

Novel Transparent Machine Learning Algorithm - Predicting Sepsis from Routine ED Data.

Alf Timney

`alf.timney@icr.ac.uk`

Institute of Cancer Research

Kate Honeyford

University College London

Céire Costello

Imperial College London

Article

Keywords: Evidence Based Emergency Medicine, Machine Learning, Sepsis, NHS

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10443491/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

We developed a novel methodology to create a transparent machine learning algorithm using Shapley value approximations. This is the first study to transform local Shapley explanations into a globally transparent algorithm. We trialled and validated our technique on the problem of predicting a sepsis diagnosis from routine ED patient data. This was a retrospective cohort study, using electronic patient records from two NHS trusts. We developed three models in our study: NNS1 was a shallow neural network representative of a “black box” algorithm. ShAprox was a transparent approximation of NNS1. ShAproxRef was a further refinement on ShAprox. Our study included 28,072 patient spells – 8.6% included a sepsis diagnosis. ShAproxRef achieved 0.74 AUC on internal testing and 0.82 AUC on external testing, both comparable to NNS1’s results. We offer a feasible approach to detect sepsis at the earliest time point, using a fully transparent algorithm improving clinical trust, and thereby aids implementation.

Introduction

Sepsis is a common and potentially life-threatening condition the NHS is working to achieve more consistent early identification and treatment of, as every hour delay in treatment increases mortality risk by 7.6%.^[1] The “early treatment of sepsis is not complicated or expensive”,^[2] however 48,000 patients in the UK die from sepsis each year.^[3]

To aid with the early detection of sepsis NHS trusts use a range of algorithms, often linked to digital systems. We have previously shown that digital sepsis alerting algorithms are associated with reduced mortality in NHS trusts.^[4] Implemented algorithms include NEWS2,^[5] and the St. Johns Sepsis Agent.^[6] These are expert systems^[7] which try to emulate the decision-making process of medical experts. These algorithms can effectively alert clinicians to a patient’s deterioration.^{[8],[9]}

The next step in research is the identification of patients with sepsis before deterioration. Elderly patients who have survived severe sepsis are at increased risk of long-term cognitive impairment and functional disability.^[10] Earlier treatment of sepsis would reduce severe cases, improve patient’s long-term outcomes, and reduce hospitals’ cost for sepsis treatment.^{[11],[12]} While expert systems can accurately predict mortality in deteriorating patients,^[13] machine learning offers potential benefits in earlier prediction.

Machine learning is well suited to the problem of predicting sepsis during emergency room triage^[14] however, standard interpretability tools such as SHapley Additive exPlanations (SHAP)^{[15],[16]} & Local Interpretable Model-agnostic Explanations (LIME)^[17], do not provide enough insight to increase clinicians trust in existing “black box” algorithms. Without sufficient interpretability, these algorithms^[18] are unlikely to be implemented and therefore cannot help patients.

Shapley values^[19] are a game theory concept which provide optimal distribution of profit/loss between players in a game, based on the players contributions. This approach has been implemented for machine learning explanations with (SHAP) and applied to recent sepsis machine learning model publications.^{[14],}

^[38] Our study builds on this work. We use a neural network to predict a patient's future sepsis diagnosis only from the patient's initial vitals, demographic, and comorbidity information. We then demonstrate our novel ShAprox methodology for creating globally explainable, transparent algorithms informed by Shapley values.

Methods

Study design and data source

This was a retrospective cohort study using data from two acute NHS healthcare trusts. The data is derived from patients who attended A&E (6/3/2019–12/8/2020) and were subsequently admitted to hospital. 80% of the data from Site 1 was used as training data, while the remaining 20% was used for internal testing. Data from Site 2 was solely used for external testing.

Study population

Patients aged 18 or older, who were admitted through the emergency department (ED) and who received a diagnosis included in Inada-Kim *et al*'s Suspicion of Sepsis (SoS) list.^[20] This list can be used to identify patients with a serious bacterial infection.^[12] Each attendance is a separate event in the data, and every eligible event was included.

Outcome

A diagnosis of sepsis at any point in the patient's stay, extracted from the patients ICD-10 codes. (Supplementary Values 2).

Predictor variables

Measurements of a patient's first recorded vital signs: heart rate, respiratory rate, temperature, systolic blood pressure, and Glasgow Coma Scale (GCS) score. Patient's demographic characteristics: age, gender, and ethnicity. Patient's comorbidities: patient's Elixhauser comorbidity measures,^[21] extracted from the patients ICD10 diagnosis codes using the comorbidity R package.^[22] We excluded fluid and electrolyte disorders, and coagulopathy from the Elixhauser components, as these are more likely to represent complications arising after the diagnosis of sepsis rather than pre-existing comorbidities.

Missing and extreme values

Patients with a missing age or gender were excluded from the study. Patients with a missing ethnicity had their ethnicity coded as "not known". Missing and extreme vitals were imputed as a random healthy value (Supplementary Values 1), for example a heart rate < 35 would be imputed as a random value between 60 and 100. We used the Python random module^[23] with a fixed seed for consistency.

