scientific reports

OPEN Predicting bloodstream infection outcome using machine learning

Yazeed Zoabi[1,2,6], Orli Kehat[3,6], Dan Lahav[1,2,4], Ahuva Weiss-Meilik[3,7✉], Amos Adler[1,5,7✉] & Noam Shomron[1,2,7✉]

Bloodstream infections (BSI) are a main cause of infectious disease morbidity and mortality worldwide. Early prediction of BSI patients at high risk of poor outcomes is important for earlier decision making and effective patient stratification. We developed electronic medical record-based machine learning models that predict patient outcomes of BSI. The area under the receiver-operating characteristics curve was 0.82 for a full featured inclusive model, and 0.81 for a compact model using only 25 features. Our models were trained using electronic medical records that include demographics, blood tests, and the medical and diagnosis history of 7889 hospitalized patients diagnosed with BSI. Among the implications of this work is implementation of the models as a basis for selective rapid microbiological identification, toward earlier administration of appropriate antibiotic therapy. Additionally, our models may help reduce the development of BSI and its associated adverse health outcomes and complications.

Bloodstream infections (BSI) can lead to prolonged hospital stays, and life-threatening and aggressive complications, in addition to high costs to health care systems[1–4]. Increasing rates of antimicrobial-resistant pathogens, particularly gram-negative bacteria, limit treatment options; this often prompts empirical use of broad-range antibiotics[5]. Therefore, timely and critical assessment of available microbiology results are necessary to ensure that individuals with BSI receive prompt, effective, and targeted treatment, for optimal clinical outcomes[5]. However, the current standard-of-care, which mostly depends on blood culture-based diagnosis, is often extremely slow[6].

Sepsis is a life-threatening medical condition, defined as the body's systemic immunological response to an infectious process, which may cause end-stage organ dysfunction and eventually death[7]. Several studies have utilized electronic medical records (EMRs) to construct prediction models for mortality from sepsis[8,9] and sepsis onset[10–13]. From a clinical perspective, these types of early warning systems may be useful in detecting patients at risk of BSI, and typically provide information at a certain point in time (for example, preoperatively[14]) or in a certain time-window before deterioration (see review in[15]). Identifying a patient at risk can trigger early goal-directed therapy regarding confirmation of infection, administration of antimicrobial therapy, and transition to the intensive care unit (ICU).

In contrast to studies that aimed to detect patients at risk of BSI, the current study focused on patients with confirmed infections. Based on EMRs of patients hospitalized with positive blood cultures, we constructed machine learning models that predict poor outcomes of hospitalized patients with BSI, which was confirmed by identifying a bacterial morphology on direct gram stain from a positive blood culture (Fig. 1; see "Methods"). Our prediction occurs at a certain point of time: just after confirming a BSI, yet well before performing specific pathogen identification. At this point of time, early goal-directed clinical practice can be diverted to patients at risk for poor clinical outcomes. As part of that, the prediction can facilitate the use of costly diagnostic procedures, such as the use of rapid microbiological identification techniques that are costly and not widely available instead of the traditional lengthy methods[16]. A forewarning system could direct the required resources to the patients with BSI in the greatest need.

[1]Sackler Faculty of Medicine, Tel Aviv University, 6997801 Tel Aviv, Israel. [2]Edmond J Safra Center for Bioinformatics, Tel Aviv University, 6997801 Tel Aviv, Israel. [3]I-Medata AI Center, Tel-Aviv Sourasky Medical Center, 6423906 Tel Aviv, Israel. [4]The Blavatnik School of Computer Science, Tel Aviv University, 6997801 Tel Aviv, Israel. [5]Clinical Microbiology Laboratory, Tel Aviv Sourasky Medical Center, 6423906 Tel Aviv, Israel. [6]These authors contributed equally: Yazeed Zoabi and Orli Kehat. [7]These authors jointly supervised this work: Ahuva Weiss-Meilik, Amos Adler and Noam Shomron. ✉email: ahuvawm@tlvmc.gov.il; amosa@tlvmc.gov.il; nshomron@tauex.tau.ac.il
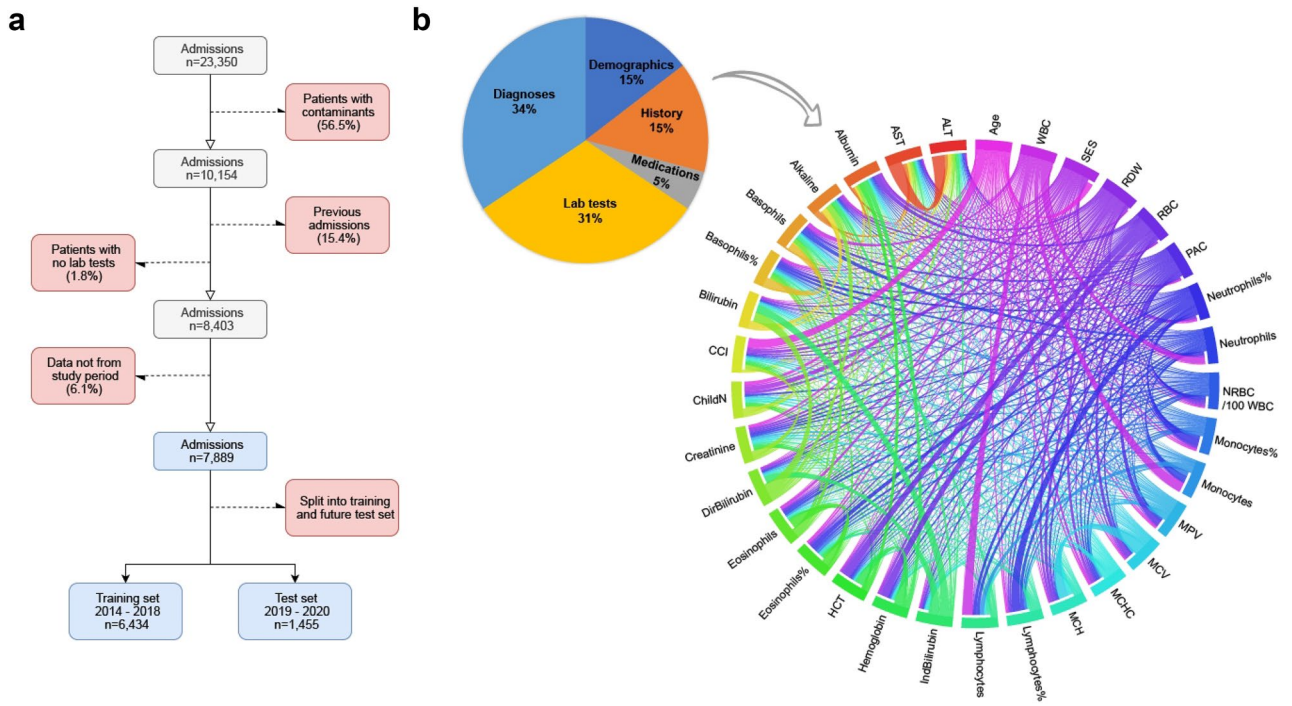
Figure 1. Data and cohort characteristics. (**a**) Cohort selection. Bloodstream infection in this cohort was confirmed by identifying a bacterial morphology on direct gram stain from a positive blood culture. Next, patients with only contaminants were excluded (a list of bacteria classified as contaminants is available as a supplementary file). Subsequently, previous admissions for each patient and patients with no lab tests information were excluded. Finally, the cohort was divided into training and validation sets (see "Methods"). (**b**) Feature modality distribution. Pie charts are divided according to the sum of data points in each feature set. A substantial proportion of the data originates from laboratory test results during current or previous admissions. The Circos plot shows the correlation between continuous features from the entire cohort (test and training sets). Correlation strength is determined by Pearson correlation, thicker bands correspond to a stronger Pearson correlation coefficient.

