

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/166778/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Burton, Ross J., Raffray, Loïc, Moet, Linda M., Cuff, Simone M. , White, Daniel A., Baker, Sarah E. , Moser, Bernhard , O'Donnell, Valerie B. , Ghazal, Peter , Morgan, Matt P., Artemiou, Andreas and Eberl, Matthias 2024. Conventional and unconventional T cell responses contribute to the prediction of clinical outcome and causative bacterial pathogen in sepsis patients. *Clinical & Experimental Immunology* , uxae019. 10.1093/cei/uxae019

Publishers page: <http://dx.doi.org/10.1093/cei/uxae019>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



**Conventional and unconventional T cell responses
contribute to the prediction of clinical outcome
and causative bacterial pathogen in sepsis patients**

**Ross J. Burton^{1,2}, Loïc Raffray^{1,3}, Linda M. Moet¹, Simone M. Cuff¹,
Daniel A. White¹, Sarah E. Baker¹, Bernhard Moser^{1,4}, Valerie B. O'Donnell^{1,4},
Peter Ghazal^{1,4}, Matt P. Morgan², Andreas Artemiou⁵, Matthias Eberl^{1,4,*}**

¹Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff CF14 4XN, United Kingdom; ²Adult Critical Care, University Hospital of Wales, Cardiff and Vale University Health Board, Cardiff CF14 4XW, United Kingdom; ³Department of Internal Medicine, Félix Guyon University Hospital of La Réunion, Saint Denis, Réunion Island, France; ⁴Systems Immunity Research Institute, Cardiff University, Cardiff CF14 4XN, United Kingdom; ⁵School of Mathematics, Cardiff University, Cardiff CF24 4AG, United Kingdom

*** Correspondence:** Matthias Eberl, Division of Infection and Immunity, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK. Phone: +44-29206-87011; Email: eberlm@cf.ac.uk

