
Review

Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review

Melissa Y. Yan ¹, Lise Tuset Gustad ^{2,3}, and Øystein Nytrø ¹

¹Department of Computer Science, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, Trondheim, Norway, ²Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway, and ³Department of Medicine, Levanger Hospital, Clinic of Medicine and Rehabilitation, Nord-Trøndelag Hospital Trust, Levanger, Norway

Corresponding Author: Melissa Y. Yan, MBI, Department of Computer Science, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway; melissa.yan@ntnu.no

Received 4 July 2021; Revised 11 September 2021; Editorial Decision 29 September 2021; Accepted 11 October 2021

ABSTRACT

Objective: To determine the effects of using unstructured clinical text in machine learning (ML) for prediction, early detection, and identification of sepsis.

Materials and methods: PubMed, Scopus, ACM DL, dblp, and IEEE Xplore databases were searched. Articles utilizing clinical text for ML or natural language processing (NLP) to detect, identify, recognize, diagnose, or predict the onset, development, progress, or prognosis of systemic inflammatory response syndrome, sepsis, severe sepsis, or septic shock were included. Sepsis definition, dataset, types of data, ML models, NLP techniques, and evaluation metrics were extracted.

Results: The clinical text used in models include narrative notes written by nurses, physicians, and specialists in varying situations. This is often combined with common structured data such as demographics, vital signs, laboratory data, and medications. Area under the receiver operating characteristic curve (AUC) comparison of ML methods showed that utilizing both text and structured data predicts sepsis earlier and more accurately than structured data alone. No meta-analysis was performed because of incomparable measurements among the 9 included studies.

Discussion: Studies focused on sepsis identification or early detection before onset; no studies used patient histories beyond the current episode of care to predict sepsis. Sepsis definition affects reporting methods, outcomes, and results. Many methods rely on continuous vital sign measurements in intensive care, making them not easily transferable to general ward units.

Conclusions: Approaches were heterogeneous, but studies showed that utilizing both unstructured text and structured data in ML can improve identification and early detection of sepsis.

Key words: sepsis, natural language processing, machine learning, electronic health records, systematic review

INTRODUCTION

Sepsis is a life-threatening illness caused by the body's immune response to an infection that leads to multi-organ failure.¹ Annually, there are 31.5 million sepsis cases, 19.4 million severe sepsis cases, and 5.3 million sepsis deaths estimated in high-income countries.²

Studies have shown that early identification of sepsis following rapid initiation of antibiotic treatment improves patient outcomes,³ and 6 h of treatment delay is shown to increase the mortality risk by 7.6%.⁴ Unfortunately, sepsis is commonly misdiagnosed and mistreated because deterioration with organ failure is also common in

other diseases.⁵⁻⁸ The heterogeneity in infection source, immune responses, and pathophysiological changes make identification and therefore sepsis treatment difficult. Additionally, the diversity in age, gender, and comorbidities affect the symptoms and outcome of septic patients.⁷

Machine learning (ML) has been employed to improve sepsis outcomes through early detection. ML can utilize structured and unstructured data from electronic health records (EHRs).⁹⁻¹⁴ Structured clinical data come in a fixed format, such as age, vital signs, and laboratory data, which make data preprocessing easier. In contrast, clinical notes are in unstructured free-text form, such as progress notes, nursing notes, chief complaints, or discharge summaries. Clinical notes contain abbreviations, grammatical errors, and misspellings. Using clinical text is a complex, time-consuming process because it requires using natural language processing (NLP) to extract features that transform text into a machine-understandable representation.¹⁵⁻²² This usually requires assistance from clinical experts to convert text into machine-interpretable representations that capture clinical knowledge for specific clinical domains. The effort required to utilize unstructured clinical text can deter researchers; however, unstructured clinical text contains valuable information.^{16,22-25} Multiple studies and a review²⁵ have shown that using unstructured clinical text has increased model performance to detect or predict colorectal surgical complications,²⁶ postoperative acute respiratory failure,²⁷ breast cancer,²⁸ pancreatic cancer,²⁹ fatty liver disease,³⁰ pneumonia,³¹ inflammatory bowel disease,^{32,33} rheumatoid arthritis,³⁴⁻³⁶ multiple sclerosis,³⁷ and acute respiratory infection.^{38,39}

Prior reviews related to sepsis detection and prediction include: sepsis detection using Systemic Inflammatory Response Syndrome (SIRS) screening tools,⁴⁰ sepsis detection using SIRS and organ dysfunction criteria with EHR vital signs and laboratory data,⁴¹ clinical perspectives on the use of ML for early detection of sepsis in daily practice,¹⁴ ML for diagnosis and early detection of sepsis patients,⁹⁻¹³ infectious disease clinical decision support,⁴² and healthcare-associated infections mentioning sepsis.⁴³⁻⁴⁵ However, to the best of our knowledge, no reviews focus on the effect of utilizing unstructured clinical text for sepsis prediction, early detection, or identification; this makes it challenging to assess and utilize text in future ML and NLP sepsis research.

OBJECTIVE

The review aims to gain an overview of studies utilizing clinical text in ML for sepsis prediction, early detection, or identification.

MATERIALS AND METHODS

This systematic review follows the Preferred Reporting Items for Systematic review and Meta-Analyses guidelines.⁴⁶

Search strategy

Relevant articles were identified from 2 clinical databases (PubMed and Scopus) and 3 computer science databases (ACM DL, dblp, and IEEE Xplore) using defined search terms. The 3 sets of search terms included: (1) “sepsis,” “septic shock,” or “systemic inflammatory response syndrome”; (2) “natural language processing,” “machine learning,” “artificial intelligence,” “unstructured data,” “unstructured text,” “clinical note,” “clinical notes,” “clinical text,” “free-text,” “free text,” “record text,” “narrative,” or

“narratives”; and (3) detect, identify, recognize, diagnosis, predict, prognosis, progress, develop, or onset. Searches on clinical databases were performed using all 3 sets of search terms and excluded animal-related terms. Whereas searches on computer science databases only used the first set of search terms. No additional search restrictions, such as date, language, and publication status, were included. Additional articles were identified from relevant review articles or backward reference and forward citation searches of eligible articles. Complete search strategies are in [Supplementary Table S1](#).

The search was initially conducted using only computer science databases on December 10, 2019 and was updated to include clinical databases on December 14, 2020. The first search found that 4 of 454 articles met inclusion criteria,⁴⁷⁻⁵⁰ and the second search uncovered 2 more articles that met inclusion criteria (6 of 1335 articles).^{51,52} Those 2 searches did not contain the search terms: “systemic inflammatory response syndrome,” “artificial intelligence,” identify, recognize, diagnosis, prognosis, progress, develop, and onset. Hence, a search on May 15, 2021, including those terms, found 2 additional articles.^{53,54} To ensure inclusion of other relevant articles, a broader search was conducted on September 3, 2021 to include the following terms: “unstructured data,” “unstructured text,” “clinical note,” “clinical notes,” “clinical text,” “free-text,” “free text,” “record text,” “narrative,” or “narratives.” This resulted in 1 additional article.⁵⁵

Study selection

Titles, abstracts, and keywords were screened using Zotero v5.0.96.3 (Corporation for Digital Scholarship, Vienna, VA) and Paperpile (Paperpile LLC, Cambridge, MA). Screening removed duplicates and articles that did not contain the following terms: (1) text, (2) notes, or (3) unstructured. Full-text articles were evaluated to determine if the study used unstructured clinical text for the identification, early detection, or prediction of sepsis onset in ML. Thus, selected articles had to rely on methods that automatically improve based on what they learn and not rely solely on human-curated rules. Additionally, articles solely focusing on predicting sepsis mortality were excluded as these articles are based on already established sepsis cases. Reviews, abstract-only articles, and presentations were removed. Additionally, a backward and forward search was performed on eligible full-text articles.

Data extraction

One author independently extracted data, which a second author verified. Any discrepancies were resolved either through discussion with the third author by assessing and comparing data to evidence from the studies or by directly communicating with authors from included articles. The following information was extracted: (1) general study information including authors and publication year, (2) data source, (3) sample size, (4) clinical setting, (5) sepsis infection definition, (6) task and objective, (7) characteristics of structured and unstructured data, (8) underlying ML and NLP techniques, and (9) evaluation metrics.

RESULTS

Selection process

The initial search identified 2268 articles from 5 databases and 5 additional articles⁵⁶⁻⁶⁰ from 2 relevant review articles ([Figure 1](#)).^{43,44} From the 1817 unique articles, 1620 articles were excluded based on eligibility criteria described in the methods. After assessing the

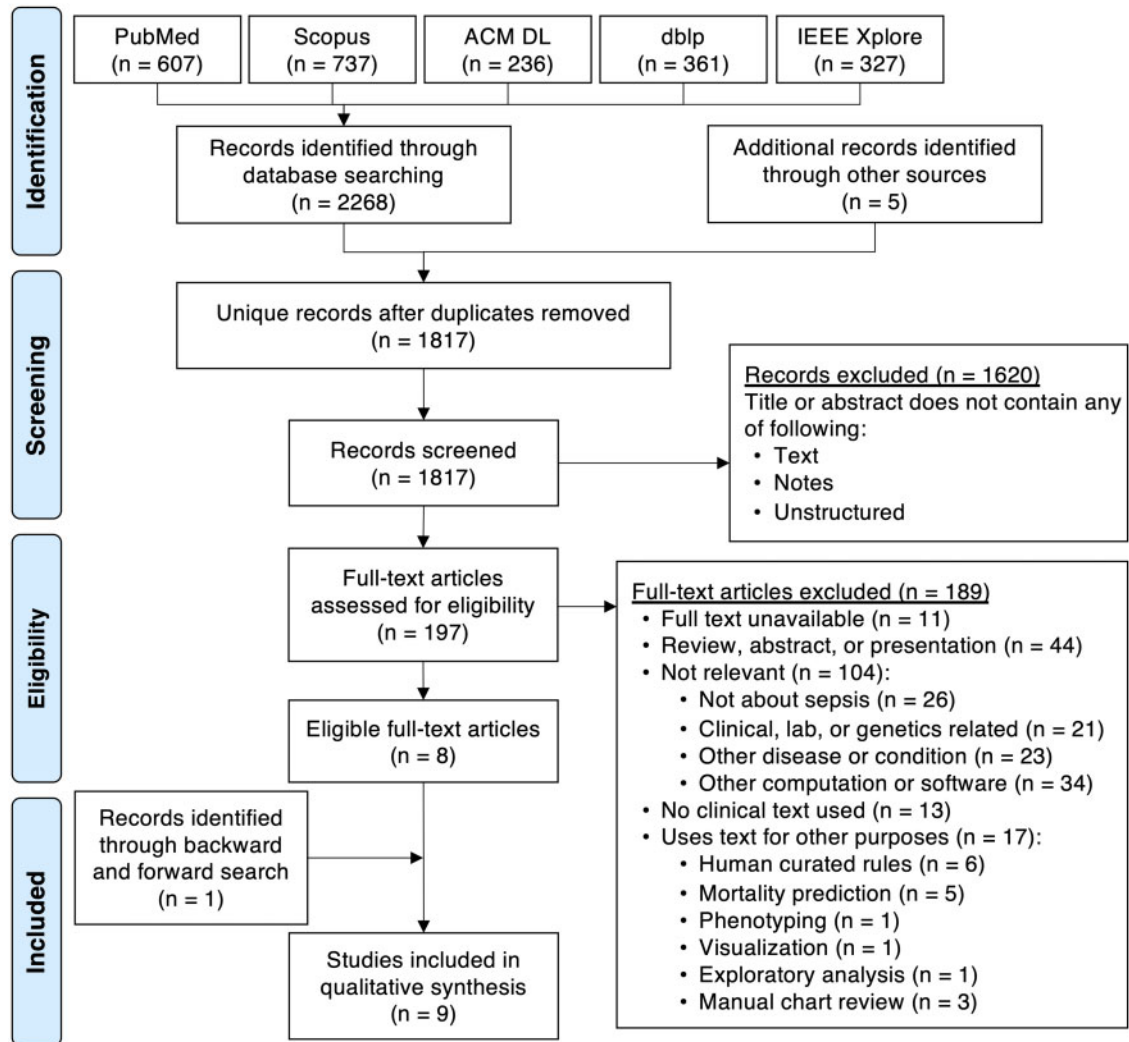


Figure 1. PRISMA (Preferred Reporting Items for Systemic reviews and Meta-Analyses) flowchart for study selection.

remaining 197 articles, most studies (189 of 197, ie, 96%) were excluded because they had not used or attempted to use unstructured clinical text in their ML models to identify, detect, or predict sepsis onset. For instance, there were sepsis-related studies that used text but for other purposes such as mortality prediction,^{61–65} phenotyping,⁶⁶ visualization,⁶⁷ exploratory data analysis,⁶⁸ and manual chart review.^{69–71} Additionally, 6 articles about infection detection,⁶⁰ central venous catheter adverse events,⁵⁸ postoperative sepsis adverse events,^{72–74} and septic shock identification⁷⁵ were excluded because they used manually human-curated rules instead of ML methods that automatically learn from data. The remaining 8 eligible articles were used to perform backward and forward searches,^{47–50,52–55} which led to the inclusion of 1 additional article.⁵¹ This resulted in 9 articles for synthesis.

