Machine learning in infection management using routine electronic health records

Luz, C. F.; Vollmer, M.; Decruyenaere, J.; Nijsten, M. W.; Glasner, C.; Sinha, B.

Published in:
Clinical Microbiology and Infection

DOI:
10.1016/j.cmi.2020.02.003

Publication date:
2020

Clinical Microbiology and Infection 26 (2020) 1291–1299

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Review

Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies

C.F. Luz [1,*], M. Vollmer [2], J. Decruyenaere [3], M.W. Nijsten [4], C. Glasner [1], B. Sinha [1]

[1] University of Groningen, University Medical Center Groningen, Department of Medical Microbiology and Infection Prevention, Groningen, the Netherlands
[2] Institute of Bioinformatics, University Medicine Greifswald, Greifswald, Germany
[3] Ghent University, Ghent University Hospital, Department of Intensive Care, Ghent, Belgium
[4] University of Groningen, University Medical Center Groningen, Department of Critical Care, Groningen, the Netherlands

ARTICLE INFO

Article history:
Received 19 November 2019
Received in revised form
1 February 2020
Accepted 3 February 2020
Available online 13 February 2020

Editor: L Leibovici

Keywords:
Algorithms
Artificial intelligence
Electronic health records
Infection
Inpatient
Machine learning
Methods
Review

ABSTRACT

Background: Machine learning (ML) is increasingly being used in many areas of health care. Its use in infection management is catching up as identified in a recent review in this journal. We present here a complementary review to this work.
Objectives: To support clinicians and researchers in navigating through the methodological aspects of ML approaches in the field of infection management.
Sources: A Medline search was performed with the keywords artificial intelligence, machine learning, infection*, and infectious disease* for the years 2014–2019. Studies using routinely available electronic hospital record data from an inpatient setting with a focus on bacterial and fungal infections were included.
Content: Fifty-two studies were included and divided into six groups based on their focus. These studies covered detection/prediction of sepsis ($n = 19$), hospital-acquired infections ($n = 11$), surgical site infections and other postoperative infections ($n = 11$), microbiological test results ($n = 4$), infections in general ($n = 2$), musculoskeletal infections ($n = 2$), and other topics (urinary tract infections, deep fungal infections, antimicrobial prescriptions; $n = 1$ each). In total, 35 different ML techniques were used. Logistic regression was applied in 18 studies followed by random forest, support vector machines, and artificial neural networks in 18, 12, and seven studies, respectively. Overall, the studies were very heterogeneous in their approach and their reporting. Detailed information on data handling and software code was often missing. Validation on new datasets and/or in other institutions was rarely done. Clinical studies on the impact of ML in infection management were lacking.
Implications: Promising approaches for ML use in infectious diseases were identified. But building trust in these new technologies will require improved reporting. Explainability and interpretability of the models used were rarely addressed and should be further explored. Independent model validation and clinical studies evaluating the added value of ML approaches are needed. **C.F. Luz, Clin Microbiol Infect 2020;26:1291**

Introduction

Artificial intelligence (AI), defined as computer algorithms exhibiting cognitive-like features such as learning capabilities, is already impacting our lives in various areas. In the medical field AI-supported image analysis has already gained an important role in radiology, dermatology, and pathology [1]. In another data-intense discipline, genomics, AI helps to predict phenotypes from genotypes [2].

Computer algorithms for AI rely largely on machine learning (ML) techniques in a broad sense, including natural language processing and computer vision [3]. A recent review on ML in healthcare epidemiology defined ML as the study of tools and methods for identifying patterns in data [4]. ML techniques

* Corresponding author. C.F. Luz, Department of Medical Microbiology and Infection Prevention, University Medical Center Groningen, Hanzeplein 1, 9713 GZ Groningen, the Netherlands.
E-mail address: c.f.luz@umcg.nl (C.F. Luz).

https://doi.org/10.1016/j.cmi.2020.02.003
1198-743X/

constitute a diverse set of algorithms (e.g. logistic regression, decision trees, or deep learning) that can be categorized into supervised, unsupervised, and reinforcement learning techniques (Fig. 1). While unsupervised learning provides methods for clustering data, supervised learning is focused on classification. A detailed introduction to the background of ML in health care has recently been published [5].

Patient medical data are increasingly being stored as electronic health records (EHRs) at healthcare institutions worldwide. For example, hospitals in high-income countries use EHRs containing basic functionalities such as patient demographics, physician notes, nursing assessments, patient problem lists, patient medication lists, discharge summaries, radiology reports, diagnostic test results, and order entries for medications [6]. However, EHRs can be inconsistent and noisy, may contain many missing values, and frequently include unstructured text fields. Nevertheless, the very fact that these data are electronically available in large volumes provides the potential for applying ML, including in the field of infection management. For example, life-threatening conditions like sepsis require immediate diagnostic and therapeutic actions at a time when the causative pathogen is often still unknown [7]. Early identification of septic patients through ML-derived prediction models could improve and facilitate patient care in situations where 'time is life' [8].

A recent review listed current applications of ML for clinical decision support in infectious diseases and identified different aims such as support of diagnosis, severity prediction, and choice of appropriate antimicrobial treatment [9]. This resource provides a comprehensive overview of objectives and characteristics of ML systems with a special focus on variable selection for ML approaches. Our aim is complementary to this work in identifying and describing different methodological approaches, performance measures, and future methodological requirements for optimal use of ML in infection management.

Methods

We queried the MEDLINE database for articles published within the last 5½ years (2014-01-01 until 2019-08-20) using the terms *machine learning* or *artificial intelligence* and *infection** or *infectious disease**. The time span was selected to focus on recent developments that are in line with current technologies available to researchers and clinicians. Articles were included and assessed if they used routinely available EHR data in a retrospective manner with a focus on bacterial and fungal infection management. Exclusion criteria were pure laboratory or image data analyses (e.g. whole-genome sequencing data, CT scans), free text-based analyses (e.g. natural language processing), reviews and commentaries, or other non-routine data collection (e.g. phone follow-up questionnaire data as part of a clinical trial). This search strategy and the inclusion criteria are similar to those of a recent, complementary review [9]. Yet, while this review included search terms related to decision-making, we took a broader approach without such terms to also include studies with more experimental study designs that might not yet focus on direct clinical application. This is important to capture and assess different methodological approaches in this early phase of ML applications in infection management.