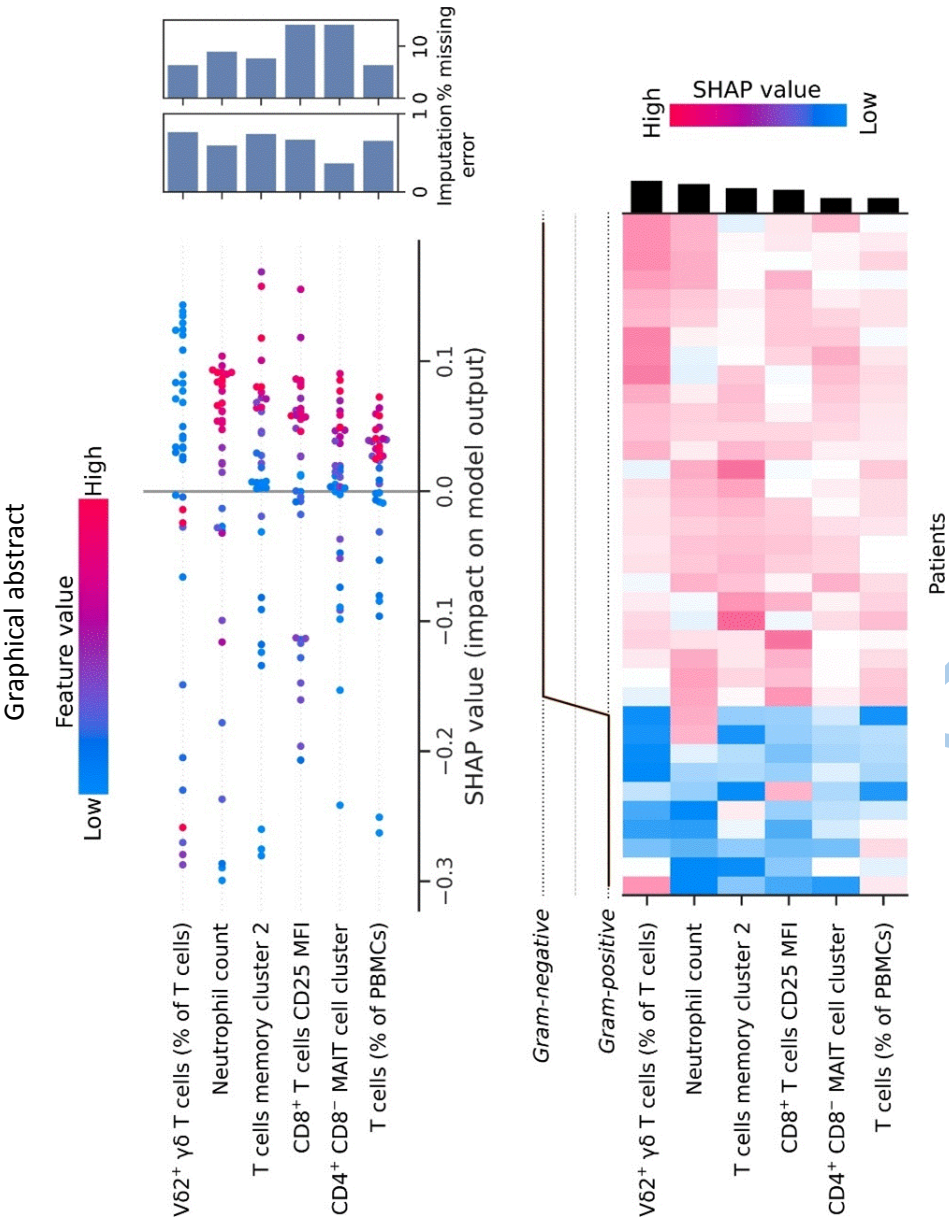
© The Author(s) 2024. Published by Oxford University Press on behalf of the British Society for Immunology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Sepsis is characterised by a dysfunctional host response to infection culminating in life-threatening organ failure that requires complex patient management and rapid intervention. Timely diagnosis of the underlying cause of sepsis is crucial, and identifying those at risk of complications and death is imperative for triaging treatment and resource allocation. Here, we explored the potential of explainable machine learning models to predict mortality and causative pathogen in sepsis patients. By using a modelling pipeline employing multiple feature selection algorithms, we demonstrate the feasibility to identify integrative patterns from clinical parameters, plasma biomarkers and extensive phenotyping of blood immune cells. Whilst no single variable had sufficient predictive power, models that combined five and more features showed a macro area under the curve (AUC) of 0.85 to predict 90 day mortality after sepsis diagnosis, and a macro AUC of 0.86 to discriminate between Gram-positive and Gram-negative bacterial infections. Parameters associated with the cellular immune response contributed the most to models predictive of 90 day mortality, most notably, the proportion of T cells among PBMCs, together with expression of CXCR3 by CD4⁺ T cells and CD25 by mucosal-associated invariant T (MAIT) cells. Frequencies of Vδ2⁺γδ T cells had the most profound impact on the prediction of Gram-negative infections, alongside other T cell-related variables and total neutrophil count. Overall, our findings highlight the added value of measuring the proportion and activation patterns of conventional and unconventional T cells in the blood of sepsis patients in combination with other immunological, biochemical and clinical parameters.

Keywords: Endotoxin Shock, Sepsis, Unconventional T Cells, Cytokines, Inflammation



Abbreviations:

- area under the curve (AUC)
- mucosal-associated invariant T (MAIT)
- Acute Physiology and Chronic Health Evaluation (APACHE)
- Sequential Organ Failure Assessment (SOFA)
- C-reactive protein (CRP)
- procalcitonin (PCT)
- out-of-bag (OOB)
- normalised mean root squared error (NMRSE)
- Minimum Redundancy-Maximum Relevance (MRMR)
- Support Vector Machines (SVMs)
- Leave-one-out cross-validation (LOOCV)
- 5-fold cross-validation (5-fold CV)
- SHapely Additive exPlanations (SHAP)

Accepted Manuscript

INTRODUCTION

Sepsis is a life-threatening syndrome characterised by organ failure arising from a dysfunctional host response to infection [1]. Timely diagnosis of sepsis is crucial, and identifying those at risk of complications and death is imperative for triaging treatment and resource allocation. While severity scores such as Acute Physiology and Chronic Health Evaluation (APACHE) and the Sequential Organ Failure Assessment (SOFA) can be used for audits and to direct care, these tools rely on routinely collected clinical data and observations, and their performance for predicting in-hospital mortality is relatively poor [2, 3]. Most importantly, they do not capture the complex maladaptive immune and metabolic mechanisms that are now known to contribute significantly to sepsis pathology [1].