	Training set	Test set
General		
N	6434	1455
Age [median (±)]	74 (24)	73 (22)
Females [%]	47.7	46.25
CCI		
Mild [%]	20.66	20.03
Moderate [%]	40.4	37.94
Severe [%]	38.94	42.03
Admission type		
Elective [%]	5.08	5.54
Emergency [%]	91.17	90.36
Urgent [%]	3.75	4.1
Other		
Infectious background [%]	9.71	9.42
Surgery during hospital stay [%]	14.45	13.33
Clinical outcomes		
Short-term mortality [%]	23.17	18.49
Long hospital stay [%]	12.17	9.07
Mechanical ventilation [%]	1.23	0.97
Composite score [%]	34.24	26.6

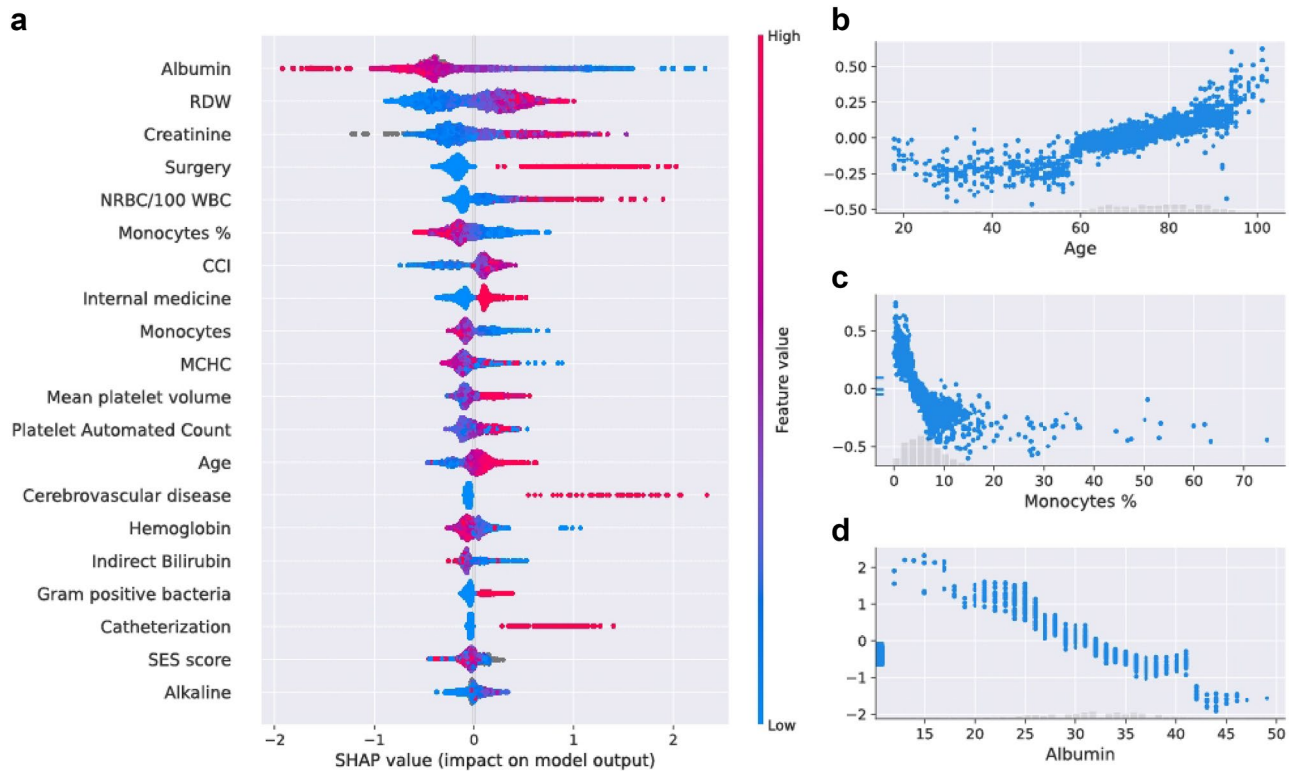**Table 1.** Population characteristics. *CCI* Charlson Comorbidity Index.

Figure 2. Feature analysis of the inclusive model. (**a**) A summary plot of the SHaply Additive exPlanations (SHAP) values for each feature. From top to bottom, the features are ordered by their overall influence on the final prediction (sum of SHAP values). In each feature (line), each point represents a specific case (individual), with colors ranging from red (high values of the predictor) to blue (low values of the predictor). Gray points signal missing values. A point's location on the X-axis represents the SHAP value—the effect the variable had on the prediction in a given individual; points further right indicate greater risk, and points to the left indicate lesser risk. The vertical line in the middle represents no change in risk. (**b**) A plot of SHAP for different values of age (years). The light histogram along the X-axis shows the density of the data. (**c**) A plot of SHAP for different values of monocytes percentage (%) in the blood. The light histogram along the X-axis shows the density of the data. (**d**) A plot of SHAP for different values of albumin (g/L). The light histogram along the x-axis shows the density of the data. (**a–d**) are based on the future test set, n = 1455 unique patients.

Results

The inclusive model.
The characteristics of the population used for the training and testing of the inclusive model are described in Table 1 (see "Methods"). From 7889 adults with a BSI, 2590 (32.8%) positive composite outcomes were recorded. The contribution of each feature of the inclusive model to predict the composite outcome was measured by SHaply Additive exPlanations (SHAP) scores, for each patient (Fig. 2). The predictive contribution of missingness (gray points) was also assessed. Accordingly, a missing value of a feature (e.g., albumin) serves as a signal regarding the patient's risk. SHAP values of three selected variables—age, monocyte %, albumin—are presented in Fig. 2b–d.

The main features that influenced model prediction of a poor outcome included: low albumin, high red cell distribution width (RDW), and high creatinine. RDW and albumin have been associated with mortality, and have been used as prognostic markers in a number of studies[17–24]. This ranking shown partly in Fig. 2 contributed to the creation of the compact model described in the following subsection. The top 20 features ranked by the SHAP scores were also calculated (see Supplementary Fig. S1).

Performance of the inclusive prediction model on the test set showed area under the receiver-operating characteristics curve (auROC) of 0.82 (95% confidence interval (CI): 0.80–0.845), which indicates good discrimination; and an area under the precision-recall curve (auPRC) of 0.65 (95% CI 0.61–0.70) (Fig. 3). The calibration plot, which runs very close to the diagonal, shows excellent calibration (Fig. 3c).

