Computer Methods and Programs in Biomedicine 170 (2019) 1–9

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Prediction of sepsis patients using machine learning approach: A meta-analysis

Md. Mohaimenul Islam[a,b], Tahmina Nasrin[a,b], Bruno Andreas Walther[c], Chieh-Chen Wu[a,b], Hsuan-Chia Yang[b], Yu-Chuan Li (Jack)[a,b,d,e,*]

[a] Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan
[b] International Center for Health Information Technology, Taipei Medical University, Taipei, Taiwan
[c] Department of Biological Sciences, National Sun Yat-sen University, Gushan District, Kaohsiung City, 804, Taiwan
[d] Department of Dermatology, Wan Fang Hospital, Taipei, Taiwan
[e] TMU Research Center of Cancer Translational Medicine, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 6 November 2018
Revised 28 November 2018
Accepted 24 December 2018

Keywords:
Area under receiver operating curve
Sepsis
Machine learning
Diagnostic odd ratio

ABSTRACT

Study objective: Sepsis is a common and major health crisis in hospitals globally. An innovative and feasible tool for predicting sepsis remains elusive. However, early and accurate prediction of sepsis could help physicians with proper treatments and minimize the diagnostic uncertainty. Machine learning models could help to identify potential clinical variables and provide higher performance than existing traditional low-performance models. We therefore performed a meta-analysis of observational studies to quantify the performance of a machine learning model to predict sepsis.

Methods: A comprehensive literature search was conducted through the electronic database (e.g. PubMed, Scopus, Google Scholar, EMBASE, etc.) between January 1, 2000, and March 1, 2018. All the studies published in English and reporting the sepsis prediction using machine learning algorithms were considered in this study. Two authors independently extracted valuable information from the included studies. Inclusion and exclusion of studies were based on the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guidelines.

Results: A total of 7 out of 135 studies met all of our inclusion criteria. For machine learning models, the pooled area under receiving operating curve (SAUROC) for predicting sepsis onset 3 to 4 h before, was 0.89 (95%CI: 0.86–0.92); sensitivity 0.81 (95%CI:0.80–0.81), and specificity 0.72 (95%CI:0.72–0.72) whereas the pooled SAUROC for SIRS, MEWS, and SOFA was 0.70, 0.50, and 0.78. Additionally, diagnostic odd ratio for machine learning, SIRS, MEWS, and SOFA was 15.17 (95%CI: 9.51–24.20), 3.23 (95%CI: 1.52–6.87), 31.99 (95% CI: 1.54–666.74), and 3.75(95%CI: 2.06–6.83).

Conclusion: Our study findings suggest that the machine learning approach had a better performance than the existing sepsis scoring systems in predicting sepsis.

1. Introduction

1.1. Background

Sepsis is a common and life-threatening syndrome, and a leading cause of morbidity and mortality globally [1]. It has already become a major global health burden due to higher treatment cost, and excessive hospital stay [2,3]. However, correct and accurate identification of the risk factors, and appropriate antibiotic selection would play a significant role in the overall mortality reduction and cost burden of sepsis treatment. Currently, available screening methods such as the modified early warning system (MEWS), systemic inflammatory response syndrome (SIRS), etc., are not enough to clearly identify sepsis patients and transfer their treatment into a higher level of care [4].

Machine learning has been emerging as a promising tool to decrease diagnostic uncertainty, select appropriate antibiotics, and identify proper sepsis patients [5,6]. Real-time clinical variables have been used to generate a suitable prediction model that can accurately predict the onset of sepsis in an intensive care unit (ICU) before clinical recognition [7–9]. Several published epidemi-

Abbreviations: SAUROC, Summarized area under receiver operating curve; SIRS, Systemic inflammatory response syndrome; MEWS, Modified early warning system; SOFA, Sequential organ failure assessment; QSOFA, Quick sequential organ failure assessment; DOR, Diagnostic odd ratio; ICU, Intensive care unit.

* Corresponding author at: College of Medicine Science and Technology (CoMST), Taipei Medical University, Chair, Dept. of Dermatology, Wan Fang Hospital, 250-Wuxing Street, Xinyi District, Taipei 11031, Taiwan.

https://doi.org/10.1016/j.cmpb.2018.12.027
0169-2607/

ological studies that have already revealed the risk factors of sepsis. Indeed, a significant number of studies have also reported machine learning can detect and predict the onset of sepsis using these potential variables [10–12].

1.2. Goal of this investigation

To our knowledge, no meta-analysis evaluated the performance of machine learning models for predicting sepsis so far. Therefore, we conducted a meta-analysis to investigate the potential of machine learning for identifying sepsis patients. Correctly identification and prediction of sepsis patients has a clinical implication for prevention and treatment of sepsis patients.

2. Materials and methods

2.1. Research protocol

This study was conducted in accordance with the preferred Reporting Items for Systematic Reviews and Meta-Analyses. We have used similar method in our previously published systematic review and meta-analysis [13–15] (**Supplementary Table S5**).

2.2. Literature review

A comprehensive search strategy was developed in consultation with various professionals such as pharmacists, physicians, data scientists in the International Center for Health Information Technology. The most popular and reliable electronic databases such as PubMed, EMBASE, Google Scholar, Scopus were searched between January 1, 2000, and March,1, 2018 to find out relevant studies. The study reported machine learning approach for predicting sepsis was included in our present study. The following MeSH terms were used to search the articles: ((“Machine learning*”), OR (“Machine learning model*”), OR (“Machine learning algorithms*”) OR (“Algorithms”) OR (“Artificial Intelligence*”) OR (“Deep learning”), (“Computational Approach”) OR (“Automated- computer aided”)) And ((“Sepsis”), OR (“Sepsis prediction”) OR (“Sepsis onset”) OR (“Sepsis identification”), OR (“Sepsis detection”) OR (“Sepsis risk prediction”) OR (“Sepsis development”), (“Severe sepsis prediction”) AND ((“Intensive care unit”), OR (“Emergency department”)) **Supplementary Table S1**. Additionally, the bibliography of each study was also scrutinized to find any missing study in the initial search. A widely implemented commercial referencing software EndNote X7 (Thomson Reuters) was further used to compile the studies and duplication checking.