Study characteristics

Of the 9 identified articles, 2 studies aimed at identifying infection,^{47,48} 6 studies focused on early detection of sepsis,^{51,53,55} severe sepsis,⁴⁹ or septic shock,^{50,54} and 1 study considered both identification and early detection for a combination of sepsis, severe sepsis, and septic shock.⁵² Most studies focused on intensive care unit (ICU)^{48,50,52–55} or emergency department (ED)^{47,51} data; only 1 used inpatient care data.⁴⁹ Four studies

utilized data from hospitals,^{47,49,51,52} 1 utilized MIMIC-II⁵⁴ and 4 utilized MIMIC-III.^{48,50,53,55} MIMIC-II and MIMIC-III are publicly available ICU datasets created from Boston's Beth Israel Deaconess Medical Center; MIMIC-II contains data from 2001–2007⁷⁶ and MIMIC-III contains data from 2001–2012.⁷⁷ Eight studies used data from the United States^{47–51,53–55} and 1 study used data from Singapore.⁵² Sample sizes varied greatly in terms of the number of patients or notes used. To select patient cohorts or notes associated with sepsis, 3 studies used International Statistical Classification of Diseases and Related Health Problems (ICD) codes,^{47,49,52} 5 applied sepsis definition criteria,^{49–51,53,55} 1 utilized descriptions of antibiotics usage,⁴⁸ and another⁵⁴ applied criteria from Henry et al⁷⁸ that include ICD codes, sepsis criteria, and notes mentioning sepsis or septic shock. Table 1 summarizes the study characteristics and additional details are in Supplementary Table S2 (for Culliton et al,⁴⁹ the 8 structured variables for the Modified Baystate clinical definition of severe sepsis and 29 structured variables used in models were provided through personal communications with the corresponding author of Culliton et al,⁴⁹ Steve Gallant, on June 4, 2021).

Clinical text used in models

The 9 studies utilized narrative notes written by nurses,^{47–50,53–55} physicians,^{49–53,55} or specialists^{49–51,54,55} to document symptoms,

Table 1. Study characteristics

| Study (year) | Clinical setting and data source | Sample size ^a | Cohort criteria infection definition | Task and objective |
|---|---|---|---|--|
| Hornig et al. ⁴⁷ (2017) | <ul style="list-style-type: none"> ED Beth Israel Deaconess (Boston, MA, United States) Dec 17, 2008—Feb 17, 2013 | 230 936 patient visits <ul style="list-style-type: none"> Infection: 32 103 P; 14% No infection: 198 833 P; 86% Train : 147 799 P; 64% Validation: 46 187 P; 20% Test: 36 950 P; 16% | Angus Sepsis ICD-9-CM abstraction criteria ⁷⁹ | Identify patients with suspected infection to demonstrate benefits of using clinical text with structured data for detecting ED patients with suspected infection. |
| Apostolova and Velez ⁴⁸ (2017) | <ul style="list-style-type: none"> ICU MIMIC-III 2001–2012 | 634 369 nursing notes <ul style="list-style-type: none"> Infection presence: 186 158 N; 29% Possible infection: 3262 N; 1% No infection: 448 211 N; 70% Train: 70% Test: 30% | Notes describing patient taking or being prescribed antibiotics for treating infection | Identify notes with suspected or presence of infection to develop a system for detecting infection signs and symptoms in free-text nursing notes. |
| Culliton et al. ⁴⁹ (2017) | <ul style="list-style-type: none"> Inpatient care Baystate hospitals (Springfield, MA, United States) 2012–2016 | 203 000 adult inpatient admission encounters <ul style="list-style-type: none"> Used 68 482 E Severe sepsis: 1427 E; 2.1% 3-fold cross validation: only text data Model construction: 2012–2015 data Test set: 2016 data: <ul style="list-style-type: none"> Used 13 603 E Severe sepsis: 425 P; 3.1% | Modified Baystate clinical definition of severe sepsis (8 structured variables) and severe sepsis ICD codes | Predict severe sepsis 4, 8, and 24 h before the earliest time structured variables meet the severe sepsis definition to compare accuracy of predicting patients that will meet the clinical definition of sepsis when using unstructured data only, structured data only, or both types. |
| Delahanty et al. ⁵¹ (2019) | <ul style="list-style-type: none"> ED Tenet Healthcare Hospitals (Nashville, TN, United States) January 1, 2016—October 31, 2017 | 2 759 529 patient encounters <ul style="list-style-type: none"> Sepsis: 54 661 E; 2% No Sepsis: 2 704 868 E; 98% Train: 1 839 503 E; 66.7% <ul style="list-style-type: none"> Sepsis: 36 458 E; 2% No sepsis: 1 803 045 E; 98% Test: 920 026 E; 33.3% <ul style="list-style-type: none"> Sepsis: 18 203 E; 2% No sepsis: 901 823 E; 98% | Rhee's modified Sepsis-3 definition ⁸⁰ | Predict sepsis risk in patients 1, 3, 6, 12, and 24 h after the first vital sign or laboratory result is recorded in the EHR to develop a new sepsis screening tool comparable to benchmark screening tools. |
| Liu et al. ⁵⁰ (2019) | <ul style="list-style-type: none"> ICU MIMIC-III 2001–2012 | 38 645 adult patients Train: 70% P Test: 30% P Applied model to: 15 930 P with suspected infection and at least 1 physiological EHR data | Sepsis-3 definition ¹ | Predict septic shock in sepsis patients before the earliest time septic shock criteria are met to demonstrate an approach using NLP features for septic shock prediction. |
| Amrollahi et al. ⁵³ (2020) | <ul style="list-style-type: none"> ICU MIMIC-III 2001–2012 | 40 175 adult patients <ul style="list-style-type: none"> Sepsis: 2805 P; ~7% Train 80% P Test 20% P | Sepsis-3 definition ¹ | Predict sepsis onset hours in advance using a deep learning approach to show a pre-trained neural language representation model can improve early sepsis detection. |

(continued)

Table 1. continued

| Study (year) | Clinical setting and data source | Sample size ^a | Cohort criteria infection definition | Task and objective |
|-------------------------------------|---|--|--|--|
| Hammoud et al. ⁵⁴ (2020) | <ul style="list-style-type: none"> ICU MIMIC-II 2001–2007 | 17 763 patients <ul style="list-style-type: none"> Sepsis: 6097 P Severe sepsis: 3962 P Septic shock : 1469 P 5-fold cross validation | Sepsis definition based on what Henry et al ⁷⁸ used | Predict early septic shock in ICU patients using a model that can be optimized based on user preference or performance metrics. |
| Goh et al. ⁵² (2021) | <ul style="list-style-type: none"> ICU Singapore government-based hospital (Singapore, Singapore) Apr 2, 2015—Dec 31, 2017 | 5317 patients (114 602 notes) Train and validation: 3722 P (80 162 N) <ul style="list-style-type: none"> Sepsis: 6.45% No sepsis: 93.55% Test: 1595 P (34 440 N) <ul style="list-style-type: none"> Sepsis: 5.45% No sepsis: 94.55% | ICU admission with an ICD-10 code for sepsis, severe sepsis, or sepsis shock | Identify if a patient has sepsis at consultation time or predict sepsis 4, 6, 12, 24, and 48 h after consultation to develop an algorithm that uses structured and unstructured data to diagnose and predict sepsis. |
| Qin et al. ⁵⁵ (2021) | <ul style="list-style-type: none"> ICU MIMIC-III 2001–2012 | 49 168 patients Train: 33 434 P <ul style="list-style-type: none"> Sepsis: 1353 P No Sepsis: 32 081 P Validation: 8358 P <ul style="list-style-type: none"> Sepsis: 338 P No Sepsis: 8020 P Test: 7376 P <ul style="list-style-type: none"> Sepsis: 229 P No Sepsis: 7077 P | PhysioNet Challenge restrictive Sepsis-3 definition ⁸¹ | Predict if a patient will develop sepsis to explore how numerical and textual features can be used to build a predictive model for early sepsis prediction. |

ED: emergency department; ICU: intensive care unit; ICD: International Classification of Diseases; ICD-9 CM: ICD Clinical Modification, 9th revision; ICD-10: ICD 10th revision; MIMIC-II: Multiparameter Intelligent Monitoring in Intensive Care II database; MIMIC-III: Medical Information Mart for Intensive Care dataset.

^aSample size unit abbreviations: P: patients; N: notes; E: encounters.

signs, diagnoses, treatment plans, care provided, laboratory test results, or reports. EHRs contain various types of clinical notes. A note covers an implicit time period or activity and describes events, hypotheses, interventions, and observations within the health care provider's responsibilities. The note's form depends on its function: an order, a plan, a prescription, an investigation or analysis report, a narrative or log of events, information for the next shifts, or a requirement for legal, medical, or administrative purposes. An episode of care begins when a patient is admitted to the hospital and ends when the patient is discharged. Throughout a patient's hospital stay, documentation can include chief complaints, history-and-physical notes, progress notes, reports, descriptions of various laboratory tests, procedures, or treatments, and a discharge summary. Chief complaints are the symptoms or complaints provided by a patient for why they are seeking care.⁸² History-and-physical notes can include history about the current illness, medical history, social history, family history, a physical examination, a chief complaint, probable diagnosis, and a treatment plan.⁸³ Progress notes document care provided and a description of the patient's condition to convey events to other clinicians.⁸⁴ Free-text reports can include interpretations of echocardiograms, electrocardiograms (ECGs), or imaging results such as X-rays, computerized tomography scans, magnetic resonance imaging scans, and ultrasounds. At discharge, the health care personnel write a discharge summary note comprised of patient details, hospital admittance reason, diagnosis, conditions, history, progress, interventions, prescribed medications, and follow-up plans.^{85–87} The discharge summary letter is a formal document used to transfer patient care to another provider for further treatment and follow-up care.^{88–90}

Studies have shown that nursing documentation differs from physician documentation.^{91,92} Nurses document more about a patient's functional abilities than physicians,⁹¹ and the information from notes used and the frequency of viewing and documenting differs between health care personnel.⁹² Additionally, documentation varies between hospitals,^{93,94} hospitals have different resources and practices,^{95–97} and communicative behavior differs among professions in different wards.⁹⁸ Hence, the type of notes used, who wrote the notes, and purpose of the note will play a role in how the documentation is interpreted.⁹⁹

Table 2 provides information regarding documentation types, author of the note, time content of the data, time latency between documentation and availability in records, and the documentation frequency. In Figure 2, the relationship between hospital events and longitudinal data used to train models is shown. As sepsis develops in a patient over time, it shows there are typically delays between a patient's actual state, clinical observations, and recorded documentation, such as ICU vital signs, narrative notes, and ICD codes.