The calibrated model was also tested on two subsets of the future test set, which comprised only patients hospitalized in the ICU (N = 91), or patients tested in the ER (N = 268), and where the prevalence of the composite adverse outcome was 56.7% and 27.75%, respectively. Results for these subsets showed an auROC curve of 0.72 (95% CI 0.614–0.812) for ICU and 0.82 (95% CI 0.758–0.889) for ER (see full results in Supplementary Figs. S2 and S3).

The Charlson Comorbidity Index (CCI) was found as a predictor of poor outcome (Fig. 2a). CCI is a well-known mortality predictor in various medical situations[25,26]. Thus, we compared the ROC curve we receive from CCI, and from the inclusive and compact model. auROC 0.62 (95% CI 0.585–0.648), which shows that
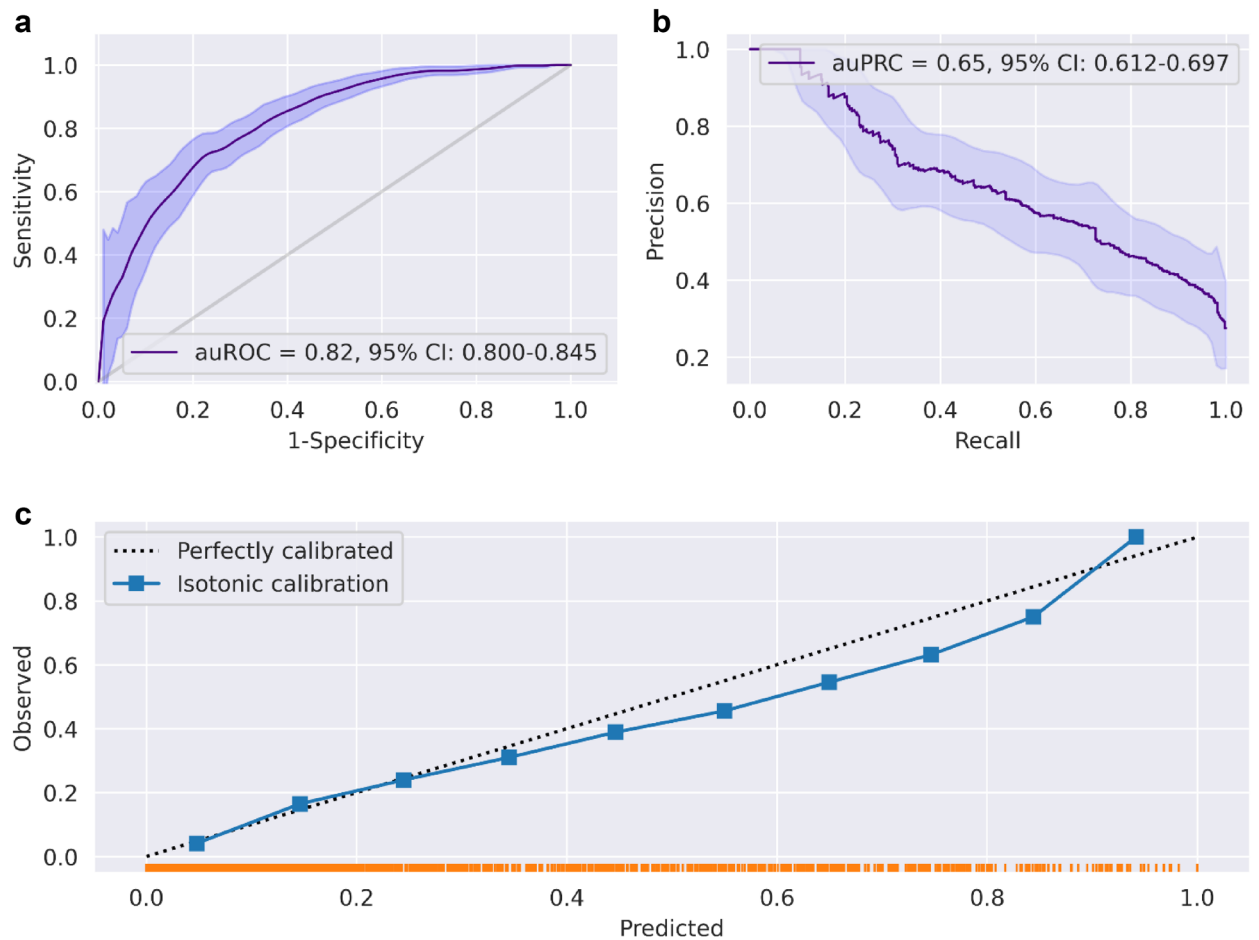
Figure 3. Performance of the inclusive model (**a**). Receiver-operating characteristics (ROC) curves for predictions of the inclusive model on the prospective test set. The light band around the curve represents pointwise 95% confidence intervals derived by bootstrapping. (**b**) A plot of the precision (positive predictive value, PPV) against the recall (sensitivity) of the predictor for different thresholds. The light band around the curve represents pointwise 95% confidence intervals derived by bootstrapping. (**c**) Calibration plot, plotting the observed outcome against the predicted probabilities. The diagonal gray line represents perfect calibration. A smoothed line is fit to the curve, and points are drawn to represent the averages in ten discretized bins. The rug under the plot illustrates the distribution of predictions.

the inclusive model is much better at predicting a poor outcome compared to CCI. ROC curves can be found in Supplementary Fig. S4a.

The compact model. The numerous EMR features, more than 600, incorporated in the inclusive model render its external validation and the reproducibility of results very difficult. This limits its applicability to other hospitals. Hence, we established a simpler and more compact prediction model that incorporates the features with the greatest impact on outcome, and that are most commonly listed in EMR datasets from other hospitals. To this end, we trained and evaluated the performance of a compact model with only 25 features, including simple demographic information, a single blood test, and a brief medical history (Fig. 4d). We used the same cohort of patients of the training and test sets from the previous analysis (see the inclusive model). The compact model achieves an auROC of 0.81 (95% CI 0.78–0.83) (Fig. 4a), and an auPRC of 0.63 (95% CI 0.58–0.68) (Fig. 4b), which are only slightly lower than the values of the inclusive model, with good calibration (Fig. 4c). SHAP scores are shown in Fig. 5. ROC curves comparing compact model performance and CCI predictions can be found in Supplementary Fig. S4b.

The code to test this model is available in our GitHub repository.