2.3. Inclusion and exclusion criteria

We carefully screened all the studies in the initial search and checked their relevant titles and abstracts. Two authors (MMI, TNP) examined these studies independently. To be included, all the studies had to fulfill the following criteria: 1) study must be in English and be peer-reviewed, 2) provided an outcome of the machine learning algorithms and sepsis patients prediction, 3) provided information regarding sensitivity and specificity, 4) provided the total number of sepsis patients, 5) provided a clear definition of sepsis, 6) clearly described machine learning models and predictor variables used in the sepsis prediction, 7) provided explicit overview of dataset used in the study and origin of data source (emergency/ intensive care unit), 8) provided clear information of sepsis identification.

Editorials, short reports, traditional methods for predicting sepsis were excluded. All the studies meeting inclusion criteria at this stage were additionally reviewed by the same two authors to ensure the appropriateness of the final analysis. All disagreement between two authors for selecting potential studies were then resolved by chief investigator (YC, L). Studies providing the most detailed information regarding algorithms and clinical variables were kept for references.

2.4. Data extraction and quality assessment

Data abstraction was conducted by the same two authors who used a predefined, standardized protocol and data collection instrument. All the data were entered into Review Manager software (RevMan-5) and checked for accuracy. Finally, they selected appropriated seven studies that reported machine learning and sepsis prediction. The following data were extracted from these included studies.

a) Author information.
b) Year of publication.
c) Clinical variables for sepsis prediction.
d) Machine learning prediction model name.
e) Area under receiving operating curve (AUROC), sensitivity, and specificity value.
f) Number of patients admitted in emergency department (ED)/intensive care unit (ICU) and number of sepsis patients.
g) Database information.
h) sepsis identification method.

However, systematic reviews and meta-analysis of diagnostic accuracy studies are often characterized by some sorts of heterogeneous findings that is originated from differences in the design and conduct of included studies. We therefore carefully assessed the quality of included studies in this meta-analysis. The Quality Assessment of Diagnostic Accuracy Studies–2 (QUADAS) tool [16] was used to assess the quality of included studies **(Supplementary Table S6).**

2.5. Gold standards

In this meta-analysis, we considered the studies where machine learning performance were tested according to various gold standards (clinical indications). We considered only the studies classified as positive or negative for having acquired in-hospital sepsis with SIRS, MEWS, and SOFA, or at least one of these gold standards. However, included studies had to identified each septic if

(1) The patient record was recorded in-hospital contraction of sepsis with an ICD9 code (995.9)/ ICD-10.
(2) The patient met the 1991 systemic inflammatory response syndrome (SIRS) criteria for sepsis for a persistent 5-h period of time. The beginning of the patient’s first 5-h SIRS event was considered as the zero hour.

2.6. Definition of sepsis

The sepsis gold standard is defined by using the 2001 consensus sepsis definition. However, a new Sepsis-3 definition is introduced in 2016 [4], and discussions on the effectiveness of the newly proposed definition are continuing. The patients are suspected of infection if they meet two or more SIRS criteria. The onset time is defined as the first time two or more SIRS criteria observed in the same hour.

SIRS criteria [17] are given below:

a) Temperature $> 38^{\circ}C$ or $< 36^{\circ}C$.
b) Heart rate > 90/min.
c) Respiratory rate > 20/min or $PaCO2 < 32$ mmHg (4.3 kPa).
d) White blood cell count $< 12{,}000/mm3$ or $4000/mm3$ or 10% immature bands.

2.7. Primary and subgroup analysis

The included studies that investigated the performance of the machine learning models for predicting sepsis patients were considered in the primary analysis. The primary analysis was divided into two parts:

a) Sepsis prediction 3 to 4 h prior to onset
b) Sepsis detection (0 h)

In addition, the commonly and widely used traditional models such as systemic inflammatory response syndrome (SIRS), modified early warning system (MEWS), sequential organ failure assessment (SOFA), and quick sequential organ failure assessment (QSOFA) were also evaluated as a subgroup analysis (at least 3 to 4 h prior to sepsis onset).

2.8. Statistical analysis

For each included study, we obtained the sensitivity, specificity, and receiving operating curve (ROC) value with 95% confidence interval (CI) for predicting sepsis patients in ICU. ROC curve is plotted the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The value of ROC curve 0.9–1, 0.8–0.9, 0.7–0.8, 0.6–0.7, 0.5–0.6 defines as excellent, good, fair, poor and fail. ROC, sensitivity and specificity were measured with 95% CI in the final analysis. An I^2 value was used to assess the statistical heterogeneity which provided an estimate of the percentage of variability among the included studies. An I^2 value 0∼25%, 25∼50%, 50∼75%, more than 75% represents very low, low, medium and high heterogeneity. The results from all included studies were pooled, and an overall estimate of effect size was evaluated using a random effect model which help to reduce heterogeneity among studies. The I^2 statistic describes the percentage of variation across studies that is due to heterogeneity rather than chance [18]. I^2 was calculated as follows:

$$I^2 = 100\% \times (Q - df)/Q$$

Where Q is Cochrane's heterogeneity statistic and df is the degree of freedom. Negative values of I^2 were set a zero; the I^2 results are between 0% (no observed heterogeneity) and 100% (maximum heterogeneity)

In our meta-analysis, we used the symmetric method because of the assumed heterogeneity among the included studies. MetaDiSc (version 1.4) was used for pooled estimate of AUROC, sensitivity, specificity and diagnostic odds ratio. It helps to a) summarize data from each individual study, b) investigate the homogeneity of studies graphically and statistically, c) compute the pooled estimate, d) explore heterogeneity. Likelihood ratios were assessed to express how much more frequent the respective result is among the studies with sepsis disease than among subjects without sepsis disease.

$$LR+ = (Sensitivity/1 - Specificity) \text{ and}$$
$$LR- = (1 - Sensitivity/Specificity)$$

In addition, diagnostic odd ratio (DOR) was calculate to provide, how much greater the odds of having the sepsis disease are for the people with a positive test result than for the people with a negative test result. In mathematically, DOR = LR + /LR-. Area under receiver operating curve, sensitivity, specificity, diagnostic odd ratio, likelihood ration was used to evaluate the performance of each method **(Supplementary Table S3).**

The confidence intervals of overall sensitivity and specificity was also analyzed using the F distribution method to compute the exact confidence limits for the binomial proportion [19]. However, Meta-DiSc was used to compute them using over dispersion correction. In this case, it used the normal approximation to binomial, i.e.

