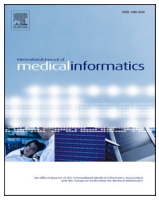
International Journal of Medical Informatics 167 (2022) 104878

Contents lists available at ScienceDirect

International Journal of Medical Informatics

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



Decision support system and outcome prediction in a cohort of patients with necrotizing soft-tissue infections

Sonja Katz [a,c,1], Jaco Suijker [b,d,e,1], Christopher Hardt [a], Martin Bruun Madsen [f], Annebeth Meij-de Vries [b,g], Anouk Pijpe [b,d], Steinar Skrede [h,i], Ole Hyldegaard [j], Erik Solligård [k,l], Anna Norrby-Teglund [m], Edoardo Saccenti [c], Vitor A.P. Martins dos Santos [a,c,*]

[a] LifeGlimmer GmbH, Berlin, Germany
[b] Burn Centre, Red Cross Hospital, Beverwijk, the Netherlands
[c] Laboratory of Systems and Synthetic Biology, Wageningen University and Research, Wageningen, the Netherlands
[d] Department of Plastic, Reconstructive and Hand Surgery, Amsterdam Movement Sciences Amsterdam UMC, Amsterdam, the Netherlands
[e] Association of Dutch Burn Centers, Beverwijk, the Netherlands
[f] Department of Intensive Care, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
[g] Pediatric Surgical Centre, Emma Children's Hospital, Amsterdam UMC, Amsterdam, the Netherlands
[h] Department of Medicine, Haukeland University Hospital, Bergen, Norway
[i] Department of Clinical Science, University of Bergen, Bergen, Norway
[j] Department of Anesthesia, Hyperbaric Unit, University Hospital of Copenhagen, Rigshospitalet, Copenhagen, Denmark
[k] Clinic of Anaesthesia and Intensive Care, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
[l] Gemini Center for Sepsis Research, Department of Circulation and Medical Imaging, NTNU Norwegian University of Science and Technology, Trondheim, Norway
[m] Centre for Infectious Medicine, Department of Medicine Huddinge, Karolinska Institute, Stockholm, Sweden

ARTICLE INFO

Keywords:
Necrotizing soft-tissue infections
Clinical decision support system
Intensive care unit
Machine learning
Random forest
Mortality

ABSTRACT

Introduction: Necrotizing Soft Tissue Infections (NSTI) are severe infections with high mortality affecting a heterogeneous patient population. There is a need for a clinical decision support system which predicts outcomes and provides treatment recommendations early in the disease course.

Methods: To identify relevant clinical needs, interviews with eight medical professionals (surgeons, intensivists, general practitioner, emergency department physician) were conducted. This resulted in 24 unique questions. Mortality was selected as first endpoint to develop a machine learning (Random Forest) based prediction model. For this purpose, data from the prospective, international INFECT cohort (N = 409) was used.

Results: Applying a feature selection procedure based on an unsupervised algorithm (Boruta) to the > 1000 variables available in INFECT, including baseline, and both NSTI specific and NSTI non-specific clinical data yielded sixteen predictive parameters available on or prior to the first day on the intensive care unit (ICU). Using these sixteen variables 30-day mortality could be accurately predicted (AUC = 0.91, 95% CI 0.88–0.96). Except for age, all variables were related to sepsis (e.g. lactate, urine production, systole). No NSTI-specific variables were identified. Predictions significantly outperformed the SOFA score($p < 0.001$, AUC = 0.77, 95% CI 0.69–0.84) and exceeded but did not significantly differ from the SAPS II score ($p = 0.07$, AUC = 0.88, 95% CI 0.83–0.92). The developed model proved to be stable with AUC > 0.8 in case of high rates of missing data (50% missing) or when only using very early (<1 h) available variables.

Conclusions: This study shows that mortality can be accurately predicted using a machine learning model. It lays the foundation for a more extensive, multi-endpoint clinical decision support system in which ultimately other outcomes and clinical questions (risk for septic shock, AKI, causative microbe) will be included.

* Corresponding author at: Wageningen University and Research, Systems and Synthetic Biology, Stippeneng 4, 6708WE Wageningen, the Netherlands.
E-mail address: vitor.martinsdossantos@wur.nl (V.A.P. Martins dos Santos).

[1] Authors contributed equally to this work

https://doi.org/10.1016/j.ijmedinf.2022.104878
Received 6 June 2022; Received in revised form 6 September 2022; Accepted 19 September 2022
Available online 24 September 2022

1. Introduction

Necrotizing soft-tissue infections (NSTI) are rare, fulminant infections, which affect a heterogeneous population in regards to age, sex, and the presence of comorbidities [1]. Both a single pathogen (monomicrobial type) or multiple pathogens acting synergistically (polymicrobial type) may be responsible, with different pathogenic mechanisms [2]. Besides local tissue destruction these pathogens cause systemic toxicity leading to sepsis, and in many cases septic shock (28–50%) [3,4]. If left untreated, NSTI will be fatal within days, making timely recognition an essential prerequisite for successful disease management. Current reported mortality is 10–29% [3,5–9]. Besides mortality, long-term morbidity is extensive: functionally due to scars, amputations, fatigue, as well as psychosocially, which may including fear for recurrence, post-traumatic stress, depression and changes in social activities [10–13].

A major challenge in the treatment of NSTI is to provide adequate, individualized treatment as early as possible, while preventing over-treatment or under-treatment. Support for treatment decisions has until now been limited to simple algorithmic procedures in guidelines [14], and overall substantial differences exist internationally, which include optimal time to second surgery [15–17], and whether or not to use Hyperbaric Oxygen Therapy (HBOT) [17,16] or Intravenous Immunoglobulin (IVIG) [18,15,16]). To better tailor treatment on individual patients, various prediction scores for NSTI (or sepsis in general) can be used. Diagnostically, the LRINEC score can be used to discern between patients with high and low risk of NSTI, although its usefulness is under debate [19,20]. Also, Acute Kidney Injury (AKI) [21,22] or risk of developing septic shock [23] may be predicted. The prediction of mortality, which is important for both treatment allocation as well as in the communication with patients and their families, is currently performed using general ICU mortality predictions (SOFA score [24], SAPS II [25], SAPS III [26], APACHE IV score [27]). It has however not been sufficiently studied how well these general mortality prediction scores perform among patients with NSTI. For example, one study found the SAPS II mortality prediction to be less in case of sepsis compared to other disorders for which patients were admitted to the ICU [28]. Also, the requirement of different predictive scoring systems to estimate different outcomes is not very practical. Ideally, a more comprehensive overview of expected disease characteristics, disease progression, and outcomes would be obtained early after admission.