Discussion

In this study, we examined the ability to utilize EMRs to predict a composite poor outcome of patients with BSI, which may promote both rapid interventions and patient stratification. Several scoring systems have been developed in recent years for stratifying the risk of patients with sepsis[8,10–13,15], but not for outcomes of patients with BSI. In addition, none are commonly used in routine practice nor recommended according to current guidelines.
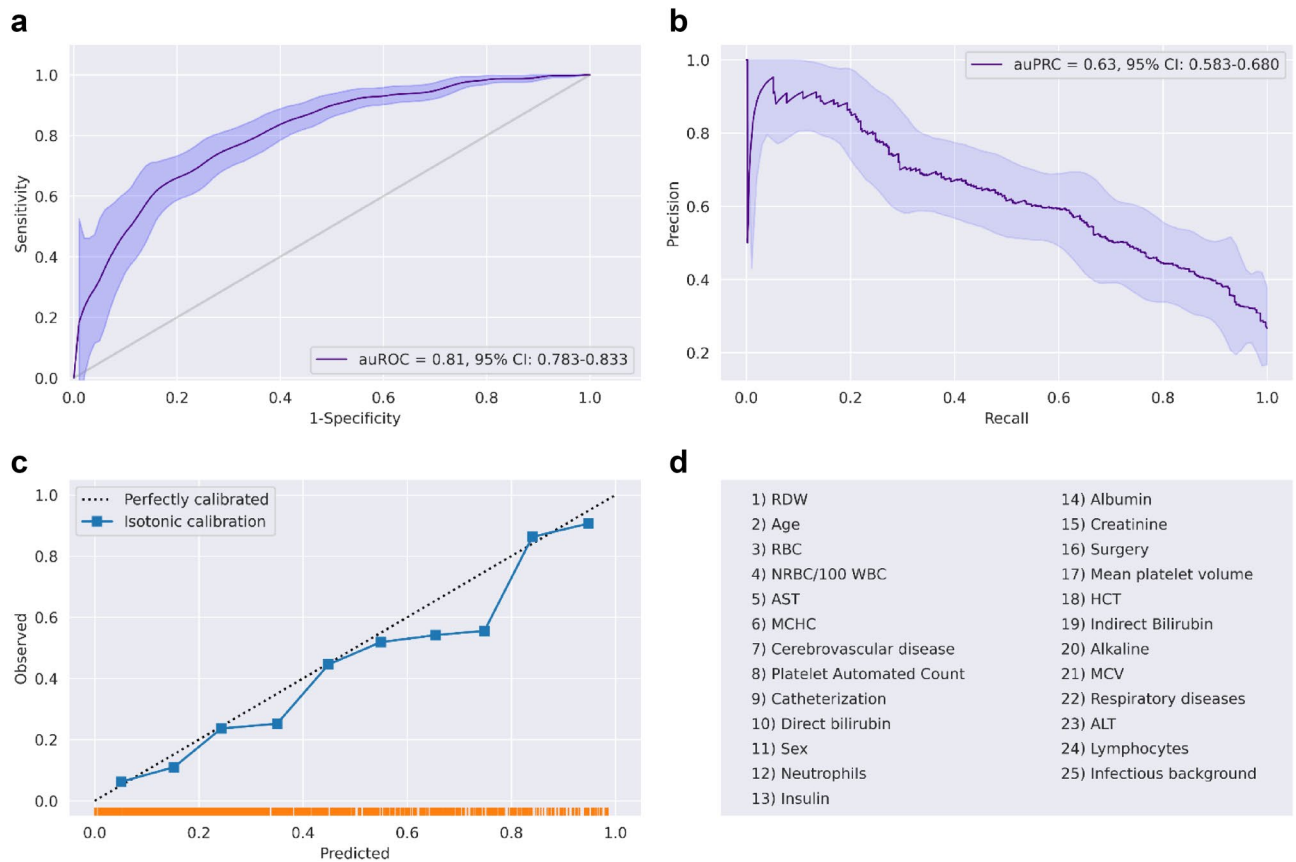
4

Figure 4. Performance of the compact model. (**a**) Receiver-operating characteristics (ROC) curves for predictions of the compact model on the prospective test set. The light band around the curve represents pointwise 95% confidence intervals derived by bootstrapping. (**b**) A plot of the precision (positive predictive value, PPV) against the recall (sensitivity) of the predictor for different thresholds. The light band around the curve represents pointwise 95% confidence intervals derived by bootstrapping. (**c**) Calibration plot, plotting the observed outcome against the predicted probabilities. The diagonal gray line represents perfect calibration. A smoothed line is fit to the curve, and points are drawn to represent the averages in ten discretized bins. The rug under the plot illustrates the distribution of predictions. (**d**) All 25 features used by the compact model.

Our results show that EMRs can be used to produce accurate predictions of BSI outcomes. Our retrospective analysis demonstrates the feasibility of accurate prediction of BSI outcomes using data available in EMRs. In the inclusive model, the prediction was after a gram-stain, and yielded an auROC of 0.82, and an auPRC of 0.65. In the compact model, the prediction was based on only 25 features available at the time of culture, not including gram-stain results; and yielded an auROC of 0.81, and an auPRC of 0.63. Notably, these models outperformed CCI, which is the standard scoring system for mortality. The high performance of machine learning models demonstrates their potential in contributing to infection management by providing accurate information at very early stages of the infection[27–29]. Recent studies have shown machine learning models to outperform standard scoring systems[27,30,31]. In line with these findings, our results show enhanced performance in comparison to the CCI score. It should be noted that CCI predicts mortality, rather than composite poor outcome used in the current study. While our clinical outcome was composed of several variables, it relies heavily on mortality, as 68% of all adverse outcome patients had shown short-term mortality. The substantial improvement in prediction ability (auROC of 0.62 vs. 0.83 for the full model and auROC of 0.62 vs. 0.81 for the compact model) demonstrates the advantage in using machine learning generated scores over the standard conventional scores.

In addition to the well-established risk factors for complications from BSI, such as age and previous infections, our analysis revealed less-known factors as highly predictive of a poor outcome from BSIs. The main factors that were identified as increasing risk included: red cell distribution width, albumin, and creatinine. Red cell distribution width and albumin have been associated with mortality, and have been used as prognostic markers in a number of studies[17–24]. The Charlson Comorbidity Index has also been found to predict mortality in various medical situations[25,26]. However, other factors revealed as central by our analysis, such as serum creatinine values and monocyte counts, are less well recognized and used as predictors of a poor outcome in patients with BSI.