Neural Network architecture and training.

Initially a shallow neural network (NNS1)^[24] was used to predict whether a patient received a sepsis diagnosis. The neural network was made using TensorFlow version 2 in Python.^[25] The training data was weighted to account for the higher proportion of non-septic patients in the dataset, and the values were normalized. From the input layer of 55 variables, we selected a dropout rate of 40% to reduce overfitting.^[26] The model had 3 hidden layers of width 30, 15 and 5 all using the Leaky-ReLU^[27] activation function. Finally, the output layer used a sigmoid activation function to output a probability between 0 and 1. To train the model we minimised the Binary Cross Entropy loss function and used early stopping callback on the training loss.^[28] NNS1 was trained for 106 epochs, and the previous best weights were restored.

ShAprox construction

After evaluating NNS1, we calculated Shapley Values for the training data using SHAP. The Shapley values for each feature were graphed showing the calculated contribution of a feature to each patient's predicted risk of sepsis, and regressions modelled. For the vitals and age these are curves of best fit (of the 3rd degree), and for binary features they are the mean values. Our ShAprox algorithm returns a prediction of a sepsis diagnosis $F(X)$, for patient X , with recorded features of value $\{x_i\}$. This prediction is calculated as the sum of the approximate Shapley values $f_i(x_i)$ for the patient, plus the base rate C . That is: $F(X) = \sum_i f_i(x_i) + C$. We then constructed a refinement of ShAprox, called ShAproxRef. We removed features which were not well represented in the training data (less than 50 cases) or had Shapley values considered too small (between -0.04 , and 0.04).

Model evaluation

We evaluated models using the Area Under the Receiver-operating characteristic curve (ROC AUC) which summarises the model's performance across all possible model thresholds. We also use the Sensitivity at Specificity (SAS) metric,^[29] which assesses the model's performance at an optimal threshold when the model achieves our chosen specificity of 50%. The SAS is the highest point on the ROC curve left of a false positive rate of 0.5. We chose to include the SAS metric to aid the understanding of the clinical implications of these models.

Results

This study includes 28,072 patient spells split between Site 1 ($n = 13,094$) and Site 2 ($n = 14,978$). Site 1 has a higher proportion of patients with a sepsis diagnosis (10.4%) than Site 2 (7.0%). 44.5% of patient spells were missing blood count test results. Table 1 summarises patient characteristics by site and sepsis diagnosis. Site 1 has an older and less ethnically diverse patient population, compared to Site 2. Site 1 also has a higher prevalence of comorbidities than Site 2, potentially because of the age difference. Both sites see a greater prevalence of elderly, male, and comorbid patients diagnosed with sepsis. Details on specific comorbidities are available in the appendix. The most common comorbidity in both sites is uncomplicated hypertension.

Table 1
Demographic characteristics of patients with and without sepsis, by site.

Site		Site 1				Site 2			
Sepsis Diagnosis		Non-Sepsis		Sepsis		Non-Sepsis		Sepsis	
Counts	n	11735	89.6%	1359	10.4%	13932	93.0%	1046	7.0%
Age Mean (SD)	Mean (SD)	63.3	(21.4)	70.1	(16.5)	49.6	(22.4)	64.3	(18.7)
Male Gender		5524	47.1%	703	51.7%	6722	48.2%	583	55.7%
White		6557	55.9%	704	51.8%	7889	56.6%	682	65.2%
Mixed		148	1.3%	10	0.7%	300	2.2%	16	1.6%
Asian		1784	15.2%	220	16.2%	1157	8.2%	91	8.7%
Black		441	3.8%	48	3.5%	1235	8.9%	91	8.7%
Other Ethnic Groups - Any other ethnic group		812	6.9%	96	7.1%	1654	11.8%	90	8.6%
Ethnicity not known or not stated		1993	17.0%	279	20.5%	1697	12.1%	74	7%

Clinical data from the first set of vital signs are shown in Table 2. There is a large amount of missing vitals data from both sites, but in particular Site 2 has more missing respiratory rate, systolic blood pressure, and GCS score measures for non-septic patients. Patients with sepsis had increased respiratory rates, heart rates and temperatures, as well as lower GCS scores, and greater standard deviation across all these features. The systolic blood pressure in non-septic patients differed between Sites, with Site 1's non-septic patients' blood pressure averaging higher than the septic patients, while Site 2's averaged lower.

Missing or extreme vitals were removed, with random healthy values used instead. In Site 1 the inclusion of imputed temperatures slightly reduced the mean measurement in both the sepsis and non-sepsis sub-groups. In Site 2 the respiratory rate and heart rate were slightly reduced for both groups, however the systolic blood pressure in the non-septic patients were more substantially reduced (Supplementary Table 1).