We believe the outlined shortcomings in efficacy (accuracy, applicability) of general clinical scoring systems might be addressed by a so-called clinical decision support system (CDSS). A CDSS is a framework aiming to link health observations (e.g. clinical data, patient-reported outcomes) with health knowledge, thereby supporting the decision-making process by providing consultation to medical personnel. The predictive models underlying a CDSS can be trained for identification of disease, disease stages, patient stratification as well as recognising patient-specific patterns. They have shown promising results in the diagnosis of sepsis [29,30], and gain popularity as personalised medicine tools [31]. A CDSS holds many advantages compared to the current scoring systems in use. They can be designed to be highly versatile, addressing different health endpoints at the same time, such as treatment choices or prognostic outcomes. Creation of such a multiple-endpoint CDSS abrogates the need for individual scoring systems used in parallel. Furthermore, CDSS are flexible regarding the use of available information and can be tailored towards one specific disease. Another advantage is their circular, iterative design. By including a database management system in the CDSS framework new patients can easily be added or available information updated in an automated manner. This, in turn, makes predictions more and more accurate - so the system learns as it grows.

A CDSS is expected to improve evidence based treatment, both by being used complementary to local guidelines, or by use on its own. Here, we present the first two steps towards the realization of a multiple-endpoint CDSS for improving NSTI patient care. Firstly, an overview of clinical needs was acquired by means of interviews with clinicians. Secondly, one of many identified relevant outcomes (mortality) was selected as the first to base our predictive system on, which provides proof of how a AI based CDSS could assist the decision-making process. In an attempt to compare our CDSS to current systems in use, we benchmarked its performance on the SAPS II and SOFA score at admission.

2. Methods

2.1. Ethics

The INFECT study used in this work was approved by the national or regional ethics committees and data protections agencies in all participating countries. Written informed consent was obtained from every patient or their legal surrogate as soon as possible. In all cases, consent was obtained from the patient when possible. The data collection protocol has been published previously [41]. The INFECT project is registered at ClinicalTrials.gov, number NCT01790698.

2.2. INFECT study cohort

The INFECT data set utilized in this study is the result of an international, multicentre, prospective, cohort study of adult patients with NSTI included prospectively at five Scandinavian hospitals, which were referral centers for NSTI (INFECT study: ClinicalTrials.gov, number NCT01790698, posted on February 13, 2013 with the last update on April 23, 2018). A total of 409 patients above the age of 18 and with surgically confirmed NSTI cases were enrolled during February 2013 and June 2017 in five different referral centres for patients with NSTI. Participating centres included: Rigshospitalet, Copenhagen University Hospital, Denmark; Karolinska University Hospital, Solna, Sweden; Blekinge Hospital, Karlskrona, Sweden; Sahlgrenska University Hospital, Gothenburg, Sweden, and Haukeland University Hospital, Bergen, Norway [41]. Data recorded in the INFECT study include patient demographics, clinical data (blood samples, clinical findings), daily ICU data for a period of up to 7 days (fluid administration, medication, observed parameters), information regarding specific treatments and samples (surgical procedures, microbiological findings, HBO treatments), as well as follow-up data (90-day follow-up, 365-day follow-up). All of this information is encoded in a total of approximately 2,400 variables, making the INFECT study the largest prospective study in patients with NSTI to date. A detailed description on the INFECT cohort, including subject demographics, is offered by Madsen et al. [41].

2.3. Semi-structured interviews

To create an overview of the various relevant clinical questions, semi-structured interviews were conducted (Fig. 1, Supplementary Note 9). An interview guide was constructed with the PerMIT project group, an international consortium dedicated to demonstrating the potential benefits of personalized medicine approaches for NSTI and sepsis patients (https://permedinfect.com/, Grant No. 8113-00009B). Eight clinicians (3 ICU specialists, 2 surgeons, 1 microbiologist, 1 ER specialist, 1 general practitioner) were interviewed by one of the authors (JS). During the interviews, participants were asked which clinical questions they believed were most relevant in the various phases (pre-hospital, pre-ICU, ICU). When new questions emerged, those were added. Each participant was asked to attribute a score for relevance to each question; (3) highly relevant, (2) relevant, or (1) interesting but not relevant. In case of insufficient knowledge on a topic, it could be left blank. Some questions were added later in the process, in which case fewer participants scored it for relevance. The average score for the relevance of each of the questions was calculated for those that attributed a score to a question.