Although maximal model explainability requires using the patient's entire EMR, we demonstrated that a subset of features, available from only simple demographic information and a single blood test, enables accurate prediction with only a slight decrease in auROC, from 0.82 in the inclusive model to 0.81 in the compact model. This may enable accurate BSI outcome estimation by embedded systems in emergency departments of hospitals.
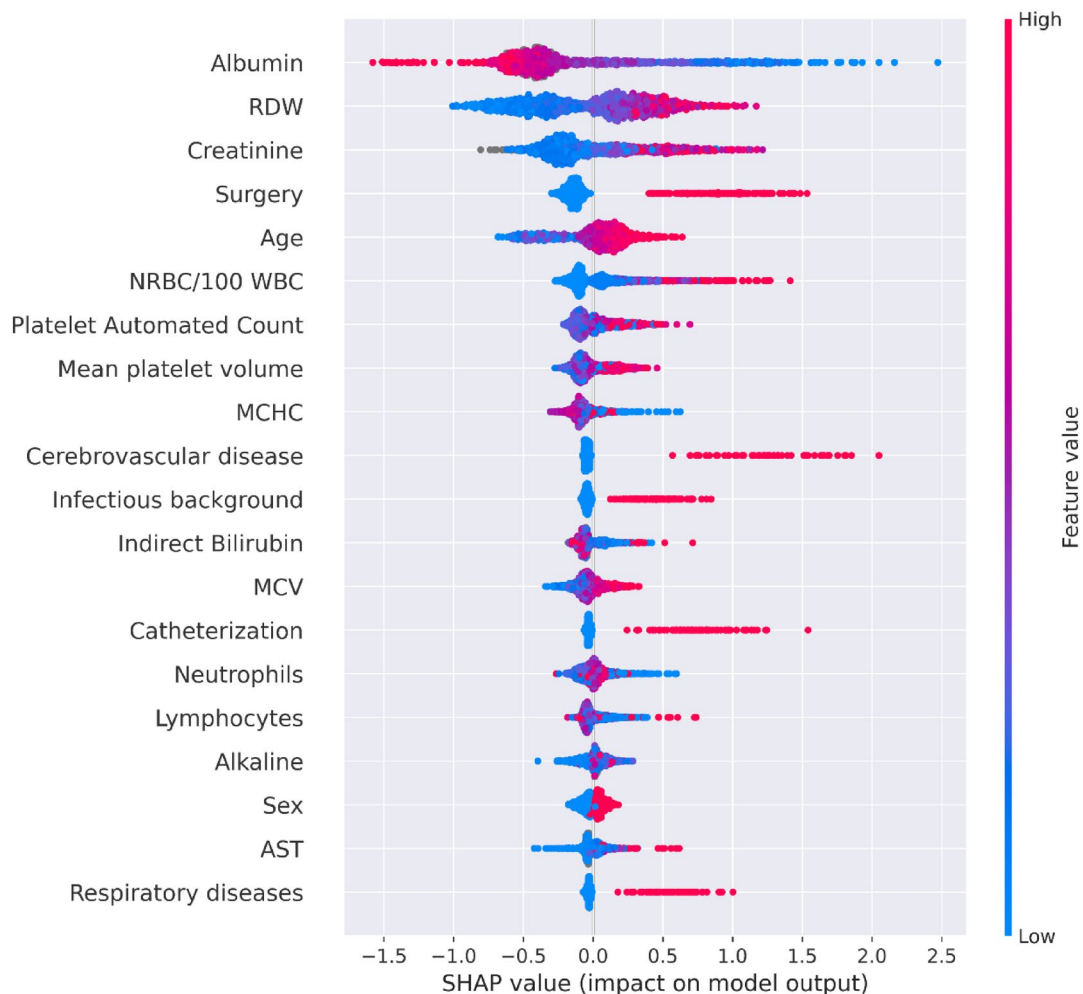
Figure 5. A summary plot of the SHAP values for each feature of the compact model. From top to bottom, features are ordered by their overall influence on the final prediction (sum of SHAP values). In each feature (line), each point represents a specific case (individual), with colors ranging from red (high values of the predictor) to blue (low values of the predictor). Gray points signal missing values. A point's location on the X-axis represents the SHAP value—the effect the variable had on the prediction in a given individual. The points further to the right indicate that for the given individual, the covariate contributed to increasing the risk. Points to the left indicate that the covariate contributed to decreasing the risk. The vertical line in the middle represents no change in risk. Values for the feature 'Sex' are 0 for female, and 1 for male.

Our work has several clinical applications. For instance, it can help select individuals at high-risk for BSI, for whom hospitalization in the ICU, treatment with broad-spectrum antibiotics, or effective early-stage rapid microbiological identification should be considered.

The benefit of early microbiological identification for patient outcomes has been well described in the medical literature[5,32] and the *time-to-administration* of appropriate antibiotics to treat BSI is an important predictor of outcomes[33]. The current study therefore paves the route for future randomized control trials to further study the effectiveness of implementing a model for early prediction of BSI outcomes, possible preventive interventions, and more efficient selection of patients for advanced microbiological diagnostic testing, thus reducing the time-to-administration of appropriate antibiotics.

Our study has several limitations. First, our prediction model is based on retrospective EMR data, which have inherent biases and are influenced by the interaction of the patient with the health system[34]. However, these biases are partially mitigated in this study, since the data contain information originating from a public hospital serving a very large population, and since the outcome of the model is based on information that is accurately and comprehensively documented in the EMRs. Another limitation of the study is that we assessed BSI outcomes only in patients who already have a confirmed BSI (having a positive blood culture). This makes it more difficult to generalize to other patient cohorts. Moreover, the models presented in the current study were designed for the purpose of alerting the clinical staff about a patient's potential risk from BSI. To perform this at an early stage, we relied on data available up to the blood culture collection and did not consider the final microbiological results or treatment effects. As these factors also effect the patients' clinical outcome, they should also be incorporated into future studies to complete the picture. Finally, the predictor was trained and validated on EMRs from Tel

Aviv Sourasky Medical Center (TASMC), composed of patients from and around Tel Aviv, Israel. Nonetheless, TASMC is a public hospital, and the medical system in Israel is accessible to the entire population.

The applicability of the model to other hospital populations needs to be shown. However, the large size of the data, and the comprehensive validation process and its result, namely, validation of the utility of established risk factors for a poor BSI outcome, all support the model's generalizability to other hospitals. Given the additional complexities introduced by the machine learning algorithms, we sought to ensure that sufficient information would be provided to enable our model to undergo external validation[35,36]. This drove us to develop the compact model, which is based on the features with the greatest influence on the overall prediction, and that are easily accessible in EMR datasets from other hospitals. This compact model achieved only a slightly reduced auROC of 0.81. We made the compact model available in our GitHub repository (see code availability) and we encourage researchers with similar data from other hospitals to test it.

In conclusion, our work demonstrates that accurate and calibrated predictions of BSI outcomes early in a hospital admission can be achieved. Earlier and better characterization of patients with BSI could potentially reduce the development of BSI and its associated adverse health outcomes and complications. Our predictive model could become the basis of selective, rapid microbiological identification, and contribute to various decisions such as ICU hospitalization and administration of broad-spectrum antibiotics. Future prospective studies, as well as those on populations from other hospitals, are needed to evaluate the clinical impact of the model.

Methods

Study design, population, and definition of outcome.
The study was designed as a retrospective cohort study for the development and validation of a clinical prediction model. The study was performed at the TASMC, a 1500-bed public tertiary care center, and the only general hospital serving the population of Tel-Aviv, the most populous city in Israel, of all socioeconomic backgrounds. Data processing, model training and analyses were performed at the TASMC Data Science Department and the Faculty of Medicine at Tel Aviv University.

The study included EMRs of adults hospitalized with a positive blood culture (bacterial only) in the period between January 2014 and January 2020. The year 2014 was determined as the starting point since frequent changes in variable identifiers occurred in the preceding years. Patients' EMRs included demographics, laboratory test results, previous diagnoses recorded at TASMC, recorded medical history, and initial gram-stain morphology of positive blood cultures that are reported by phone.

The models were developed according to features extracted from various modalities available in EMRs of patients hospitalized with a positive blood culture at the TASMC, and predicted a composite poor outcome, defined as *at least one* of the following:

- Short term in-hospital mortality within 10 days of a culture.
- Mechanical ventilation in the 10 days after the culture.
- Prolonged length of stay (> 6 weeks).