Table 2

Clinical characteristics of patients, by diagnosis and site. GCS – Glasgow Coma Score. Extreme values were identified by Respiratory rate: (< 10 or > 60), Heart rate: (< 35 or > 195), Systolic blood pressure: (< 60 or > 240), Temperature: (< 25 or > 45), Glasgow coma scale score: (< 3 or > 15)

		Site 1				Site 2			
		Non-sepsis		Sepsis		Non-sepsis		Sepsis	
Respiratory Rate	Mean (SD)	20.04	5.22	22.16	-6.63	18.57	4.71	20.8	5.72
	Missing	54	0.5%	< 5	n/a	1733	12.4%	< 5	n/a
	Extreme	23	0.2%	< 5	n/a	9	0.1%	< 5	n/a
Heart Rate	Mean (SD)	92.49	88.28	100.3	25.5	92.43	19.78	102.52	23.65
	Missing	55	0.5%	< 5	n/a	560	4.0%	< 5	n/a
	Extreme	32	0.30%	5	0.4%	10	0.10%	< 5	n/a
Systolic Blood Pressure	Mean (SD)	133.55	27.41	124.36	28.6	132.44	25.55	124.43	25.51
	Missing	55	0.50%	< 5	n/a	6183	44.4%	16	1.5%
	Extreme	14	0.10%	9	0.7%	8	0.1%	< 5	n/a
Temperature	Mean (SD)	36.87	1.03	37.18	1.41	36.95	0.95	37.4	1.14
	Missing	3274	27.9%	397	29.2%	581	4.20%	< 5	n/a
	Extreme	< 5	n/a	< 5	n/a	< 5	n/a	< 5	n/a
GCS Score	Mean (SD)	14.7	1.19	14.28	2.01	14.48	1.61	14.21	1.91
	Missing	615	5.20%	37	2.7%	8329	59.8%	235	22.5%
	Extreme	< 5	n/a	< 5	n/a	< 5	n/a	< 5	n/a

Shapley Values

The ten most predictive features, ranked by their Shapley values as used in ShAproxRef are shown in Fig. 1. Temperature was the most important feature in the model, with very high or low temperatures increasing the model's prediction by 0.4. Systolic blood pressure, heart rate, respiratory rate, and GCS score were the 2nd, 3rd, 4th, and 7th most important features in the model. All the patient's initial vitals were in the top 10 most important features. Age was the fifth most important feature. Younger patients had a model output 0.2 lower, while elderly patients had a model output 0.05 higher. Lymphoma, metastatic cancer and liver disease were the 6th, 8th, and 9th most important features. They were the most important comorbidities,

all increasing the model’s prediction by more than 0.1. The tenth most predictive feature in the model is Pakistani ethnicity, which reduces the model output by 0.11.

Figure 2 shows the Shapley values for systolic blood pressure and Lymphoma across all patient spells in the training dataset. The dashed black line represents the curve of best fit, which is used to approximate a patient’s Shapley value from their measured systolic blood pressure. A curve of best fit was used for all continuous patient features (Supplementary Figs. 1). The cross represents the mean value, which was used to approximate the Shapley value from a reported Lymphoma comorbidity. The mean was use d for all binary features.

Model performance

Table 3 shows the three models’ performances on the training, internal testing and external testing data. The ShAprox and ShAproxRef models demonstrated similar performance to one another throughout both training and testing phases, and achieved similar performance to NNS1 on both testing datasets. They also performed similarly to NNS1 on both internal and external testing datasets. The ShAprox algorithms achieved identical outcomes on the training and internal testing data, suggesting stronger generalisation capabilities than NNS1, which showed a marked drop in performance from training to testing. While sensitivity analysis of a simpler neural network improved generalisability, it came at the cost of reduced testing performance. All three models performed better on the external testing data than the internal testing data. Training versus validation loss, ROC curves and calibration curves where drawn (Supplementary Figs. 2)

Table 3
ML Model Performances

Training Results	AUC	SAS
NNS	0.845	0.934
ShAprox	0.747	0.822
ShAproxRef	0.744	0.830
Internal Testing Results		
NNS1	0.756	0.814
ShAprox	0.735	0.807
ShAproxRef	0.740	0.831
External Testing Results		
NNS1	0.818	0.911
ShAprox	0.824	0.906
ShAproxRef	0.819	0.906

NNS1: Small neural network

ShAprox: Shapley approximation of NNS1

ShAproxRef: Reduced features refinement of ShAprox

Discussion

Main results

We successfully applied our novel, transparent, algorithm to predict sepsis, from the earliest possible time point - when patients arrive at the emergency department. The ShAprox algorithm combines Shapley values within standard regression models and achieves sepsis prediction performance comparable to a neural network approach.

The results show that both the ShAprox algorithms performed comparably to NNS1 on both the internal and external testing data while having a completely transparent decision-making process. They also performed consistently between the training and internal testing data suggesting strong generalisation of results. All three algorithms performed better on the external testing data than the internal testing data, this emphasises the need to externally verify medical models. Training and testing performance were more similar in the ShAprox models compared to NNS1, suggesting potentially, that the methodology automatically reduces some of the neural networks over training. We were able to further simplify the algorithm by identifying features which contribute most to the prediction, something which is not possible in a neural network, without retraining. When we reduced the features included in the algorithm (ShAproxRef), there was no clinically important change in the performance.

