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

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. 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.

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). 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. 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.

Multiple studies and a review have shown that using unstructured clinical text has increased model performance to detect or predict colorectal surgical complications, postoperative acute respiratory failure, breast cancer, pancreatic cancer, fatty liver disease, pneumonia, inflammatory bowel disease, rheumatoid arthritis, pneumonia, inflammatory bowel disease, and acute respiratory infection.

Prior reviews related to sepsis detection and prediction include: sepsis detection using Systemic Inflammatory Response Syndrome (SIRS) screening tools, sepsis detection using SIRS and organ dysfunction criteria with EHR vital signs and laboratory data, clinical perspectives on the use of ML for early detection of sepsis in daily practice, ML for diagnosis and early detection of sepsis patients, infectious disease clinical decision support, and healthcare-associated infections mentioning sepsis. 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.

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, and the second search uncovered 2 more articles that met inclusion criteria (6 of 1335 articles). 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.

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.

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 articles from 2 relevant review articles (Figure 1). From the 1817 unique articles, 1620 articles were excluded based on eligibility criteria described in the methods. After assessing the
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,78 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, 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)47,50,52–55 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-II74 and 4 utilized MIMIC-III69,50,53,55 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 States47–51,53–55 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, and another48 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 specialists49–51,54,55 to document symptoms,
| Study (year) | Clinical setting and data source | Sample size* | Cohort criteria infection definition | Task and objective |
|-------------|---------------------------------|--------------|--------------------------------------|-------------------|
| Horng et al.47 (2017)  • ED  • Beth Israel Deaconess (Boston, MA, United States)  • Dec 17, 2008—Feb 17, 2013 | 230,936 patient visits  • 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 criteria79 | Identify patients with suspected infection to demonstrate benefits of using clinical text with structured data for detecting ED patients with suspected infection. |
| Apostolova and Velez48 (2017)  • ICU  • MIMIC-III  • 2001–2012 | 634,369 nursing notes  • 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)  • Inpatient care  • Baystate hospitals (Springfield, MA, United States)  • 2012–2016 | 203,000 adult inpatient admission encounters  • 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:  • Used 13,603 E  • Severe sepsis: 425 P; 3.1%  Train: 70%  Test: 30% | 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)  • ED  • Tenet Healthcare Hospitals (Nashville, TN, United States)  • January 1, 2016—October 31, 2017 | 2,759,529 patient encounters  • Sepsis: 54,661 E; 2%  • No Sepsis: 2,704,868 E; 98%  Train: 1,839,503 E; 66.7%  • Sepsis: 36,458 E; 2%  • No sepsis: 1,803,045 E; 98%  Test: 920,026 E; 33.3%  • Sepsis: 18,203 E; 2%  • No sepsis: 901,823 E; 98% | Rhee’s modified Sepsis-3 definition80 | 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)  • ICU  • MIMIC-III  • 2001–2012 | 38,645 adult patients  • Sepsis: 793 P; 2%  • No Sepsis: 37,852 P; 98%  Train: 70% P  Test: 30% P  Applied model to: 15,930 P with suspected infection and at least 1 physiological EHR data  | Sepsis-3 definition1 | 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)  • ICU  • MIMIC-III  • 2001–2012 | 40,175 adult patients  • Sepsis: 2,805 P; ~7%  • No Sepsis: 37,370 P; 93%  Train: 80% P  Test: 20% P | Sepsis-3 definition1 | 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. |
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 management or performance reference or prediction.89 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,47,50,52,53,55 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 model that can be optimized based on user preference or performance metrics. Studies have shown that nursing documentation differs from physician documentation.89,91