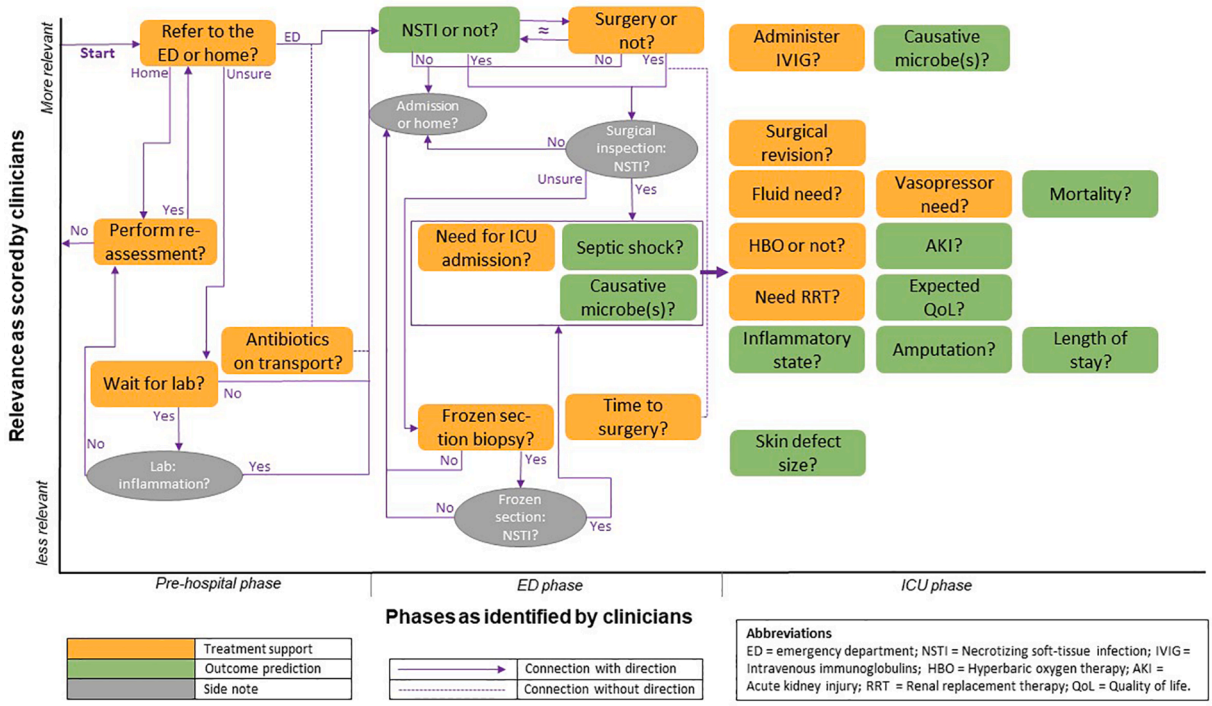


Fig. 1. A graphical display of the various (24 unique) treatment support questions (orange) and outcomes (green) which resulted from interviews with 8 clinicians from different involved specialties. The questions and outcomes are abbreviated, the full questions can be found in Supplementary Note 9. The questions and outcomes are distributed according to phase of the diagnostic and treatment process on the x-axis, and clinical relevance as attributed by the interviewed clinicians on the y-axis. In the first two phases (Pre-hospital phase and ED phase) arrows are placed which indicate how various questions are connected. Grey boxes are added where needed to improve the interdependence of various questions.

2.4. Data pre-processing

2.4.1. Time-dissection of data set

To identify the earliest possible time-point for mortality predictions, the data set was split into 11 subsets. Respective subsets comprised: ENTRY (upon hospital admission), PRE-SURGERY (prior to first surgical procedure), POST-SURGERY (posterior to first surgical procedure and prior to ICU admission), BASELINE (BL; first 24 h of ICU admission), ICU-day1 (first day in ICU), ICU-day2 (second day in ICU), ICU-day3, ICU-day4, ICU-day5, ICU-day6, ICU-day7. The ICU day follows the fluid charts, typically from 06.00 to 06.00. Day 1 is from admission time to the start of the fluid chart the next day, giving variable lengths of day 1 from 0 to 24 h. To address this discrepancy between patients, the BASELINE data set was included, as variables therein pertain to time of ICU admittance and 24 h forth.

2.4.2. Data cleaning & imputation

The selected (binary) target was 30-day mortality (upon first admission to the study centre). All other clinical endpoints were omitted, as were patients with missing mortality data (n = 4), resulting in a sample of 349 alive and 56 deceased patients. Only data available upon ICU admission were included. Irrelevant or potentially biasing variables (e.g. PatientIDs, dates), duplicates, or variables with disputable accuracy due to subjectivity of the data (i.e. estimated Glasgow Coma Scale) were additionally removed from the analysis. Imputation was performed if the total number of missing entries per variable did not exceed 5%, while variables with higher percentages of missing data were discarded. Discarded variables included mostly cytokine measurements, few preoperative lab values (hemoglobin, glucose, lactate, natrium), skin anesthesia and crepitus upon presentation, gas upon radiology, alcohol, and smoking status. To account for the mixed data types present in the data sets, we differentiated between continuous (numerical), binary, and categorical features during imputation. Continuous features were treated using the *IterativeImputer* from scikit-learn [42] and scaled through min–max normalisation. Missing binary information was completed using k-Nearest Neighbours method. Categorical features, such as hospital names, were imputed by the most frequently occurring value and encoded to numerical representation by an ordinal encoder. For the total numbers of variables and patients included in each subset after data cleaning and imputation see Supplementary Note 3.

2.4.3. Variable selection

Relevant variables were selected using a combination of unsupervised filtering and manual curation in a two-step process. Firstly, the preprocessed, time-dissected data sets underwent an unsupervised feature selection step, in which the full feature space was filtered using the python implementation of the Boruta algorithm [43]. Secondly, during optimization of the best-performing model, the filtered variables were manually curated, removing variables that were deemed impractical to use in clinical practice. A more detailed description on the feature selection procedure can be found in Supplementary Note 2. The original variable names in the INFECT data set, abbreviations used in this publication, and a more detailed clinical description of curated variables can be found in Supplementary Note 4.

2.5. Classification

2.5.1. Model development and validation

Random Forest Classifiers (RFC) were utilized to predict patient mortality [44]. Robust internal validation was achieved by an iterative 5-fold double cross-validation (DCV) approach (100 iterations). To quantify the quality of the classification models, we calculated and compared several different metrics, including areas under the receiver operating characteristic curve (ROC AUC), the F_1 score, and the F_2 score. Conversion of SAPS II and SOFA scores to probabilities of mortality was

done using the relationship established by Le Gall et al. [32] and Moreno et al. [45], respectively. Confidence intervals of ROC curves were computed using bootstrapping. For the p-values, a two-sided test for difference in AUC was performed. For additional information on model development, validation and metrics please refer to Supplementary Note 2. Model calibration curves can be found in Supplementary Note 8 (Supplementary Fig. 7).

2.5.2. Assessing model stability

To assess the robustness of our systems towards missing variables, iterative random removal of variables was conducted. Therefore, a pre-specified number of variables (between 1 and $m-1$ where m is the total number of variables in the subset) were removed from the data set and the reduced model was trained and validated using the same approach as during model development.

2.5.3. Selection of surrogate variables

Surrogate variables were determined by calculating the absolute Pearson correlation for selected (primary) variables with the whole feature space, excluding cross-correlation amongst primary variables themselves. The effects on model performance were assessed by replacing missing variables through surrogates, random variables, or removing them from the data set for all patients. Random variables were defined as features with absolute correlations of less than 5% with any primary variable (correlation < 0.05). Model training and validation were carried out using the iterative DCV described above (100 iterations).

2.6. Software

For all classification algorithms the implementations available in the scikit-learn Python library (version 0.24.1) [42] were used. ROC curve bootstrapping and p-value calculation was done using the R package *pROC* [46].

3. Results

3.1. Interviews

The interviews yielded a total of 24 unique questions that were deemed relevant in the diagnostic and treatment process of patients with NSTI. All emerged during the first five interviews. Of these questions, 14 were treatment support questions, and 10 were predictions (Fig. 1). Most questions (14) concerned the ICU phase, but relevant questions were also identified in the pre-ICU phase (7) and the pre-hospital phase (4). One question, regarding expected microbial etiology, was deemed relevant in both the ICU and pre-ICU phase, and therefore mentioned in both phases with a different score for relevance (Fig. 1). As can be observed in the table (Supplementary Note 9, Supplementary Table 4) and figure (Fig. 1), there was substantial variation (1.5–3.0) in the average scores for relevance attributed to the different questions. Although all questions are relevant to varying degrees, those that can potentially be answered by the available INFECT data set are of most relevance for the initial development of a CDSS on NSTI. Among the most relevant (score >2) of these questions were the prediction of causative microbes, chance of developing septic shock, chance of mortality, and chance of developing Acute Kidney Injury (AKI). Of these relevant endpoints, the prediction of mortality was selected as the first to develop an artificial intelligence based approach within the CDSS.