$$SE\ (Sen_T) = \sqrt{\frac{Sen_T(1 - Sen_T)}{\sum_i Di}}$$
$$SE\ (Spe_T) = \sqrt{\frac{Spe_T(1 - Spe_T)}{\sum_i ND_i}}$$

the confidence interval [20] corrected by over dispersion were:

$$Sen_T \pm Z_{\propto/2}\varphi_{Sen}SE(Sen_T) \qquad Spe_T \pm Z_{\propto/2}\varphi_{Spe}SE(Spe_T)$$

If there was any evidence of diagnostic threshold variation among studies, the best summary of study results would be an ROC curve rather than a single point. The shape of the ROC curve actually depends on the underlying distribution of test results in individuals with and without the disease. Diagnostic tests where the DOR was constant regardless of the diagnostic threshold have symmetrical curves around the "Sen = Spe" line. In these situations, it is possible to combine DOR's by the Mantel–Haenszel or the Der-Simonian Laird methods to estimate the overall DOR and hence to determine the best-fitting ROC curve [21]. The equation of curve was calculated:

$$Sen = \frac{1}{1 + \frac{1}{DOR_T} \times \left(\frac{1-Spe}{Spe}\right)}$$

However, the AUC was computed by numeric integration of the curve equation by the trapezoidal method. A Q* index is another useful statistic which was defined by the point where sensitivity and specificity are equal, which was the point closest to the ideal top-left corner of the ROC space. A Q* value calculated by-

$$Q* = \frac{\sqrt{DOR_T}}{1 + DOR_T}$$

Moreover, the standard error of the area under the symmetrical ROC curve was calculated by

$$SE\ (AUCsym) = \frac{DOR_T}{(DOR_T - 1)3}$$
$$[(DOR_T + 1)\ln DOR_T - 2(DOR_T - 1)]SE(lnDOR_T)$$

The standard error of Q* was calculated by the following equation-

$$SE\ (Q*) = \frac{\sqrt{DOR_T}}{2\left(1 + \sqrt{DOR_T}\right)2}\ SE\ (\ln DOR_T)$$

The confidence interval of symmetrical ROC curve was analyzed to provide the upper and lower limits of CI of overall DOR in the equation of curve.

3. Results

3.1. Study selection

A total of 135 unique titles and abstracts were identified initial search. Among those 121 articles were excluded based on our study inclusion and exclusion criteria's (described in method part). Furthermore, 14 studies had gone for the full-text review, and of these, 7 studies met all inclusion criteria [10,22–27]. Fig. 1 shows inclusion and exclusion criteria of the study process.

Study characteristics: The baseline characteristics of the eight included studies are shown in Table 1.

The year of publication ranged from 2016 to 2018. All studies included patients from an intensive care unit (ICU). The minimum and maximum inclusion of sepsis patients were 140 and 32,103. All the studies included sepsis patients who were older than 18 years or above. Five studies compared the machine learning approach with other golden scoring systems in current use for the

Table 1
Baseline characteristic of included studies.

Author	Publication year	Data collection period	Sepsis patients	Prediction models	Data source	Sepsis definition/ identification	Department	Prediction before sepsis	AUROC	External validation	Model discrimination
Desautels	2016	2001–08	2577	Insight	MIMIC-III	SIRS criteria	ICU	0 and 4 h before	ML = 0.88 SIRS = 0.61 SOFA = 0.73 QSOFA = 0.77 MEWS = 0.80	N	NR
Horng	2017	2008–13	32,103	SVM	MIMIC-III	ICD-9 code	ICU	4 h	ML = 0.86	N	NR
Mao	2017	2011–16	140	InSight	University of California, San Francisco and MIMIC-III	SIRS criteria, and ICD-9 code: 995.91	ICU	0 and 4 h before	ML = 0.92 MEWS = 0.76 SOFA = 0.63 SIRS = 0.75	Y	NR
Nemati	2017	2013–15	2,375	APeX	Emory University Hospitals And MIMIC-III	Third International Consensus Definitions	ICU	4 h before	ML = 0.85 SOFA = 0.87	Y	NR
Calvert	2016	2001–08	159	InSight	MIMIC-III	SIRS criteria, and ICD9 code: 995.9	ICU	3 h before	ML = 0.83	N	NR
Kam	2017	2001–12	360	Deep neural network	MIMIC-III	SIRS criteria, and ICD9 code: 995.9	ICU	3 h before	ML = 0.92	N	NR
Faisal	2018	2014–15	4,909	Logistic regression	York hospital, and Northern Lincolnshire and Google Hospital	ICD-10	ICU	4 h before	ML = 0.78	Y	0.186

- N.B: Y = Yes, N = No, NR = Not Reported SVM = Support vector machine, ML = Machine learning, N A = $^{-1}$ Not applicable, SIRS = Systemic inflammatory response syndrome, MEWS = Modified early warning system, SOFA = Sequential organ failure assessment, q SOFA = Quick Sequential organ failure assessment, ICU = Intensive care unit ED = Emergency department.
- **Note: SIRS** criteria are defined as: 1. Heart rate > 90 beats/ min, 2. body temperature > 38 °C or < 36 °C, 3. respiratory rate > 20 breaths/min 4. white blood cell count > 12,000 cells/μL or < 4,000 cells/μL. Organ dysfunction criteria are defined as: 1. Lactate > 2 mmol/L 2. Systolic blood pressure < 90 mmHg 3. Urine output < 0.5 mL/kg, over two hours 4. Creatinine > 2 mg/dL 5. Bilirubin > 2 mg/dL 6. Platelet count < 100,000 μL 7. International normalized ratio > 1.5 8. PaO2 > 0.5.
- **MIMIC-III** = The Multi-Parameter Intelligent Monitoring in Intensive Care.
- **InSight** is a machine learning classification system that uses multivariable combinations of easily obtained patient data (vitals, peripheral capillary oxygen saturation, Glasgow Coma Score, and age). Included studies compared ***InSight*** predictions for each gold standard to three common patient deterioration scoring systems: SIRS, SOFA and MEWS.
- **APeX** is developed by Epic Systems (Verona, Wisconsin, USA), and the prediction algorithm was developed by Dascena (Hayward, California, USA).