The ShAprox algorithms achieved a high ROC AUC suggesting potential clinical utility. SAS results of 0.8 signify that the ShAprox algorithms can be used to identify 80% of patients who will receive a sepsis diagnosis, with only a 50% false positive rate. This could prompt far earlier treatment of patients, reducing the negative side-effects of sepsis. ShAproxRef also identified age, lymphoma, metastatic cancer, and liver disease alongside vital signs, as key factors predicting a sepsis diagnosis, suggesting more sepsis flagging/predicting models should include these variables. The identification of Pakistani ethnicity as reducing the likelihood of a sepsis diagnosis should not be interpreted causally and may reflect confounding variables or coding differences at Site 1. We reported this result to highlight the value of transparency in machine learning models.

The ShAprox methodology appears well suited to handle ED sepsis triage data. The curve of best fit from Fig. 2 visibly follows the trend of the data while making the model's prediction uniform. For example, with the ShAprox methodology if two patients have the same systolic blood pressure, their blood pressure measurements contribute equally to the prediction. This is not the case for neural networks. The curve of

best fit also means a small change in a patient's systolic blood pressure creates a small change in the ShAprox models prediction. With score based models, such as NEWS2, a small change in a vital sign can cause a multiple point difference in the total score, thus changing the clinical response.

Clinical implications

Our motivation for creating a sepsis triage algorithm specific to ED was to inform the early treatment of patients with clear markers of sepsis. The algorithm was designed to complement existing, implemented sepsis algorithms by focussing on flagging sepsis early with an acceptable specificity, rather than being a continual sensitive alert. For the algorithm to be as effective as possible in a busy ED setting it is important that it does not add to the workload of healthcare workers, therefore we utilised only routine data. Other studies have shown good performance^[30] after including blood test results in their models, but in our datasets of SoS patients, blood tests were not consistently available. Our model is therefore better suited for use with patients who are not suspected of having sepsis and therefore have lower chance of having the full range of blood tests^[31], and in situations where blood test results are not yet available and in low resource settings where blood tests are not always possible.^[32]

We worked towards a transparent algorithm to increase trust in the algorithm's decision-making process. Clinicians basing their decisions on black-box algorithms, such as neural networks, need to trust the equation of the model, the database used to train the model and the cleaning and labelling of the data^[33]. In addition, clinicians are expected to be able to justify their recommendations to patients and colleagues, which is challenging if the driver for the decision is a black-box algorithm. It has been shown that clinicians rate the quality of the AI advice higher if the advice was accompanied with an explanation^[34]. If clinicians do not understand and agree with an AI algorithm's decision-making process, they are less likely to be adopted – increasing the understanding of clinicians through an interpretable AI model would lead to increased adoption and subsequent improvements in treatment of patients.

Currently models are often evaluated using the ROC AUC, which is a useful metric for comparing the performance of prediction models. However, models which perform better using this metric may not be the most clinically useful, as highlighted by Halligan et al, 'different tests can have identical AUC but different performance at clinically important thresholds. In addition, the positive predictive value may be more important to clinicians than sensitivity^[35]. We additionally utilised a specificity at sensitivity metric to improve the understanding of how our algorithm could change patient care, something not clear from ROC AUC alone. We selected a minimum specificity of 50% and can then directly compare the model's sensitivity. A minimum specificity of 50% means that a positive prediction is probably right, therefore any patients identified requires urgent attention and possibly treatment. This is different to implemented algorithm which can cause "alert fatigue"^[36] through excessive false positive alerts.

Comparison with previous literature

Many machine learning algorithms predicting sepsis/septic shock have been published.^[30] Most of these use lab test results and are not comparable to our algorithm. However, other researchers have also

identified the potential benefit of using machine learning to predict sepsis from only routine ED data. Liu et al's^[14] model achieved 0.83 AUC, while Kijpaisalratana et al's^[37] model achieved 0.93 AUC.

Liu et al's model achieved strong results in terms of prediction metrics, with their inclusion of the SHAP explainer the model is locally explainable. However, this does not explain the model's decision-making process or provide a way of altering the model if a prejudice is identified. These issues mean the model may be less likely to be accepted in an emergency room^[34]. Liu et al also used LIME to construct interpretable surrogate models, however these did not capture complex feature interactions and could not replicate the parent models performance. Yang et al^[38] have also used SHAP values to improve interpretability in a black box algorithm predicting mortality in ICU patients and demonstrated good performance. Although they plot the SHAP values, and consider feature importance, they don't develop this to improve explainability, as we have done. Our approach of construction a surrogate model informed by Shapley values combines both the SHAP and LIME interpretability methods.

The St. John Sepsis Surveillance Agent, and Quick Sequential [Sepsis-Related] Organ Failure Assessment are two implemented algorithms used to flag sepsis. Both algorithms are moderately accurate at predicting patient mortality or ICU stay^[8]. Both algorithms make limited use of comorbidities or demographic information. We have found that a patients age, history of cancer and liver disease are key predictors to whether a patient will be diagnosed with Sepsis. Sepsis predicting algorithms could benefit from the use of additional patient information to more accurately identify patients with severe and non-severe sepsis.