Articles were assessed on their research focus (e.g. sepsis, hospital-acquired infections) and mapped to research areas. Definitions of the predicted/labelled outcome (e.g. sepsis defined by diagnostic code) were evaluated. Furthermore, information on the used ML model techniques and model performance—area under the receiver operating characteristic curve (AUROC), sensitivity, specificity—were extracted. Additionally, study size, number of

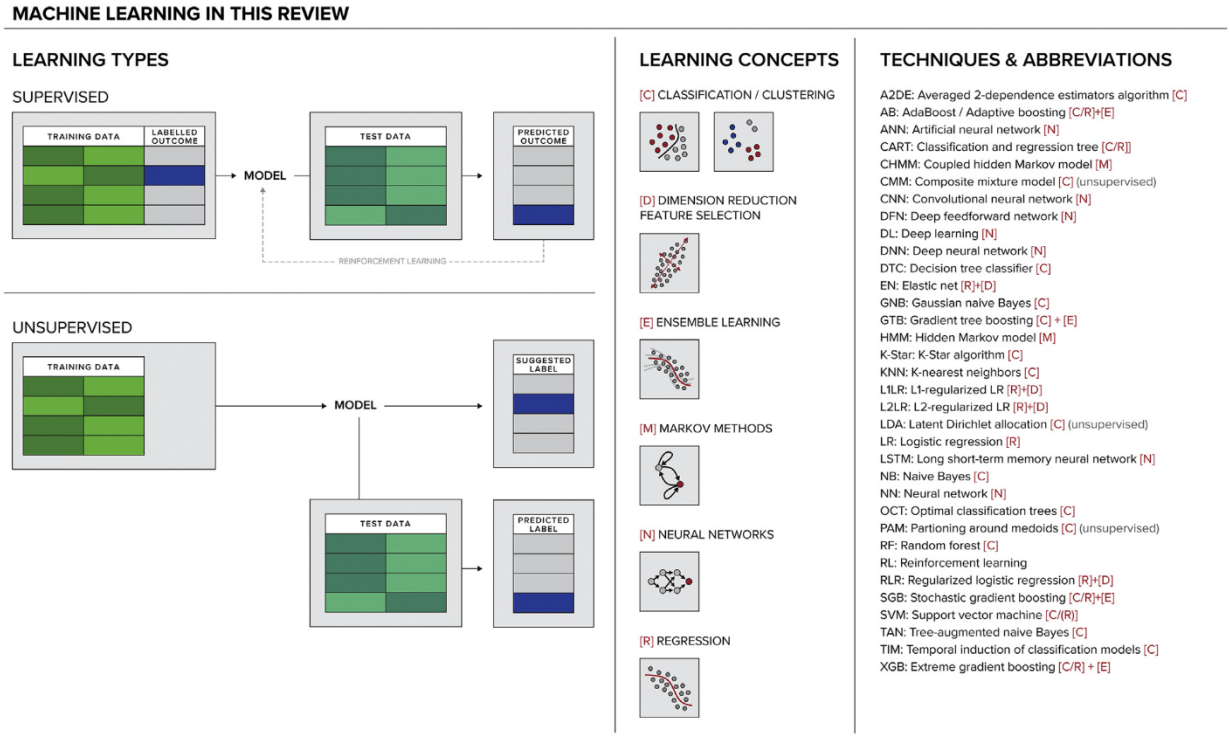

Fig. 1. Machine learning types and concepts of included studies. Supervised learning requires labelled outcomes ('ground truth', e.g. sepsis = TRUE or FALSE) in the training phase whereas unsupervised learning can suggest/predict labels. Icons display simplified algorithm setups/learning concepts. Letters in red behind abbreviations and names indicate the learning concept(s) behind an algorithm/technique. Multiple concepts per algorithm/technique are possible.

variables/features, availability of software code, missing data handling, and information on external model validation or trials were assessed.

Results

We identified 52 articles (12 from 2019, 17 from 2018, seven from 2017, eight from 2016, four from 2015, and four from 2014). After assessing the research focus of the articles, six research areas could be determined: sepsis ($n = 19$), hospital-acquired infections (HAIs; $n = 11$), surgical site infections (SSIs) and other postoperative infections ($n = 11$), microbiological test results ($n = 4$), infections in general ($n = 2$), musculoskeletal infections ($n = 2$), and other focuses: urinary tract infections (UTIs), deep fungal infections, antimicrobial prescriptions, each $n = 1$ (Table 1).

Data underlying identified machine learning studies

The study size (i.e. number of patients) of all included studies ranged from 148 to 500,000 with a median of 5471. The median number of features/variables included was 30 (range: 2–23,968). Summary characteristics of the identified studies per research area are described in Table 1.

Data were extracted from local EHR systems in 42 studies (81%). Eleven studies [10–20] used a publicly available critical care database (MIMIC version II and III) [21].

Pre-processing of data was reported to varying degrees. Missing data handling was not described in 39% of the studies ($n = 20$). When reported, different strategies were applied: e.g. transformation to binary variables indicating missingness [22], carry-forward of last observation [12,14,20,23–26], including complete cases only [18,27–30], or applying multiple imputation [11,17,31–35]. Two studies assessed the effect of missing data on model performance through a stepwise introduction of missing variables [36,37]. Class imbalance of the labelled outcome variable was explicitly mentioned if applicable in 39% ($n = 18$) of the studies.

Of the 41 studies (79%) reporting on the software used for modelling, 30 (73.2%) used open-source programming languages such as R or Python. However, only six (11.5%) studies made their code publicly available [38–43].

Machine learning techniques in use

Overall, 35 different ML techniques were used in the 52 included studies. Logistic regression (LR) was used in 18 studies (51.4%) but usually as a reference to reflect traditional model approaches. Random forest (RF), support vector machines (SVM), and artificial neural networks (ANN) were used in 18 (51.4%), 12 (34.3%), and seven (20.0%) studies, respectively.

Forty-eight studies (92%) used supervised learning approaches, i.e. developing the model by training using input data with the corresponding output label (Fig. 1). Three studies used an unsupervised clustering approach for clustering sepsis patients [31,42,44]. Data from sepsis patients were also used to learn optimal treatment by reinforcement learning [19].

In 43 studies (82.7%) an AUROC was reported, of which 27 (62.8%) achieved an AUROC >0.80, which is considered as an excellent discrimination performance [45]. Sensitivity and specificity, however not relevant for all ML tasks, were reported by 32 (61.5%) studies (Fig. 2).

Thirty studies (57.7%) compared multiple ML techniques. Among these studies, the best performing techniques per research area were long short-term memory networks (LSTM) in the sepsis group [12,46], ANN in the HAI group [28], L1-regularized logistic regression (L1LR) in the SSI and other postoperative infections group [26], SVM in the infections (general) group [37], classification and regression tree (CART) in the microbiological test results group [47], and stochastic gradient boosting (SGB) in the musculoskeletal infections group [35]. However, as outlined below, the definition of the predicted outcome can be very heterogenous. Thus, comparability of the different approaches per research area is limited and should be interpreted with great caution.

Predicted outcomes per research area

When applying supervised ML techniques, the outcome of interest—i.e. the event that should be predicted—needs to be labelled (e.g. sepsis onset defined by ICD code together with a time stamp). The predicted outcome in the identified studies are very diverse and differ greatly even within a given research area (Table 2).