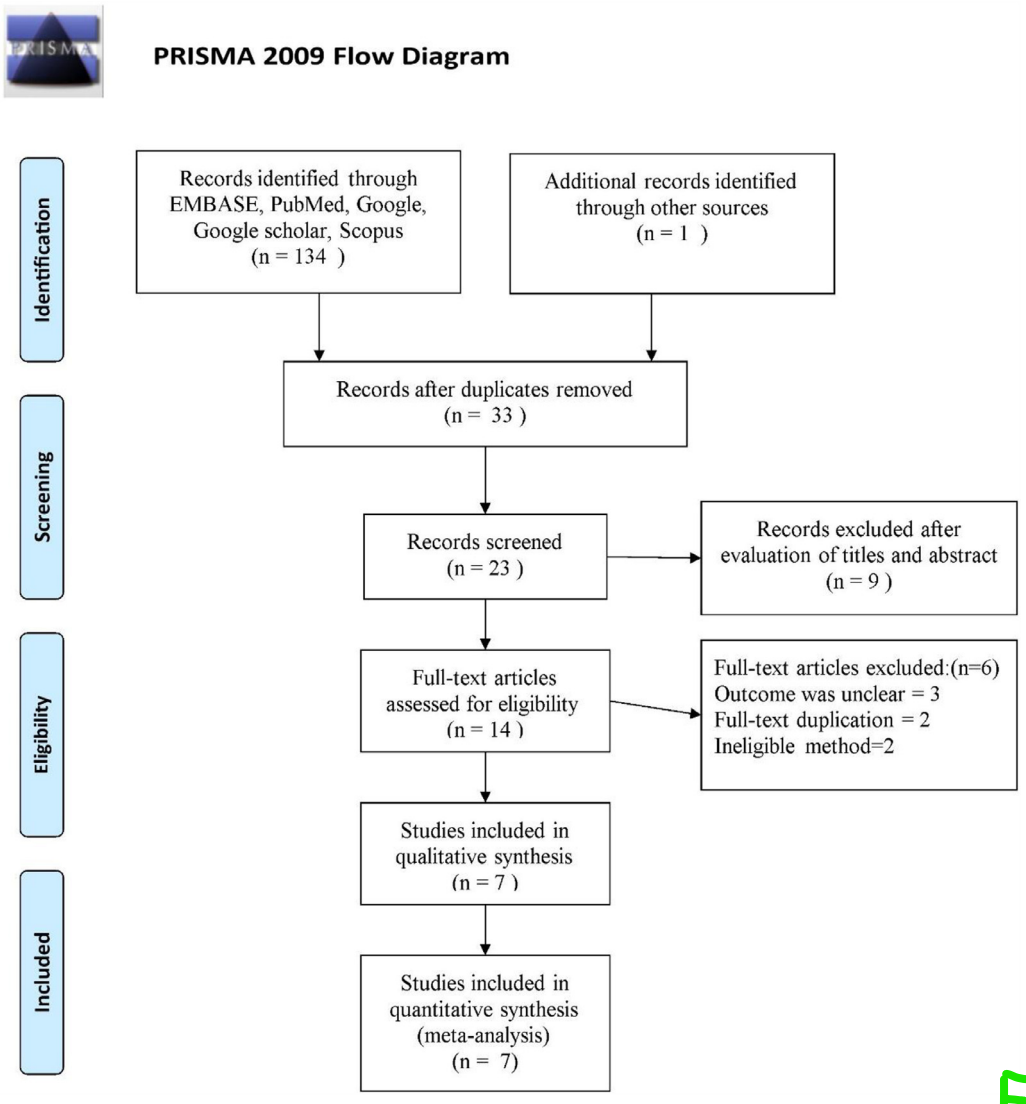

Fig. 1. PRISM flow diagram for study selection.

determination or prediction of sepsis, including the SOFA, qSOFA, MEWS, and SIRS score, but four studies only investigated the machine learning model for predicting sepsis patients. Patients were confirmed with sepsis if they met the golden standard definition of sepsis: "the presence of two or more SIRS criteria paired with a suspicion of infection".

All patients fulfilled the criteria for SIRS during the first 4 h of hospital admission or the first 1 h of the first ICU admission (Table 2).

However, each study also confirmed sepsis patients by using the International Classification of Diseases (ICD) code. Patient data of included studies were de-identified to comply with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. All the studies used all of the most similar features for the prediction of sepsis onset **(Supplementary Table S4).**

3.2. Machine learning model and sepsis prediction

Seven studies investigated the performance of the machine learning approaches for the prediction of sepsis patients. The machine learning models showed higher performance to accurately predict sepsis, and identified sepsis 3 to 4 h prior to its onset compared with the other traditional scoring systems. The overall pooled AUROC for machine learning to predict sepsis 3 to 4 h prior to onset was 0.89 (Fig. 2A).

Additionally, sensitivity, specificity, and diagnostic odds ratio were 0.81 (95%CI: 0.80–0.81, $p < 0.0001$, $I^2 = 99.6\%$), 0.72 (95%CI:0.72–0.72, $p < 0.0001$, $I^2 = 100\%$), and 15.17 (95%CI: 9.51–24.20, $p < 0.001$, $I^2 = 99.1\%$), respectively **(**Fig. 2, B-D**) (Supplementary Table S2)**.

Furthermore, the positive likelihood ratio and negative likelihood ratio were 3.31 (95% CI: 1.46–7.48, P < 0.0001, $I^2 = 100$) and 0.23 (95%CI:0.20–0.27, $I^2 = 94.9\%$), respectively **(Sup. Fig. S1).**

Three studies evaluated the performance of machine learning to accurately detect the onset of sepsis. The overall pooled SAUROC for machine learning to detect sepsis onset (0 hr.) was 0.98. However, sensitivity, specificity, and diagnostic odds ratio were 0.87 (95% CI: 0.86–0.88, $p < 0.0001$, $I^2 = 99.4\%$), 0.93 (95%CI: 0.92–0.93, $p < 0.0001$, $I^2 = 100\%$), and 1240.11 (95%CI: 28.85–53,297.69, $p < 0.0001$, $I^2 = 99.6\%$), respectively. Additionally, the overall pooled positive likelihood ratio and negative likelihood ratio were 32.45 (95%CI: 6.56–160.57, $p < 0.0001$, $I^2 = 100\%$) and 0.06 (95%CI: 0.01–0.56, $p < 0.0001$, $I^2 = 99.3\%$), respectively **(Supplementary Fig. S2)**.

