	29	48
Device failure	0.895	1000	100	10	45	15
PIVC	0.877	10000	87	8	39	50
Catheters	1.000	1000	37	1	31	50
Merged Infections	0.843	4072	40	5	23	22
Merged Device Fails	0.743	1050	26	9	43	27

[a] N most frequent occurring words. [b] K number of topics. [c] n_{min} lower bound for number of words in a note. [d] n_{max} upper bound for number of words in a note. [d] l_{min} lower bound for number of characters in a note.

Hyperparameters chosen by the Bayesian search also differed between tasks, but the results had some patterns. Optimal performance on individual tasks was achieved using different sets of hyperparameters. Hence, using a single model for all tasks would result in overall degraded performance on individual tasks. Having a large number of topics K and a large number of words N were beneficial for detecting rarer and likely more challenging tasks.

VI. Discussion

This study presents a new AE dataset, a corresponding annotated dataset for PIVC-related BSIs, and preliminary data characterization results. The dataset is currently in development, but the plan is to make it openly available in the future. Initial experiments using a machine learning technique on a selection of tasks showed promising results.

The original 18 555 Norwegian AE reports dataset from a large representative university hospital are intended to be processed by hospital administration for quality improvement instead of responsibility, legal, or commercial reasons. The partially structured reports are written by health care personnel in complete sentences; this differs greatly from EHR clinical text, which are grammatically incomplete and brief [26]. The AE data may thus highlight patient safety issues that require addressing at an organizational or local level as well as drive national policy. Hence, this clinical dataset differs in quality, purpose, and structure from EHR clinical text.

To prepare for PIVC-related BSI studies using reasoning tasks and supervised machine learning, this study developed an annotation guideline and a corresponding annotated corpus which represents and captures PIVCs and BSIs documented in AE notes. Further work is required to develop an ontology based on the guideline as a framework to test the representation and reasoning about PIVC-related BSI. There are plans to develop PIVC-related BSI classifiers using word- and phrase-level annotations. Additionally, a previous study predicted central venous catheter events using sentences from clinical text with limited training data [26]. Thus, the annotated synthetic

[6]https://github.com/andreped/adverse-events

data will be preprocessed further to easily use sentence-level annotations for detecting sepsis-related events.

Due to the limited dataset size, unsupervised methods like LDA were preferred, as they tend to be more robust on smaller datasets. However, hyperparameter selection in the classification pipeline was tuned on the test set. It was not possible to use the training set for tuning, as the note-level annotations were only present in the test set. Using only unsupervised objectives for tuning is challenging, as topics might be distributed in numerous ways. Therefore, to obtain appropriate classification performance, guiding hyperparameter selection in a supervised manner was necessary. However, as the model was tuned on the same data used for evaluation, the model might have overfitted. In future work, trained models should be evaluated on independent test data.

To increase data accessibility, the AE dataset can be translated into other languages. Additional prospective work includes cross-lingual annotations, such that word- or phrase-level annotations and insights can be used in other languages.

VII. Conclusion

The Norwegian AE dataset is a resource for quality control improvement in hospitals. In addition to AEs, the dataset contains honest and open reporting about clinically relevant events and improvement suggestions which offers insight for quality assurance and patient safety in healthcare. This differs vastly from other available datasets focusing on adverse drug events and faulty devices malfunctioning. We want to collaborate with other research groups in order to use this dataset to improve patient safety and care quality.

Acknowledgment

We would like to thank the annotators. This research was funded by the Computational Sepsis Mining and Modelling project through the Norwegian University of Science and Technology (NTNU) Health Strategic Area and approved by the Norwegian Regional Committees for Medical and Health Research Ethics (REK), approval no. 26814. Technical support and the platform for performing analyses were provided by NTNU's HUNT Cloud.

References

[1] C. Deutschman and K. Tracey, "Sepsis: Current Dogma and New Perspectives," *Immunity*, vol. 40, pp. 463–475, Apr. 2014.

[2] C. Rhee, R. Dantes, L. Epstein, D. J. Murphy, C. W. Seymour, T. J. Iwashyna, S. S. Kadri, D. C. Angus, R. L. Danner, A. E. Fiore, and et al., "Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014," *JAMA: the J. of the Amer. Med. Assoc.*, vol. 318, no. 13, p. 1241–1249, Oct. 2017.

[3] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, and et al., "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)," *JAMA: the J. of the Amer. Med. Assoc.*, vol. 315, no. 8, p. 801–810, Feb. 2016.

[4] L. E. Huerta and T. W. Rice, "Pathologic Difference between Sepsis and Bloodstream Infections," *The J. of Appl. Lab. Med.*, vol. 3, no. 4, pp. 654–663, Oct. 2019.

[5] J. Paulsen, A. Mehl, Å. Askim, E. Solligård, B. O. Åsvold, and J. K. Damås, "Epidemiology and outcome of Staphylococcus aureus bloodstream infection and sepsis in a Norwegian county 1996-2011: an observational study," *BMC infectious diseases*, vol. 15, p. 116, Mar. 2015.

[6] L. A. Mermel, "Short-term Peripheral Venous Catheter-Related Bloodstream Infections: A Systematic Review," *Clin. Infectious Diseases*, vol. 65, no. 10, p. 1757–1762, Oct. 2017.