The Shapley approximation offers two key advantages over blackbox models: it provides feature level results across all patients, making potential biases immediately visible, and this transparency supports clinician confidence in the model^[38]. Unlike other interpretable approaches, it also preserves some interaction information from the parent blackbox model. Because it predicts a probability rather than a binary label, it is better suited to inconsistent clinical events. Conceptually, the Shapley model is closest to the NEWS2 algorithm, where each feature contributes an independent score and the overall prediction is the sum of these contributions. The Shapley based approach also benefits from producing a direct probability estimate, which is more intuitive for patients and clinicians.

Strengths and weaknesses of our study

Our study benefits from the inclusion of multiple hospitals within two NHS Trusts with different patient populations, and different approaches to diagnosing sepsis.^[4] Many published novel algorithms are limited to a single site^[30]; this limitation makes the models less generalisable. Published model accuracy could be very different when the model is adopted in other settings. Our study had a notably smaller and messier dataset than is ideal for the creation of a deep neural network, that would have had more power to predict possible feature interactions. However, the shallow neural network we constructed performed well, and the results are more generalisable to messy real-world data.

A second limitation was the higher frequency of missing vital signs data in non-sepsis patients compared with those who had sepsis. We theorized that the absence of recorded vital signs reflected nurses'

perception that the patients would have measured within normal limits. Therefore, we imputed the missing data using values consistent with a healthy profile.

A final limitation is that our dataset is limited to the SoS patients. We know SoS patients make up a large subset of the ED patients, however this means the training and testing of our models is not immediately, directly applicable to an NHS trusts emergency ward. We theorise that non-SoS patients would have vitals and characteristics noticeably distinct from septic patients, however this is uncertain. Before an algorithm is implemented into an emergency ward it should be re-trained and validated on all the patients.

Another way to increase the ShAprox model's power would be to train a more complex surrogate model. A more complex model could learn feature interactions and better approximate the ShApley values to each individual patient. While this would reduce the model's interpretability, it should certainly be considered and used when feature interactions have a large impact on the parent models predictions

The ShAprox methodology can be applied to any machine learning algorithm. Future iterations should be used to approximate trusted and validated black box models, trained on large and diverse datasets. A more powerful parent model would create a more powerful ShAprox model. A larger dataset would also allow for a more confident analysis of the impact of less prevalent ethnicities and comorbidities on the model's prediction.

Our study was informed by discussions with patients, nurses, and clinicians^[39]. Healthcare workers found that digital sepsis alerts could be useful for both junior and senior staff. They found the iterative adjustment of the algorithms based on their feedback during implementation useful, something easy to do with the ShAprox methodology that a black box method could not easily replicate. They also mentioned the alerts tended to be unspecific and oversensitive, hence our research on improving specificity through including additional patient features. Finally, it was strongly agreed that the alerts should support clinical decision making, rather than being a substitution. Although AI could be deployed in many ways across the NHS, our work focused on developing a model that aligns with how AI is currently used in practice. We therefore prioritised simplicity and transparency, enabling clinicians to understand the basis of each prediction and to feel confident disagreeing with the model when their clinical judgement indicates an alternative diagnosis. Our findings also add to the evidence that incorporating additional patient information can improve sepsis prediction performance. Finally, we restricted the model to routinely collected data to help reduce the workload placed on nursing staff.

Conclusion

The Shapley approximating algorithm matches the predictive power of the neural network, accurately predicting a future sepsis diagnosis for ED patients. The model has high transparency and adjustability to aid clinical implementation, and further algorithm development.

Ethics

The DiALS Study was approved by the UK NHS Health Research Authority (Project ID – 288328, approved 20th February 2020), who confirmed that the study does not need formal review by a research ethics committee. The data used was routinely collected and pseudo-anonymised, therefore formal consent was not required. This study was conducted in accordance with the 1964 Helsinki Declaration and its later amendments. The data was stored and analysed securely in the NIHR Biomedical research centre (BRC) Informatics Digital Environment (BRIDgE) secure data environment, and anonymised in line with UK GDPR. Patients were able to opt-out of having their data shared for research purposes through the NHS National Data Opt-Out.

Data Availability

We cannot make the data publicly available, however the data can be accessed by eligible researchers with the appropriate permissions and approval from NHS trusts discussed through BRIDgE. Requests for access to the algorithms should be addressed to the corresponding author.

Declarations

Department of Health and Social Care Disclaimer

This study has been delivered through the National Institute for Health and Care Research (NIHR) Royal Marsden (BRC) and Imperial College (BRC). The views expressed are those of the authors' and not necessarily those of the NIHR or Department of Health and Social Care.

Funding

This manuscript presents independent research funded by the National Institute for Health and Care Research (NIHR) Health and Social Care Delivery Research programme as award number NIHR129082.

Author Contribution

A.T. and K.H. wrote the manuscript text. A.T. developed the algorithm, wrote the code, and produced figures. All authors reviewed the manuscript.