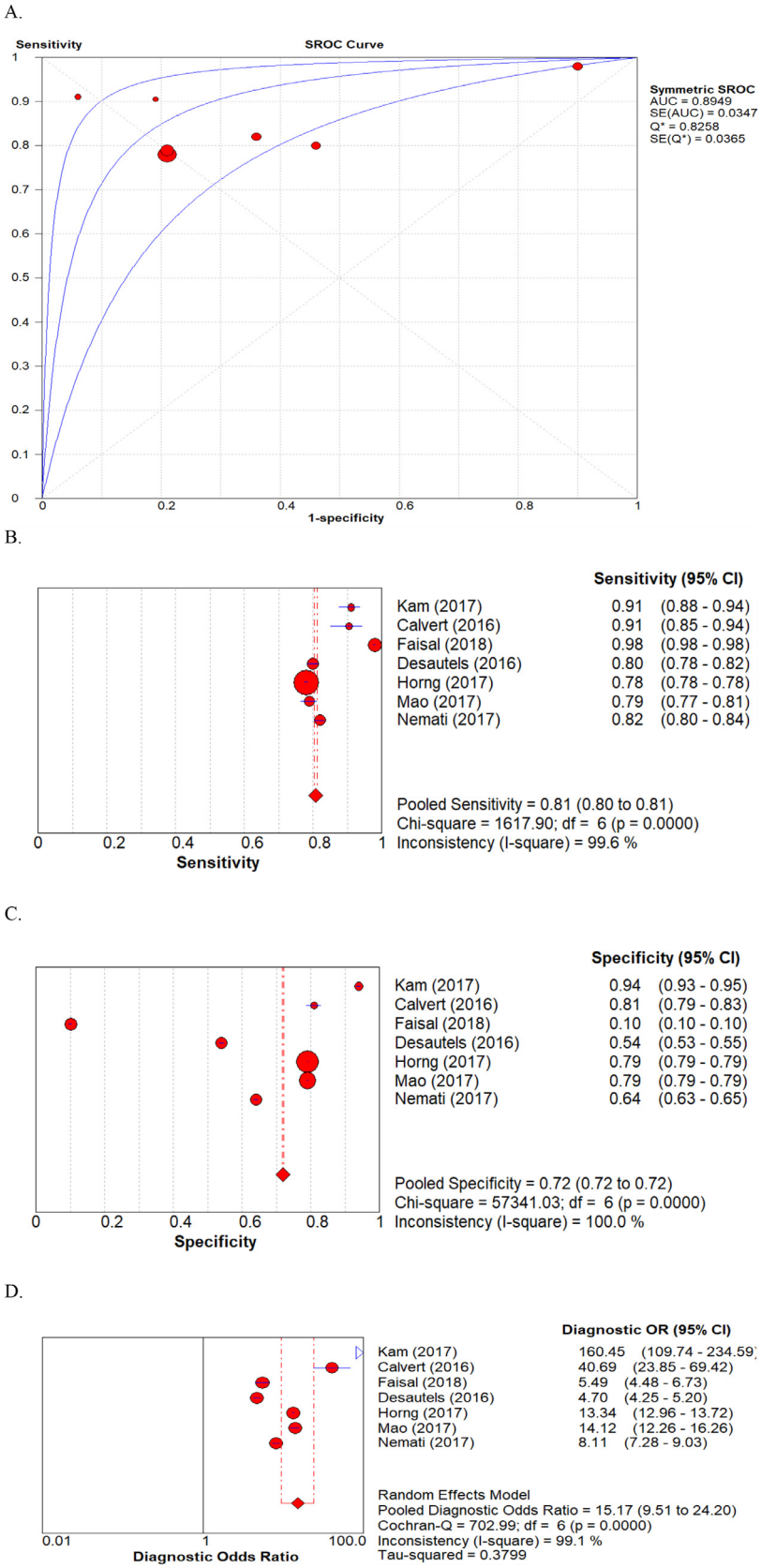

Fig. 2. Performance of machine learning model for predicting sepsis (A. SAUROC, B. Sensitivity C. Specificity D. Diagnostic odd ratio). The color circle represents proportion of the number of patients with sepsis and the number of patients without sepsis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Definition of systemic inflammatory response syndrome, sepsis and severe sepsis.

Culture-positive populations	Culture-negative populations
Systemic inflammatory response syndrome (SIRS): Systemic inflammatory response to a variety of severe clinical injuries. The response is manifested by two or more of the following conditions: 1. temperature > 38 °C or < 36 °C. 2. heart rate > 90 bpm. 3. respiratory rate > 20 breath min^{-1}, or PaCO2 < 32 mmHg. 4. white blood cell count > 12 000 mm^{-3}, < 44000 mm^{-3}, or > 10% immature (band) forms. **Culture-positive sepsis:** SIRS with confirmed evidence of infection. **Culture-positive severe sepsis:** Culture positive sepsis associated with organ dysfunction, hypo-perfusion, or hypotension. The perfusion abnormalities may include, but are not limited to, lactic acidosis, oliguria, or an acute alteration in mental status.	**Culture-negative sepsis**: SIRS without confirmed evidence of infection. **Culture-negative severe sepsis:** Culture-negative sepsis associated with organ dysfunction, hypo perfusion or hypotension. The perfusion abnormalities may include, but are not limited to, lactic acidosis, oliguria or an acute alteration in mental status.

Table 3
Performance of predicting sepsis by commonly used methods.

Name	No. study	AUROC	Sensitivity (95%CI)	Specificity (95%CI)	PLR(95%CI)	NLR(95%CI)	Diagnostic odd ratio
ML	7	0.89	0.81 (0.80–0.81)	0.72 (0.72–0.72)	3.31(1.46–7.48)	0.23(0.20–0.27)	15.17 (9.51–24.20)
SIRS	2	0.70	0.75 (0.74–0.77)	0.50 (0.50–0.50)	1.50(1.19–1.88)	0.46 (0.27–0.80)	3.23 (1.52–6.87)
MEWS	2	0.50	0.80 (0.79–0.81)	0.73 (0.73–0.73)	3.30(2.69–4.04)	0.10(0.00–2.43)	31.99 (1.54–666.77)
SOFA	3	0.78	0.77 (0.76–0.78)	0.42 (0.42–0.42)	1.68(1.13–2.52)	0.45 (0.37–0.54)	3.75 (2.06–6.83)
QSOFA	1	0.77	0.56 (0.54–0.57)	0.84 (0.83–0.84)	3.50 (3.34–3.67)	0.52 (0.50–0.55)	N/A

Note: ML = Machine learning, PLR = Positive Likelihood Ratio, NLR = Negative Likelihood Ratio, SIRS = Systemic inflammatory response syndrome, MEWS = Modified early warning system, SOFA = Sequential organ failure assessment, QSOFA = Quick sequential organ failure assessment.