A detailed list of all studies included in this review can be found in the Appendix.

Discussion

Even within each research area remarkable differences were found. The following sections highlight the great heterogeneity between the identified studies, with a focus on challenges in predicted outcome definitions, outcome and data complexity, reporting standards, model performance and validation, and model interpretability. For a detailed description and assessment of patient variables used for ML models, readers are referred to the complementary review of Peiffer-Smadja et al. in this journal [9].

Table 1
Summary characteristics per research area

Research area	*n*	Features/variables (*n*)		Study size (n)		Best model	Max AUROC
		Median	Range	Median	Range		
Sepsis	19	17	2–3058	5803	242–122,670	LSTM	0.96
Hospital-acquired infections	11	54	19–23,968	11,251	148– 256,732	ANN	0.92
SSI and other postoperative infections	11	50	10–9828	5214	1005–483,686	L1LR	0.96
Microbiological test results	4	92	50–134	1234	376–3327	CART	0.77
Infection (general)	2	6	6	330,154	160,307–500,000	SVM	0.84
Musculoskeletal infections	2	73	68–78	710	367–1053	SGB	0.89
Other (antimicrobial prescription)	1	–	–	2,442	–	TIM	–
Other (deep fungal infection)	1	19	–	696	–	ANN	0.83
Other (urinary tract infections)	1	211	–	80,387	–	XGB	0.90

AUROC, area under the receiver operating characteristic curve; LSTM, long short-term memory neural network; AN, artificial neural network; L1LR, L1-regularized logistic regression; CART, classification and regression tree; SVM, support vector machines; SGB, stochastic gradient boosting; TIM, temporal induction of classification models; XGB, extreme gradient boosting.

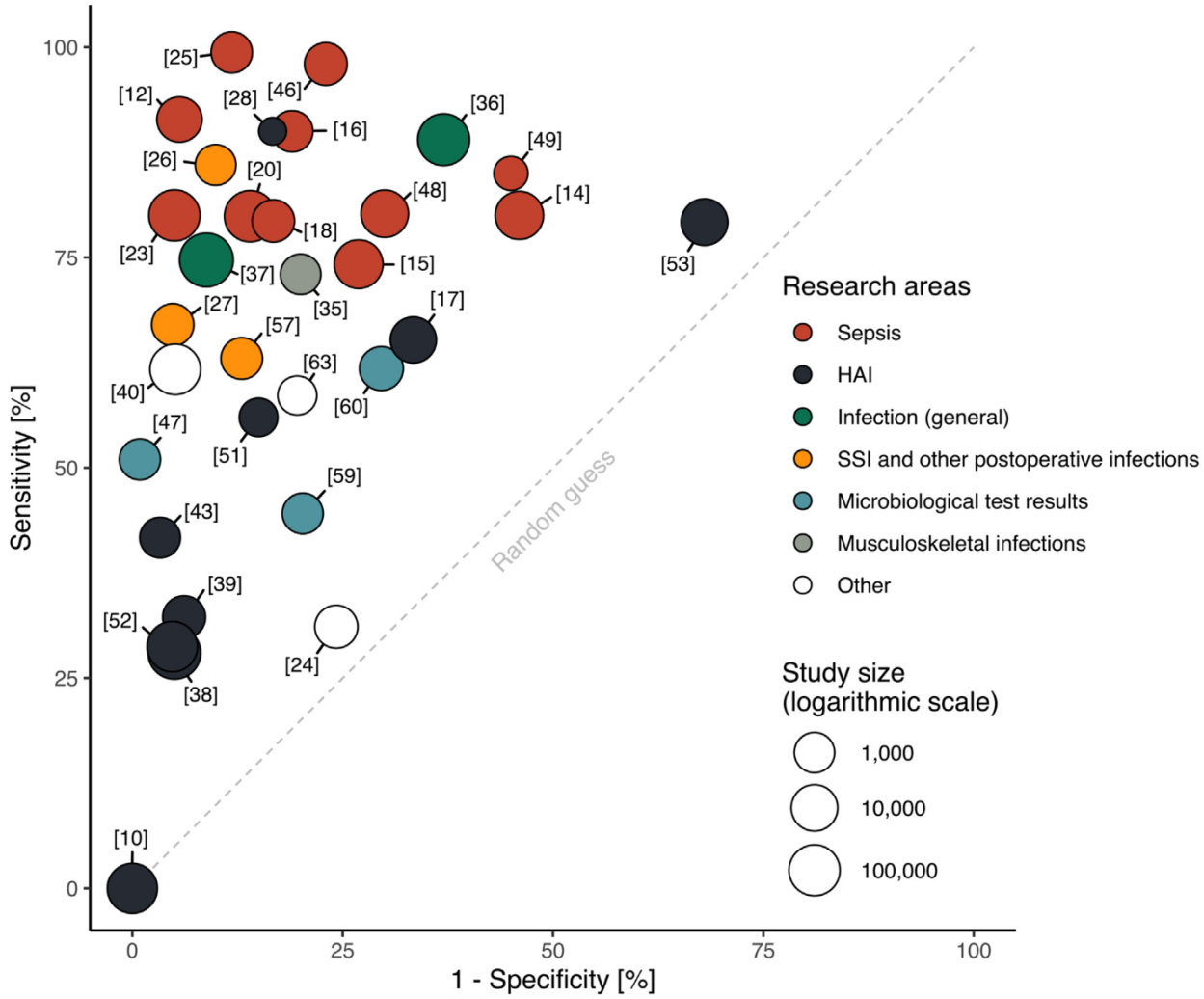

Fig. 2. Receiver operating characteristic space of reviewed articles with reference number in brackets; model performance in relation to the research area and study size for studies reporting performance measures ($n = 32$). Sensitivity and specificity of the applied models are plotted for the predicted outcome (e.g. sepsis). Due to very heterogeneous studies, outcomes, and outcome definitions, comparability is limited and should be done with great caution. HAI, hospital-acquired infection; SSI, surgical site infection.

Challenges in predicted outcome definitions

The identified studies could be grouped into several research areas. Studies on the identification/prediction of sepsis predominate, underscoring a special interest in this field. A similar relevance of this group was found in the complementary review [9]. Severely ill sepsis patients are usually admitted to ICUs with extensive data collection through monitoring and testing that can be used for ML approaches. Thirteen (68%) of the 19 studies in the sepsis research area were based on ICU data only, of which nine (69%) used publicly accessible data [21]. Although similar data were used, the predicted outcomes used by the 19 sepsis studies—bacteraemia, ICD codes, clinical sepsis definitions, clusters of septic patients, or sepsis-associated mortality—were very heterogeneous and thus difficult to compare. This heterogeneity was also observed in the group of SSI and other postoperative infections. National guidelines for these complications were used by some studies to define the predicted outcome [29,56,64]. Nonetheless, different complications differ in their predictability, as seen in one study using optimal classification trees (OCTs) with performances ranging from an AUROC of 0.67 for superficial SSI to 0.93 for septic shock [56]. The authors hypothesized in their openly available response to reviewers that performance values for specific outcomes, such as superficial SSI, tend to be much lower due to the collected data being less suitable for predicting less severe or local complications. Instead the data better reflect patients' general state and major complications.