[7] E. Alexandrou, G. Ray-Barruel, P. J. Carr, S. A. Frost, S. Inwood, N. S. Higgins, F. Lin, L. Alberto, L. A. Mermel, and C. M. Rickard, "Use of short peripheral intravenous catheters: Characteristics, management, and outcomes worldwide," *J. of Hospital Med.*, vol. 13 5, 2018.

[8] E. Alexandrou, G. Ray-Barruel, P. J. Carr, S. Frost, S. Inwood, N. Higgins, F. Lin, L. Alberto, L. Mermel, and C. M. Rickard, "International prevalence of the use of peripheral intravenous catheters," *J. of Hospital Med.*, vol. 10, no. 8, p. 530–533, Aug. 2015.

[9] W. Zingg and D. Pittet, "Peripheral venous catheters: an under-evaluated problem," *Int. J. of Antimicrobial Agents*, vol. 34 Suppl. 4, p. S38–42, 2009.

[10] L. Zhang, S. Cao, N. Marsh, G. Ray-Barruel, J. Flynn, E. Larsen, and C. M. Rickard, "Infection risks associated with peripheral vascular catheters," *J. of Infection Prevention*, vol. 17, no. 5, pp. 207–213, Sep. 2016.

[11] J. M. Rohde, A. J. Odden, C. Bonham, L. Kuhn, P. N. Malani, L. M. Chen, S. A. Flanders, and T. J. Iwashyna, "The epidemiology of acute organ system dysfunction from severe sepsis outside of the intensive care unit," *J. of Hospital Med.*, vol. 8, no. 5, p. 243–247, May 2013.

[12] E. N. de Vries, M. A. Ramrattan, S. M. Smorenburg, D. J. Gouma, and M. A. Boermeester, "The incidence and nature of in-hospital adverse events: a systematic review," *Qual. & Saf. in Health Care*, vol. 17, no. 3, p. 216–223, Jun. 2008.

[13] P. D. Hibbert, C. J. Molloy, T. D. Hooper, L. K. Wiles, W. B. Runciman, P. Lachman, S. E. Muething, and J. Braithwaite, "The application of the global trigger tool: a systematic review," *Int. J. for Qual. in Health Care*, vol. 28, no. 6, p. 640–649, Dec. 2016.

[14] H. J. Murff, V. L. Patel, G. Hripcsak, and D. W. Bates, "Detecting adverse events for patient safety research: a review of current methodologies," *J. of Biomed. Inform.*, vol. 36, no. 1-2, p. 131–143, Feb. 2003.

[15] F. A. Griffin and R. K. Resar, *IHI Global Trigger Tool for Measuring Adverse Events*, 2nd ed. Cambridge, MA: Inst. for Healthcare Improvement, 2009.

[16] S. N. Musy, D. Ausserhofer, R. Schwendimann, H. U. Rothen, M.-M. Jeitziner, A. W. Rutjes, and M. Simon, "Trigger tool-based automated adverse event detection in electronic health records: Systematic review," *J. of Med. Internet Res.*, vol. 20, no. 5, p. e198, May 2018.

[17] E. Apostolova and T. Velez, "Toward Automated Early Sepsis Alerting: Identifying Infection Patients from Nursing Notes," in *BioNLP 2017*, Aug. 2017, pp. 257–262.

[18] I. Larizgoitia, M.-C. Bouesseau, and E. Kelley, "Who efforts to promote reporting of adverse events and global learning," *J. of Public Health Res.*, vol. 2, no. 3, p. e29, Dec. 2013.

[19] "Japan council for quality health care (JQ) project to collect medical near-miss/adverse event information." [Online]. Available: https://www.med-safe.jp/contents/english/index.html

[20] R. Tikkanen, R. Osborn, E. Mossialos, A. Djordjevic, and G. Wharton, "2020 international health care system profiles," Dec. 2020. [Online]. Available: https://www.commonwealthfund.org/international-health-policy-center/system-profiles

[21] P. Bjark, E. Hansen, and E. Lingaas, "Helsetjenesteassosiert infeksjon som årsak til død i sykehus," *Tidsskrift for den Norske legeforening*, vol. 140, Apr. 2020.

[22] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, and J. Tsujii, "BRAT: a Web-based Tool for NLP-assisted Text Annotation," in *Proc. Demonstrations at 13th Conf. EACL 2012*, Apr. 2012, pp. 102–107.

[23] N. Halko, P.-G. Martinsson, and J. Tropp, "Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions," *SIAM Review*, vol. 53, pp. 217–288, Jan. 2011.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine Learning in Python," *J. of Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[25] T. Head, M. Kumar, H. Nahrstaedt, G. Louppe, and I. Scherbatyi, "scikit-optimize," Sep. 2020, v0.8.1. [Online]. Available: http://doi.org/10.5281/zenodo.4014775

[26] T. B. Røst, C. R. Tvedt, H. Husby, I. A. Berg, and Ø. Nytrø, "Identifying catheter-related events through sentence classification," *Int. J. of Data Mining and Bioinf.*, vol. 23, no. 3, p. 213–233, Jan. 2020.