The included studies utilized the following types of notes: 6 studies used unstructured nursing-related documentation,^{47,48,50,53–55} 4 used physician notes,^{50,52,53,55} 3 used radiology reports,^{50,54,55} 3 used respiratory therapist progress notes,^{50,54,55} 2 used ED chief complaints,^{47,51} 2 used ECG interpretations,^{50,54} 2 used pharmacy reports,^{50,54} 2 used consultation notes,^{50,52} 1 used discharge summaries,⁵⁰ 1 included mostly progress notes and history-and-physical notes,⁴⁹ and 3 used additional unspecified notes.^{49,50,54} Not all notes used are listed. Liu et al⁵⁰ used all MIMIC-III notes to build a vocabulary of unique words, and discharge summaries were likely not used in predictions because they are unlikely to occur before

Table 2. Clinical documentation from electronic health records

| Documentation types | Author | Description | Temporal perspective | Record latency ^a | Frequency |
|----------------------------|---|--|--|------------------------------|-------------------------|
| Chief complaints | <ul style="list-style-type: none"> Physician Nurse Specialist | Symptoms or complaints provided by a patient at start of care for why they are seeking care. | Current | Seconds to days | One per episode |
| History-and-physical notes | <ul style="list-style-type: none"> Physician Nurse | Past medical history, family history, developmental history of present illness, problems about present illness, past medications or immunizations, allergies, or habits. | Retrospective | Immediately | One per episode |
| Progress notes | <ul style="list-style-type: none"> Physician Nurse Specialist (eg, respiratory therapist) | <p>Observations of patient status and care provided to document progress and response to treatment plans.</p> <p>For physician, it includes determining diagnosis, prescriptions, and laboratory orders.</p> | <ul style="list-style-type: none"> Retrospective Prospective | 4–8 h | One per shift |
| Reports | Specialist | Radiologist results and cardiology results. | Retrospective | Days | One to many per episode |
| Discharge summary notes | Health care personnel | Episode of care summary and follow-up plans. | <ul style="list-style-type: none"> Retrospective Prospective | At discharge or days after | One per episode |
| Discharge summary letter | Physician | Formal required letter containing follow-up treatment plans. | <ul style="list-style-type: none"> Retrospective Prospective | Days to months after episode | One per episode |
| Laboratory results | Laboratory technician | Laboratory test analysis results from provided samples (eg, blood, urine, skin, and device) based on the physician's order. | Retrospective | Days | One to many per episode |
| ICD codes | <ul style="list-style-type: none"> Physician Professional ICD coder ICD data aggregator organization | Diagnosis classification for billing. | Retrospective | Days to months | One per episode |
| Administrative | Administration | Patient information such as name, age, gender, address, contact information, and occupation. | <ul style="list-style-type: none"> Retrospective Current | Immediately | One per episode |

^aRecord latency is defined as time between measurement/observation and the availability of the results in electronic health records.

observations. Additionally, Hammoud et al⁵⁴ used all MIMIC-II notes except discharge summaries.

These 9 studies utilized clinical notes differently. For the unit of analysis, 6 studies used a single note,^{47,48,50,52–54} 1 used a set of many notes from a patient encounter,⁴⁹ 1 used a set of many notes within a specific hour of consideration,⁵⁵ and 1 used keywords from notes.⁵¹ To identify infection signs, Hornig et al⁴⁷ and Apostolova

and Velez⁴⁸ processed individual notes. While Goh et al⁵² used notes at each patient consultation instance to identify sepsis patients. For early detection, 5 studies defined onset time as the earliest time when definition criteria are met^{49,50,53–55} and 1 defined sepsis onset time as ICU ward admission time.⁵² Studies for early detection used varying windows with different durations. A window decides how and where to obtain longitudinal data, and duration is the length of

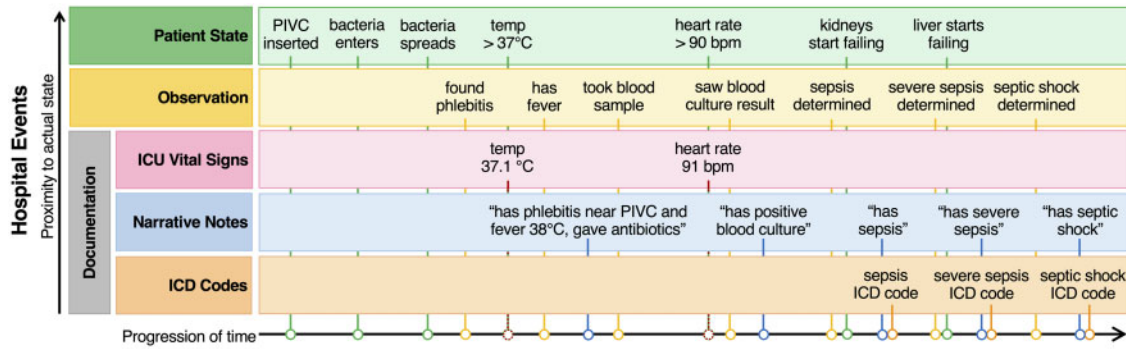


Figure 2. Overview of data from a patient timeline used to create models. The proximity of events toward a patient’s actual state and the actual documentation recorded in the electronic health records typically has delays. Green represents patient states as sepsis develops in a patient. Yellow are observations made by clinicians. Documentation includes ICU vital signs^a in pink, narrative notes in blue, and ICD codes in orange. ICU vital sign^a documentation can be instantaneous, narrative notes can be written after observations are made, and ICD codes are typically registered after a patient is discharged. PIVC: peripheral intravenous catheter. ^aVital signs include temperature, pulse, blood pressure, respiratory rate, oxygen saturation, and level of consciousness and awareness.

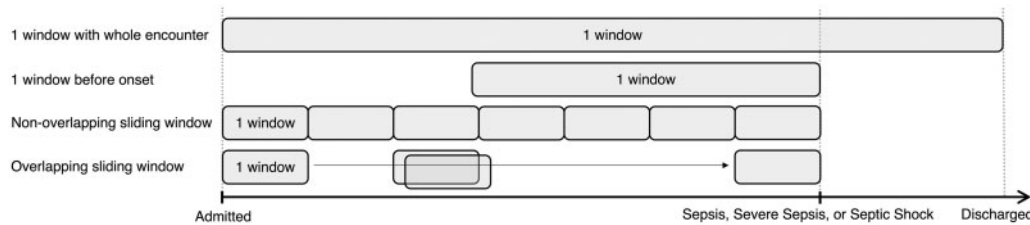


Figure 3. Different types of windows were used to obtain longitudinal data. Each gray box represents a single window, which can vary in duration (length of time) depending on the study. One window with the whole encounter means the study used a single window containing data with a duration of the whole encounter from admittance until discharge. One window before onset signifies data from a window with a duration of time before sepsis, severe sepsis, or septic shock onset. Sliding windows are consecutive windows until before sepsis, severe sepsis, or septic shock onset; this includes non-overlapping and overlapping sliding windows. Non-overlapping sliding windows indicate that data within one window of a fixed duration does not contain data in the next window. In contrast, overlapping sliding windows indicate windows of a fixed duration overlap, and data within one window will be partially in the next window.

time. As shown in Figure 3, studies can use windows differently, such as a window with the duration of the whole encounter, a window with a duration of hours before onset, non-overlapping sliding windows with a fixed duration until onset, or overlapping sliding windows with a fixed duration until onset. Culliton et al⁴⁹ used a 4-, 8-, or 24-h duration window before severe sepsis, and concatenated all text within a window. Goh et al⁵² used a 4-, 6-, 12-, 24-, or 48-h duration window of before sepsis, severe sepsis, or septic shock onset. Liu et al⁵⁰ used 10 data points within a 1-h duration window spanning 2 h before septic shock, and used the most recently entered note for a data point to predict septic shock. Hammoud et al⁵⁴ binned data in 15-minute duration non-overlapping sliding windows to update septic shock predictions every 15 minutes, and used the last note within the window. Amrollahi et al⁵³ binned data into 1-h duration non-overlapping sliding windows to provide hourly sepsis predictions, and used sentences within a note to capture the semantic meanings. Qin et al⁵⁵ used 6-h duration overlapping sliding windows with 6 data points to predict sepsis; a data point was generated from each hour within the window and all clinical notes within the hour were concatenated in random-order. Delahanty et al⁵¹ used a 1-, 3-, 6-, 12-, or 24-h duration window after the first vital sign or laboratory result was documented in the EHR to identify patients at risk for sepsis, and utilized keywords.

First 2 columns in Table 3 show the type of text and unit of analysis used. Additional details about variables and specific notes used are listed in Supplementary Table S3 (the types of notes and usage for Liu et al⁵⁰ was confirmed through personal communications

with Ran Liu on June 2, 2021, for Hammoud et al⁵⁴ by Ibrahim Hammoud on May 29, 2021, and for Qin et al⁵⁵ by Fred Qin on September 9, 2021. Additionally, the structured variables used in models for Culliton et al⁴⁹ were provided through personal communications with Steve Gallant on June 4, 2021). In Figure 4, single notes or a set of many notes are preprocessed and represented to extract features, whereas keywords are used as is. Then structured data can be added, and the data are used to train ML models.

As shown in Figures 3 and 4 and listed in Tables 1 and 3 and Supplementary Tables S2 and S3, although all studies are related to sepsis, there are varying sample sizes, data types, inclusion criteria, and objectives. This heterogeneity makes it challenging to compare results for a meta-analysis.

Natural language processing and machine learning study outcomes

To utilize text in ML, it must be transformed into a representation understandable by computers. In order to do that, Bag-of-words (BoW),¹⁰⁰ n-gram, term frequency-inverse document frequency (tf-idf), and paragraph vectors (PV)¹⁰¹ representations can be used. These representations can be improved using additional NLP techniques, such as stop word removal, lemmatization, and stemming. In addition, other useful features can be extracted from text using part-of-speech (POS) tagging, named entity recognition, or Latent Dirichlet Allocation (LDA) topic modeling.¹⁰² In recent years, neural networks (NNs) have shown high predictive performance. As a result, many state-of-the-art results have been achieved using NNs to learn