This phenomenon is also underscored by a study in the HAI group predicting central line placement and central-line-associated bloodstream infections (CLABSIs) [10]. The ML model achieved an AUROC of only 0.64 when modelling CLABSI but 0.82 for modelling the placement of a central line. Given the logic above (data being more predictive for general states and major complications), it appears reasonable to argue that the placement of a central line is a proxy for the patient's worsening condition and thus is easier to model more accurately. It also needs to be kept in mind that modelling the placement of a line may, in fact, be modelling physicians' behaviour. This behaviour is likely to be very much informed by the available data and thus be of better use for modelling.

While antimicrobial resistance (AMR) is a major concern in infection management, only three studies addressed this topic [24,60,61], one of which falls into the area of antimicrobial stewardship [24]. Since AMR constitutes an increasing threat to patients

Table 2
Predicted outcomes per research area

Research area	Definition of the predicted outcome	Reference
Sepsis	Bacteraemia (positive blood culture result)	[32,46,48]
	Sepsis based on ICD codes	[11–13,15,16,20,23]
	Clinical definition of sepsis (e.g. the third international consensus definitions for sepsis and septic shock—Sepsis-3)	[14,19,25,49]
	Clustering septic patients (unsupervised)	[42,44]
	Sepsis-associated mortality	[18,22,31]
HAI	Line-associated infections	[10,28,39,50]
	HAI defined by ICD codes	[17]
	Clinical definition of HAI (clinical manifestation plus microbiological confirmation)	[51]
	Clostridium difficile infections	[38,41,43,52,53]
SSI and other postoperative infections	"Any case of opening a wound or use of antibiotics" (original quote from the authors)	[33]
	"Any superficial, deep, organ space SSI and wound dehiscence" according to national guidelines (NSQIP/NSQIP-P)	[29,54]
	ICD-10 or NOMESCO Classification of Surgical Procedures (NCSP) codes	[26,34,55]
	18 individual postoperative complications according to national guidelines (ACS-NSQIP)	[56]
	SSI according to CDC and clinical diagnosis	[27,30,57]
	Postoperative infection on the basis of local sets of quality indicators	[58]
Microbiological test results	Infections with microorganisms with an extended-spectrum β-lactamase (ESBL) in bacteraemic patients	[47]
	Gram-stain of isolates from a positive blood culture	[59]
	Colonization with carbapenem-resistant organisms (CREs)	[60]
	Infections with methicillin-resistant *Staphylococcus aureus* (MRSA) in bacteraemic patients	[61]
Infections (general)	"Pre-determined clinical case definition" (original quote from the authors)	[36]
	"Any type of infection"; positive microbiological culture (original quote from the authors)	[37]
Musculoskeletal infections	"Failure of non-operative management" (original quote from the authors)	[62]
	Spinal epidural abscess-associated mortality	[35]
Other	Identification of inappropriate antimicrobial prescriptions	[24]
	Deep fungal infections based on "clinical manifestations, laboratory tests and/or identification of fungi" (original quote from the authors)	[63]
	Urinary tract infections (positive urine culture defined by $>10^4$ colony forming units/high-powered field)	[40]

ICD, International Classification of Diseases; HAI, hospital-acquired infection; SSI, surgical site infection; CDC, Centers for Disease Control and Prevention.

and healthcare systems worldwide [65,66], it would be worthwhile to further explore the use of ML as a component of innovative solutions. Due to the use of outpatient data, a recent study on the prediction of AMR in urinary tract infections based on personal clinical history was not included in this review [67]. However, it presented an interesting approach using a gradient-tree-boosting (GTB) algorithm that enabled personalized drug-specific predictions of AMR in patients with urinary tract infections. Moreover, this study retrospectively evaluated the rate of mismatched empirical treatment recommendations by physicians compared to algorithmic drug recommendations based on the ML predictions. The latter performed significantly better overall (30% lower mismatched treatment rate). In general, studies on ML approaches tend to report model performance measure only. A study design such as that described above can help to put model development into clinical perspective and to highlight the added value of the developed model.

Outcome and data complexity

Infection management is highly complex. The cumulative nature of information acquisition in clinical reasoning (e.g. turnaround time of diagnostics from bed to bench and back) creates scenarios where no specific/definite information might be available on the causative pathogen. At the same time, increasing AMR demands a responsible use of anti-infective therapies.

Modelling these medical processes is challenging. Machine learning techniques like LSTM can analyse complex time series data and try to follow a similar logic to that clinicians use. LSTM models constitute a specific set of ML techniques that have been shown to be successful when predicting non-infection patient outcomes in EHR [68]. LSTM models might be better suited to identifying/predicting sepsis and other infection management processes with their complexities as described above. Predicting positive blood cultures 12 h in advance in a group of ICU patients receiving blood culture tests resulted in an AUROC of 0.96 using this modelling technique [46]. Early sepsis detection based on defining the time of sepsis onset using ICD codes and timestamps achieved an AUROC of 0.93 [12].

The use of deep learning techniques (a special set of ML techniques) for EHR analysis has been described recently [69]. The authors provide useful technical background details about the techniques and assess their use for EHR data extraction, outcome prediction, or de-identification of EHR data. Although different in scope, similar aspects to this review could be identified as challenges: data heterogeneity, benchmarks, and model interpretability.

In contrast to supervised ML models, unsupervised approaches for cluster analysis were used by only three studies identified in our review [31,42]. All three focused on sepsis patients. Given the high complexity of these patients, unsupervised methods can provide promising potential to better understand this population. The studies aimed at identifying treatment patterns, clinical phenotypes, and mortality-associated patient groups (methods used: latent Dirichlet allocation [42], partitioning around medoids methods [44], and composite mixture models [31] respectively) (Fig. 1). It can be speculated that more studies on unsupervised methods for heterogeneous patient populations will follow, and it would be highly interesting to explore the potential of these methods further.