### Table 1. continued

| Study (year) | Clinical setting and data source | Sample size<sup>a</sup> | Cohort criteria infection definition | Task and objective |
|--------------|---------------------------------|--------------------------|--------------------------------------|-------------------|
| Hammoud et al.54 (2020) | • ICU | 17 763 patients | Sepsis definition based on what Henry et al30 used | Predict early septic shock in ICU patients using a model that can be optimized based on user preference or performance metrics. |
| | • MIMIC-II | | | |
| | • 2001–2007 | 6097 P | | |
| | • Severe sepsis: 3962 P | | | |
| | • Septic shock: 1469 P | 5-fold cross validation | | |
| | Goh et al.52 (2021) | • ICU | 5317 patients (114 602 notes) | 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. |
| | • Singapore government-based hospital (Singapore, Singapore) | Train and validation: 3722 P (80 162 N) | | |
| | • Apr 2, 2015—Dec 31, 2017 | Sepsis: 6.45% | | |
| | | No sepsis: 93.55% | | |
| | | Test: 1595 P (34 440 N) | | |
| | | Sepsis: 5.45% | | |
| | | No sepsis: 94.55% | | |
| | Qin et al.55 (2021) | • ICU | 49 168 patients | PhysioNet Challenge restrictive Sepsis-3 definition<sup>81</sup> | 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. |
| | • MIMIC-III | Train: 33 434 P | | |
| | • 2001–2012 | Sepsis: 1353 P | | |
| | | No Sepsis: 32 081 P | | |
| | | Validation: 8358 P | | |
| | | Sepsis: 338 P | | |
| | | No Sepsis: 8020 P | | |
| | | Test: 7376 P | | |
| | | Sepsis: 229 P | | |
| | | No Sepsis: 7077 P | | |

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.

<sup>a</sup>Sample size unit abbreviations: P: patients; N: notes; E: encounters.
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, Horng et al \(^47\) and Apostolova and Velez \(^48\) processed individual notes. While Goh et al \(^52\) 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

| Documentation types          | Author                  | Description                                                                 | Temporal perspective | Record latency | Frequency          |
|------------------------------|-------------------------|-----------------------------------------------------------------------------|----------------------|---------------|--------------------|
| Chief complaints             | Physician               | Symptoms or complaints provided by a patient at start of care for why they are seeking care. | Current             | Seconds to days | One per episode    |
|                              | Nurse                   |                                                                             |                      |               |                    |
|                              | Specialist              |                                                                             |                      |               |                    |
| History-and-physical notes   | Physician               | 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    |
|                              | Nurse                   |                                                                             |                      |               |                    |
|                              | Specialist (eg, respiratory therapist) |                                                                      |                      |               |                    |
| Progress notes               | Physician               | Observations of patient status and care provided to document progress and response to treatment plans. For physician, it includes determining diagnosis, prescriptions, and laboratory orders. | Retrospective, Prospective | 4–8 h          | One per shift      |
|                              | Nurse                   |                                                                             |                      |               |                    |
|                              | Specialist (eg, respiratory therapist) |                                                                      |                      |               |                    |
| 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.                               | Retrospective, Prospective | At discharge or days after | One per episode    |
| Discharge summary letter     | Physician               | Formal required letter containing follow-up treatment plans.                | 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                    | Physician               | Diagnosis classification for billing.                                       | Retrospective        | Days to months | One per episode    |
|                              | Professional ICD coder  |                                                                             |                      |               |                    |
|                              | ICD data aggregator organization |                                                                     |                      |               |                    |
| Administrative               | Administration          | Patient information such as name, age, gender, address, contact information, and occupation. | Retrospective, Current | Immediately | One per episode    |