Table 3. Text used in studies

| Study (year) | Free-text document type | Unit of analysis | Text processing |
|---|---|------------------------------------|---|
| Hornig et al. ⁴⁷ (2017) | <ul style="list-style-type: none"> ED chief complaints Nursing triage assessments | One note | Representation: <ul style="list-style-type: none"> Bi-gram BoW (15 240-word vocabulary) LDA topic modeling (500 topics) Techniques: <ul style="list-style-type: none"> Convert to lowercase Remove rare tokens and punctuation Negation |
| Apostolova and Velez ⁴⁸ (2017) | Nursing notes | One note | Representation: <ul style="list-style-type: none"> BoW CBoW (200 vector size with window size of 7 = 441-term vocabulary of antibiotics usage and rules for negation and speculations) tf-idf PV (600 vector size for document-level representation) Techniques: <ul style="list-style-type: none"> Convert to lowercase Remove frequent tokens and non-alphanumeric characters Negation |
| Culliton et al. ⁴⁹ (2017) | Clinical notes (mostly progress notes and history-and-physical notes) | One patient encounter = many notes | Representation: <ul style="list-style-type: none"> GloVe (300-dimensional vector) + summing word vectors Techniques: <ul style="list-style-type: none"> Concatenated all notes for an encounter into a single text block |
| Delahanty et al. ⁵¹ (2019) | ED chief complaints | Keywords | Other: <ul style="list-style-type: none"> Keywords extracted by experts |
| Liu et al. ⁵⁰ (2019) | All MIMIC-III clinical notes, such as but not limited to: <ul style="list-style-type: none"> Nursing notes Physician notes | One note | Representation: <ul style="list-style-type: none"> BoW (8907 unique term vocabulary and 832 predictive terms) GloVe (300-dimensional vector for each unique term) Techniques: <ul style="list-style-type: none"> Convert to lowercase Remove rare tokens, frequent tokens, and non-alphanumeric characters |
| Amrollahi et al. ⁵³ (2020) | <ul style="list-style-type: none"> Nursing notes Physician notes | One note | Representation: <ul style="list-style-type: none"> tf-idf (2227 vector size features = 2187 text features + 40 structured features) ClinicalBERT (808 vector size features = 768 text features + 40 structured features) Techniques: <ul style="list-style-type: none"> Remove rare tokens, frequent tokens, stop words, dates, and special characters |
| Hammoud et al. ⁵⁴ (2020) | All MIMIC-II notes except discharge summaries, such as but not limited to: <ul style="list-style-type: none"> Nursing progress notes Respiratory therapist progress notes | One note | Representation: <ul style="list-style-type: none"> BoW tf-idf Techniques: <ul style="list-style-type: none"> Remove rare and frequent tokens |

(continued)

Table 3. continued

| Study (year) | Free-text document type | Unit of analysis | Text processing |
|---------------------------------|--|------------------|--|
| Goh et al. ⁵² (2021) | Physician notes: <ul style="list-style-type: none"> • Admission notes • Progress notes • ICU consultations • Pharmacy notes • Allied health notes | One note | Representation: <ul style="list-style-type: none"> • tf-idf • LDA topic modeling (100 topics) Techniques: <ul style="list-style-type: none"> • Remove rare tokens, punctuation, and stop words • Lemmatization • POS tagging • Manual classification of topics into categories |
| Qin et al. ⁵⁵ (2021) | <ul style="list-style-type: none"> • Nursing notes • Physician notes • Radiology notes • Respiratory notes | Many notes | Representation: <ul style="list-style-type: none"> • tf-idf (1000 vector size = 1000 most common term vocabulary) • ClinicalBERT (768 vector size features^a = either by concatenating all text first as input or using individual notes as input and concatenating output of individual notes) Techniques: <ul style="list-style-type: none"> • Random-order concatenation of all clinical notes within the hour of consideration.^a • Named entity recognition |

BoW: Bag-of-words; CBoW: Continuous bag-of-words; ClinicalBERT: Clinical Bidirectional Encoder Representations from Transformers; ED: emergency department; GloVe: Global Vectors for Word Representation; ICU: intensive care unit; LDA: Latent Dirichlet Allocation; POS tagging: Part-of-speech tagging; PV: paragraph vectors; tf-idf: term frequency-inverse document frequency.

^aRepresentation and technique details for Qin et al.⁵⁵ were provided through personal communications (with Fred Qin on September 7, 2021).

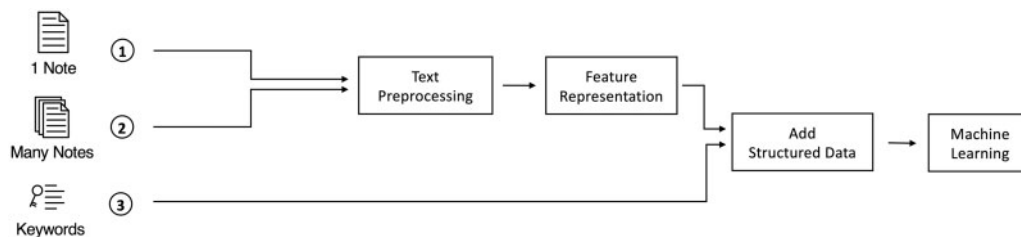


Figure 4. The unit of analysis used to train machine learning models for the included studies was either (1) a single note, (2) a set of many notes, or (3) keywords. In general, text was preprocessed and represented as features interpretable by a computer, then structured data were added, and the data were used to fit machine learning models.

a suitable representation of texts, often known as embeddings.¹⁰³ Embedding techniques include Global Vectors for Word Representation (GloVe),¹⁰⁴ Word2Vec as a continuous bag-of-words (CBoW) model or skip-gram model,¹⁰⁵ Bidirectional Encoder Representations from Transformers (BERT),¹⁰⁶ and ClinicalBERT.¹⁰⁷ The advantage of using embeddings is that it retains the sequential information lost in a BoW representation and does feature extraction automatically.¹⁰³

Utilized text processing operations are in Table 3. One study used keyword extraction instead of text processing operations.⁵¹ Six studies used tokenization of words for word-level representation,^{47–50,52,54} 1 also tried PV for document-level representation,⁴⁸ and another used the first 40 tokens in a sentence to get sentence-level representation and averaged sentence-level representations to provide

document-level representation.⁵³ The most common technique for improving representation was token removal, such as removing rare tokens,^{47,50,52–54} frequent tokens,^{48,50,53,54} punctuation or special characters,^{47,48,50,52,53} and stop words.^{52,53} The most frequently used representation was tf-idf,^{48,52–55} followed by BoW,^{47,48,50,54} LDA,^{47,52} GloVe,^{49,50} ClinicalBERT,^{53,55} bi-gram,⁴⁷ CBoW,⁴⁸ and PV.⁴⁸ Three studies created a vocabulary of unique terms using BoW,⁵⁰ CBoW,⁴⁸ and tf-idf.⁵³ Apostolova and Velez⁴⁸ found that using structured data was inadequate for identifying infection in nursing notes, so they used antibiotic usage and word embeddings to create a labeled dataset of notes with infection, suspected infection, and no infection. Additionally, Horng et al⁴⁷ and Liu et al⁵⁰ listed predictive terms in their models, and Goh et al⁵² provided a list of categories used to classify the top 100 terms. Examples of predictive features

are: (1) For sepsis, severe sepsis, or septic shock, Goh et al⁵² classified the top 100-topics into 7 categories: clinical condition or diagnosis, communication between staff, laboratory test order or results, non-clinical condition updates, social relationship information, symptoms, and treatments or medication. (2) Liu et al's⁵⁰ most predictive NLP terms for the pre-shock versus non-shock state include "tube," "crrt," "ards," "vasopressin," "portable," "failure," "shock," "sepsis," and "dl." (3) Horng et al's⁴⁷ most predictive terms or topics for having an infection in the ED include "cellulitis," "sore_throat," "abscess," "uti," "dysuria," "pneumonia," "redness_swelling," "erythema," "swelling," "redness, cellulitis, left, leg, swelling, area, rle, arm, lle, increased, erythema," "abscess, buttock, area, drainage, axilla, groin, painful, thigh, left, hx, abscesses, red, boil," and "cellulitis, abx, pt, iv, infection, po, keflex, antibiotics, leg, treated, started, yesterday." Whereas the least predictive terms or topics for not having an infection include "motor vehicle crash," "laceration," "epistaxis," "pancreatitis," "etoh"(ethanol for drunkenness), "etoh, found, vomiting, apparently, drunk, drinking, denies, friends, trauma_neg, triage," and "watching, tv, sitting, sudden_onset, movie, television, smoked, couch, pt, pot, 5pm, theater."

ML methods for detecting sepsis using clinical text included: ridge regression,⁴⁹ lasso regression,⁵⁴ logistic regression,^{47,48,52} Naïve Bayes (NB),⁴⁷ support vector machines (SVMs),^{47,48} K-nearest neighbors (KNNs),⁴⁸ random forest (RF),^{47,52} gradient boosted trees (GBTs),^{50-52,55} gated recurrent unit (GRU),⁵⁰ and long short-term memory (LSTM).⁵³ Although the methods are listed separately, 2 studies combined different ML methods^{48,52} (see [Supplementary Table S4](#) for details). Ridge and lasso regression are linear regression methods that constrain the model parameters. A linear regression model is represented as $\hat{y} = \beta_1 x + \beta_0$, where \hat{y} is the predicted value, x is the input variable and β_1 and β_0 are model parameters. Model parameters are estimated by minimizing $\sum_{i=1}^N (y_i - \hat{y}_i)^2$, where y_i is the label and N is the number of training samples. In ridge and lasso regression, $\sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^2 f(\beta_j)$ is minimized instead, where λ is a hyperparameter that trades-off between fitting the data and model complexity, and $f(z) = z^2$ for ridge regression or $f(z) = |z|$ for lasso regression. Logistic regression is a classification method that models $P(y|x)$, which is the probability of a class y given the feature x . The logistic regression model is defined as $f(x) = \frac{1}{1+e^{-(\beta_1 x + \beta_0)}}$. NB is a Bayesian network that eases computation by assuming all input variables are independent given the outcome.¹⁰⁸ SVM is an extension of a support vector classifier that separates training data points into 2 class regions using a linear decision boundary and classifies new data points based on which region they belong to. To accommodate for non-linearity in the data, SVM enlarges the feature space by applying kernels.¹⁰⁹ KNNs assume similar data points are close together and use similarity measures to classify new data based on "proximity" to points in the training data.¹¹⁰ RF and GBT are ensemble models that use a collection of decision trees to improve the predictive performance of the models. RF classification takes the majority vote of a collection of trees to reduce the decision tree variance.¹¹¹ GBT trains decision trees sequentially so that each tree trains based on information from previously trained trees.^{112,113} To avoid overfitting, each tree is scaled by a hyperparameter λ , often known as the shrinkage parameter or learning rate that controls the rate the model learns. Recurrent neural networks (RNNs) are a type of NN with recurrent connections and assume that the input data have an ordering, for example, words in a sentence.¹¹⁴⁻¹¹⁶ RNN can be seen as a feed-forward NN with a connection from output to input.¹¹⁵ GRU¹¹⁷ and LSTM¹¹⁸ are improved variations of RNN with gating mechanisms to combat the

vanishing gradient problem. The improvements help the models to better model long-term temporal dependencies. To tune hyperparameters, grid-search and Bayesian optimization were used in the studies.^{47,48,50,53,54} The grid-search method iterates exhaustively through all hyperparameter values within a pre-defined set of values to find the optimal hyperparameter with respect to a validation set. In contrast, the Bayesian optimization method makes informed choices on which values to evaluate using the Bayes formula. The goal of using Bayesian optimization for hyperparameter tuning is to minimize the number of values to evaluate.