Reporting standards

It was recently pointed out that "*methodological, ethical, and data security standards*", when investigating ML and its application in healthcare, are greatly needed [5]. Looking at the heterogeneity of the articles included in our review we fully support this statement. For example, we found only six articles providing openly available code. This is particularly worrisome not only because of the debate on reproducibility [70] but also as ML tools are often being criticized as black boxes. The maintainers of the public MIMIC data repository lead the way in providing open source code

alongside the available data giving a good example for enabling reproducibility [21]. There has been work on the standardization of REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) [71]. The authors state that "*A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided.*" Furthermore, reports should include a "[...] *discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time* [...]" [71]. Furthermore, the guideline for transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) lists several checkpoints relevant to ML research [72]. We found only three studies (6%) that acknowledge following reporting guidelines such as TRIPOD [22,35,40].

Among the 52 articles in our review, 20 (39%) did not report how missing data were handled. This greatly hinders comparability, reproducibility, further use of the reported approaches, and trust in ML technologies. Routinely available data are not collected focusing on research. Thus, this can create challenges in accuracy and completeness [73]. Accounting for this fact through transparent methodology and reporting is even more important when using routine data for research and informing decision-makers based on the derived prediction models. Cleaning and transforming data are often said to be 80% of the entire work. This should be described in detail in scientific reports using EHR data.

Model performance and validation

The performance of supervised ML models is most often evaluated using measures such as sensitivity, specificity, positive and negative predictive values, and the AUROC in particular, which was also found in the majority of the studies in this review. However, the clinical usefulness of a model identified by a single best AUROC (of potentially hundreds of models) cannot be assessed by these standard traditional measures [74]. Furthermore, the clinical picture (e.g. in sepsis [75]) and the diagnosis of an infection itself is very heterogeneous and not always binary. Any binary approach could, thus, lead to overestimation of the effect of a diagnostic tool. In addition, staff availability, constrained workflows, and cost define the potential clinical impact of a given model. Additional analysis would be required and should be incorporated when evaluating clinical ML models. One good example was a study in this review that included financial aspects (costs of blood tests) in the model evaluation [26].

Essential next steps, prior to any application of ML models in clinical practice, are external model validation and clinical trials. While all of the identified studies in this review used some rules for splitting training and test data during model development (e.g. 80% training and 20% test data), only three studies validated the developed models on an independent, external dataset [19,23,25]. This hinders statements about generalizability or transferability. Furthermore, this could also introduce model bias. An ML model trained on a particular group of patients could perform less well in other subgroups or could even be systematically discriminating against specific populations (e.g. racial discrimination) [76]. First studies addressing health disparities using AI exist [77,78]. Studies such as a randomized controlled trial with a group of patients treated without the ML model information compared to a group of patients with this information available in real time would further validate the clinical effectiveness and usefulness of ML models. While clinical outcomes are important to establish (e.g. survival, length of stay in hospital, duration of antimicrobial therapy, rate of complications, readmission rate), insights into how clinicians will integrate ML in clinical practice are also highly needed. No randomized controlled trial could be identified in our review.

Model interpretability

With three exemptions, studies in our review did not explicitly highlight ML model explainability or interpretability in their objectives. However, efforts are also being made in other fields to increase the interpretability of ML models (Fig. 3) [79,80].

It is of great importance to further explore perceptions, interpretation, and trust in ML decision support with the potential users (i.e. physicians, other healthcare providers, and patients) and how to avoid the black box notion of ML applications. This could improve the 'algorithm literacy' of clinicians and help them to understand when to trust a model and when and why errors might occur [81]. In our review one study in the sepsis group that used reinforcement learning applied a random forest model to explain the developed model and reported these findings in their supplementary materials [19]. Two studies provided a web interface to study variable importance in the developed ML models [11,35]. However, just recently a debate has unfolded as to whether interpretable predictive models, such as logistic regression, should be preferred over explainable models (demonstrated in Fig. 3) when profound background knowledge is available (e.g. established risk factors for a specific disease) [82]. The author argues that using an explanatory model on top of a black box model cannot be as correct as the original model, otherwise the model could be used in the first place. Such an approach could lead to the risk that any explanatory method for a black box model could be an inaccurate representation of the original model [82]. The use of ML models should thus be based on thorough reasoning and not only on the availability of new and sophisticated technology.

Limitations

For this review, only the MEDLINE database was queried for relevant articles. The search term for infections might not detect all relevant articles. Systematically focusing on specific topics, such as sepsis, extended to additional databases (e.g. EMBASE, Google Scholar) will most likely capture additional studies that were not included in the umbrella term of infection. Nevertheless, it has

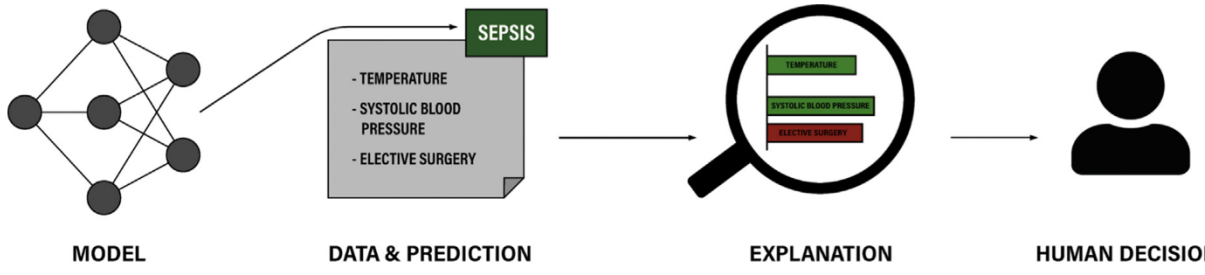

Fig. 3. Approach for explaining individual predictions highlighting the variables that led to the prediction. Adapted from [80] with the authors' permission.

Table 3
Core issues and recommendations for the development of machine learning models for infection management based on this narrative review

Core issue	Recommendation
Data and modelling	Clear definition of the predicted outcome
	Inclusion and evaluation of model interpretability and explainability with the end-user in mind (e.g. nurses, physicians)
Reporting	Adhering to reporting guidelines (e.g. RECORD, TRIPOD)
	Making software code openly available
	Using open-source programming languages
Study design	Model validation with external data (multicentre studies)
	Clinical evaluation of machine learning models
	Evaluation of the adoption and integration of machine learning models into clinical workflows by end-users

already become apparent that the large heterogeneity within the body of literature hindered a more detailed comparison of the included studies. Viral infections were not part of this review. The complementary work by Peiffer-Smadja et al. identified ten studies which, however, did not meet our inclusion criteria for study year, patient population, study data, or MEDLINE index [9]. They identified a special focus on ML and HIV infections. A systematic survey on this topic can provide further information [83]. Our review used a grouping approach to stratify the included studies. These groups of research areas are limited as SSI could also be grouped under HAI. However, for this review we decided to keep surgery-related infection separate from other infections based on our search results.