4. Subgroup analysis

Subgroup analysis was also performed to compared machine learning prediction performance with widely used scoring systems. Table 3 shows the performance of ML, SIRS, MEWS, SOFA and QSOFA.

4.1. Sequential organ failure assessment (SOFA)

Three studies also assessed the performance of SOFA for predicting sepsis patients. The overall pooled AUROC for identifying sepsis patients was 0.78 (AUROC = 0.78, q = 0.7231). Additionally, the sensitivity, specificity, diagnostic odd ratio was 0.77 (95%CI: 0.76–0.78, $p < 0.0001$, $I^2 = 96.8\%$), 0.42 (95%CI: 0.42–0.42, $p < 0.0001$, $I^2 = 100\%$), 3.75 (95% CI:2.06–6.83, $p < 0.0001$, $I^2 = 98.9\%$).

4.2. Systematic inflammatory response syndrome (SIRS)

Two studies examined the performance of SIRS for the prediction of sepsis patients. The overall pooled AUROC for predicting sepsis was 0.50. However, the sensitivity, specificity, and diagnostic odd ratio was 0.75 (95%CI: 0.74–0.77, $p < 0.0001$, $I^2 = 97.8\%$), 0.50 (95%CI: 0.50–0.50, $p < 0.0001$, $I^2 = 99.4\%$), and 3.23 (95%CI: 1.52–6.87, $p < 0.0001$, $I^2 = 98.7\%$).

4.3. Modified early warning systems (MEWS)

Two studies evaluated the performance of MEWS for the prediction of sepsis patients. The overall pooled AUROC for predicting sepsis was 0.500. Moreover, the sensitivity, specificity and diagnostic odd ratio was 0.80 (95%CI: 0.79–0.81, $p < 0.0001$, $I^2 = 99.8\%$), 0.73 (95%CI: 0.73–0.73, $p < 0.0001$, $I^2 = 99.5\%$), and 31.99 (95%CI: 1.54–666.77, $p < 0.001$, $I^2 = 99.5\%$).

4.4. Quick sequential organ failure assessment (QSOFA)

One study investigated the performance of QSOFA for predicting sepsis patients. The overall pooled AUROC for predicting sepsis was 0.77. Furthermore, the sensitivity, specificity and diagnostic odd ratio was 0.56(95% CI: 0.54-

5. Discussion

5.1. Principle findings

We conducted a meta-analysis to investigate the performance of machine learning for predicting sepsis three to four hours before onset. The overall pooled estimation showed that machine learning performance for early recognition of sepsis and non-sepsis patients performed better when compared with the performance of traditional sepsis scoring tools such as SIRS, MEWS, SOFA, and QSOFA. Additionally, the ability of machine learning models was also higher for sepsis detection. In order to determine the generalizability of machine learning algorithm for sepsis prediction to different settings, machine learning also demonstrated a higher performance when used with different datasets with varying types and frequencies of patient's measurements. The findings of our study suggest that the strong predictive performance of machine learning would be helpful to decrease sepsis-related in-hospital mortality, and sepsis-related length of hospital stay. Since the diagnosis of sepsis patients is always challenging due to preexisting organ dysfunction, treatment prior to admission, and concurrent organ support. But higher sensitivity and specificity of the machine learning approach could correctly and accurately identify the patients, provide supportive treatment, and improve patient outcomes. Implementation of machine learning prediction tools may create immense opportunity to measure patient's criteria quickly and easily and assessed repeatedly over time in patients at risk of sepsis.

5.2. Public health implications and clinical practice

Sepsis is one of the major public health concerns because it is lethal, prevalent, and costly [28]. It has emerged as a global burden due to higher morbidity and mortality [29,30]. In the USA alone, each year 700 000 people are affected by severe sepsis, accounting for 20 million dollars per year [31,32]. Additionally, lengthy hospital stays due to sepsis are twice as common as for any other condition. However, early and accurate diagnosis and treatment of severe sepsis has shown an association with improved patients outcome, reduced mortality rate, and decreased cost of care [33,34].

Despite many attempts to identify sepsis patients, it is still difficult for healthcare providers to correctly recognize and diagnose this condition because of the heterogeneous nature of possible infection. The traditional definition of sepsis is the presence of at least two criteria of the Systemic Inflammatory Response Syndrome (SIRS), together with known or suspected infection, and organ dysfunction [35]. However, a new definition of sepsis (Sepsis-3) has created an opportunity for accurately identifying sepsis patients in clinical and preclinical settings [4,36,37]. An ideal situation is needed where minimal data is required, and data is routinely collected. A machine learning risk score model has the potential to use routine vital sign data to recognize sepsis patients hours before its onset [28]. Several studies have already used *InSight,* a novel version of the machine learning algorithm, which uses the most common variables (vital signs and other easily assessed bedside measurements, plus age) obtained from electronic health records to correctly identify and predict sepsis patients in the ICU [10,11,22,23]. Features for predicting sepsis patients among the included machine learning prediction models are known to be associated with the risk of sepsis onset. In addition, these machine learning models helped to identify well-known risk factors for sepsis even among the noise of many unrelated variables.

A machine learning based sepsis prediction system is designed to assist physicians in diagnosis, treatment, and patient's management in the emergency and intensive care units. Automated machine learning tools may be beneficial for physicians with a complex and difficult diagnosis of sepsis onset by analyzing the current trends and correlations between vital sign measurements. As widely available traditional tools often suffer from low sensitivity or specificity and fail to predict patients with a higher risk of sepsis [38,39], machine learning models with higher accuracy may provide early warning signals which identify sepsis patients, help supportive treatment, and open the door to prevent the progression of the condition [34]. Several studies already reported that early treatments could improve the patient's outcome by confirmation of a positive microbiology and appropriate antibiotic therapy.