Record latency is defined as time between measurement/observation and the availability of the results in electronic health records.
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 al. used a 4-, 8-, or 24-h duration window before severe sepsis, and concatenated all text within a window. Goh et al. used a 4-, 6-, 12-, 24-, or 48-h duration window of before sepsis, severe sepsis, or septic shock onset. Liu et al. 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 al. 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 al. 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 al. 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 al. used a 1-, 3-, 6-, 12-, or 24-h duration window after the first vial sign or laboratory result was documented in the EHR to identify patients at risk for sepsis, and utilized keywords. 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), n-gram, term frequency-inverse document frequency (tf-idf), and paragraph vectors (PV) 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. 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
| Study (year) | Free-text document type | Unit of analysis | Text processing |
|-------------|-------------------------|-----------------|----------------|
| Horng et al.47 (2017) | • ED chief complaints • Nursing triage assessments | One note | Representation: • Bi-gram • BoW (15 240-word vocabulary) • LDA topic modeling (500 topics) Techniques: • Convert to lowercase • Remove rare tokens and punctuation • Negation |
| Apostolova and Velez48 (2017) | Nursing notes | One note | Representation: • 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: • 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: • GloVe (300-dimensional vector) + summing word vectors Techniques: • Concatenated all notes for an encounter into a single text block |
| Delahanty et al.51 (2019) | ED chief complaints | Keywords | Other: • Keywords extracted by experts |
| Liu et al.50 (2019) | All MIMIC-III clinical notes, such as but not limited to: • Nursing notes • Physician notes | One note | Representation: • BoW (8907 unique term vocabulary and 832 predictive terms) • GloVe (300-dimensional vector for each unique term) Techniques: • Convert to lowercase • Remove rare tokens, frequent tokens, and non-alphanumeric characters |
| Amrollahi et al.53 (2020) | • Nursing notes • Physician notes | One note | Representation: • tf-idf (2227 vector size features = 2187 text features + 40 structured features) • ClinicalBERT (808 vector size features = 768 text features + 40 structured features) Techniques: • 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: • Nursing progress notes • Respiratory therapist progress notes | One note | Representation: • BoW • tf-idf Techniques: • Remove rare and frequent tokens |
a suitable representation of texts, often known as embeddings. Embedding techniques include Global Vectors for Word Representation (GloVe), Word2Vec as a continuous bag-of-words (CBOW) model or skip-gram model, Bidirectional Encoder Representations from Transformers (BERT), and ClinicalBERT. The advantage of using embeddings is that it retains the sequential information lost in a BoW representation and does feature extraction automatically.

Utilized text processing operations are in Table 3. One study used keyword extraction instead of text processing operations. Six studies used tokenization of words for word-level representation, 47–50,52,54 I 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, frequent tokens, punctuation or special characters, and stop words. The most frequently used representation was tf-idf, followed by BoW, LDA, GloVe, ClinicalBERT, bi-gram, CBOV, and PV. Three studies created a vocabulary of unique terms using BoW, CBOV, and tf-idf. Apostolova and Velez 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 al and Liu et al listed predictive terms in their models, and Goh et al 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 al\textsuperscript{52} 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\textsuperscript{49} 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\textsuperscript{53} most predictive terms or topics for having an infection in the ED include “cellulitis,” “sore_throat,” “abcess,” “uti,” “dysuria,” “pneumonia,” “redness_swelling,” “erythma,” “swelling,” “redness, cellulitis, left, leg, swelling, area, rle, arm, lle, increased, erythema,” “abcess, buttock, area, drainage, axilla, groin, painful, thigh, left, hfx, abcesses, red, boil,” and “cellulitis, abx, pt, iv, infection, po, kefex, antibiotics, leg, treated, started, yesterday.” Whereas the least predictive terms or topics for not having an infection include “motor vehicle crash,” “laceration,” “epistaxis,” “etoh”(ethanol for drunkenness), “etoh, found, vomition include “motor vehicle crash,” “laceration,” “epistaxis,” “etoh”(ethanol for drunkenness), “etoh, found, vomition” are: (1) For sepsis, severe sepsis, or septic shock, Goh et al\textsuperscript{52} classified

\[
\begin{align*}
\text{Journal of the American Medical Informatics Association, 2022, Vol. 29, No. 3}
\end{align*}
\]
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). Each figure contains the study and year, machine learning model, and natural language processing technique.