The need to identify the cause and target treatment has driven an interest in identifying diagnostic and prognostic biomarkers [4, 5], and in developing algorithms that leverage electronic health records [6, 7, 8]. Whereas single biomarker studies have shown mixed results, the value of multiple biomarkers in combination is increasingly being recognised [5, 9, 10]. In particular, prognostic biomarkers derived from the pathophysiology of sepsis may help guide treatment and monitoring of the disease, the most widely studied ones being C-reactive protein (CRP) and procalcitonin (PCT). However, elevated plasma levels of CRP on admission are not a reliable predictor of mortality [11], and although early levels of plasma PCT differ between survivors and non-survivors, high heterogeneity between study populations puts the general applicability of these findings into question [12, 13]. Other promising biomarkers include pro-adrenomedullin [14], IL-6 [15], lymphopenia [16], neutrophil to lymphocyte ratio [17], amphiregulin [18], and CD64 expression on neutrophils [19], among others.

Most biomarker research in sepsis focuses on diagnosis rather than prognosis, given the importance of early interventions, such as anti-microbial treatments, on survival [20]. In this respect, Gram-negative organisms are often associated with poorer outcomes in first-hit infections [21]. However, the reality of sepsis is that the causative pathogens and the best way to target them are unknown at the time of presentation, and broad-spectrum antibiotics are often administered empirically, with arguments for [22] and against [23] their rapid use. In a significant proportion of patients with sepsis, the causative pathogen is never identified, and there are contradictory findings as to how severity of illness, length of stay and in-hospital mortality compare between culture-negative and culture-positive sepsis [24]. In fact, it is still uncertain whether culture-negative sepsis represents a separate clinical entity, with the possibility of the absence of infection entirely. Earlier identification of the causative pathogen would undoubtedly lead to better targeted therapy and improve antibiotic stewardship [25].

In this respect, multiple molecular methods for pathogen identification have come to market, potentially reducing the time needed to identify the causative pathogen by up to 30 hours [26]. However, many technologies still require a positive blood culture, and sensitivities for direct detection of bacteria in blood by PCR are only modest [27]. Encouragingly, plasma levels of PCT appear to be higher amongst patients with Gram-negative infections compared to Gram-positive infections [28]. Other biomarkers that have shown promise are soluble CD14 ('presepsin'), which appears to be increased in Gram-negative bacteraemia [29], and the cytokines IL-1 β , IL-6, and IL-18, with elevated concentrations in patients with Gram-positive infection [30].

Of particular interest for identifying the aetiology in sepsis are unconventional T cells such as mucosal associated invariant T (MAIT) cells and $\gamma\delta$ T cells, which are capable of microbial pattern recognition. MAIT cells are characterised by a semi-invariant T cell antigen receptor with

specificity for microbial riboflavin (vitamin B2) derivatives found in fungi and most bacteria except *Streptococcus* spp., *Enterococcus* spp. and *Listeria monocytogenes* [31]. Human V γ 9/V δ 2⁺ $\gamma\delta$ T cells respond to the microbial isoprenoid precursor (*E*)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate (HMB-PP), a molecule produced by most Gram-negative and some Gram-positive pathogens, but notably absent from *Streptococcus* spp., *Enterococcus* spp., *Staphylococcus* spp. and fungi [31]. The innate functionality and specificity of these unconventional T cells are likely to contribute to pathogen-specific ‘immune fingerprints’, with proof of concept already shown in patients with acute peritonitis [32].