Conclusion

The use of AI/ML in the field of infection management is still in its infancy [84]. Our review revealed several promising approaches such as the use of LSTM for early sepsis detection. Three identified studies validated their results using external data. However, no clinical trial or clinical assessment of an ML model could be identified. The clinical utility of AI/ML, preferably in randomized controlled trials, has still to be demonstrated (Table 3). Moreover, in the future it will be crucial to explore the best ways to integrate AI/ML into clinical workflows as this was not yet part of the identified studies in this review. The global challenges of AI/ML in infectious diseases have also been recently described in a complementary review [9]. In addition, we have detected and highlighted the necessity for improved reporting, a stronger focus on explainability and interpretability, and clinical validation of developed approaches to secure the next steps towards a happy 'childhood' and a positive clinical impact.

Author contributions

CFL: conceptualization, writing original draft, reviewing, and editing; CFL, MV, JD, MWN, CG, and BS: writing, reviewing, and editing.

Transparency declaration

This work is part of a project funded by the European Commission Horizon 2020 Framework Marie Skłodowska-Curie Actions (grant agreement number: 713660-PRONKJEWAIL-H2020-MSCA-COFUND-2015).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cmi.2020.02.003.

References

[1] Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018;18:500–10. https://doi.org/10.1038/s41568-018-0016-5.
[2] Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A. A primer on deep learning in genomics. Nat Genet 2019;51:12–8. https://doi.org/10.1038/s41588-018-0295-5.
[3] Kaplan A, Haenlein M. Siri, Siri, in my hand: who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Bus Horiz 2019;62:15–25. https://doi.org/10.1016/j.bushor.2018.08.004.
[4] Wiens J, Shenoy ES. Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. Clin Infect Dis 2018;66:149–53. https://doi.org/10.1093/cid/cix731.
[5] Roth JA, Battegay M, Juchler F, Vogt JE, Widmer AF. Introduction to machine learning in digital healthcare epidemiology. Infect Control Hosp Epidemiol 2018;39:1457–62. https://doi.org/10.1017/ice.2018.265.
[6] Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. J Am Med Inform Assoc 2017;24:1142–8. https://doi.org/10.1093/jamia/ocx080.
[7] Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. Crit Care Med 2017;45:486–552. https://doi.org/10.1097/CCM.0000000000002255.
[8] Rello J, Valenzuela-Sánchez F, Ruiz-Rodriguez M, Moyano S. Sepsis: a review of advances in management. Adv Ther 2017;34:2393–411. https://doi.org/10.1007/s12325-017-0622-8.
[9] Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Pantelis G, Lescure F-X, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clin Microbiol Infect 2020;26:584–95. https://doi.org/10.1016/j.cmi.2019.09.009.
[10] Parreco JP, Hidalgo AE, Badilla AD, Ilyas O, Rattan R. Predicting central line-associated bloodstream infections and mortality using supervised machine learning. J Crit Care 2018;45:156–62. https://doi.org/10.1016/j.jcrc.2018.02.010.
[11] Saqib M, Sha Y, Wang MD. Early prediction of sepsis in EMR records using traditional ML techniques and deep learning LSTM networks. Conf Proc IEEE Eng Med Biol Soc 2018;2018:4038–41. https://doi.org/10.1109/EMBC.2018.8513254.
[12] Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. Comput Biol Med 2017;89:248–55. https://doi.org/10.1016/j.compbiomed.2017.08.015.
[13] Ghosh S, Li J, Cao L, Ramamohanarao K. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. J Biomed Inform 2017;66:19–31. https://doi.org/10.1016/j.jbi.2016.12.010.
[14] Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Med Inform 2016;4:e28. https://doi.org/10.2196/medinform.5909.
[15] Calvert J, Desautels T, Chettipally U, Barton C, Hoffman J, Jay M, et al. High-performance detection and early prediction of septic shock for alcohol-use disorder patients. Ann Med Surg (Lond) 2016;8:50–5. https://doi.org/10.1016/j.amsu.2016.04.023.
[16] Calvert JS, Price DA, Chettipally UK, Barton CW, Feldman MD, Hoffman JL, et al. A computational approach to early sepsis detection. Comput Biol Med 2016;74:69–73. https://doi.org/10.1016/j.compbiomed.2016.05.003.
[17] Warner JL, Zhang P, Liu J, Alterovitz G. Classification of hospital acquired complications using temporal clinical information from a large electronic health record. J Biomed Inform 2016;59:209–17. https://doi.org/10.1016/j.jbi.2015.12.008.
[18] Ribas Ripoll VJ, Vellido A, Romero E, Ruiz-Rodríguez JC. Sepsis mortality prediction with the quotient basis kernel. Artif Intell Med 2014;61:45–52. https://doi.org/10.1016/j.artmed.2014.03.004.
[19] Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018;24:1716–20. https://doi.org/10.1038/s41591-018-0213-5.

[20] Calvert J, Saber N, Hoffman J, Das R. Machine-learning-based laboratory developed test for the diagnosis of sepsis in high-risk patients. Diagnostics (Basel) 2019;9. https://doi.org/10.3390/diagnostics9010020.

[21] Johnson AEW, Pollard TJ, Shen L, Lehman L-WH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016;3:160035. https://doi.org/10.1038/sdata.2016.35.

[22] Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. Acad Emerg Med 2016;23:269–78. https://doi.org/10.1111/acem.12876.

[23] Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open 2018;8: e017833. https://doi.org/10.1136/bmjopen-2017-017833.

[24] Beaudoin M, Kabanza F, Nault V, Valiquette L. Evaluation of a machine learning capability for a clinical decision support system to enhance antimicrobial stewardship programs. Artif Intell Med 2016;68:29–36. https://doi.org/10.1016/j.artmed.2016.02.001.

[25] van Wyk F, Khojandi A, Kamaleswaran R. Improving prediction performance using hierarchical analysis of real-time data: a sepsis case study. IEEE J Biomed Health Inform 2019;23:978–86. https://doi.org/10.1109/JBHI.2019.2894570.

[26] Kocbek P, Fijacko N, Soguero-Ruiz C, Mikalsen KØ, Maver U, Povalej Brzan P, et al. Maximizing interpretability and cost-effectiveness of surgical site infection (SSI) predictive models using feature-specific regularized logistic regression on preoperative temporal data. Comput Math Methods Med 2019;2019:2059851. https://doi.org/10.1155/2019/2059851.

[27] Kuo P-J, Wu S-C, Chien P-C, Chang S-S, Rau C-S, Tai H-L, et al. Artificial neural network approach to predict surgical site infection after free-flap reconstruction in patients receiving surgery for head and neck cancer. Oncotarget 2018;9:13768–82. https://doi.org/10.18632/oncotarget.24468.

[28] Habibi Z, Ertiaei A, Nikdad MS, Mirmohseni AS, Afarideh M, Heidari V, et al. Predicting ventriculoperitoneal shunt infection in children with hydrocephalus using artificial neural network. Childs Nerv Syst 2016;32:2143–51. https://doi.org/10.1007/s00381-016-3248-2.