A. AUC values for infection identification. Horng et al. (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 al. AUC values are from detecting 4 h before sepsis onset, and Qin et al. AUC values are from detecting 0 to 6 h before sepsis onset. (C) AUC values for early septic shock detection. Hammoud et al. AUC values are from detecting 30.64 h before septic shock onset, and Liu et al. 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. Goh et al. Different symbols separate data types.

E. AUC values for early septic shock detection for Culliton et al. using results from the test set. (F) AUC values for early septic shock detection for Culliton et al. using results from 3-fold validation.

Disclaimer: 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.

Machine 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. Natural 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.
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.\(^4\) 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.\(^4\) 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).

### 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\(^i\)     |                  |                             |           |
| Horng et al.\(^4\) (2017)  | Identify    | DV - -            | CC + NN RF (BoW)            | 0.87      |
| Apostolova and Velez\(^4\) | Identify    | - - - -          | NN SVM (BoW + tf-idf)       | –         |
| Culliton et al.\(^4\) (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.\(^4\) (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.\(^4\) (2019)   | -7          | -VLM -           | CN GBT (GloVe)             | 0.92      |
|                            | -7.3        | -VLM -           | CN GBT (BoW)               | 0.91      |
|                            | -6          | -VLM -           | – GBT                      | 0.85      |
| Amrollahi et al.\(^4\) (2020)| -4\(^b\)   | -VL - -          | PN + NN LSTM (ClinicalBERT) | 0.84      |
| Hammoud et al.\(^4\) (2020)| -30.6       | DVLM -           | CN Lasso regression (BoW + tf-idf) | 0.89      |
| Goh et al.\(^4\) (2021)   | Identify    | DVLM -           | PN Logistic regression + RF (LDA) | 0.94      |
|                            | -4          | DVLM -           | – Logistic regression + RF  | 0.93      |
|                            | -6          | DVLM -           | PN Logistic regression + RF (LDA) | 0.85      |
|                            | -12         | DVLM -           | PN Logistic regression + RF (LDA) | 0.92      |
|                            | -24         | DVLM -           | – Logistic regression + RF  | 0.79      |
|                            | -48         | DVLM -           | PN Logistic regression + RF (LDA) | 0.90      |
| Qin et al.\(^4\) (2021)   | -6 to 0\(^i\) | -VL - -          | CN GBT (ClinicalBERT-sf)    | 0.89      |

\(^a\) Hours: Identify: not detecting hours before or after; < hours before; +/- hours after an event.

\(^b\) Data types: D: demographics; V: vitals; L: laboratory; M: medications; C: codes; T: text; ‘-’ position in DVLMC indicates which is not used.

\(^c\) Text data types: CC: chief complaints; CN: various types of clinical notes; NN: nursing notes; PN: physician notes; –: no notes.

\(^d\) Machine 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.

\(^e\) Natural 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.

\(^f\) Area under the curve (AUC). Apostolova and Velez\(^4\) did not provide metrics for AUC.

\(^g\) Culliton et al.\(^4\) performed 2 experiments, these results are from using a test set instead of 3-fold validation.

\(^h\) Number of hours before onset for Amrollahi et al.\(^4\) was confirmed through personal communications (with Shamim Nemati on May 27, 2021 and Fatemeh Amrollahi on June 13, 2021).

\(^i\) Qin et al.\(^4\) AUC values are an average from 0 to 6 h before sepsis, not the specified hours.