Studies have only begun to incorporate a combination of clinically available data with novel biomarkers to create predictive models for sepsis [33, 34, 35]. With the advent of multi-omics technology, there is a growing abundance of data, with the promise that a multi-layered approach to phenotyping the immunological response to sepsis may help identify diagnostic and prognostic signatures with direct application to the clinic [36, 37]. This diverse feature space presents the challenge of analysing extensive high-dimensional data from which informative biomarker combinations are to be found, in a task that is analogous to feature selection in machine learning [38]. The minimal yet optimal variables are identified to help reduce model complexity, avert overfitting and improve performance. Numerous feature selection methodologies already exist, each with benefits and disadvantages [39]. Since no single machine learning algorithm will be optimal for every task [40, 41], it is advised to search across multiple solutions and make conclusions based on the performance of observed data. Experimenting with multiple methodologies will reduce the risk of overlooking an informative signature or focusing on a single suboptimal solution.

Data-driven pattern recognition with feature selection has successfully identified predictive signatures in the pathogenic cause of peritonitis [32], prognosis and treatment response in traumatic injury [42], prognosis in COVID-19 [43], and vaccine response [44]. The application of model agnostic methods for the measure of feature importance was recently demonstrated for predicting multiple organ dysfunction in paediatric sepsis [45] and the identification of risk factors in COVID-19 [46]. We here set out to combine routine clinical data with immunological profiling and develop machine learning models describing composite biomarker patterns predictive of mortality or underlying cause in sepsis patients, aiming to create valuable tools to prioritise and direct care in a resource-limited environment. Our findings underscore the power of comprehensive models integrating clinical parameters with an analysis of acute phase proteins, cytokines and lipids, and an extensive immune phenotyping of the cellular compartment. While perhaps unsurprisingly no variable on its own was informative enough to help guide clinical decisions, models combining five or more features predicted 90 day mortality after sepsis diagnosis, and discriminated between Gram-positive and Gram-negative infections. A particularly compelling observation from this study was the importance attributed to MAIT cells and V δ 2⁺ $\gamma\delta$ T cells, providing *in vivo* evidence for an involvement of unconventional T cell subsets in the early immune response in many microbial infections.

MATERIALS & METHODS

Subjects. A total of 77 patients were enrolled between 2018 and 2021 who were over 18 years old with a diagnosis of sepsis according to the Third International Consensus Definitions for Sepsis and Septic Shock ('Sepsis-3') [47]. They were cared for in the intensive care unit at the University Hospital of Wales in Cardiff, and were recruited within 36 hours of the presumed onset of

infective illness when they already had or would require arterial cannulation as part of standard treatment [48]. Patients were excluded if they were pregnant or breastfeeding, or were females of childbearing age in whom a pregnancy test had not been performed; had severe immune deficiency (*e.g.* diagnosis of AIDS, or treatment with anti-rejection transplant drugs or high dose corticosteroids); had haematologic malignancy or ongoing chemotherapy; had pre-existing severe liver failure; were adjudged by the admitting clinician to be unlikely to survive for the duration of the study period; were admitted post-cardiac arrest; or had an underlying impairment of higher function that would make it impossible for informed consent to be given upon recovery (*e.g.* severe learning disability).

Blood samples and clinical records. Whole blood was obtained within the first 36 hours of sepsis from patients with a SOFA score >2 and a suspected infection; cell-free plasma was obtained and frozen within the first hour of sample collection. Mortality rates of 22.1% and 27.3% at 30 and 90 days after sepsis diagnosis, respectively (Supplementary Tables S1 and S2), were in line with a recent multi-centre prevalence study of sepsis in Wales [49]. 52 patients (67.5% of the cohort) had a microbiologically confirmed infection. Three patients had a mixed culture result with undefined causative pathogen, two had influenza A with no bacterial isolates, and one had candidiasis; these patients were excluded from pathogen-specific analyses. No significant differences in patient demographics, severity scores, therapeutic interventions or mortality were observed between patients with Gram-negative, Gram-positive and culture-negative sepsis (Supplementary Table S3). A further differential analysis according to bacterial species and location of the infection was not attempted given the very small group sizes (Supplementary Table S4). Routine clinical data such as full blood count, liver profile and blood gas data were

recorded; variables captured for fewer than five patients were removed from the subsequent analysis, leaving 63 routinely collected variables (Supplementary Table S5).