The study flow chart is presented in Fig. 1. Exclusion criteria were: the absence of laboratory data of medical history information, age younger than 18 years or older than 100 years, and patients' explicit objection to the use of their medical data for research purposes. Before initiating any analysis, the study population was divided into a training-validation set from years 2014–2018 (included 6434 admissions, of whom 2203 had a composite poor outcome) and a test set of admissions from 2019 and the first month of 2020 (included 1455 admissions, of whom 387 had a composite poor outcome). The training-validation set was further divided to training and validation sets at a ratio of 4:1. The model was also tested on two subsets of the test set, which comprised only patients from ICUs or only patients from the emergency room. Both these subsets posed a high challenge to the model regarding its generalization. The prospective test cross-sections were performed to emulate the model's use in practice and in real world situations.

Variable and feature selection.
To evaluate whether EMR-derived information might accurately predict outcomes of patients with BSI, we compiled a set of 606 features. All these features were available at the time a blood sample was sent for culture, except for the gram-stain information (used only in the inclusive model). With all these features, we trained a gradient-boosting model, the inclusive model, to predict the probability that each held-out sample (patients not included in the training set) would have a poor outcome. Distributions of the various modalities of the features used are depicted in Fig. 1b. In addition, we trained and tested a compact model, comprising 25 features, for application on EMRs from other hospitals.

Each of the 606 features was assigned a category. For more comprehensive representation, some features within a category were combined, such that all the features could be represented using 96 variables. These variables were used to create the pie chart in Fig. 1.b. The following list describes the mechanism for generating the features, and for grouping them:

Demographics (238 features, 14 variables when grouped)
- Contains features such as age, sex, and number of children. A total of 228 categorical features, which included birth country and nationality, were grouped to five variables.

History (108 features, 14 when grouped)
- Contains medical history that is not documented as diagnosis history. This includes unit information (ICU, emergency room, etc.), surgery information, and chest pain. A total of 97 categorical features were grouped to three variables.
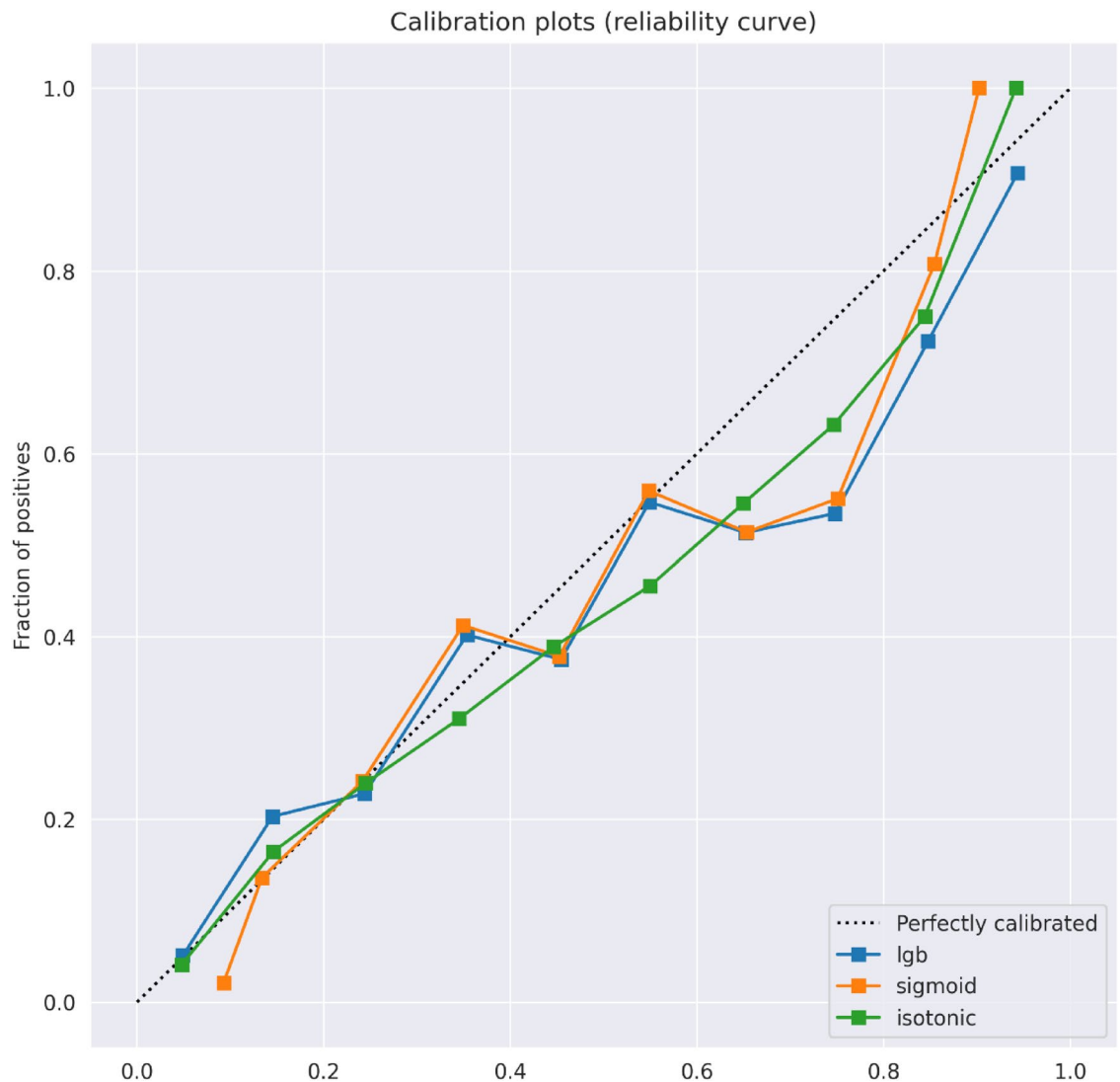
Figure 6. Calibration plots of the observed outcome against the predicted probabilities. The diagonal gray line represents perfect calibration. A smoothed line is fit to the curve, and points are drawn to represent the mean values in ten discretized bins. Blue, orange, and green lines correspond to the original uncalibrated model, the model after sigmoid calibration, and after isotonic calibration, respectively.

Medications (five features)
- Contains binary information about five medications: general diabetes drugs, insulin, anti-coagulants, anti-aggregants, and coumadin on admission.

Laboratory (33 features, 30 when grouped)
- Contains laboratory test results. Four categorical features, describing gram staining results, were combined into one variable. All other features in this category had continuous numeric values.

Diagnoses (222 features, 33 when grouped)
- Diagnoses history is recorded in TASMC as ICD-9 codes. The ICD-9 code hierarchy was used to group these features into 33 variables represented by ICD-9 codes of higher hierarchy.

Analysis platform. All computational analyses were performed on a secure compute cluster environment located at TASMC. Python 3, with numpy, pandas, and scikit-learn formed the backbone of the data-processing pipeline.