All studies reported evaluation results for different algorithms or data types and almost all reported area under the receiver operating characteristic curve (AUC) values except 1.⁴⁸ [Figure 5](#) shows differences in AUC values for infection ([Figure 5A](#)), sepsis ([Figure 5B](#)), septic shock ([Figure 5C](#)), and severe sepsis ([Figure 5E](#)) when using structured data only, text data only, or a combination of structured and text data. Studies that compared their methods for different hours prior to onset are also included ([Figure 5D and F](#)), the lines connecting the points are to visually separate the methods and do not indicate changing AUC values over time. This figure compares data type usage and model performance within an individual study; it should not be used to compare AUC values between subfigures and studies because the studies used different cohorts, sepsis definitions, and hours before onset. Additionally, sepsis, severe sepsis, and septic shock have different manifestations.^{119,120} [Table 4](#) summarizes the best and worst AUC values for each study; a full table with additional evaluation metrics is available in [Supplementary Table S4](#) (number of hours before onset for Amrollahi et al⁵³ was confirmed through personal communications with Shamim Nemati on May 27, 2021 and Fatemeh Amrollahi on June 13, 2021). GBT was the most widely used ML method,^{50-52,55} followed by logistic regression,^{47,48,52} SVMs,^{47,48} RF,^{47,52} ridge regression,⁴⁹ lasso regression,⁵⁴ NB,⁴⁷ KNNs,⁴⁸ GRU,⁵⁰ and LSTM.⁵³ For hyperparameter tuning, 3 studies used the grid-search method^{47,48,54} and 2 used the Bayesian optimization method^{50,53} (hyperparameter tuning was provided by personal communication with Ran Liu on September 7, 2021 and Fatemeh Amrollahi on September 7, 2021). Delahanty et al,⁵¹ Hammoud et al,⁵⁴ Goh et al,⁵² and Qin et al⁵⁵ compared their algorithm to scoring systems used in clinical practice, such as SIRS,¹²¹ sequential organ failure assessment (SOFA),¹²² quick SOFA (qSOFA),¹²³ modified early warning system (MEWS),¹²⁴ or a targeted real-time early warning score (TREWScore).⁷⁸ In addition, Apostolova and Velez⁴⁸ evaluated their model on a ground truth set with 200 nursing notes that were manually reviewed by a qualified professional, and Goh et al⁵² compared their model with the Rhodes et al¹²⁵ sepsis guidelines used by physicians. Furthermore, Horng et al⁴⁷ performed additional tests on different patient cohorts for error analysis. Although results are difficult to compare directly because of study heterogeneity, most results suggest that utilizing both structured data and text generally results in better performance for sepsis identification and early detection.

DISCUSSION

Identification, early detection, prediction, and method transferability

Nine studies utilized clinical text for sepsis identification, early detection, or prediction. As all identified studies focus on the identification or early detection of sepsis within a fixed time frame, this

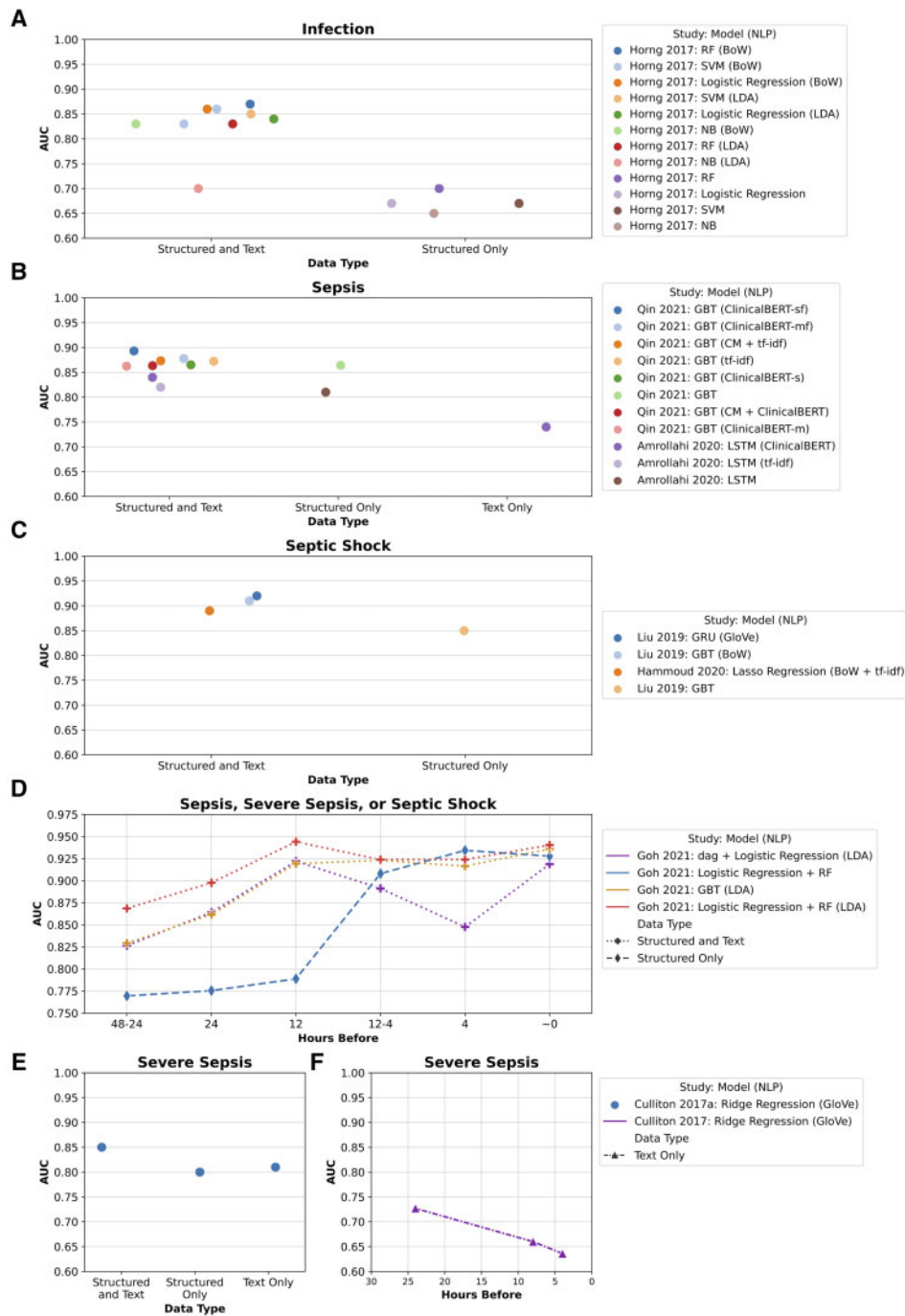


Figure 5. Overview of area under the curve (AUC) values for identification or early detection of infection, sepsis, septic shock, and severe sepsis using different data types (structured data and text, structured data only, and text only).^a Each figure contains the study and year, machine learning model,^a and natural language processing technique.^b (A) AUC values for infection identification. Horng et al⁴⁷ 2017: SVM (BoW) has 2 AUC values; 0.86 when using chief complaints and nursing notes and 0.83 when using only chief complaints. (B) AUC values for early sepsis detection. Amrollahi et al⁵³ AUC values are from detecting 4 h before sepsis onset, and Qin et al⁵⁵ AUC values are the average from detecting 0 to 6 h before sepsis onset. (C) AUC values for early septic shock detection. Hammoud et al⁵⁴ AUC values are from detecting 30.64 h before septic shock onset, and Liu et al⁵⁰ AUC values are from detecting 6.0 to 7.3 h before septic shock onset. (D) AUC values for early sepsis, severe sepsis, or septic shock detection and sepsis identification in Goh et al.⁵² Different symbols separate data types. (E) AUC values for early septic shock detection for Culliton et al⁴⁹ using results from the test set. (F) AUC values for early septic shock detection for Culliton et al⁴⁹ using results from 3-fold validation. ^aDisclaimer: AUC values should not be directly compared between studies and different figures for infection, sepsis, severe sepsis, and septic shock. Additionally, the lines connecting points do not indicate AUC values changing over time (Figure 5D and 5F); lines only separate the different methods visually. ^bMachine learning models: dag: dagging (partition data into disjoint subgroups); GBT: gradient boosted trees; GRU: gated recurrent unit; LSTM: long short-term memory; NB: Naïve Bayes; RF: random forest; SVM: support vector machines. ^bNatural language processing techniques: BoW: Bag-of-words; ClinicalBERT: Clinical Bidirectional Encoder Representations from Transformers; ClinicalBERT-m: ClinicalBERT from merging all textual features to get embeddings; ClinicalBERT-sf: finetuned ClinicalBERT from concatenating individual embeddings of each textual feature; CM: Amazon Comprehend Medical service for named entity recognition; GloVe: Global Vectors for Word Representation; LDA: Latent Dirichlet Allocation; tf-idf: term frequency-inverse document frequency.

Table 4. Study outcome overview of best and worst area under the curve values

| Study (year) | Hours ^a | Data types ^b | | Models ^d (NLP) ^e | AUC ^f | |
|---|----------------------|-------------------------|----------------|--|---------------------------------|------|
| | | DVLMC | T ^c | | | |
| Horng et al. ⁴⁷ (2017) | Identify | DV- - - | CC + NN | RF (BoW) | 0.87 | |
| | | DV- - - | - | NB | 0.65 | |
| Apostolova and Velez ⁴⁸ (2017) | Identify | - - - - - | NN | SVM (BoW + tf-idf) | - | |
| | | - - - - - | NN | Logistic regression + KNN + SVM (PV) | - | |
| Culliton et al. ⁴⁹ (2017) | -4 | - - - - - | CN | Ridge regression (GloVe) | 0.64 | |
| | -8 | - - - - - | CN | Ridge regression (GloVe) | 0.66 | |
| | -24 | - - - - - | CN | Ridge regression (GloVe) | 0.73 | |
| | -24 ^g | -V- -C | CN | Ridge regression (GloVe) | 0.85 | |
| | | -V- -C | - | Ridge regression (GloVe) | 0.80 | |
| Delahanty et al. ⁵¹ (2019) | +1 | -VL- - | - | GBT | 0.93 | |
| | +3 | -VL- - | - | GBT | 0.95 | |
| | +6 | -VL- - | - | GBT | 0.96 | |
| | +12 | -VL- - | - | GBT | 0.97 | |
| | +24 | -VL- - | - | GBT | 0.97 | |
| Liu et al. ⁵⁰ (2019) | -7 | -VLM- | CN | GRU (GloVe) | 0.92 | |
| | -7.3 | -VLM- | CN | GBT (BoW) | 0.91 | |
| | -6 | -VLM- | - | GBT | 0.85 | |
| Amrollahi et al. ⁵³ (2020) | -4 ^h | -VL- - | PN + NN | LSTM (ClinicalBERT) | 0.84 | |
| | | - - - - - | PN + NN | LSTM (ClinicalBERT) | 0.74 | |
| Hammoud et al. ⁵⁴ (2020) | -30.6 | DVL- - | CN | Lasso regression (BoW + tf-idf) | 0.89 | |
| Goh et al. ⁵² (2021) | Identify | DVLM- | PN | Logistic regression + RF (LDA) | 0.94 | |
| | | DVLM- | PN | dag + Logistic regression (LDA) | 0.92 | |
| | | -4 | DVLM- | - | Logistic regression + RF | 0.93 |
| | | | DVLM- | PN | dag + Logistic regression (LDA) | 0.85 |
| | | -6 | DVLM- | PN | Logistic regression + RF (LDA) | 0.92 |
| | | | DVLM- | PN | dag + Logistic regression (LDA) | 0.89 |
| | | -12 | DVLM- | PN | Logistic regression + RF (LDA) | 0.94 |
| | | | DVLM- | - | Logistic regression + RF | 0.79 |
| | | -24 | DVLM- | PN | Logistic regression + RF (LDA) | 0.90 |
| | | | DVLM- | - | Logistic regression + RF | 0.78 |
| | -48 | DVLM- | PN | Logistic regression + RF (LDA) | 0.87 | |
| | | DVLM- | - | Logistic regression + RF | 0.77 | |
| Qin et al. ⁵⁵ (2021) | -6 to 0 ⁱ | -VL- - | CN | GBT (ClinicalBERT-sf) | 0.89 ⁱ | |
| | | -VL- - | - | GBT (ClinicalBERT-m) | 0.86 ⁱ | |

^aHours: Identify: not detecting hours before or after; -: hours before; +: hours after an event.

^bData types: D: demographics; V: vitals; L: laboratory; M: medications; C: codes; T: text; -'s position in DVLMC indicates which is not used.

^cText data types: CC: chief complaints; CN: various types of clinical notes; NN: nursing notes; PN: physician notes; -: no notes.