Acknowledgement

Acknowledgments to the DiALS co-investigators.

Data Availability

We cannot make the data publicly available, however the data can be accessed by eligible researchers with the appropriate permissions and approval from NHS trusts discussed through BRIDGe. Requests for access to the algorithms should be addressed to the corresponding author.

References

1. Kumar, A. M. D. et al. Aseem PhD; Cheang, Mary MSc. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*. *Critical Care Medicine* 34(6):p 1589–1596, June | (2006). 10.1097/01.CCM.0000217961.75225.E9
2. NHS England. <https://www.england.nhs.uk/ourwork/clinical-policy/sepsis/sepsis-frequently-asked-questions/#q1-what-happens-when-someone-gets-sepsis-and-how-do-you-treat-it>
3. The Uk Sepsis Trust. <https://sepsistrust.org/> sourced from 2018 Lancet Journal of Respiratory Medicine by Prof Sir Brian Jarman, corroborated by IHME study.
4. Honeyford, K. et al. Digital innovation in healthcare: quantifying the impact of digital sepsis screening tools on patient outcomes—a multi-site natural experiment. *BMJ Health Care Inf.* **32**, e101141. <https://doi.org/10.1136/bmjhci-2024-101141> (2025).
5. NHS England Guidance. <https://www.england.nhs.uk/wp-content/uploads/2017/09/sepsis-guidance-implementation-advice-for-adults.pdf> page 8
6. Amland, R. C. & Hahn-Cover, K. E. Clinical Decision Support for Early Recognition of Sepsis. *Am. J. Med. Qual.* **31** (2), 103–110. 10.1177/1062860614557636 (2014).
7. Juri Yanase, E. & Triantaphyllou A systematic survey of computer-aided diagnosis in medicine: Past and present developments, *Expert Systems with Applications*, 138,2019,112821,ISSN 0957–4174, <https://doi.org/10.1016/j.eswa.2019.112821>
8. Amland, R. C. & Sutariya, B. B. Quick Sequential [Sepsis-Related] Organ Failure Assessment (qSOFA) and St. John Sepsis Surveillance Agent to Detect Patients at Risk of Sepsis: An Observational Cohort Study. *Am. J. Med. Qual.* **33** (1), 50–57 (2018 Jan/Feb). Epub 2017 Feb 1. PMID: 28693336; PMCID: PMC5774614.
9. Redfern, O. C. et al. A Comparison of the Quick Sequential (Sepsis-Related) Organ Failure Assessment Score and the National Early Warning Score in Non-ICU Patients With/Without Infection. *Crit Care Med.* ;46(12):1923–1933. (2018). 10.1097/CCM.0000000000003359. PMID: 30130262.
10. Iwashyna, T. J., Ely, E. W., Smith, D. M. & Langa, K. M. Long-term Cognitive Impairment and Functional Disability Among Survivors of Severe Sepsis. *JAMA* **304** (16), 1787–1794. 10.1001/jama.2010.1553 (2010).
11. Vincent, X. et al. The Timing of Early Antibiotics and Hospital Mortality in Sepsis, *American Journal of Respiratory and Critical Care Medicine*, Volume 196, Issue 7, October Pages 856–863, (2017). <https://doi.org/10.1164/rccm.201609-1848OC>
12. Stephen, L. et al. Wray. Reductions in Sepsis Mortality and Costs After Design and Implementation of a Nurse-Based Early Recognition and Response Program. *The Joint Commission Journal on Quality and*