3.2. Earliest time point for prediction of mortality

Comparison of prediction performances showed distinct differences between different time-dissected data (Fig. 2). Prediction with ICU data sets (BASELINE (BL) - ICU-day7) revealed that performance peaks using data acquired within the first 24 h in the ICU (BL), constituting the earliest time point for satisfactory mortality predictions. Therefore, the BL data set was selected to act as the base for further analysis.

3.3. Model optimisation

The BL model included a total of 20 variables derived through unsupervised feature selection. To further refine our model, we conducted manual curation, leading to the removal of the variables ‘total blood product administration’ (impracticality), ‘total fluid administration’, ‘systolic blood pressure’ (duplicate entries), ‘Glasgow Coma Score (GCS)’ (disputable accuracy due to subjectivity of the data), after which 16 variables remained (Fig. 3A, Fig. 3B, Supplementary Note 4). Comparison of the predictive power of this model to the SAPS II and SOFA score (Fig. 3C) revealed excellent discriminatory power (AUC = 0.91 (95% CI, 0.88–0.96)) of our system, outperforming the SOFA score (AUC = 0.77 (95% CI, 0.69–0.84)), p-value=$5.07E-05$) and showing

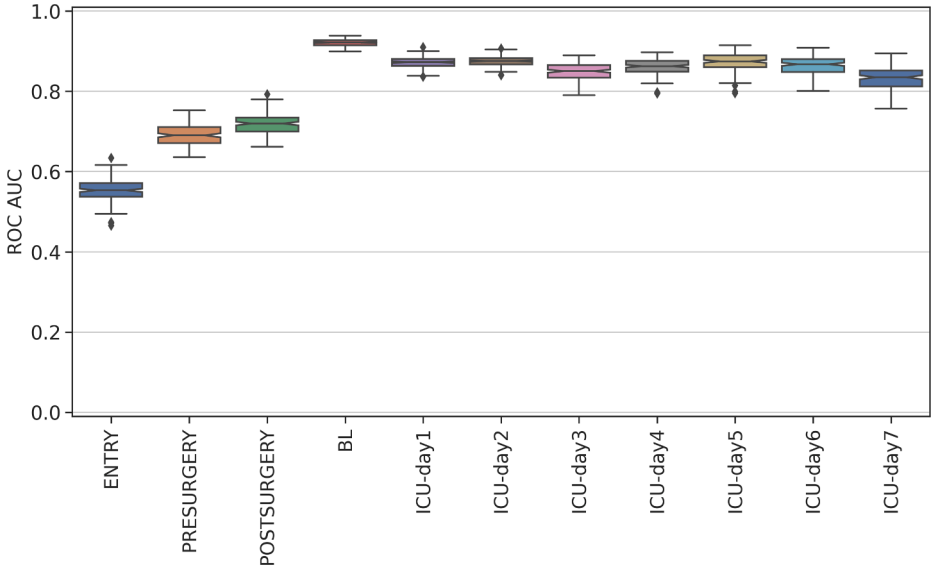


Fig. 2. Comparison prediction performances of time-dissected data sets. Scores displayed as average F1-score (*macro* averaging) over 100 double cross-validation iterations. Notches represent 95% confidence interval around the median.

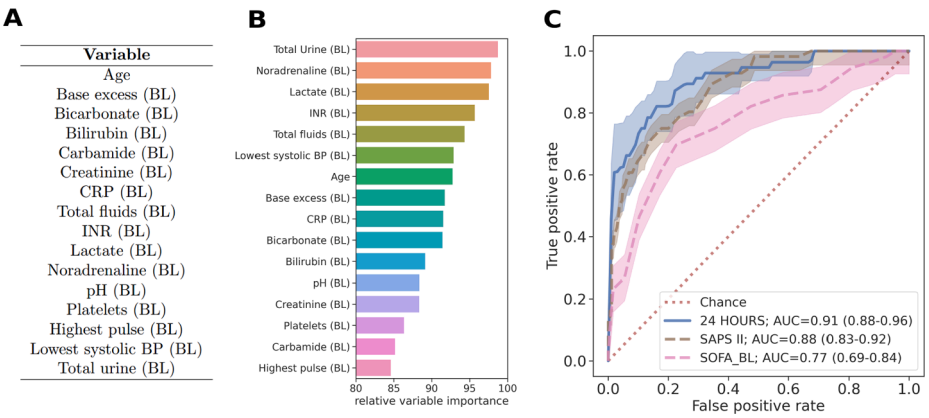


Fig. 3. Overview (A) manually curated variables (ordered alphabetically), (B) respective feature importance, and (C) prediction performances of manually curated variables displayed as ROC curve with SAPS/SOFA score as comparison. The relative importance displayed is the number of iterations (100) minus the (geometric) mean ranking of variables according to their importance during mortality predictions; rank 1 indicated the most important variable and rank 16 the least important.

slightly, yet not statistically significantly, higher AUC than the SAPS II score (AUC = 0.88 (95% CI, 0.83–0.92), p-value=0.07).

Probabilities of death were derived and examined for all patients (Fig. 4A). The calculated likelihood of mortality for patients that are known to be alive (blue) ranges from 0% to around 40% with a dominating peak at around 10%, illustrating the ability of our system to proficiently identify non-critically ill patients. In the probability distributions of patients known to have deceased (orange), the picture is less distinct. Three subgroups of patients can be distinguished, with peaks around 20, 60, and 80% respectively. The lowest-performing subgroup constituted more often of individuals dying at later time points (after day 10), indicating that predictions are better for early deaths (Fig. 4B). It is evident from Fig. 4A that using an intuitive threshold of 50% for identifying patients with higher probabilities of death is not optimal and results in elevated numbers of false-negative predictions (Table 1). Determination of the optimal threshold (through trying to minimize the number of false negatives) instead yielded an ideal cutoff value of 30% (26%, red line), resulting in a good trade-off between the number of false-positive (FP) and false-negative (FN) predictions.

3.4. *Improving flexibility and usability*

To maximize the flexibility of the developed system the 16 selected variables were grouped into subsets according to their availability (Fig. 5A). Performance comparison showed that not all identified predictors needed to be included for the model to perform well (Fig. 5B). Using only variables obtainable within an hour in the ICU (BLOOD set) resulted in a satisfactory mortality prediction. Comparison to the SAPS II and SOFA score proved good discriminatory power with AUC < 0.8 (Supplementary Note 5).

Table 1
Summary of model accuracy at different threshold levels. Highlighted in bold is the threshold identified as optimal trade-off between model precision and recall when aiming to reduce false-negatives as much as possible.