^dMachine learning models: dag: dagging (partition data into disjoint subgroups); GBT: gradient boosted trees; GRU: gated recurrent unit; KNN: K-nearest neighbors; LSTM: long short-term memory; NB: Naïve Bayes; RF: random forest; SVM: support vector machines.

^eNatural language processing (NLP) techniques: BoW: Bag-of-words; ClinicalBERT: Clinical Bidirectional Encoder Representations from Transformers; ClinicalBERT-m: ClinicalBERT from merging all textual features to get embeddings; ClinicalBERT-sf: finetuned ClinicalBERT from concatenating individual embeddings of each textual feature; GloVe: Global Vectors for Word Representation; LDA: Latent Dirichlet Allocation; PV: paragraph vectors; tf-idf: term frequency-inverse document frequency.

^fArea under the curve (AUC). Apostolova and Velez⁴⁸ did not provide metrics for AUC.

^gCulliton et al.⁴⁹ performed 2 experiments, these results are from using a test set instead of 3-fold validation.

^hNumber of hours before onset for Amrollahi et al.⁵³ was confirmed through personal communications (with Shamim Nemati on May 27, 2021 and Fatemeh Amrollahi on June 13, 2021).

ⁱQin et al.⁵⁵ AUC values are an average from 0 to 6 h before sepsis, not the specified hours.

indicates much work is still needed before sepsis prediction can use text from complete patient histories. Studies from this review focus mainly on the ICU and ED, and the addition of continuous measurements of vital signs for sepsis makes generalizability to the ward units limited. However, Culliton et al.⁴⁹ was successful in detecting sepsis early utilizing only the text from EHR clinical notes, which is a promising approach for all inpatients. Additionally, Horng et al.⁴⁷ showed that their ML model performed on subsets of specific patient cohorts like pneumonia or urinary tract infection. The different ML methods and NLP techniques from each study may be applicable for

different retrospective cohort or case-control studies. Though the studies have varying sepsis definitions, cohorts, ML methods, and NLP techniques, overall, they show that using clinical text and structured data can improve sepsis identification and early detection. Unstructured clinical text predicts sepsis 48–12 h before onset, while structured data predicts sepsis closer to onset (<12 h before).

Sepsis definition impact

In ML, many studies rely heavily on sepsis definitions and ICD-codes to identify patient cohort datasets for sepsis studies.^{9,11,13}

Among changing sepsis definitions over time are the 2001 Angus Sepsis ICD-9 abstraction criteria,⁷⁹ 2012 Surviving Sepsis Campaign Guidelines,¹²⁶ 2016 Sepsis-3 consensus definition,¹ and 2017 Rhee's modified Sepsis-3 definition.⁸⁰ Although a consensus sepsis definition exists,¹ not all definition elements will be present in a sepsis patient because sepsis is a very heterogeneous syndrome¹²⁷ and the infection site is difficult to identify correctly.¹²⁸ Many patients with sepsis are often misdiagnosed with other diseases such as respiratory failure¹²⁹ and pneumonia.^{129,130} In practice, hospitals also have varying sepsis coding methods.^{131–135} As the sepsis definitions change, studies also tend to use the most current definition in their study. A recent study that used different sepsis definitions to generate patient cohorts found significant heterogeneous characteristics and clinical outcomes between cohorts.¹³⁶ Similarly, previous work by Liu et al¹³⁷ demonstrated that using different infection criteria resulted in a different number of patients and slightly different outcomes. Similar to how changes in the definition and varying coding methods can affect sepsis mortality outcomes,¹³⁸ the sepsis definition and codes used in ML studies will likely change the outcome, results, and reporting methods. Thus, future studies should acknowledge that sepsis is a syndrome and clearly characterize each sign of sepsis to reflect the heterogeneity in the definition.

Suggestions for future studies

Predicting sepsis earlier than 12 h prior to sepsis onset can reduce treatment delays and improve patient outcomes.^{3,4} Because predictions 48–12 h before sepsis onset appear to rely more on clinical text than structured data, additional NLP techniques should be considered for future ML studies. Additionally, since the sepsis definition used will change the cohort, this indicates opportunities to expand the cohort. Like Apostolova and Velez,⁴⁸ who determined their cohort by finding notes describing the use of antibiotics. It should be possible to determine cohorts by using notes describing infection signs (eg, fever, hypotension, or deterioration in mental status), indicators of diseases that sepsis is misdiagnosed with (eg, pulmonary embolism, adrenal insufficiency, diabetic ketoacidosis, pancreatitis, anaphylaxis, bowel obstruction, hypovolemia, colitis, or vasculitis), or medication effect and toxin ingestion, overdose, or withdrawal.¹³⁹ NLP methods from infectious diseases known to trigger sepsis can be incorporated to extract infection signs and symptoms from the text for determining potential sepsis signs, patient groups, and risk factors. For instance, many sepsis patients are often admitted with pneumonia, and there are several studies about identifying pneumonia from radiology reports using NLP.^{23,140,141} Additionally, heterogeneous sepsis signs or symptoms might be identified by utilizing NLP features for detecting healthcare-associated infections risk patterns⁵⁹ or infectious symptoms.¹⁴² Information from other NLP related reviews about using clinical notes can also be applied, such as: challenges to consider,¹⁶ clinical information extraction tools and methods,¹⁸ methods to overcome the need for annotated data,²² different embedding techniques,^{143,144} sources of labeled corpora,¹⁴³ transferability of methods,¹⁴⁵ and processing and analyzing symptoms.¹⁴⁶ Moreover, heterogeneous or infectious diseases, with overlapping signs and symptoms of other diseases, can utilize similar sepsis ML and NLP methods to improve detection. The identified studies did not utilize complete patient history data. Thus, future research utilizing complete patient history data can study if sepsis risk can be predicted earlier than 48 h by incorporating sepsis risk factors, such as comorbidities,⁷ chronic diseases,¹⁴⁷ patient trajectories,¹⁴⁸ or prior infection incidents.¹⁴⁹

Limitations

This review has several limitations. The narrow scope of including only studies about utilizing clinical text for sepsis detection or prediction could have missed studies that use other types of text for sepsis detection or prediction. For example, search terms did not include “early warning system,” “feature extraction,” and “topic modeling.” Additionally, search terms did not include possible sources of infection for sepsis, such as bloodstream infection, catheter-associated infection, pneumonia, and postoperative surgical complications. Further, the sensitivity to detect sepsis in text, structured data, or the combined data from these will depend on the timestamps these data recordings have in the EHR. These timestamps may vary depending on the data used to inform the study or the different systems implemented at different hospitals. The articles identified in this review had a homogenous choice of structured data (ie, demographics, vital signs, and laboratory measurements). Of those, laboratory test results have the largest time lag, around 1–2 h to obtain the blood test results.¹⁵⁰ Thus, the good performance of text to detect sepsis in these articles are unlikely explained fully by the time lag between measurement and recording of the structured data. This review thus shows that it is possible to detect sepsis early using text, with or without the addition of structured data.

CONCLUSION

Many studies about sepsis detection exist, but very few studies utilize clinical text. Heterogeneous study characteristics made it difficult to compare results; however, the consensus from most studies was that combining structured data with clinical text improves identification and early detection of sepsis. There is a need to utilize the unstructured text in EHR data to create early detection models for sepsis. The lack of utilizing the complete patient history in early prediction models for sepsis is an opportunity for future ML and NLP studies.

FUNDING

Financial support for this study was provided by the Computational Sepsis Mining and Modelling project through the Norwegian University of Science and Technology Health Strategic Area.

AUTHOR CONTRIBUTIONS

MYY and ØN conceptualized the study and design with substantial clinical insight from LTG. MY Y conducted the literature search and initial analysis, LTG verified results, and ØN resolved discrepancies. All authors participated in data analysis and interpretation. MY Y drafted the manuscript, which LTG and ØN critically revised.

SUPPLEMENTARY MATERIAL

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

ACKNOWLEDGMENTS

We thank those from the Gemini Center for Sepsis Research group for valuable discussions and recommendations related to clinical databases, missing search terms, and presenting results. Specifically, Ms Lise Husby Høvik (RN), Dr Erik Solligård, Dr Jan Kristian Damås, Dr Jan Egil Afset, Dr Kristin Vardheim Liyanarachi, Dr Randi Marie Mohus, and Dr Anuradha Ravi.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

The data underlying this article are available in the article and in its online [supplementary material](#).

REFERENCES

- Singer M, Deutschman CS, Seymour CW, *et al*. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315 (8): 801–10.
- Fleischmann C, Scherag A, Adhikari NKJ, *et al*.; International Forum of Acute Care Trialists. Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. *Am J Respir Crit Care Med* 2016; 193 (3): 259–72.
- Rivers E, Nguyen B, Havstad S, *et al*. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 2001; 345 (19): 1368–77.
- Kumar A, Roberts D, Wood KE, *et al*. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* 2006; 34: 1589–96.
- Polat G, Ugan RA, Cadirci E, *et al*. Sepsis and septic shock: current treatment strategies and new approaches. *Eurasian J Med* 2017; 49 (1): 53–8.
- Arnold C. News feature: the quest to solve sepsis. *Proc Natl Acad Sci USA* 2018; 115 (16): 3988–91.
- Iskander KN, Osuchowski MF, Stearns-Kurosawa DJ, *et al*. Sepsis: multiple abnormalities, heterogeneous responses, and evolving understanding. *Physiol Rev* 2013; 93 (3): 1247–88.
- Jawad I, Lukšić I, Rafnsson SB. Assessing available information on the burden of sepsis: global estimates of incidence, prevalence and mortality. *J Glob Health* 2012; 2 (1): 010404.
- Islam MM, Nasrin T, Walther BA, *et al*. Prediction of sepsis patients using machine learning approach: a meta-analysis. *Comput Methods Programs Biomed* 2019; 170: 1–9.
- Schinkel M, Paranjape K, Nannan Panday RS, *et al*. Clinical applications of artificial intelligence in sepsis: a narrative review. *Comput Biol Med* 2019; 115: 103488.
- Wulff A, Montag S, Marschollek M, *et al*. Clinical decision-support systems for detection of systemic inflammatory response syndrome, sepsis, and septic shock in critically ill patients: a systematic review. *Methods Inf Med* 2019; 58 (S 02): e43–57.
- Teng AK, Wilcox AB. A review of predictive analytics solutions for sepsis patients. *Appl Clin Inform* 2020; 11 (3): 387–98.
- Fleuren LM, Klausch TLT, Zwager CL, *et al*. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* 2020; 46 (3): 383–400.
- Giacobbe DR, Signori A, Del Puente F, *et al*. Early detection of sepsis with machine learning techniques: a brief clinical perspective. *Front Med (Lausanne)* 2021; 8: 617486.
- Assale M, Dui LG, Cina A, *et al*. The revival of the notes field: leveraging the unstructured content in electronic health records. *Front Med (Lausanne)* 2019; 6: 66.
- Tayefi M, Ngo P, Chomutare T, *et al*. Challenges and opportunities beyond structured data in analysis of electronic health records. *Wiley Interdiscip Rev Comput Stat* 2021; 13: e1549. doi:10.1002/wics.1549
- Sheikhalishahi S, Miotto R, Dudley JT, *et al*. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform* 2019; 7 (2): e12239.
- Wang Y, Wang L, Rastegar-Mojarad M, *et al*. Clinical information extraction applications: a literature review. *J Biomed Inform* 2018; 77: 34–49.
- Datta S, Bernstam EV, Roberts K. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. *J Biomed Inform* 2019; 100: 103301.
- Jackson RG, Patel R, Jayatilleke N, *et al*. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. *BMJ Open* 2017; 7 (1): e012012.
- Kreimeyer K, Foster M, Pandey A, *et al*. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *J Biomed Inform* 2017; 73: 14–29.
- Spasic I, Nenadic G. Clinical text data in machine learning: systematic review. *JMIR Med Inform* 2020; 8 (3): e17984.
- Elkin PL, Froehling D, Wahner-Roedler D, *et al*. NLP-based identification of pneumonia cases from free-text radiological reports. *AMIA Annu Symp Proc* November 8–12, 2008; 2008: 172–6; Washington, DC.
- Jensen K, Soguero-Ruiz C, Oyvind Mikalsen K, *et al*. Analysis of free text in electronic health records for identification of cancer patient trajectories. *Sci Rep* 2017; 7: 46226.
- Ford E, Carroll JA, Smith HE, *et al*. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc* 2016; 23 (5): 1007–15.
- Soguero-Ruiz C, Hindberg K, Mora-Jiménez I, *et al*. Predicting colorectal surgical complications using heterogeneous clinical data and kernel methods. *J Biomed Inform* 2016; 61: 87–96.
- Huddar V, Desiraju BK, Rajan V, *et al*. Predicting complications in critical care using heterogeneous clinical data. *IEEE Access* 2016; 4: 7988–8001.
- Ribelles N, Jerez JM, Rodriguez-Brazzazola P, *et al*. Machine learning and natural language processing (NLP) approach to predict early progression to first-line treatment in real-world hormone receptor-positive (HR+)HER2-negative advanced breast cancer patients. *Eur J Cancer* 2021; 144: 224–31.
- Friedlin J, Overhage M, Al-Haddad MA, *et al*. Comparing methods for identifying pancreatic cancer patients using electronic data sources. *AMIA Annu Symp Proc* November 13–17, 2010; 2010: 237–41; Washington, DC.
- Van Vleck TT, Chan L, Coca SG, *et al*. Augmented intelligence with natural language processing applied to electronic health records for identifying patients with non-alcoholic fatty liver disease at risk for disease progression. *Int J Med Inform* 2019; 129: 334–41.
- DeLisle S, Kim B, Deepak J, *et al*. Using the electronic medical record to identify community-acquired pneumonia: toward a replicable automated strategy. *PLoS One* 2013; 8 (8): e70944.
- Gundlapalli AV, South BR, Phansalkar S, *et al*. Application of natural language processing to VA electronic health records to identify phenotypic characteristics for clinical and research purposes. *Summit Transl Bioinform* March 10–12, 2008; 2008: 36–40; San Francisco, CA.
- Ananthakrishnan AN, Cai T, Savova G, *et al*. Improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. *Inflamm Bowel Dis* 2013; 19 (7): 1411–20.
- Carroll RJ, Thompson WK, Eyler AE, *et al*. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. *J Am Med Inform Assoc* 2012; 19 (e1): e162–9.
- Liao KP, Cai T, Gainer V, *et al*. Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2010; 62 (8): 1120–7.
- Carroll RJ, Eyler AE, Denny JC. Naïve electronic health record phenotype identification for Rheumatoid arthritis. *AMIA Annu Symp Proc* October 22–26, 2011; 2011: 189–96; Washington, DC.
- Xia Z, Secor E, Chibnik LB, *et al*. Modeling disease severity in multiple sclerosis using electronic health records. *PLoS One* 2013; 8 (11): e78927.
- DeLisle S, South B, Anthony JA, *et al*. Combining free text and structured electronic medical record entries to detect acute respiratory infections. *PLoS One* 2010; 5 (10): e13377.
- Zheng H, Gaff H, Smith G, *et al*. Epidemic surveillance using an electronic medical record: an empiric approach to performance improvement. *PLoS One* 2014; 9 (7): e100845.