[29] Bartz-Kurycki MA, Green C, Anderson KT, Alder AC, Bucher BT, Cina RA, et al. Enhanced neonatal surgical site infection prediction model utilizing statistically and clinically significant variables in combination with a machine learning algorithm. Am J Surg 2018;216:764–77. https://doi.org/10.1016/j.amjsurg.2018.07.041.

[30] Gowd AK, Agarwalla A, Amin NH, Romeo AA, Nicholson GP, Verma NN, et al. Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. J Shoulder Elbow Surg 2019. https://doi.org/10.1016/j.jse.2019.05.017.

[31] Mayhew MB, Petersen BK, Sales AP, Greene JD, Liu VX, Wasson TS. Flexible, cluster-based analysis of the electronic medical record of sepsis with composite mixture models. J Biomed Inform 2018;78:33–42. https://doi.org/10.1016/j.jbi.2017.11.015.

[32] Ratzinger F, Haslacher H, Perkmann T, Pinzan M, Anner P, Makristathis A, et al. Machine learning for fast identification of bacteraemia in SIRS patients treated on standard care wards: a cohort study. Sci Rep 2018;8:12233. https://doi.org/10.1038/s41598-018-30236-9.

[33] Weller GB, Lovely J, Larson DW, Earnshaw BA, Huebner M. Leveraging electronic health records for predictive modeling of post-surgical complications. Stat Methods Med Res 2018;27:3271–85. https://doi.org/10.1177/0962280217696115.

[34] Soguero-Ruiz C, Fei WME, Jenssen R, Augestad KM, Álvarez J-LR, Jiménez IM, et al. Data-driven temporal prediction of surgical site infection. AMIA Ann Symp Proc 2015;2015:1164–73.

[35] Karhade AV, Shah AA, Bono CM, Ferrone ML, Nelson SB, Schoenfeld AJ, et al. Development of machine learning algorithms for prediction of mortality in spinal epidural abscess. Spine J 2019;19:1950–9. https://doi.org/10.1016/j.spinee.2019.06.024.

[36] Rawson TM, Hernandez B, Moore LSP, Blandy O, Herrero P, Gilchrist M, et al. Supervised machine learning for the prediction of infection on admission to hospital: a prospective observational cohort study. J Antimicrob Chemother 2018. https://doi.org/10.1093/jac/dky514.

[37] Hernandez B, Herrero P, Rawson TM, Moore LSP, Evans B, Toumazou C, et al. Supervised learning for infection risk inference using pathology data. BMC Med Inform Decis Mak 2017;17:168. https://doi.org/10.1186/s12911-017-0550-1.

[38] Oh J, Makar M, Fusco C, McCaffrey R, Rao K, Ryan EE, et al. A generalizable, data-driven approach to predict daily risk of *Clostridium difficile* infection at two large academic health centers. Infect Control Hosp Epidemiol 2018;39: 425–33. https://doi.org/10.1017/ice.2018.16.

[39] Savin I, Ershova K, Kurdyumova N, Ershova O, Khomenko O, Danilov G, et al. Healthcare-associated ventriculitis and meningitis in a neuro-ICU: incidence and risk factors selected by machine learning approach. J Crit Care 2018;45: 95–104. https://doi.org/10.1016/j.jcrc.2018.01.022.

[40] Taylor RA, Moore CL, Cheung K-H, Brandt C. Predicting urinary tract infections in the emergency department with machine learning. PLoS One 2018;13: e0194085. https://doi.org/10.1371/journal.pone.0194085.

[41] Pak TR, Chacko KI, O'Donnell T, Huprikar SS, van Bakel H, Kasarskis A, et al. Estimating local costs associated with *Clostridium difficile* infection using machine learning and electronic medical records. Infect Control Hosp Epidemiol 2017;38:1478–86. https://doi.org/10.1017/ice.2017.214.

[42] Fohner AE, Greene JD, Lawson BL, Chen JH, Kipnis P, Escobar GJ, et al. Assessing clinical heterogeneity in sepsis through treatment patterns and machine learning. J Am Med Inform Assoc 2019. https://doi.org/10.1093/jamia/ocz106.

[43] Li BY, Oh J, Young VB, Rao K, Wiens J. Using machine learning and the electronic health record to predict complicated *Clostridium difficile* infection. Open Forum Infect Dis 2019;6:ofz186. https://doi.org/10.1093/ofid/ofz186.

[44] Guilamet MCV, Bernauer M, Micek ST, Kollef MH. Cluster analysis to define distinct clinical phenotypes among septic patients with bloodstream infections. Medicine (Baltimore) 2019;98:e15276. https://doi.org/10.1097/MD.0000000000015276.

[45] Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression. John Wiley & Sons; 2013.

[46] Van Steenkiste T, Ruyssinck J, De Baets L, Decruyenaere J, De Turck F, Ongenae F, et al. Accurate prediction of blood culture outcome in the intensive care unit using long short-term memory neural networks. Artif Intell Med 2019;97:38–43. https://doi.org/10.1016/j.artmed.2018.10.008.

[47] Goodman KE, Lessler J, Cosgrove SE, Harris AD, Lautenbach E, Han JH, et al. A clinical decision tree to predict whether a bacteremic patient is infected with an extended-spectrum β-lactamase-producing organism. Clin Infect Dis 2016;63:896–903. https://doi.org/10.1093/cid/ciw425.

[48] Ratzinger F, Dedeyan M, Rammerstorfer M, Perkmann T, Burgmann H, Makristathis A, et al. A risk prediction model for screening bacteremic patients: a cross sectional study. PLoS One 2014;9:e106765. https://doi.org/10.1371/journal.pone.0106765.

[49] Shashikumar SP, Stanley MD, Sadiq I, Li Q, Holder A, Clifford GD, et al. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. J Electrocardiol 2017;50:739–43. https://doi.org/10.1016/j.jelectrocard.2017.08.013.

[50] Beeler C, Dbeibo L, Kelley K, Thatcher L, Webb D, Bah A, et al. Assessing patient risk of central line-associated bacteremia via machine learning. Am J Infect Control 2018;46:986–91. https://doi.org/10.1016/j.ajic.2018.02.021.

[51] Chen J, Pan Q-S, Hong W-D, Pan J, Zhang W-H, Xu G, et al. Use of an artificial neural network to predict risk factors of nosocomial infection in lung cancer patients. Asian Pac J Cancer Prev 2014;15:5349–53.

[52] Wiens J, Campbell WN, Franklin ES, Guttag JV, Horvitz E. Learning data-driven patient risk stratification models for *Clostridium difficile*. Open Forum Infect Dis 2014;1:ofu045. https://doi.org/10.1093/ofid/ofu045.