Threshold [%]	TN [%]	FP [%]	FN [%]	TP [%]	TPR	TNR
0	0	86	0	14	1.00	0.00
10	62	24	1	13	0.93	0.72
20	75	12	3	11	0.79	0.86
30	**80**	**6**	**4**	**10**	**0.71**	**0.93**
40	84	2	6	8	0.57	0.98
50	85	1	7	7	0.50	0.99
60	86	0	9	5	0.36	1.00
70	86	0	11	3	0.21	1.00
80	86	0	12	2	0.14	1.00
90	86	0	14	0	0.00	1.00
100	86	0	14	0	0.00	1.00

TN - true negatives FP - false positives FN - false negatives.
TP - true positives TPR - true positive rate TNR - true negative rate.

In an everyday clinical setting, some variables may be missing. To simulate the effect of missing data a specified number of random variables from our data sets were iteratively removed and subsequently the predictive power of the perturbed set was measured (Fig. 6). This highlights the stability of our system towards missing information, with

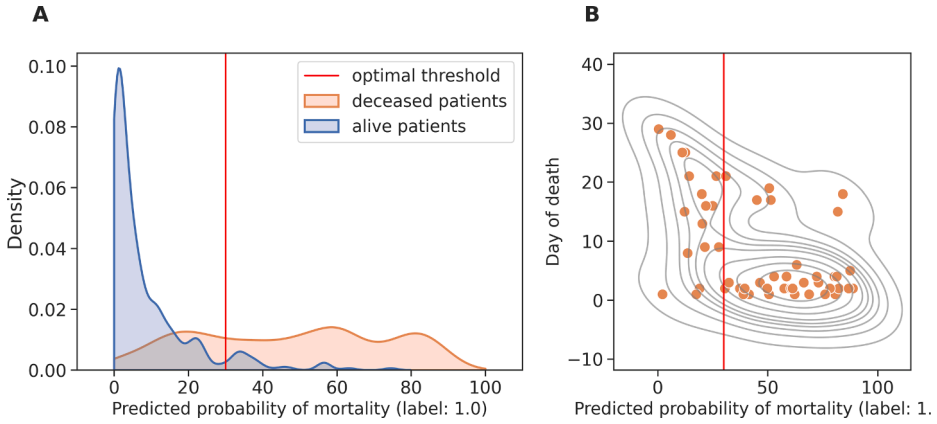


Fig. 4. Calculated probabilities of patient mortality. (A) CDSS predictions for patients known to be alive (blue) or deceased (orange). (B) Patients known to be deceased, plotted with regard to their time-point of death. Red line indicates the optimal decision threshold when aiming to minimise false-negative predictions. Densities calculated from the average probability of mortality for each patient over 100 iterations.

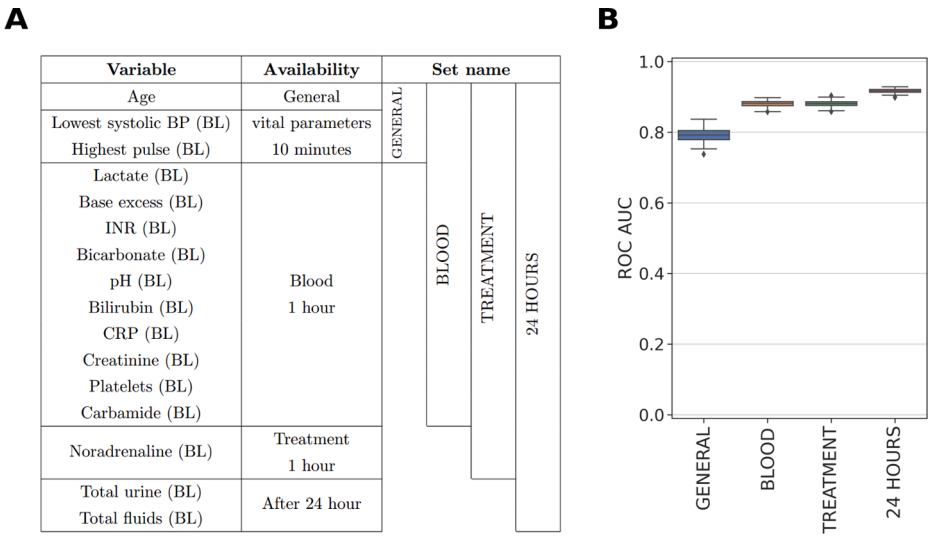


Fig. 5. Taking variable availability into account. (A) Variables included in subsets of BL model. Time points indicate the approximate time of availability in the ICU. (B) Prediction performances of models trained on selected variable subsets.

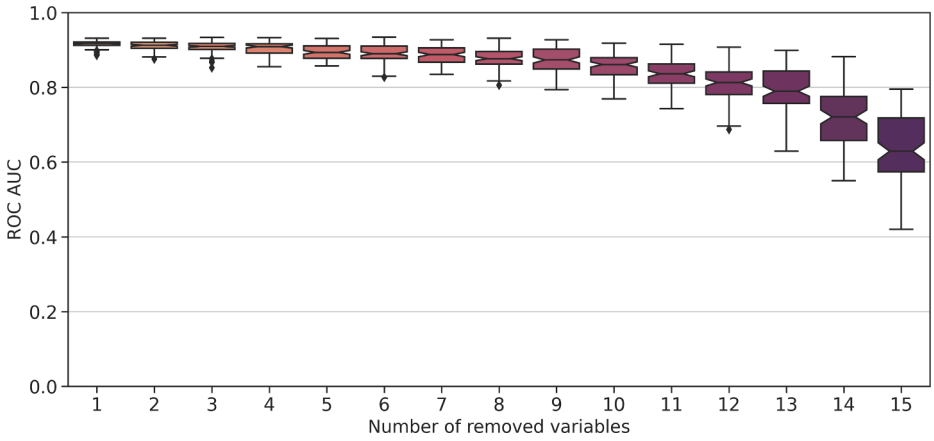


Fig. 6. Assessment of model stability through iterative removal of random variables. 24 HOUR set with a total number of 16 variables. Displayed as average ROC-AUC score over 100 iterations. Notches represent 95% confidence interval around the median.

ROC AUC scores remaining high (AUC 0.9) even when removing more than half of the data (e.g. removing variables: AUC = 0.89 (95% CI, 0.88–0.89)). Similar results were obtained when using the early available BLOOD set (Supplementary Note 6). The gradual increase of standard deviation can be accredited to differences in the importance of variables (Fig. 3B). With higher missingness rates, important features in the successful prediction of patient mortality are more likely to be absent, and subsequently prediction performance decreases.