40. Bhattacharjee P, Edelson DP, Churpek MM. Identifying patients with sepsis on the hospital wards. *Chest* 2017; 151 (4): 898–907.
41. Despina LA. Automated detection of sepsis using electronic medical record data: a systematic review. *J Healthc Qual* 2017; 39 (6): 322–33.
42. Peiffer-Smadja N, Rawson TM, Ahmad R, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin Microbiol Infect* 2020; 26 (5): 584–95.
43. Freeman R, Moore LSP, García Álvarez L, et al. Advances in electronic surveillance for healthcare-associated infections in the 21st century: a systematic review. *J Hosp Infect* 2013; 84 (2): 106–19.
44. de Bruin JS, Seeling W, Schuh C. Data use and effectiveness in electronic surveillance of healthcare associated infections in the 21st century: a systematic review. *J Am Med Inform Assoc* 2014; 21 (5): 942–51.
45. Luz CF, Vollmer M, Decruyenaere J, et al. Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies. *Clin Microbiol Infect* 2020; 26 (10): 1291–9.
46. Moher D, Liberati A, Tetzlaff J, et al.; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; 6 (7): e1000097.
47. Horng S, Sontag DA, Halpern Y, et al. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS One* 2017; 12 (4): e0174708.
48. Apostolova E, Velez T. Toward automated early sepsis alerting: identifying infection patients from nursing notes. In: BioNLP 2017. Association for Computational Linguistics August 4, 2017: 257–62; Vancouver, Canada. doi:10.18653/v1/W17-2332
49. Culliton P, Levinson M, Ehresman A, et al. Predicting Severe Sepsis Using Text from the Electronic Health Record. In: Workshop on Machine Learning For Health at the Conference on Neural Information Processing Systems (NIPS ML4H 2017), December 8, 2017. arXiv: 1711.11536; Long Beach, CA.
50. Liu R, Greenstein JL, Sarma SV, et al. Natural language processing of clinical notes for improved early prediction of septic shock in the ICU. *Annu Int Conf IEEE Eng Med Biol Soc* July 23–27, 2019; 2019: 6103–8; Berlin, Germany. doi:10.1109/EMBC.2019.8857819
51. Delahanty RJ, Alvarez J, Flynn LM, et al. Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. *Ann Emerg Med* 2019; 73 (4): 334–44.
52. Goh KH, Wang L, Yeow AYK, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat Commun* 2021; 12 (1): 711.
53. Amrollahi F, Shashikumar SP, Razmi F, et al. Contextual embeddings from clinical notes improves prediction of sepsis. *AMIA Annu Symp Proc* November 14–18, 2020; 2020: 197–202; Virtual event.
54. Hammoud I, Ramakrishnan IV, Henry M, et al. Multimodal early septic shock prediction model using lasso regression with decaying response. In: 2020 IEEE International Conference on Healthcare Informatics (ICHI). IEEE November 30–December 3, 2020; Virtual conference in Oldenburg, Germany. doi:10.1109/ichi48887.2020.9374377
55. Qin F, Madan V, Ratan U, et al. Improving early sepsis prediction with multi modal learning, 2021. arXiv: 2107.11094
56. Bouam S, Girou E, Brun-Buisson C, et al. An intranet-based automated system for the surveillance of nosocomial infections: prospective validation compared with physicians' self-reports. *Infect Control Hosp Epidemiol* 2003; 24 (1): 51–5.
57. Koller W, Blacky A, Bauer C, et al. Electronic surveillance of healthcare-associated infections with MONI-ICU—a clinical breakthrough compared to conventional surveillance systems. *Stud Health Technol Inform* 2010; 160 (Pt 1): 432–6.
58. Penz JFE, Wilcox AB, Hurdle JF. Automated identification of adverse events related to central venous catheters. *J Biomed Inform* 2007; 40 (2): 174–82.
59. Proux D, Marchal P, Segond F, et al. Natural language processing to detect risk patterns related to hospital acquired infections. In: Proceedings of the Workshop on Biomedical Information Extraction: Association for Computational Linguistics; September 18, 2009: 35–41; Borovets, Bulgaria. doi:10.5555/1859776.1859782
60. Bouzbid S, Gicquel Q, Gerbier S, et al. Automated detection of nosocomial infections: evaluation of different strategies in an intensive care unit 2000-2006. *J Hosp Infect* 2011; 79 (1): 38–43.
61. Jo Y, Loghmanpour N, Rosé CP. Time series analysis of nursing notes for mortality prediction via a state transition topic model. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM). Association for Computing Machinery October 18–23, 2015: 1171–80; Melbourne, Australia. doi:10.1145/2806416.2806541
62. Wang T, Velez T, Apostolova E, et al. Semantically Enhanced Dynamic Bayesian Network for Detecting Sepsis Mortality Risk in ICU Patients with Infection, 2018. arXiv: 1806.10174
63. Baghdadi Y, Bourrée A, Robert A, et al. Automatic classification of free-text medical causes from death certificates for reactive mortality surveillance in France. *Int J Med Inform* 2019; 131: 103915.
64. Guo W, Xu Z, Ye X, et al. A time-critical topic model for predicting the survival time of sepsis patients. *Sci Program* 2020; 2020: 1. doi:10.1155/2020/8884539
65. Ribas Ripoll VJ, Vellido A, Romero E, et al. Sepsis mortality prediction with the Quotient Basis Kernel. *Artif Intell Med* 2014; 61: 45–52.
66. Halpern Y, Horng S, Choi Y, et al. Electronic medical record phenotyping using the anchor and learn framework. *J Am Med Inform Assoc* 2016; 23 (4): 731–40.
67. Vilic A, Petersen JA, Hoppe K, et al. Visualizing patient journals by combining vital signs monitoring and natural language processing. *Annu Int Conf IEEE Eng Med Biol Soc* August 16–20, 2016; 2016: 2529–32; Orlando, FL. doi:10.1109/EMBC.2016.7591245
68. Zhu X, Plasek JM, Tang C, et al. Embedding, aligning and reconstructing clinical notes to explore sepsis. *BMC Res Notes* 2021; 14 (1): 136.
69. Marks M, Pollock E, Armstrong M, et al. Needles and the damage done: reasons for admission and financial costs associated with injecting drug use in a Central London Teaching Hospital. *J Infect* 2013; 66 (1): 95–102.
70. Ippolito P, Larson EL, Furuya EY, et al. Utility of electronic medical records to assess the relationship between parenteral nutrition and central line-associated bloodstream infections in adult hospitalized patients. *JPEN J Parenter Enteral Nutr* 2015; 39 (8): 929–34.
71. Liu YZ, Chu R, Lee A, et al. A surveillance method to identify patients with sepsis from electronic health records in Hong Kong: a single centre retrospective study. *BMC Infect Dis* 2020; 20 (1): 652.
72. Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA* 2011; 306 (8): 848–55.
73. FitzHenry F, Murff HJ, Matheny ME, et al. Exploring the frontier of electronic health record surveillance: the case of postoperative complications. *Med Care* 2013; 51 (6): 509–16.
74. Piscitelli A, Bevilacqua L, Labella B, et al. A keyword approach to identify adverse events within narrative documents from 4 Italian institutions. *J Patient Saf*. doi:10.1097/PTS.0000000000000783
75. Vermassen J, Colpaert K, De Bus L, et al. Automated screening of natural language in electronic health records for the diagnosis septic shock is feasible and outperforms an approach based on explicit administrative codes. *J Crit Care* 2020; 56: 203–7.
76. Saeed M, Villarreal M, Reisner AT, et al. Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database. *Crit Care Med* 2011; 39 (5): 952–60.
77. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; 3: 160035.
78. Henry KE, Hager DN, Pronovost PJ, et al. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* 2015; 7 (299): 299ra122.
79. Angus DC, Linde-Zwirble WT, Lidicker J, et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med* 2001; 29 (7): 1303–10.