Development of the models. Predictions were generated using a gradient-boosting machine model built with decision-tree base-learners[37]. Such models have demonstrated efficacy in prediction, using tabular data[38], and have been incorporated in several successful algorithms in the field of machine learning[39]. We implemented the gradient-boosting predictor trained with the LightGBM[40] Python package. LightGBM has shown effectiveness on clinical and patient tabular data in particular, and was adopted by many recently published models[41–46]. Missing values were inherently handled by the LightGBM predictor[40,47,48]. The validation set was used for early stopping[49], with auROC as the performance measure. Hyperparameters were chosen after a cross-validated grid

search while training the inclusive model, and were also used to train the compact model. Hyperparameters are available at our GitHub repository. Two LightGBM classifiers with differing complexity were developed: the inclusive and compact models. The compact model is available at our GitHub repository (see Code Availability).

The mechanistic basis of the models.   To identify the principal features driving model prediction, SHAP values[50] were calculated. These values are suited for complex models such as artificial neural networks and gradient-boosting machines[51]. Originating in game theory, SHAP values partition the prediction result of every sample into the contribution of each constituent feature value. This is done by estimating differences between models with subsets of the feature space. By averaging across samples, SHAP values estimate the contribution of each feature to overall model predictions. A higher value indicates that a feature has a larger impact on the model, which indicates that the feature is more important.

Calibration of the models.   We analyzed the calibration (observed risk versus raw prediction score) of our proposed inclusive and compact models. The raw prediction scores produced by the machine learning model (LightGBM) were calibrated and evaluated on the test-set. We used isotonic regression, which fits a rank-preserving transformation between the original scores and transformed scores; and minimizes the deviation between the target label and the prediction score. We used the scikit-learn library (version 0.20.0) for fitting the isotonic regression model. Ten prediction score bins were used, with regular spacing between the minimum/maximum prediction score produced by a model. We noticed that the raw scores were already well-calibrated. Curves for isotonic and sigmoid calibrations, as well as for the raw model are presented in Fig. 6.

Evaluation of the models.   The models were scored on the test set using the auROC. In addition, plots of positive predictive value (PPV) against sensitivity (precision-recall curve) were drawn across different thresholds. For all the thresholds from all the ROC curves, metrics were calculated, including sensitivity, specificity, PPV, negative predictive value, false positive rate, false negative rate, false discovery rate, and overall accuracy (Supplementary Dataset 1). Confidence intervals for the various performance measures were derived through resampling, using the bootstrap percentile method[52] with 1000 repetitions.

Ethics declarations.   This study (TLV-0684-18) was approved by the TASMC Institutional Review Board (IRB). All methods were performed in accordance with the IRB policy, guidelines, and regulations. As this is retrospective study, informed consent was waived by the Tel Aviv Sourasky Medical Center IRB, as all identifying details of the participants were removed before any computational analysis.

Data availability
The data that support the findings of this study are from TASMC. Access restrictions apply to these data and they are therefore not publicly available. Due to these restrictions, these data can be accessed only by request to TASMC or the authors.

Code availability
Hyperparameters for the models and the analytic code of the compact model are available at: https://github.com/nshomron/infecpred.

References
1. Goto, M. & Al-Hasan, M. N. Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe. *Clin. Microbiol. Infect.* **19**, 501–509. https://doi.org/10.1111/1469-0691.12195 (2013).
2. Pittet, D., Tarara, D. & Wenzel, R. P. Nosocomial bloodstream infection in critically III patients: Excess length of stay, extra costs, and attributable mortality. *JAMA* **271**, 1598–1601. https://doi.org/10.1001/jama.1994.03510440058033 (1994).
3. Rudd, K. E. *et al.* Global, regional, and national sepsis incidence and mortality, 1990–2017: Analysis for the Global Burden of Disease Study. *The Lancet* **395**, 200–211. https://doi.org/10.1016/S0140-6736(19)32989-7 (2020).
4. Angus, D. C. *et al.* Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Crit. Care Med.* **29**, 1303–1310 (2001).
5. Seymour, C. W. *et al.* Time to treatment and mortality during mandated emergency care for sepsis. *N. Engl. J. Med.* **376**, 2235–2244. https://doi.org/10.1056/NEJMoa1703058 (2017).
6. MacVane, S. H. & Nolte, F. S. Benefits of adding a rapid PCR-based blood culture identification panel to an established antimicrobial stewardship program. *J. Clin. Microbiol.* **54**, 2455–2463. https://doi.org/10.1128/JCM.00996-16 (2016).
7. Gyawali, B., Ramakrishna, K. & Dhamoon, A. S. Sepsis: The evolution in definition, pathophysiology, and management. *SAGE Open Med.* https://doi.org/10.1177/2050312119835043 (2019).
8. Taylor, R. A. *et al.* Prediction of In-hospital mortality in emergency department patients with sepsis: A local big data-driven, machine learning approach. *Acad. Emerg. Med.* **23**, 269–278. https://doi.org/10.1111/acem.12876 (2016).
9. Gultepe, E. *et al.* 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**, 315–325. https://doi.org/10.1136/amiajnl-2013-001815 (2014).
10. Nemati, S. *et al.* An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit. Care Med.* **46**, 547–553. https://doi.org/10.1097/CCM.0000000000002936 (2018).
11. Islam, Md. M. *et al.* Prediction of sepsis patients using machine learning approach: A meta-analysis. *Comput. Methods Programs Biomed.* **170**, 1–9. https://doi.org/10.1016/j.cmpb.2018.12.027 (2019).
12. Delahanty, R. J. *et al.* Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. *Ann. Emerg. Med.* **73**, 334–344. https://doi.org/10.1016/j.annemergmed.2018.11.036 (2019).