Despite the robustness of our system, excluding (key) variables like urine, lactate, or amount of administered noradrenaline, has a negative impact on model performance. Therefore, alternative measurements replacing potentially missing variables were identified (Table 2). Exchanging variables with their highly correlated alternatives revealed no noticeable loss of performance, illustrating the relevance of the approach (Supplementary Note 7, Supplementary Fig. 5). In the case of multiple missing variables, compensation for absent values performs better than simply removing the missing values from the model. This is of special importance in scenarios of high missingness rates (> 60%) (Supplementary Note 7, Supplementary Fig. 6).

Table 2
Overview on selected surrogate variables and their respective Pearson correlation to the original variable. Best performing surrogates are highlighted in bold.

Original	Surrogate	Correlation
Age	-	-
Lowest systolic BP (BL)	Skin bullae (preop)	−0.17
Highest pulse (BL)	Potassium (BL)	0.27
Lactate (BL)	Potassium (BL)	0.26
Base excess (BL)	Creatinine (preop)	−0.33
INR (BL)	Chronic liver disease	0.29
Bicarbonate (BL)	Creatinine (preop)	−0.36
pH (BL)	Creatinine (preop)	−0.37
Bilirubin (BL)	Chronic liver disease	0.26
CRP (BL)	**CRP (preop)**	**0.67**
Creatinine (BL)	**Creatinine (preop)**	**0.88**
Platelets (BL)	WBC (BL)	0.35
Carbamide (BL)	**Creatinine (preop)**	**0.63**
Noradrenaline (BL)	Potassium (BL)	0.22
Total urine (BL)	Creatinine (preop)	−0.25
Total fluids (BL)	Corticosteroid use	0.22

4. Discussion

This study provides a proof of concept for the development of a CDSS. After the identification of clinically relevant endpoints, one of the most relevant (30-day mortality) was taken as health endpoint to develop a machine learning based predictive system. The developed

model was found to be proficient in predicting mortality with similar or better accuracy as general scoring systems in use.

The variables important for prediction were diverse, ranging from demographics and vital measurements to information obtainable only after a certain time period spent in the ICU. However, all of the selected variables are well-known parameters when assessing the vital state of patients, especially in the ICU. Therefore, many of the identified variables can also be found in established scoring systems like the SAPS II (age, pulse, systolic blood pressure, bicarbonate, carbamide)[32], SOFA score (platelets, bilirubin, creatinine, urine output, vasopressor requirement)[25], or APACHE IV score (pH) [47]. Also previously described as associated with mortality were lactate [33,34], base excess [35], C-reactive protein [36], and INR [37]. Since fluid requirement increases depending on sepsis severity, this predictor was expected as well. More computationally-oriented publications comparable to the work carried out here have reported similar variables as informative, although their ranking of importance differs [38–40,29]. The fact that all selected variables were previously found to be associated with risk of mortality in sepsis patients validate the findings from this study. It is evident that the selected variables do not include information unique to NSTI, such as the type of causative micro-organism, anatomical location affected, or surgical findings. This is likely due to the fact that mortality of these patients in the ICU is primarily due to sepsis, making it logical that identified predictors are connected to the systemic disease rather than the local characteristics of NSTI. Estimation of NSTI specific endpoints, such as wound size or need for amputation, may yield a more specialised set of predictive variables. Our CDSS performed equally well as the long established SOFA and SAPS score. Reasons thereof can be various, and might include the fact that our CDSS uses more variables (16 variables versus 6 (SOFA) or 15 (SAPS II)). An observation supporting this is that our performance suffers upon including less variables (Fig. 5B). Another explanation may lie in the algorithm underlying the predictive systems; in case of complex, non-linear interactions between variables, Random Forest may be better suited to model the data structure than the more simplistic logistic regression model used by the SOFA score. The final CDSS envisioned should not merely deliver a binary prediction of patient outcomes, but rather give information on the likelihood of an event, which can subsequently be interpreted by clinicians. When examining the calculated likelihood of death for each patient, the CDSS notably seems much more capable of identifying survivors than patients at risk. This pronounced difference might be ground in the heavy imbalance observed in the data set, potentially favoring the identification of alive patients. The observed clustering of deceased patients during the first week is explainable from a clinical perspective, since patients with refractory septic shock will often die in the first day after admission. Since those early deaths represent the majority of those who died during admission, it is unsurprising that most predictors selected are related to septic shock, and may therefore lead to the most accurate predictions for early deaths.

The exceptional stability of our CDSS is of high clinical relevance, as missing measurements are frequent in a typical clinical setting. Strategies of handling missing data when working with the existing scoring systems include imputation strategies such as i) replacement by a previous measurement or ii) assuming the value to be in the normal physiological range. Although easy to use, both of these approaches may introduce erroneous data points potentially biasing predictions and should thus be applied with care. Not all variables possess equal explanatory power, leading to more and less favorable scenarios of data missingness. Suggestions of alternative measurements could mitigate the effect of missing variables, especially of those with high predictive power. Our results suggest that including alternative variables can be successfully implemented in the case of well-suited surrogates (e.g. with correlation > 0.5) and is of special importance when multiple measurements are missing.

By engaging the clinical specialists upfront through interviews, we were able to not only clearly identify clinical needs, but also rank them according to relevance. Although no classic qualitative approach was applied and the ranking system used is not a validated method, we believe the obtained overview is sufficient to act as a starting point for the design of a CDSS. However, although questions in all phases would ideally be supported, the data set utilised in this study limits the possibilities to perform predictions for questions in the pre-ICU phases, mainly because of the lack of sufficient pre-hospital data as well as the lack of non-NSTI patients for diagnostic predictions. Therefore, the initial CDSS will be designed for use in the ICU, and include the most relevant questions in the ICU phase. When adjacent data sets become available, this could be expanded in the future. During model development, we used the largest NSTI cohort available as of today. Unfortunately, currently there is no NSTI cohort of comparable quality available which could have been used as a validation cohort. However, the authors behind this study are in the process of planning a prospective study on Dutch patients to validate and improve the system. We believe that a prospective validation is the most appropriate study design to test a decision support system, as it not only gives feedback on the generalizability of the model, but also on its practicality, as well as a direct comparison to clinicians' decisions, ultimately justifying the incorporation of a CDSS. Despite the current lack of external validation we have taken care of deploying robust internal validation techniques during variable selection and model optimization phases. Imbalance of labels is a common problem when working with health care data. While resampling techniques can be used to equalize ratios, these must be used with caution, as they artificially craft patients. We thus regarded the imbalance ratio as 'natural', as it reflects the real mortality rates and took care in maintaining the same ratio of labels during data splitting. The final framework uses a small number of predictors that are readily available in most ICUs, which facilitates international use. The possibility of providing various input variables allows this CDSS to be used at different stages throughout the treatment process. Despite the potential benefits of a CDSS it must be stressed that there is an inherent risk for misuse, as with any tools of this kind. Special care must be taken when relating probabilities to outcomes, as e.g. the intuitive classification threshold of 50% does not translate to a 50% risk of mortality. Proper application and consistent monitoring of results will be essential in fostering trust in a DSS utilizing machine learning algorithms.