80. Rhee C, Dantes R, Epstein L, *et al.*; CDC Prevention Epicenter Program. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. *JAMA* 2017; 318 (13): 1241-9.
81. Reyna MA, Josef C, Seyedi S, *et al.* Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. In: 2019 Computing in Cardiology (CinC) September 8-11, 2019: 1-4; Singapore. doi:10.23919/CinC49843.2019.9005736
82. Wagner MM, Hogan WR, Chapman WW, *et al.* Chief complaints and ICD codes. *Handbook of Biosurveillance* 2006: 333.
83. Lucian J, Davis JFM. History and physical examination. *Murray and Nadel's Textbook of Respiratory Medicine* 2016: 263. doi:10.1016/B978-1-4557-3383-5.00016-6
84. Aghili H, Mushlin RA, Williams RM, *et al.* Progress notes model. In: Proceedings of AMIA Annual Fall Symposium October 25-29, 1997: 12-6; Nashville, TN.
85. Hellesø R. Information handling in the nursing discharge note. *J Clin Nurs* 2006; 15 (1): 11-21.
86. Ioanna P, Stilianoi K, Vasiliki B. Nursing documentation and recording systems of nursing care. *Health Sci J* 2007; (4): 7. <https://www.hsj.gj/medicine/nursing-documentation-and-recording-systems-of-nursing-care.php?aid=3680> Accessed June 29, 2021.
87. Flink M, Ekstedt M. Planning for the discharge, not for patient self-management at home - an observational and interview study of hospital discharge. *Int J Integr Care* 2017; 17 (6): 1.
88. Stopford E, Ninan S, Spencer N. How to write a discharge summary. *BMJ* 2015; 351: h2696.
89. Sørby ID, Nytrø Ø. Does the electronic patient record support the discharge process? A study on physicians' use of clinical information systems during discharge of patients with coronary heart disease. *Health Inf Manag* 2006; 34 (4): 112-9.
90. Kind AJH, Smith MA. Documentation of mandated discharge summary components in transitions from acute to subacute care. In: Henriksen K, Battles JB, Keyes MA, *et al.*, eds. *Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 2: Culture and Redesign)*. AHRQ Publication No. 08-0034-2. Rockville, MD: Agency for Healthcare Research and Quality (US); 2011.
91. Jensdóttir A-B, Jónsson P, Noro A, *et al.* Comparison of nurses' and physicians' documentation of functional abilities of older patients in acute care-patient records compared with standardized assessment. *Scand J Caring Sci* 2008; 22 (3): 341-7.
92. Penoyer DA, Cortelyou-Ward KH, Noblin AM, *et al.* Use of electronic health record documentation by healthcare workers in an acute care hospital system. *J Healthc Manag* 2014; 59: 130-44.
93. Furniss D, Lyons I, Franklin BD, *et al.* Procedural and documentation variations in intravenous infusion administration: a mixed methods study of policy and practice across 16 hospital trusts in England. *BMC Health Serv Res* 2018; 18 (1): 270.
94. Sohn S, Wang Y, Wi C-I, *et al.* Clinical documentation variations and NLP system portability: a case study in asthma birth cohorts across institutions. *J Am Med Inform Assoc* 2018; 25 (3): 353-9.
95. McPherson K. International differences in medical care practices. *Health Care Financ Rev* 1989; 1989: 9-20.
96. Georgopoulos BS. Organization structure and the performance of hospital emergency services. *Ann Emerg Med* 1985; 14 (7): 677-84.
97. McKinlay J, Link C, Marceau L, *et al.* How do doctors in different countries manage the same patient? Results of a factorial experiment. *Health Serv Res* 2006; 41 (6): 2182-200.
98. Sørby ID, Nytrø Ø. Analysis of communicative behaviour: profiling roles and activities. *Int J Med Inform* 2010; 79 (6): e144-51-e151.
99. Røst TB, Tvedt CR, Husby H, *et al.* Identifying catheter-related events through sentence classification. *Int J Data Min Bioinform* 2020; 23 (3): 213-33.
100. Harris ZS. Distributional Structure. *Word World* 1954; 10 (2-3): 146-62.
101. Le Q, Mikolov T. Distributed representations of sentences and documents. In: *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. JMLR.org June 21-26, 2014: II-1188-96; Beijing, China. doi:10.5555/3044805.3045025
102. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003; 3: 993-1022. doi:10.5555/944919.944937
103. Mikolov T, Sutskever I, Chen K, *et al.* Distributed Representations of Words and Phrases and their Compositionality. 2013. arXiv:1310.4546
104. Pennington J, Socher R, Manning C. GloVe: Global Vectors for Word Representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics October 25-29, 2014: 1532-43; Doha, Qatar. doi:10.3115/v1/D14-1162
105. Mikolov T, Chen K, Corrado G, *et al.* Efficient Estimation of Word Representations in Vector Space. 2013. arXiv: 1301.3781
106. Devlin J, Chang M-W, Lee K, *et al.* BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics June 2-7, 2019: 4171-86; Minneapolis, MN. doi:10.18653/v1/N19-1423
107. Alsentzer E, Murphy J, Boag W, *et al.* Publicly available clinical BERT Embeddings. In: Proceedings of the 2nd Clinical Natural Language Processing Workshop. Association for Computational Linguistics June 7, 2019: 72-8; Minneapolis, MN. doi:10.18653/v1/W19-1909
108. Koller D, Friedman N. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA and London, England: MIT Press, 2009.
109. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995; 20 (3): 273-97.
110. Fix E, Hodges JL. Nonparametric discrimination: consistency properties. *Randolph Field, Texas, Project* 1951: 21-49.
111. Breiman L. Random forests. *Mach Learn* 2001; 45 (1): 5-32.
112. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 1999; 29: 1189-232. <http://www.jstor.org/stable/2699986> Accessed May 3, 2021.
113. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* 2002; 38 (4): 367-78.
114. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986; 323 (6088): 533-6.
115. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation In: Rumelhart DE, McClelland JL, eds. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. MIT Press 1986: 318-62. <https://ieeexplore.ieee.org/document/6302929> Accessed May 3, 2021.
116. Mikolov T, Karafiat M, Burget L, *et al.* Recurrent neural network based language model. In: *INTERSPEECH* 2010. ISCA September 26-30, 2010; Makuhari, Chiba, Japan. doi: 10.21437/Interspeech.2010-343
117. Cho K, van Merriënboer B, Gulcehre C, *et al.* Learning phrase representations using RNN Encoder-Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics October 25-29, 2014: 1724-34; Doha, Qatar. doi:10.3115/v1/D14-1179
118. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997; 9 (8): 1735-80.
119. Cawcutt KA, Peters SG. Severe sepsis and septic shock: clinical overview and update on management. *Mayo Clin Proc* 2014; 89 (11): 1572-8.
120. Hotchkiss RS, Moldawer LL, Opal SM, *et al.* Sepsis and septic shock. *Nat Rev Dis Primers* 2016; 2: 16045.
121. Bone RC, Balk RA, Cerra FB, *et al.* Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992; 101 (6): 1644-55.
122. Vincent JL, Moreno R, Takala J, *et al.* The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996; 22 (7): 707-10.

123. Seymour CW, Liu VX, Iwashyna TJ, *et al.* Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315 (8): 762–74.
124. Subbe CP, Kruger M, Rutherford P, *et al.* Validation of a modified early warning score in medical admissions. *QJM* 2001; 94 (10): 521–6.
125. Rhodes A, Evans LE, Alhazzani W, *et al.* Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive Care Med* 2017; 43 (3): 304–77.
126. Dellinger RP, Levy MM, Rhodes A, *et al.*; Surviving Sepsis Campaign Guidelines Committee including the Pediatric Subgroup. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med* 2013; 41 (2): 580–637.
127. Vincent J-L. The clinical challenge of sepsis identification and monitoring. *PLoS Med* 2016; 13 (5): e1002022.
128. Abe T, Tokuda Y, Shiraishi A, *et al.*; JAAM SPICE Study Group. In-hospital mortality associated with the misdiagnosis or unidentified site of infection at admission. *Crit Care* 2019; 23 (1): 202.
129. Rothberg MB, Pekow PS, Priya A, *et al.* Variation in diagnostic coding of patients with pneumonia and its association with hospital risk-standardized mortality rates: a cross-sectional analysis. *Ann Intern Med* 2014; 160 (6): 380–8.
130. Bewick T, Simmonds M, Chikhani M, *et al.* Pneumonia in the context of severe sepsis: a significant diagnostic problem. *Eur Respir J* 2008; 32 (5): 1417–8.
131. Rhee C, Murphy MV, Li L, *et al.*; for the Centers for Disease Control and Prevention Epicenters Program. Comparison of trends in sepsis incidence and coding using administrative claims versus objective clinical data. *Clin Infect Dis* 2015; 60 (1): 88–95.
132. Tidswell R, Inada-Kim M, Singer M. Sepsis: the importance of an accurate final diagnosis. *Lancet Respir Med* 2021; 9 (1): 17–8.
133. Lopansri BK, Miller RR III, Burke JP, *et al.* Physician agreement on the diagnosis of sepsis in the intensive care unit: estimation of concordance and analysis of underlying factors in a multicenter cohort. *J Intensive Care* 2019; 7: 13.
134. Rhee C, Kadri SS, Danner RL, *et al.* Diagnosing sepsis is subjective and highly variable: a survey of intensivists using case vignettes. *Crit Care* 2016; 20: 89.
135. Rhee C, Klompas M. Sepsis trends: increasing incidence and decreasing mortality, or changing denominator? *J Thorac Dis* 2020; 12 (Suppl 1): S89–100.
136. Yu SC, Betthausen KD, Gupta A, *et al.* Comparison of sepsis definitions as automated criteria. *Crit Care Med* 2021; 49 (4): e433–43.
137. Liu R, Greenstein JL, Granite SJ, *et al.* Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU. *Sci Rep* 2019; 9 (1): 6145.
138. Walkey AJ, Shieh M-S, Liu VX, *et al.* Mortality measures to profile hospital performance for patients with septic shock. *Crit Care Med* 2018; 46 (8): 1247–54.
139. Vincent J-L. The challenge of early identification of the hospital patient at risk of septic complications. *Ann Transl Med* 2017; 5 (3): 56.
140. Mendonça EA, Haas J, Shagina L, *et al.* Extracting information on pneumonia in infants using natural language processing of radiology reports. *J Biomed Inform* 2005; 38 (4): 314–21.
141. Dublin S, Baldwin E, Walker RL, *et al.* Natural language processing to identify pneumonia from radiology reports. *Pharmacoepidemiol Drug Saf* 2013; 22 (8): 834–41.
142. Matheny ME, Fitzhenry F, Speroff T, *et al.* Detection of infectious symptoms from VA emergency department and primary care clinical documentation. *Int J Med Inform* 2012; 81 (3): 143–56.
143. Wu S, Roberts K, Datta S, *et al.* Deep learning in clinical natural language processing: a methodical review. *J Am Med Inform Assoc* 2020; 27 (3): 457–70.
144. Khattak FK, Jebblee S, Pou-Prom C, *et al.* A survey of word embeddings for clinical text. *J Biomed Informatics: X* 2019; 100: 100057.
145. Névéol A, Zweigenbaum P; Section Editors for the IMIA Yearbook Section on Clinical Natural Language Processing. Expanding the diversity of texts and applications: findings from the section on clinical natural language processing of the international medical informatics association yearbook. *Yearb Med Inform* 2018; 27 (1): 193–8.
146. Koleček TA, Dreisbach C, Bourne PE, *et al.* Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc* 2019; 26 (4): 364–79.
147. Wang HE, Shapiro NI, Griffin R, *et al.* Chronic medical conditions and risk of sepsis. *PLoS One* 2012; 7 (10): e48307.
148. Prescott HC, Carmichael AG, Langa KM, *et al.* Paths into sepsis: trajectories of pre-sepsis healthcare use. *Ann Am Thorac Soc* 2019; 16 (1): 116–23.
149. Delano MJ, Ward PA. The immune system's role in sepsis progression, resolution, and long-term outcome. *Immunol Rev* 2016; 274 (1): 330–53.
150. Faisal M, Scally A, Richardson D, *et al.* Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results. *Crit Care Med* 2018; 46 (4): 612–8.