- Patient Safety. Volume 41, Issue 11, Pages 483-AP3, ISSN 1553–7250, (2015).
[https://doi.org/10.1016/S1553-7250\(15\)41063-3](https://doi.org/10.1016/S1553-7250(15)41063-3)
13. Matt & Inada-Kim NEWS2 and improving outcomes from sepsis, *Clinical Medicine*, Volume 22, Issue 6, Pages 514–517, ISSN 1470–2118, (2022). <https://doi.org/10.7861/clinmed.2022-0450>
 14. Liu, Z. et al. Interpretable machine learning for predicting sepsis risk in emergency triage patients. *Sci. Rep.* **15**, 887. <https://doi.org/10.1038/s41598-025-85121-z> (2025).
 15. SHAP Documentation. <https://shap.readthedocs.io/en/latest/>
 16. Scott, M. & Lundberg, S. I. Lee. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Pages 4765–4774
https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
 17. LIME Documentation. https://contextual-ai.readthedocs.io/en/latest/tutorials/explainer/tutorial_lime_tabular_explainer.html
 18. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G. & King, D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* **17** (1), 195. <https://doi.org/10.1186/s12916-019-1426-2> (2019).
 19. Shapley, L. A Value for n-Person Games. In: (eds Kuhn, H. & Tucker, A.) *Contributions to the Theory of Games II*, Princeton University Press, Princeton, 307–317. <https://doi.org/10.1515/9781400881970-018> (1953).
 20. Inada-Kim, M. et al. Defining and measuring suspicion of sepsis: an analysis of routine data. *BMJ Open.* **7**, e014885. [10.1136/bmjopen-2016-014885](https://doi.org/10.1136/bmjopen-2016-014885) (2017).
 21. Elixhauser, A. P. D., Steiner, Claudia, M. D. & Harris, M. P. H. Robert PhD; Coffey, Rosanna M. PhD. Comorbidity Measures for Use with Administrative Data. *Med. Care.* **36** (1), 8–27. [10.1097/00005650-199801000-00004](https://doi.org/10.1097/00005650-199801000-00004) (January 1998).
 22. Comorbidity Package. <https://cran.r-project.org/web/packages/comorbidity/comorbidity.pdf>
 23. Random Documentation. <https://docs.python.org/3/library/random.html>
 24. <https://>
 25. TensorFlow Package. <https://www.tensorflow.org>
 26. Nitish & Srivastava Geoffrey. Hinton, Alex. Krizhevksy, Ilya. Sutskever, Ruslan. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
 27. Leaky ReLU. <https://www.sciencedirect.com/topics/computer-science/leaky-rectified-linear-unit>
 28. <https://medium.com/@piyushkashyap045/early-stopping-in-deep-learning-a-simple-guide-to-prevent-overfitting-1073f56b493e>
 29. [https://measures%20the%20proportion%20of,%2F%20\(tn%20%2B%20fp\).](https://measures%20the%20proportion%20of,%2F%20(tn%20%2B%20fp).)
 30. Zhang, H., Wang, C. & Yang, N. Diagnostic performance of machine-learning algorithms for sepsis prediction: An updated meta-analysis. *Technol. health care: official J. Eur. Soc. Eng. Med.* **32** (6), 4291–4307. <https://doi.org/10.3233/THC-240087> (2024).

31. Litell, J. M., Guirgis, F., Driver, B., Jones, A. E. & Puskarich, M. A. Most emergency department patients meeting sepsis criteria are not diagnosed with sepsis at discharge. *Acad. Emerg. Med.* **28**, 745–752. <https://doi.org/10.1111/acem.14265> (2021).
32. Akase, Iorhen Ephraim^{1,2}, Akujobi, H. C. & Oghide³; Oziegbe³; Olusegun-Joseph, Akinsanya Daniel^{2,4}; Agbarakwe, Chukwuemeka Abraham³; Opawoye, Adefolarin². Utility and Impact of Early Diagnostic Assessment of Sepsis on Mortality in an Emergency Department of a Tertiary Hospital in Lagos, Nigeria. *Nigerian Postgraduate Medical Journal* 30(1):p 70–74, Jan–Mar | (2023). 10.4103/npmj.npmj_190_22
33. Quinn, T. P., Jacobs, S., Senadeera, M., Le, V. & Coghlan, S. The Three Ghosts of Medical AI: Can the Black-Box Present Deliver? *Artificial Intelligence in Medicine*, 124, Article **102158**. (2022). <https://doi.org/10.1016/j.artmed.2021.102158>
34. Gaube, S. et al. Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays. *Sci. Rep.* **13**, 1383. <https://doi.org/10.1038/s41598-023-28633-w> (2023).
35. McGreevey, J. D. 3, Mallozzi, C. P., Perkins, R. M., Shelov, E., Schreiber, R. & rd,, & Reducing Alert Burden in Electronic Health Records: State of the Art Recommendations from Four Health Systems. *Appl. Clin. Inf.* **11** (1), 1–12. <https://doi.org/10.1055/s-0039-3402715> (2020).
36. WongA,OtlesE,DonnellyJPetal.External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients.*JAMA Intern. Med* 2021;181(8):1065–1070.doi:10.1001/jamainternmed.2021.2626.
37. Kijpaisalratana, N., Sanglertsinlapachai, D., Techaratsami, S., Musikatavorn, K. & Saoraya, J. Machine learning algorithms for early sepsis detection in the emergency department: A retrospective study. *Int. J. Med. Inf.* **160**. **104689**, 1386–5056. <https://doi.org/10.1016/j.ijmedinf.2022.104689> (2022).
38. Yang, M. et al. Development and Validation of an Interpretable Conformal Predictor to Predict Sepsis Mortality Risk: Retrospective Cohort Study*J. Med. Internet Res.* **26**, e50369. 10.2196/50369 (2024).
39. Lazzarino, R. et al. Views and Uses of Sepsis Digital Alerts in National Health Service Trusts in England: Qualitative Study With Health Care Professionals, *JMIR Hum Factors.* ;11:e56949, URL: <https://humanfactors.jmir.org/2024/1/e56949>, (2024). 10.2196/56949