To successfully deploy the comprehensive multiple-endpoint CDSS envisioned, future efforts will cover the inclusion of a data management system, the extension of the framework to cover a wider array of clinical questions identified in the interviews, and model deployment through a (free) web application. Additionally, the integration of more diverse data types will be aspired (e.g. -omics data) to refine patient stratification and deduce underlying biological disease mechanisms. The lack of NSTI-specific information needed to estimate the risk of mortality provides the opportunity of testing the developed framework on a more general sepsis cohort, thereby broadening its field of application. Simultaneously, however, the inclusion of more NSTI-specific variables will enable the creation of a specialized framework addressing health endpoints unique to NSTI, such as the prediction of suspected microbiological species.

5. Conclusion

In summary, our study lays the foundation of a comprehensive CDSS for NSTI patients. To the best of our knowledge, we have for the first time provided a qualitative assessment of the clinically relevant questions in NSTI patient care. By intertwining clinical and bioinformatic expertise, we have developed a tool proficient in predicting 30-day mortality for patients with NSTI admitted to the ICU. The possibility for users to adapt the input variables without severely affecting model performance is a major advantage compared to other clinical scoring systems currently in use. Furthermore, the framework itself can be easily expanded to other health endpoints related to NSTI diagnosis and treatment, thus creating a universal tool for improving NSTI care and

outcomes.

6. Summary Table

What was already known on the topic:

- Resource allocation for NSTI patients in ICU care is complex and multifactorial
- Not sufficient clinical tools exist for adequate NSTI patient care

What this study added to our knowledge:

- Identification and ranking of clinical problems at different stages in NSTI patient care
- Machine learning algorithms outperform clinical scoring systems currently in use
- A CDSS developed in close collaboration with clinicians has the potential to assist and improve clinical decision making

7. Author information

7.1. Contributions

JS, SK, CH and VMdS designed the study. JS conducted interviews. Interview partners included OH, SS, AMV. SK and CH developed the prediction pipeline. AP, AMV, MBM, OH, Paul van Zuijlen (PZ), SS, ES, VMdS, AMV, AP, CH aided in interpretation of results and critical revision of the manuscript. JS and SK wrote the manuscript. All authors read, commented, and approved the final manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to all members of the INFECT, PerAID and PerMIT collaborations, as extensive discussion within the consortium have greatly contributed to the finalisation of this manuscript. Beside named authors in this article, the INFECT Study Group includes: Michael Nekludov, Ylva Karlsson, Per Arnell, Morten Hedetoft, Marco B. Hansen, Peter Polzik, Daniel Bidstrup, Nina F. Bærnthsen, Gladis H. Frendø, Erik C. Jansen, Lærke B. Madsen, Rasmus B. Müller, Emilie M. J. Pedersen, Marie W. Petersen, Frederikke Ravn, Isabel F. G. Smidt-Nielsen, Anna M. Wahl, Sandra Wulffeld, Sara Aronsson, Anders Rosemar, Joakim Trogen, Trond Bruun, Torbjørn Nedrebø, Oddvar Oppegaard, Eivind Rath and Marianne Sævik. The PerAID/PerMIT Study groups, besides named authors, include: Mattias Svensson, Kristoffer Strålin, Trond Bruun, Oddvar Oppegaard, Knut Anders Mosevoll, Jan Kristian Damås, Paul van Zuijlen, Laura M. Palma Medina, Lorna Morris, and Marco Anteghini. The full list of all Study Group members, national site investigators, and their affiliation can be found in Supplementary Note 1. The authors are indebted to Stephan G.F. Papendorp, Marieke Verhaar, Fabienne A.M. Roossien, Evelien de Jong and Vincent M. de Jong for their participation in the interviews. The authors thank the supporters of this study: the European Union Seventh Framework Programme (FP7/2007–2013) under the grant agreement 305340 (INFECT project); the Swedish Governmental Agency for Innovation Systems (VINNOVA), Innovation Fund Denmark, and the Research Council of Norway under the frame of NordForsk (project No. 90456, PerAID); the Swedish Research Council, Innovation Fund Denmark, the Research Council of Norway, the Netherlands Organisation for Health Research and Development (ZonMW), and DLR Federal Ministry of Education and Research, under the frame of ERA PerMed (project 2018–151, PerMIT); the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 860895 TranSYS.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ijmedinf.2022.104878.