Flow cytometry. Neutrophils and monocytes were stained in freshly collected whole blood after red blood cell lysis; T cells were stained after Ficoll-Paque PLUS (Fisher Scientific) separation of peripheral blood mononuclear cells. Cells were acquired on a 16-colour BD LSR Fortessa flow cytometer (BD Biosciences), using the fluorescently labelled monoclonal antibodies listed in Supplementary Table S6. Live single cells were gated based on side and forward scatter area/height and exclusion of live/dead staining (fixable Aqua; Invitrogen). Exclusion of monocytes and B cells in T cell panels was ensured by using anti-CD14 and anti-CD19 in the live/dead staining channel. When identifying monocytes and neutrophils in whole blood, T cells were excluded by their appearance in side and forward scatter area/height, and B cells were excluded by using anti-CD19 in the live/dead staining channel. The flow cytometer was calibrated using BD FACSDiva CS&T research beads (BD Biosciences) prior to acquisition. Compensation for spectral overlap was accounted for using BD CompBeads (BD Biosciences) and a spillover matrix generated in FACSDiva. Compensation was checked for errors using FlowJo (TreeStar) prior to analysis.

The Harmony algorithm was applied to all cytometry data using CytoPy version 2.0 to align samples whilst reducing the risk of losing biological information [50], and geometric median clustering with weighted voting (GeoWaVe) was performed on batch-corrected data for T cells, monocytes and neutrophils as described before [51]. Ensembles were informed using multiple clustering algorithms popular for analysing cytometry data, providing diverse input for ensembles

and preventing biased analysis driven by a single method, including FlowSOM, Phenograph, and SPADE algorithms as well as K-Means and FlowSOM clustering of PHATE embeddings.

Soluble proteins in patient plasma. Cytokines and chemokines were quantified in two batches using Luminex standard sensitivity magplex assays based on the xMAP (multi-analyte profiling) technology, according to the manufacturer's instructions. Data were acquired on a Luminex 200 compact analyser. TNF- α (eBioscience), IFN- γ (eBioscience) and IL-6 (R&D Systems) were measured in a single batch by ELISA. Concentrations were obtained by fitting a standard curve using a five-parameter logistic fit using the generalised Hill equation, and data quality was assessed by observing the coefficient-of-variation and standard recovery. Where fewer than three observations fell within the standard range, the majority of observations had a CV greater than 50%, or the standard recovery was outside a range of 75-125%, analytes were deemed of poor quality and excluded from subsequent analysis. Batch effect was addressed with post-hoc correction and data alignment; data were \log_2 transformed and values replaced with a z score. 95% confidence intervals for odds ratios were approximated as previously described by Tenny and Hoffman [52].

Free fatty acid and acyl carnitine analysis. Concentrations of free fatty acids and acyl carnitines specified in Supplementary Table S7 were determined by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Lipids were extracted from thawed cell-free plasma in randomised batches by mixing 10 μ l plasma with 240 μ l methanol and 10 μ l of a mixture of deuterated internal lipid standards (Supplementary Table S8) for a description of internal lipid standards and concentrations). Samples were sonicated in iced water for 1 min, vortexed at 1,400 rpm for 10 min at 4 °C, and then centrifuged at $18,000 \times g$ for 10 min. 100 μ l supernatant was derivatised

using 3-nitrophenylhydrazine [53], by adding 50 μ l 200 mM 3-nitrophenyl-hydrazine in (50/50 methanol/H₂O) and 50 μ l 120 mM N-ethyl-N'-(3-dimethylaminopropyl) carbodiimide hydrochloride and 6% pyridine (50/50 methanol/H₂O). Samples were then vortexed and incubated for 30 min at 40°C. Excess derivatisation reagents were quenched by the addition of 0.5 % formic acid (100 μ l; 75/25 methanol/H₂O) and incubation at 40°C for 30 min. Samples were aliquoted into HPLC vials for LC-MS/MS. For every sample batch, five blanks of HPLC water were extracted and derivatised to account for background levels of lipids.