5.3. Strengths and limitations

There are several strengths of our study. First, this is the first meta-analysis which evaluated the performance of machine learning for predicting sepsis. Second, the performance of the machine learning models was compared with other traditional scoring systems which renders our study more robust. Finally, included studies used different kinds of database including the MIMIC-III (v1.3) database which is large and widely accepted standard database for critical care patients. This database is a publicly available database constructed by researchers at Massachusetts Institute of Technology's Laboratory for Computational Physiology, and the data were also de-identified in compliance with HIPAA. It includes such items as patient vital signs, hospital records, fluid information, laboratory test results, treatment orders, and free-text medical records. Additionally, multi-countries multi-database makes our results more reliable and trustworthy.

Our study has some limitations that also need to address. First, all the included studies evaluated different types of machine learning models for identifying sepsis patients. Therefore, our study does not allow to suggest which model is best for predicting sepsis patients. But all models in the included studies outperformed the traditional methods usually used in the hospital. Although, the prediction models may have to be different for each hospital/country because in each hospital/country, the reasons for sepsis may be different – again, there may be NO best overall or global model, just a best model for each hospital or each situation. As, all the models outperformed widely used scoring systems such as SOFA, QSOFA, and MEWS score. Therefore, they could be helpful in future clinical settings. Second, all the included almost common variables for the prediction of sepsis but variables were not categorized according to their importance. Although, machine learning prediction model has ability to provide which variables are most useful for predicting sepsis patients. The best machine learning model is to include precise and accurate variable selection for predicting sepsis patients. Third, only three studies provided results of external validation. It would have been better if all studies had provided external validation results. Finally, model discrimination and calibration information was not included.

6. Conclusion

The early prediction of sepsis onset is critical for providing effective medical care and intervention. Our study found that the machine learning prediction models performed better than the existing sepsis scoring systems such as SIRS, MEWS, SOFA, and qSOFA for identifying and predicting sepsis patients. Predicting sepsis patients using machine learning models could guide physicians to actively monitor and take preventive actions to improve the patients' condition. It would also identify patients most in need of medical support, reduce wasting healthcare resources, and increase the desired sensitivity or specificity, resulting in a decreased numbers of false alarms. Therefore, the findings of our study suggest that machine learning prediction models could be implemented in the hospital in order to significantly reduce the in-hospital mortality rate, unnecessary hospital stay, and cost of treatment. An increased accuracy of sepsis identification could lead to better patient safety and, at the same time, save millions of dollars in large clinical settings. However, more studies are warranted to use the various multi-center databases, and more precise clinical variables need to be included to predict sepsis.

Conflict of interest

None.

Financial disclosure

None.

Ethical standard

This study has been approved by our Institution's Ethics Committee and has, therefore, been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Author contribution

Idea generation, data collection, statistical analysis, manuscript writing: Md. Mohaimenul Islam, Tahmina Nasrin Poly;

Data curation, manuscript writing and editing: Bruno Andreas Walther, Chieh-Chen Wu, Hsuan-Chia Yang;

Idea generation and investigation: Yu-Chuan (Jack) Li.

Acknowledgment

This work was financially supported by the "TMU Research Center of Cancer Translational Medicine" from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan grant number: MOST 106-2622-8-038-001-TS1.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2018.12.027.