References

[1] M. Peetermans, et al., Necrotizing skin and soft-tissue infections in the intensive care unit, Clin. Microbiol. Infect. 26 (2020) 8–17.
[2] D.L. Stevens, A.E. Bryant, Necrotizing Soft-Tissue Infections, N. Engl. J. Med. 377 (2017) 2253–2265.
[3] M.B. Madsen, et al., Patient's characteristics and outcomes in necrotising soft-tissue infections: Results from a Scandinavian, multicentre, prospective cohort study, Intensive Care Med. 45 (2019) 1241–1251.
[4] G. Jabbour, et al., Pattern and predictors of mortality in necrotizing fasciitis patients in a single tertiary hospital, World J. Emerg. Surg. 11 (2016) 40.
[5] S.F.L. van Stigt, et al., Review of 58 patients with necrotizing fasciitis in the Netherlands, World J. Emerg. Surg. 11 (2016) 21.
[6] E. Audureau, et al., Mortality of necrotizing fasciitis: Relative influence of individual and hospital-level factors, a nationwide multilevel study, France, 2007–12, Br. J. Dermatol. 177 (2017) 1575–1582.
[7] L.K. Tom, et al., Comparison of Traditional and Skin-Sparing Approaches for Surgical Treatment of Necrotizing Soft-Tissue Infections, Surg. Infect. 21 (2020) 363–369.
[8] Z. Al-Qurayshi, R.L. Nichols, M.T. Killackey, E. Kandil, Mortality Risk in Necrotizing Fasciitis: National Prevalence, Trend, and Burden, Surg. Infect. 21 (2020) 840–852.
[9] D.L. Horn, et al., Predictors of mortality, limb loss, and discharge disposition at admission among patients with necrotizing skin and soft tissue infections, J. Trauma Acute Care Surg. 89 (2020) 186–191.
[10] Hakkarainen, T.W., Burkette Ikebata, N., Bulger, E. & Evans, H.L. Moving beyond survival as a measure of success: Understanding the patient experience of necrotizing soft-tissue infections. The Journal of Surgical Research 192, 143–149 (2014).
[11] A.-M. Fagerdahl, V.E. Knudsen, I. Egerod, A.E. Andersson, Patient experience of necrotising soft-tissue infection from diagnosis to six months after intensive care unit stay: A qualitative content analysis, Australian Critical Care: Official Journal of the Confederation of Australian Critical Care Nurses 33 (2020) 187–192.
[12] V.E. Knudsen, A.E. Andersson, A.-M. Fagerdahl, I. Egerod, Experiences of family caregivers the first six months after patient diagnosis of necrotising soft tissue infection: A thematic analysis, Intensive Crit. Care Nurs. 49 (2018) 28–36.
[13] T. Urbina, et al., Long-term quality of life in necrotizing soft-tissue infection survivors: A monocentric prospective cohort study, Annals Intensive Care 11 (2021) 102.
[14] F. Hietbrink, L.G. Bode, L. Riddez, L.P.H. Leenen, M.R. van Dijk, Triple diagnostics for early detection of ambivalent necrotizing fasciitis, World J. Emerg. Surg.: WJES 11 (2016) 51.
[15] H.J. Schünemann, et al., Grading quality of evidence and strength of recommendations for diagnostic tests and strategies, BMJ 336 (2008) 1106–1110.
[16] M. Sartelli, et al., World Society of Emergency Surgery (WSES) guidelines for management of skin and soft tissue infections, World J. Emerg. Surg. 9 (2014) 57.
[17] D.L. Stevens, et al., Practice Guidelines for the Diagnosis and Management of Skin and Soft Tissue Infections: 2014 Update by the Infectious Diseases Society of America, Clin. Infect. Dis. 59 (2014) e10–e52.
[18] Larry M Baddour, D.L.S. Necrotizing soft tissue infections. UpToDate (2021).
[19] C.-H. Wong, L.-W. Khin, K.-S. Heng, K.-C. Tan, C.-O. Low, The LRINEC (Laboratory Risk Indicator for Necrotizing Fasciitis) score: A tool for distinguishing necrotizing fasciitis from other soft tissue infections, Crit. Care Med. 32 (2004) 1535–1541.
[20] S.M. Fernando, et al., Necrotizing Soft Tissue Infection: Diagnostic Accuracy of Physical Examination, Imaging, and LRINEC Score: A Systematic Review and Meta-Analysis, Ann. Surg. 269 (2019) 58–65.
[21] X. Song, X. Liu, F. Liu, C. Wang, Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis, Int. J. Med. Informatics 151 (2021) 104484.
[22] M. Flechet, et al., AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: Development, validation and comparison to serum neutrophil gelatinase-associated lipocalin, Intensive Care Med. 43 (2017).
[23] C.R. Yee, N.R. Narain, V.R. Akmaev, & Vemulapalli, V.A Data-Driven Approach to Predicting Septic Shock in the Intensive Care Unit, Biomedical Informatics Insights 11 (2019), 1178222619885147.
[24] J.L. Vincent, et al., The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure, Intensive Care Med. 22 (1996) 707–710.
[25] J.R. Le Gall, S. Lemeshow, F. Saulnier, A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study, JAMA 270 (1993) 2957–2963.
[26] R.P. Moreno, et al., SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission, Intensive Care Med. 31 (2005) 1345–1355.
[27] W.A. Knaus, E.A. Draper, D.P. Wagner, J.E. Zimmerman, APACHE II: A severity of disease classification system, Crit. Care Med. 13 (1985) 818–829.

[28] M. Jahn, et al., The predictive performance of SAPS 2 and SAPS 3 in an intermediate care unit for internal medicine at a German university transplant center; A retrospective analysis, PLOS ONE 14 (2019) e0222164.
[29] S.M. Lauritsen, et al., Explainable artificial intelligence model to predict acute critical illness from electronic health records, Nature Communications 11 (2020) 3852.
[30] S. Nemati, et al., An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU, Critical care medicine 46 (2018) 547–553.
[31] R.T. Sutton, et al., An overview of clinical decision support systems: Benefits, risks, and strategies for success, npj Digital Medicine 3 (2020) 1–10.
[32] J.R. Le Gall, et al., Customized probability models for early severe sepsis in adult intensive care patients, Intensive Care Unit Scoring Group. JAMA 273 (1995) 644–650.
[33] M. Singer, et al., The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3), JAMA 315 (2016) 801–810.
[34] Z. Liu, et al., Prognostic accuracy of the serum lactate level, the SOFA score and the qSOFA score for mortality among adults with Sepsis, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine 27 (2019) 51.
[35] Montassier, E. et al. Base excess is an accurate predictor of elevated lactate in ED septic patients. The American Journal of Emergency Medicine 30, 184–187 (2012).
[36] S.M.A. Lobo, et al., C-reactive protein levels correlate with mortality and organ failure in critically ill patients, Chest 123 (2003) 2043–2049.
[37] C.M. Fischer, K. Yano, W.C. Aird, N.I. Shapiro, Abnormal coagulation tests obtained in the emergency department are associated with mortality in patients with suspected infection, The Journal of Emergency Medicine 42 (2012) 127–132.
[38] D. Chicco, L. Oneto, Data analytics and clinical feature ranking of medical records of patients with sepsis, BioData Mining 14 (2021) 12.
[39] Q. Mao, et al., Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU, BMJ Open 8 (2018) e017833.
[40] N. Hou, et al., Predicting 30-days mortality for MIMIC-III patients with sepsis-3: A machine learning approach using XGboost, Journal of Translational Medicine 18 (2020) 462.
[41] M.B. Madsen, et al., Necrotizing soft tissue infections - a multicentre, prospective observational study (INFECT): Protocol and statistical analysis plan, Acta Anaesthesiol. Scand. 62 (2018) 272–279.
[42] F. Pedregosa, et al., Scikit-learn: Machine Learning in Python, MACHINE LEARNING IN PYTHON 6 (2011).
[43] M.B. Kursa, W.R. Rudnicki, Feature Selection with the Boruta Package, J. Stat. Softw. 36 (2010).
[44] L. Breiman, Random Forests, Machine Learning 45 (2001) 5–32.
[45] R. Moreno, et al., The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study, Intensive Care Med. 25 (1999) 686–696.
[46] X. Robin, et al., pROC: An open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinformatics 12 (2011) 77.
[47] J. Zimmerman, A. Kramer, D. McNair, F. Malila, Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital Mortality Assessment for Today's Critically Ill Patients, Crit. Care Med. 34 (2006) 1297–1310, 5.