13. Desautels, T. *et al.* Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach. *JMIR Med. Inform.* **4**, e5909. https://doi.org/10.2196/medinform.5909 (2016).
14. Thottakkara, P. *et al.* Application of machine learning techniques to high-dimensional clinical data to forecast postoperative complications. *PLoS ONE* **11**, e0155705. https://doi.org/10.1371/journal.pone.0155705 (2016).
15. Fleuren, L. M. *et al.* Machine learning for the prediction of sepsis: A systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med.* **46**, 383–400. https://doi.org/10.1007/s00134-019-05872-y (2020).
16. Caliendo, A. M. *et al.* Better tests, better care: Improved diagnostics for infectious diseases. *Clin. Infect. Dis.* **57**(Suppl 3), S139-170. https://doi.org/10.1093/cid/cit578 (2013).
17. Yoo, J.-W. *et al.* Red cell distribution width/albumin ratio is associated with 60-day mortality in patients with acute respiratory distress syndrome. *Infect. Dis.* **52**, 266–270. https://doi.org/10.1080/23744235.2020.1717599 (2020).
18. Zhang, Z., Xu, X., Ni, H. & Deng, H. Red cell distribution width is associated with hospital mortality in unselected critically ill patients. *J. Thorac. Dis.* **5**, 730–736. https://doi.org/10.3978/j.issn.2072-1439.2013.11.14 (2013).
19. Patel, K. V. *et al.* Red blood cell distribution width and the risk of death in middle-aged and older adults. *Arch. Intern. Med* **169**, 515. https://doi.org/10.1001/archinternmed.2009.11 (2009).
20. Chen, L., Lu, X. Y. & Zhu, C. Q. Prognostic value of albumin-red cell distribution width score in patients with severe community-acquired pneumonia. *Ann. Palliat. Med.* **9**, 75965–75765. https://doi.org/10.21037/apm.2020.04.22 (2020).
21. Lee, J. H. *et al.* Red cell distribution width as a prognostic marker in patients with community-acquired pneumonia. *Am. J. Emerg. Med.* **31**, 72–79. https://doi.org/10.1016/j.ajem.2012.06.004 (2013).
22. Hannan, J. L., Radwany, S. M. & Albanese, T. In-hospital mortality in patients older than 60 years with very low albumin levels. *J. Pain Symptom Manage.* **43**, 631–637. https://doi.org/10.1016/j.jpainsymman.2011.04.009 (2012).
23. Akirov, A., Masri-Iraqi, H., Atamna, A. & Shimon, I. Low albumin levels are associated with mortality risk in hospitalized patients. *Am. J. Med.* **130**, 1465.e11-1465.e19. https://doi.org/10.1016/j.amjmed.2017.07.020 (2017).
24. Goldwasser, P. & Feldman, J. Association of serum albumin and mortality risk. *J. Clin. Epidemiol.* **50**, 693–703. https://doi.org/10.1016/s0895-4356(97)00015-2 (1997).
25. Bannay, A. *et al.* The best use of the Charlson Comorbidity Index with electronic health care database to predict mortality. *Med. Care* **54**, 188. https://doi.org/10.1097/MLR.0000000000000471 (2016).
26. Huang, Y. *et al.* Charlson comorbidity index helps predict the risk of mortality for patients with type 2 diabetic nephropathy. *J. Zhejiang Univ. Sci. B* **15**, 58–66. https://doi.org/10.1631/jzus.B1300109 (2014).
27. Tabaie, A. *et al.* Predicting presumed serious infection among hospitalized children on central venous lines with machine learning. *Comput. Biol. Med.* **132**, 104289. https://doi.org/10.1016/j.compbiomed.2021.104289 (2021).
28. Roimi, M. *et al.* Early diagnosis of bloodstream infections in the intensive care unit using machine-learning algorithms. *Intensive Care Med.* **46**, 454–462. https://doi.org/10.1007/s00134-019-05876-8 (2020).
29. Mahmoud, E. *et al.* Developing machine-learning prediction algorithm for bacteremia in admitted patients. *Infect. Drug Resist.* **14**, 757–765. https://doi.org/10.2147/IDR.S293496 (2021).
30. Zhang, G. *et al.* A machine learning approach for mortality prediction only using non-invasive parameters. *Med. Biol. Eng. Comput.* **58**, 2195–2238. https://doi.org/10.1007/s11517-020-02174-0 (2020).
31. Morgan, D. J. *et al.* Assessment of machine learning vs standard prediction rules for predicting hospital readmissions. *JAMA Netw. Open* **2**, e190348. https://doi.org/10.1001/jamanetworkopen.2019.0348 (2019).
32. Bernhard, M., Lichtenstern, C., Eckmann, C. & Weigand, M. A. The early antibiotic therapy in septic patients: Milestone or sticking point?. *Crit. Care* **18**, 671. https://doi.org/10.1186/s13054-014-0671-1 (2014).
33. Falcone, M. *et al.* Time to appropriate antibiotic therapy is a predictor of outcome in patients with bloodstream infection caused by KPC-producing Klebsiella pneumoniae. *Crit. Care* **24**, 29. https://doi.org/10.1186/s13054-020-2742-9 (2020).
34. Phelan, M., Bhavsar, N. A. & Goldstein, B. A. Illustrating informed presence bias in electronic health records data: How patient interactions with a health system can impact inference. *EGEMS* https://doi.org/10.5334/egems.243 (2017).
35. Dagan, N., Cohen-Stavi, C., Leventer-Roberts, M. & Balicer, R. D. External validation and comparison of three prediction tools for risk of osteoporotic fractures using data from population based electronic health records: retrospective cohort study. *BMJ* **356**, i6755. https://doi.org/10.1136/bmj.i6755 (2017).
36. Vollmer, S. *et al.* Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ* **368**, l6927. https://doi.org/10.1136/bmj.l6927 (2020).
37. Hastie, T., Tibshirani, R. & Friedman, J. Boosting and additive trees. In *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (eds Hastie, T. *et al.*) 337–387 (Springer, 2009).
38. Fernández-Delgado, M., Cernadas, E., Barro, S. & Amorim, D. Do we need hundreds of classifiers to solve real world classification problems?. *J. Mach. Learn. Res.* **15**, 3133–3181 (2014).
39. Omar, K. B. A. *XGBoost and LGBM for Porto Seguro's Kaggle challenge : A comparison Semester Project* (2018).
40. Ke, G. *et al.* LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30* (eds Guyon, I., Luxburg, U. V., Bengio, S. *et al.*) 3146–3154 (Curran Associates Inc, 2017).
41. Wang, T., Liu, G. & Lin, H. A machine learning approach to predict intravenous immunoglobulin resistance in Kawasaki disease patients: A study based on a Southeast China population. *PLoS ONE* **15**, e0237321. https://doi.org/10.1371/journal.pone.0237321 (2020).
42. Zoabi, Y., Deri-Rozov, S. & Shomron, N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj Digit. Med.* **4**, 1–5. https://doi.org/10.1038/s41746-020-00372-6 (2021).
43. Artzi, N. S. *et al.* Prediction of gestational diabetes based on nationwide electronic health records. *Nat. Med.* **26**, 71–76. https://doi.org/10.1038/s41591-019-0724-8 (2020).
44. Kopitar, L. *et al.* Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Sci. Rep.* **10**, 11981. https://doi.org/10.1038/s41598-020-68771-z (2020).
45. Shin, Y. *et al.* Emergency department return prediction system using blood samples with LightGBM for smart health care services. *IEEE Consum. Electron. Mag.* https://doi.org/10.1109/MCE.2020.3015439 (2020).
46. Razavian, N. *et al.* A validated, real-time prediction model for favorable outcomes in hospitalized COVID-19 patients. *npj Digit. Med.* **3**, 1–13. https://doi.org/10.1038/s41746-020-00343-x (2020).
47. Josse, J., Prost, N., Scornet, E. & Varoquaux, G. *On the consistency of supervised learning with missing values*. arXiv:190206931 [cs, math, stat] (2019).
48. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery* 785–794 (2016).
49. Raskutti, G., Wainwright, M. J. & Yu, B. Early stopping for non-parametric regression: An optimal data-dependent stopping rule. In *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 1318–1325 (2011).
50. Lundberg, S. & Lee, S.-I. *A Unified Approach to Interpreting Model Predictions*. arXiv:170507874 [cs, stat] (2017).
51. Lundberg, S. M. *et al.* Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2**, 749–760. https://doi.org/10.1038/s41551-018-0304-0 (2018).
52. Efron, B. & Tibshirani, R. J. *An Introduction to the bootstrap* (CRC Press, 1994).

Acknowledgements

Author contributions

Y.Z., O.K., A.W.M., A.A., N.S. designed the study and wrote the paper. Y.Z. developed the models. O.K. collected the data. Y.Z. and O.K. did the statistical analysis. Y.Z. and D.L. contributed to the development of the compact model.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-99105-2.

Correspondence and requests for materials should be addressed to A.W.-M., A.A. or N.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.