References

[1] C. Fleischmann, A. Scherag, N.K. Adhikari, C.S. Hartog, T. Tsaganos, P. Schlattmann, D.C. Angus, K. Reinhart, Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations, Am. J. Respir. Crit. Care Med. 193 (2016) 259–272.
[2] S.A. Novosad, Vital signs: epidemiology of sepsis: prevalence of health care factors and opportunities for prevention, MMWR Morb. Mortal. Wkly. Rep. (2016) 65.
[3] D.F. Gaieski, J.M. Edwards, M.J. Kallan, B.G. Carr, Benchmarking the incidence and mortality of severe sepsis in the United States, Crit. Care Med. 41 (2013) 1167–1174.
[4] M. Singer, C.S. Deutschman, C.W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G.R. Bernard, J.-D. Chiche, C.M. Coopersmith, The third international consensus definitions for sepsis and septic shock (sepsis-3), JAMA 315 (2016) 801–810.
[5] J.S. Bradley, R. Guidos, S. Baragona, J.G. Bartlett, E. Rubinstein, G.G. Zhanel, M.D. Tino, D.L. Pompliano, F. Tally, P. Tipirneni, Anti-infective research and development—problems, challenges, and solutions, Lancet Infect. Dis. 7 (2007) 68–78.
[6] P. Bhattacharjee, D.P. Edelson, M.M. Churpek, Identifying patients with sepsis on the hospital wards, Chest 151 (2017) 898–907.
[7] E. Gultepe, J.P. Green, H. Nguyen, J. Adams, T. Albertson, I. Tagkopoulos, From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system, J. Am. Med. Inform. Assoc. 21 (2013) 315–325.
[8] S. Mani, A. Ozdas, C. Aliferis, H.A. Varol, Q. Chen, R. Carnevale, Y. Chen, J. Romano-Keeler, H. Nian, J.-H. Weitkamp, Medical decision support using machine learning for early detection of late-onset neonatal sepsis, J. Am. Med. Inform. Assoc. 21 (2014) 326–336.
[9] S.M. Vieira, L.F. Mendonça, G.J. Farinha, J.M. Sousa, Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients, Appl. Soft Comput. 13 (2013) 3494–3504.
[10] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, M.D. Feldman, C. Barton, Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach, JMIR Med. Inform. (2016) 4.
[11] Q. Mao, M. Jay, J. Hoffman, J. Calvert, C. Barton, D. Shimabukuro, L. Shieh, U. Chettipally, G. Fletcher, Y. Kerem, Multicenter validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU, bioRxiv (2018) 243964.
[12] D.W. Shimabukuro, C.W. Barton, M.D. Feldman, S.J. Mataraso, R. Das, Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial, BMJ Open Respir. Res. 4 (2017) e000234.
[13] M.M. Islam, U. Iqbal, B. Walther, S. Atique, N.K. Dubey, P.-A. Nguyen, T.N. Poly, J.H.B. Masud, Y.-C.J. Li, S.-A. Shabbir, Benzodiazepine use and risk of dementia in the elderly population: a systematic review and meta-analysis, Neuroepidemiology 47 (2016) 181–191.
[14] T.N. Poly, M.M. Islam, B.A. Walther, H.-C. Yang, P.-A. Nguyen, C.-W. Huang, S.-A. Shabbir, Y.-C.J. Li, Exploring the Association between statin use and the risk of Parkinson's disease: a meta-analysis of observational studies, Neuroepidemiology 49 (2017) 142–151.
[15] M.M. Islam, H.-C. Yang, P.-A. Nguyen, T.N. Poly, C.-W. Huang, S. Kekade, A.M. Khalfan, T. Debnath, Y.-C.J. Li, S.S. Abdul, Exploring association between statin use and breast cancer risk: an updated meta-analysis, Arch. Gynecol. Obstet. 296 (2017) 1043–1053.
[16] P.F. Whiting, A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, M.M. Leeflang, J.A. Sterne, P.M. Bossuyt, QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, Ann. Intern. Med. 155 (2011) 529–536.
[17] R.C. Bone, R.A. Balk, F.B. Cerra, R.P. Dellinger, A.M. Fein, W.A. Knaus, R.M. Schein, W.J. Sibbald, Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis, Chest 101 (1992) 1644–1655.
[18] J.P. Higgins, S.G. Thompson, J.J. Deeks, D.G. Altman, Measuring inconsistency in meta-analyses, BMJ 327 (2003) 557.
[19] L.M. Leemis, K.S. Trivedi, A comparison of approximate interval estimators for the Bernoulli parameter, Am. Stat. 50 (1996) 63–68.
[20] C. Chatfield, J. Zidek, J. Lindsey, An Introduction to Generalized Linear Models, Chapman and Hall/CRC, 2010.
[21] L.E. Moses, D. Shapiro, B. Littenberg, Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations, Stat. Med. 12 (1993) 1293–1316.
[22] H.J. Kam, H.Y. Kim, Learning representations for the early detection of sepsis with deep neural networks, Comput. Biol. Med. 89 (2017) 248–255.
[23] J.S. Calvert, D.A. Price, U.K. Chettipally, C.W. Barton, M.D. Feldman, J.L. Hoffman, M. Jay, R. Das, A computational approach to early sepsis detection, Comput. Biol. Med. 74 (2016) 69–73.
[24] M. Faisal, A. Scally, D. Richardson, K. Beatson, R. Howes, K. Speed, M.A. Mohammed, 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. 46 (2018) 612–618.
[25] S. Horng, D.A. Sontag, Y. Halpern, Y. Jernite, N.I. Shapiro, L.A. Nathanson, Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning, PLoS One 12 (2017) e0174708.
[26] Q. Mao, M. Jay, J.L. Hoffman, J. Calvert, C. Barton, D. Shimabukuro, L. Shieh, U. Chettipally, G. Fletcher, Y. Kerem, Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU, BMJ open 8 (2018) e017833.
[27] S. Nemati, A. Holder, F. Razmi, M.D. Stanley, G.D. Clifford, T.G. Buchman, An interpretable machine learning model for accurate prediction of sepsis in the ICU, Crit. Care Med. 46 (2018) 547–553.
[28] S. Nemati, Looking into the seeds of time to say which fevers will grow and which will not, Crit. Care Med. 46 (2018) 651–653.
[29] V. Liu, G.J. Escobar, J.D. Greene, J. Soule, A. Whippy, D.C. Angus, T.J. Iwashyna, Hospital deaths in patients with sepsis from 2 independent cohorts, JAMA 312 (2014) 90–92.
[30] D.C. Angus, W.T. Linde-Zwirble, J. Lidicker, G. Clermont, J. Carcillo, M.R. Pinsky, Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care, Crit. Care Med. 29 (2001) 1303–1310.
[31] A.J. Goodwin, D.A. Rice, K.N. Simpson, D.W. Ford, Frequency, cost and risk factors of readmissions among severe sepsis survivors, Crit. Care Med. 43 (2015) 738.
[32] A. Pfuntner, L. Wier, C. Steiner, Costs for Hospital Stays in the United States, 2010. HCUP Statistical Brief# 146, Agency for Healthcare Research and Quality, Rockville, MD, 2013 2015.
[33] E. Rivers, B. Nguyen, S. Havstad, J. Ressler, A. Muzzin, B. Knoblich, E. Peterson, M. Tomlanovich, Early goal-directed therapy in the treatment of severe sepsis and septic shock, N. Engl. J. Med. 345 (2001) 1368–1377.
[34] H.B. Nguyen, S.W. Corbett, R. Steele, J. Banta, R.T. Clark, S.R. Hayes, J. Edwards, T.W. Cho, W.A. Wittlake, Implementation of a bundle of quality indicators for the early management of severe sepsis and septic shock is associated with decreased mortality, Crit. Care Med. 35 (2007) 1105–1112.
[35] M.M. Levy, M.P. Fink, J.C. Marshall, E. Abraham, D. Angus, D. Cook, J. Cohen, S.M. Opal, J.-L. Vincent, G. Ramsay, ccm/esicm/accp/ats/sis international sepsis definitions conference, Intensive Care Med. 29 (2001 s2003) 530–538.
[36] S.A. Sterling, M.A. Puskarich, A.F. Glass, F. Guirgis, A.E. Jones, The impact of the Sepsis-3 septic shock definition on previously defined septic shock patients, Crit. Care Med. 45 (2017) 1436–1442.
[37] C. Rhee, R. Dantes, L. Epstein, D.J. Murphy, C.W. Seymour, T.J. Iwashyna, S.S. Kadri, D.C. Angus, R.L. Danner, A.E. Fiore, Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014, JAMA 318 (2017) 1241–1249.
[38] V. Herasevich, M.S. Pieper, J. Pulido, O. Gajic, Enrollment into a time sensitive clinical study in the critical care setting: results from computerized septic shock sniffer implementation, J. Am. Med. Inform. Assoc. 18 (2011) 639–644.
[39] J.L. Nelson, B.L. Smith, J.D. Jared, J.G. Younger, Prospective trial of real-time electronic surveillance to expedite early care of severe sepsis, Ann. Emerg. Med. 57 (2011) 500–504.