LC-MS/MS analyses were performed using scheduled Multiple Reaction Monitoring (MRM) mode on a Nexera liquid chromatography system (Shimadzu) coupled to a QTRAP 4000 mass spectrometer with an ESI source (AB Sciex). Lipid separation was achieved using a Kinetex Polar C18 reverse phase column (100 Å, 100 \times 2.1 mm, 2.6 μ m particle size; Phenomenex) and gradient elution of two mobile phases (A: 100 % water + 0.1 % formic acid, B: 100 % methanol + 0.1 % formic acid). The flow rate was 0.2 ml/min and injection volume was 2 μ l; column temperature was 50°C. MS settings (declustering potential and collision energy) were optimised for individual lipids using standards. Fatty acids and acyl carnitines were detected in negative ion and positive ion mode, respectively. All MRMs used are provided in Supplementary Table S8. Calibration curves and a QC sample of pooled sample aliquots were quantified alongside sample analyses. QC samples were run after every ten measurements to monitor MS performance. Peak areas were integrated using MultiQuant software (AB Sciex) and concentrations calculated based on internal standard response. For lipids with an exact corresponding internal standard, concentrations were calculated directly based on the lipid to internal standard ratio and internal standard concentration, while ensuring that all peaks were within the linear dynamic range of the instrument. For other lipids, concentrations were derived from comparisons with an external

calibration curve of the ratio of the specific lipid standard to a deuterated internal standard. Limit of quantitation used signal:noise of >5:1 and at least 5 data points per peak. All further analysis was conducted using R and GraphPad Prism version 8.4.3, and MetaboAnalyst version 4.0 and higher (metaboanalyst.ca).

Data processing. A MongoDB database was populated with clinical parameters, soluble biomarker measurements and summary statistics of flow cytometry data. Data were combined into a table of 267 features (Supplementary Table S7), which could be broadly categorised into physiology, interventions, point of care testing, clinical laboratory results, protein biomarkers, lipids, proportions of immune cell populations and mean fluorescence intensities of surface markers on immune cells. Many clinical variables demonstrated class imbalance, with the majority class representing more than 70% of patients for each target. Features that were obviously redundant were removed, *e.g.* where information was duplicated or values were equal for all patients. A combination of experimental errors, issues of sample integrity and the sporadic nature of clinical data collection all contributed to missing values (Supplementary Fig. S1). Features that could not be assumed to be missing at random, such as clinical laboratory measurements that were obtained for only a few individuals, were excluded from further analysis. The remaining missing values were imputed using the MissRanger algorithm [54]. Imputation resulted in acceptable out-of-bag (OOB) for both continuous and categorical features, with a median normalised mean root squared error (NMRSE) of 0.68 and <20% categorical features misclassified. Features with an NMRSE >1.0 or >40% missing data were removed from further analysis. Multicollinear features were identified and either replaced with an estimated latent variable, or the variables with the greatest mutual information with the target variables were retained (Supplementary Fig. S2).

Feature selection. After removal of redundant and highly correlated features, 150 features remained. Of these, a set had to be chosen small enough to reduce the risk of overfitting, improve classification accuracy, and ensure models could easily be interpreted. Feature selection algorithms comprised filter and wrapper techniques popular in the biostatistics literature: univariate selection with permutation testing [55], ReliefF [56], Minimum Redundancy-Maximum Relevance (MRMR) [57], Boruta [58], and recursive feature elimination with Support Vector Machines (SVMs) [59]. After running each algorithm for all target variables, the top ten features from each algorithm were chosen to provide five independent feature sets for each target. Downstream analysis included validation of models on equivalent complete case data, and constraining the number of features increases the amount of data available during complete case analysis. Limiting the feature space to a maximum of ten parameters ensured at least five observations were available per feature and made interpretation of model decisions easier. Classifiers were generated for the top three to ten features and compared by classification performance. Overlap between feature selection methods was measured using the pairwise Jaccard index and visualised as heatmap for each target (Supplementary Fig. S3).