Sepsis definition impact

In ML, many studies rely heavily on sepsis definitions and ICD-codes to identify patient cohort datasets for sepsis studies.\(^5\)
Suggestions for future studies

Predicting sepsis earlier than 12 h prior to sepsis onset can reduce treatment delays and improve patient outcomes.1,4 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 effects 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,16 methods to overcome the need for annotated data,22 different embedding techniques,18 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,18 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 time-stamps 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 homogeneous 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.130 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. MYY conducted the literature search and initial analysis, LTG verified results, and ØN resolved discrepancies. All authors participated in data analysis and interpretation. MYY 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 Hoivik (RN), Dr Erik Solliåård, Dr Jan Kristian Damås, Dr Jan Egil Aset, Dr Kristin Vardheim Liyanarachi, Dr Randi Marie Mohus, and Dr Anuradha Ravil.
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
1. 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.
2. Fleischmann C, Scherag A, Adhikari NKJ, et al.; International Forum of Acute Care Trialsists. 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.
3. 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 (18): 1368–77.
4. 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.
5. Polat G, Ugan RA, Cadirci E, et al. Sepsis and septic shock: current treatment strategies and new approaches. Eurasian J Med Sept 2017; 49 (1): 53–8.
6. Arnold C. News feature: the quest to solve sepsis. Proc Natl Acad Sci USA 2018; 115 (16): 3988–91.
7. Iskander KN, Osuchowski MF, Stearns-Kurosawa DJ, et al. Sepsis: multiple abnormalities, heterogeneous responses, and evolving understanding. Physiol Rev 2013; 93 (3): 1247–88.
8. Jawad I, Lukić 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.
9. Islam MM, Nasrin T, Wallher BA, et al. Prediction of sepsis patients using machine learning approach: a meta-analysis. Comput Methods Programs Biomed 2019; 170: 1–9.
10. Schnikel M, Paranjape K, Nannan Panday RS, et al. Clinical applications of artificial intelligence in sepsis: a narrative review. Comput Biol Med 2019; 115: 103488.
11. 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 (5 02): e43–57.
12. Teng AK, Wilcox AB. A review of predictive analytics solutions for sepsis patients. Appl Clin Inform 2020; 11 (3): 387–98.
13. Fleuren LM, Klausch TL, 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.
14. 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.
15. 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.
16. 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
17. Sheikhshahi 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.
18. Wang Y, Wang L, Rastegar-Mojarrad M, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018; 77: 34–49.
19. Datta S, Bernstein EV, Roberts K. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. J Biomed Inform 2019; 100: 103301.
20. 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): e012122.
21. 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.
22. Spasic I, Nenadic G. Clinical text data in machine learning: systematic review. JMIR Med Inform 2020; 8 (3): e17984.
23. Elkin PL, Froehling D, Wahner-Roedder D, et al. NLP-based identification of pneumonia cases from free-text radiological reports. AMIA Annu Symp Proc 2008; 11: 172–6; Washington, DC.
24. 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.
25. 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.
26. Soguero-Ruiz C, Handberg K, Mora-Jimenez I, et al. Predicting colorectal surgical complications using heterogeneous clinical data and kernel methods. J Biomed Inform 2016; 61: 87–96.
27. Huddar V, Desiraju BK, Rajan V, et al. Predicting complications in critical care using heterogeneous clinical data. IEEE Access 2016; 4: 7988–8001.
28. Ribelles N, Jerez JM, Rodriguez-Brazzarella 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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): 162–9.
35. 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.
36. Carroll RJ, Eyler AE, Denny JC. Naïve electronic health record phenotyping applied to the electronic medical record: an empiric approach to performance improvement. Am J Med Inform 2019; 26 (5): 1035–43.
37. Xia Z, Secor E, Chibnik LB, et al. Modeling disease severity in multiple sclerosis using electronic health records. PLoS One 2013; 8 (11): e78927.
38. 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.
39. Zheng H, Gaff H, Smith G, et al. Epidemic surveillance using an electronic medical record: an empirical approach to performance improvement. PLoS One 2014; 9 (7): e100845.
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): 304–9.
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 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. Tidwell 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, Bethushein 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, Sheeh 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, Jeblee S, Pou-Prom C, et al. A survey of word embeddings for clinical text. J Biomed Informatics: X 2019; 100057.
145. Névéd 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. Koleck 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 presepsis 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.