[53] Escobar GJ, Baker JM, Kipnis P, Greene JD, Mast TC, Gupta SB, et al. Prediction of recurrent *Clostridium difficile* infection using comprehensive electronic medical records in an integrated healthcare delivery system. Infect Control Hosp Epidemiol 2017;38:1196–203. https://doi.org/10.1017/ice.2017.176.

[54] Hu Z, Simon GJ, Arsoniadis EG, Wang Y, Kwaan MR, Melton GB. Automated detection of postoperative surgical site infections using supervised methods with electronic health record data. Stud Health Technol Inform 2015;216: 706–10.

[55] Betts KS, Kisely S, Alati R. Predicting common maternal postpartum complications: leveraging health administrative data and machine learning. BJOG 2019;126:702–9. https://doi.org/10.1111/1471-0528.15607.

[56] Bertsimas D, Dunn J, Velmahos GC, Kaafarani HMA. Surgical risk is not linear: derivation and validation of a novel, user-friendly, and machine-learning-based predictive optimal trees in emergency surgery risk (POTTER) calculator. Ann Surg 2018;268:574–83. https://doi.org/10.1097/SLA.0000000000002956.

[57] Tunthanathip T, Sae-Heng S, Oearsakul T, Sakarunchai I, Kaewborisutsakul A, Taweesomboonyat C. Machine learning applications for the prediction of surgical site infection in neurological operations. Neurosurg Focus 2019;47: E7. https://doi.org/10.3171/2019.5.FOCUS19241.

[58] Mortazavi BJ, Desai N, Zhang J, Coppi A, Warner F, Krumholz HM, et al. Prediction of adverse events in patients undergoing major cardiovascular procedures. IEEE J Biomed Health Inform 2017;21:1719–29. https://doi.org/10.1109/JBHI.2017.2675340.

[59] Ratzinger F, Dedeyan M, Rammerstorfer M, Perkmann T, Burgmann H, Makristathis A, et al. Neither single nor a combination of routine laboratory parameters can discriminate between Gram-positive and Gram-negative bacteremia. Sci Rep 2015;5:16008. https://doi.org/10.1038/srep16008.

[60] Goodman KE, Simner PJ, Klein EY, Kazmi AQ, Gadala A, Toerper MF, et al. Predicting probability of perirectal colonization with carbapenem-resistant Enterobacteriaceae (CRE) and other carbapenem-resistant organisms (CROs) at hospital unit admission. Infect Control Hosp Epidemiol 2019;40:541–50. https://doi.org/10.1017/ice.2019.42.

[61] Butler-Laporte G, Cheng MP, McDonald EG, Lee TC. Screening swabs surpass traditional risk factors as predictors of MRSA bacteremia. BMC Infect Dis 2018;18:270. https://doi.org/10.1186/s12879-018-3182-x.

[62] Shah AA, Karhade AV, Bono CM, Harris MB, Nelson SB, Schwab JH. Development of a machine learning algorithm for prediction of failure of nonoperative management in spinal epidural abscess. Spine J 2019;19:1657–65. https://doi.org/10.1016/j.spinee.2019.04.022.

[63] Chen J, Chen J, Ding H-Y, Pan Q-S, Hong W-D, Xu G, et al. Use of an artificial neural network to construct a model of predicting deep fungal infection in lung cancer patients. Asian Pac J Cancer Prev 2015;16:5095–9.

[64] Hu Z, Melton GB, Moeller ND, Arsoniadis EG, Wang Y, Kwaan MR, et al. Accelerating chart review using automated methods on electronic health record data for postoperative complications. AMIA Annu Symp Proc 2016;2016:1822–31.

[65] O'Neill J. Review on antimicrobial resistance: tackling a crisis for the health and wealth of nations. London: Review on Antimicrobial Resistance; 2014.
[66] OECD. Stemming the superbug tide. OECD Health Policy Stud 2018;224. https://doi.org/10.1787/9789264307599-en.
[67] Yelin I, Snitser O, Novich G, Katz R, Tal O, Parizade M, et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. Nat Med 2019;25:1143–52. https://doi.org/10.1038/s41591-019-0503-6.
[68] Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. Npj Digital Med 2018;1:18. https://doi.org/10.1038/s41746-018-0029-1.
[69] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inform 2018;22:1589–604. https://doi.org/10.1109/JBHI.2017.2767063.
[70] Baker M. 1,500 scientists lift the lid on reproducibility. Nature 2016;533:452–4. https://doi.org/10.1038/533452a.
[71] Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The reporting of studies conducted using observational routinely-collected health data (RECORD) statement. PLoS Med 2015;12:e1001885. https://doi.org/10.1371/journal.pmed.1001885.
[72] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015;350:g7594. https://doi.org/10.1136/bmj.g7594.
[73] Cook JA, Collins GS. The rise of big clinical databases. Br J Surg 2015;102:e93–101. https://doi.org/10.1002/bjs.9723.
[74] Shah NH, Milstein A, Bagley SC. Making machine learning models clinically useful. JAMA 2019. https://doi.org/10.1001/jama.2019.10306.
[75] Seymour CW, Kennedy JN, Wang S, Chang C-CH, Elliott CF, Xu Z, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. JAMA 2019;321:2003–17. https://doi.org/10.1001/jama.2019.5791.
[76] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366:447–53. https://doi.org/10.1126/science.aax2342.
[77] Chen IY, Joshi S, Ghassemi M. Treating health disparities with artificial intelligence. Nat Med 2020;26:16–7. https://doi.org/10.1038/s41591-019-0649-2.
[78] Pfohl S, Marafino B, Coulet A, Rodriguez F, Palaniappan L, Shah NH. Creating fair models of atherosclerotic cardiovascular disease risk. In: Proceedings of the 2019 AAAI/ACM conference on AI, ethics, and society - AIES '19. New York, New York, USA: ACM Press; 2019. p. 271–8. https://doi.org/10.1145/3306618.3314278.
[79] Molnar C. Interpretable machine learning. Christoph Molnar; 2019.
[80] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. In: Proceedings of the 22Nd ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY, USA: ACM; 2016. p. 1135–44. https://doi.org/10.1145/2939672.2939778.
[81] Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA 2017;318:517–8. https://doi.org/10.1001/jama.2017.7797.
[82] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Machine Intelligence 2019;1:206–15. https://doi.org/10.1038/s42256-019-0048-x.
[83] Bisaso KR, Anguzu GT, Karungi SA, Kiragga A, Castelnuovo B. A survey of machine learning applications in HIV clinical research and care. Comput Biol Med 2017;91:366–71. https://doi.org/10.1016/j.compbiomed.2017.11.001.
[84] Rawson TM, Ahmad R, Toumazou C, Georgiou P, Holmes AH. Artificial intelligence can improve decision-making in infection management. Nat Hum Behav 2019. https://doi.org/10.1038/s41562-019-0583-9.