Figures

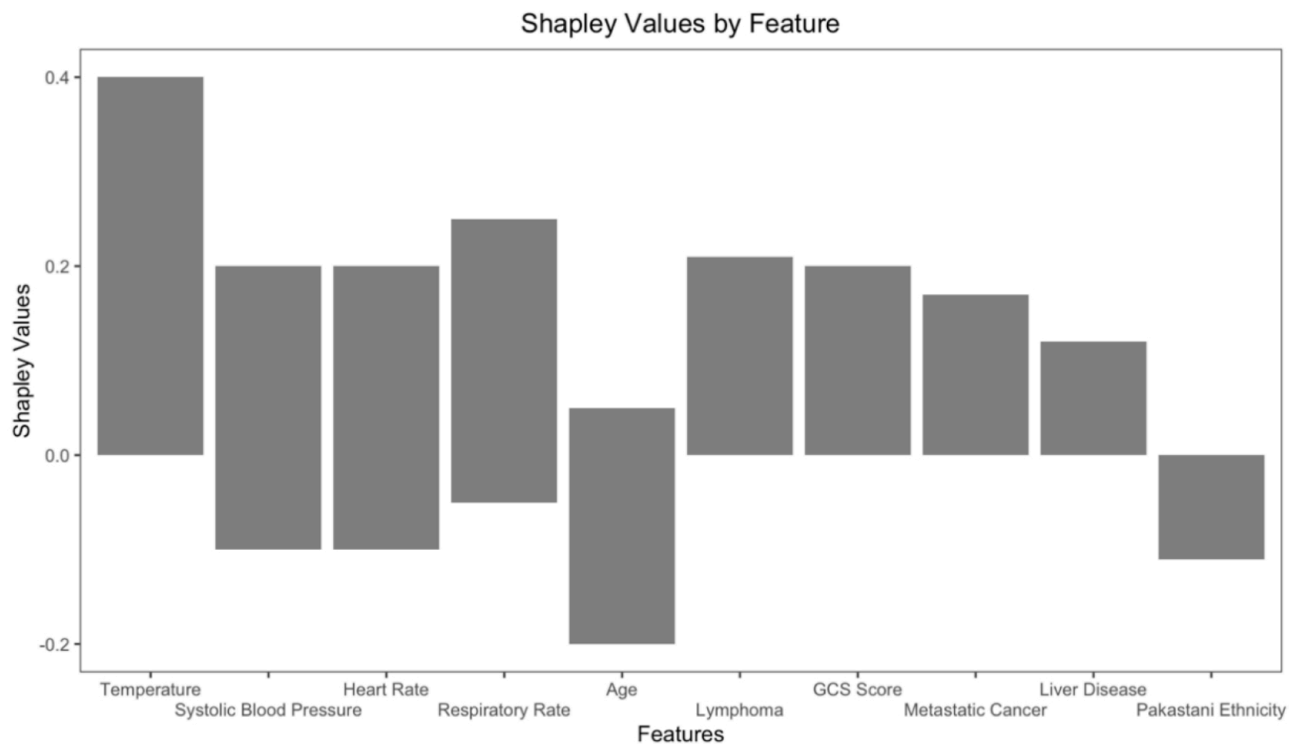


Figure 1

Shapley Values by Patient Feature

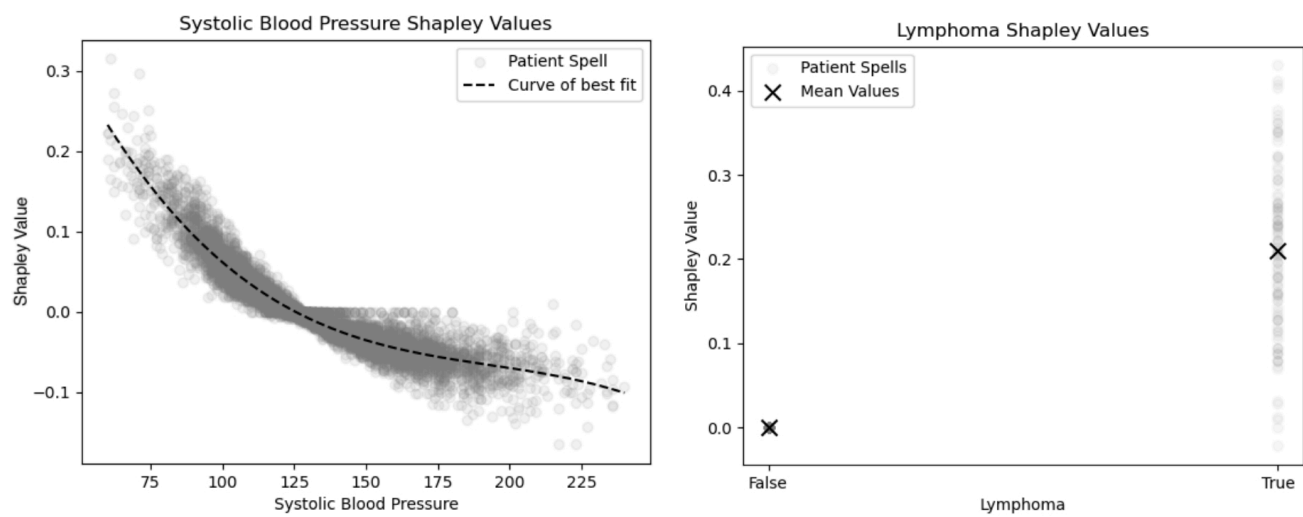


Figure 2

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ShAproxsupplementarymaterial.docx](#)