Modelling pipeline for selecting, comparing and inspecting classification algorithms. A 20% subset of the total cohort was randomly selected prior to model and feature selection and kept for independent evaluation of model performance to avoid inflated accuracy due to overfitting (Supplementary Fig. S4A). Eight groups of classifiers (referred to as ‘classifier families’ from here onward) were drawn upon for the task of binary classification: Logistic regression, SVMs with linear or non-linear kernel, Naive Bayes, K-Nearest Neighbours (KNN), Random Forest, Extra Random Forest, and Extreme Gradient Boosting (XGBoost) (Supplementary Fig. S4B). All classifiers were implemented using the Scikit-Learn library [60]. A grid search strategy was

employed to tune optimal hyperparameters. Hyperparameters included L1 and L2 regularisation of varying strengths, polynomial and radial basis function kernels for non-linear SVMs (with multiple degrees for the former and a range of γ for the latter), different distance metrics and number of nearest neighbours for KNN, and multiple hyperparameters for ensembles of tree-based learners controlling parameters such as the depth of trees, number of splits, number of features, and sampling methods. This approach resulted in 216 classification models from a diverse range of classifier families, an approach supported by the aptly named “no free lunch theorem” [40], which suggests that no single algorithm can be optimal for all problems.

Each model was trained on five independent feature sets (Supplementary Fig. S4C), chosen to increase diversity when exploring the feature space without significantly impacting computational requirements. As the optimal number of features might be less than the top ten ranked features presented by each feature selection algorithm, each classifier was trained on the top three through to the top ten features, iteratively. For each classifier, eight models were generated for each feature selection algorithm, totalling 40 models across all possible feature sets and 8,880 models trained across all possible classifiers, repeated for each target. Leave-one-out cross-validation (LOOCV) and 5-fold cross-validation (5-fold CV) were used to select the optimal model before validation on independent hold-out data (Supplementary Fig. S4D). Within each classifier family, the model and feature set combined with the highest LOOCV F1 score was chosen for evaluation and model inspection (Supplementary Fig. S4E). Cross-validation and holdout performance for the optimal model from each classifier family were compared using ROC curves, balanced accuracy, macro F1 score and macro AUC score. Models were first compared by 5-fold CV balanced accuracy using the non-parametric Friedman test and Nemenyi post-hoc testing [61]. After selecting a model from each classifier family using cross-validation, their performance was

validated on independent holdout data, and the top performing models were inspected using SHapely Additive exPlanations (SHAP) [62].

RESULTS

Insights from routine clinical data in sepsis patients. To identify possible predictors of outcomes and the underlying cause of the infection in patients presenting with sepsis, we first examined all routinely collected clinical data available within the first 36 hours. While all patients had elevated blood levels of CRP, there were no significant differences between survivors and non-survivors, or between culture-positive and culture-negative patients (Supplementary Fig. S5). However, CRP was higher in Gram-positive infections compared to Gram-negative infections. Arterial lactate [63] was increased in those who died within 90 days but was only different when considering the sample closest to enrolment time and not the average within the 48-hour window (Supplementary Fig. S6). As such, CRP and lactate were only of limited utility as biomarkers to make clinically relevant predictions. The same held true for virtually all other clinical variables (data not shown). In fact, the only routinely collected parameter that showed a significant difference between survivors and non-survivors was the fraction of inspired oxygen (FiO₂) value taken closest to the diagnosis of sepsis, with increased levels amongst non-survivors compared to survivors, corroborating findings by others [64]. No other biomarker demonstrated a significant difference relative to mortality, nor to the nature of the causative pathogen (Supplementary Fig. S7).

Soluble immune mediators in patient plasma. To gain further insight into inflammatory responses in sepsis patients, a large panel of soluble mediators in plasma were quantified (Supplementary Fig. S8). Levels of the chemokine CXCL10 were decreased in non-survivors at 30 days, while IL-15 was increased in non-survivors. However, over 80% of patients had IL-15 levels below the detection limit, and the trends seen for CXCL10 and IL-15 were diminished when observing 90 day mortality. FLT3L levels were moderately increased in culture-positive sepsis patients compared to those without a confirmed infection, and ferritin was decreased in patients with Gram-negative compared to Gram-positive infections. No other plasma analytes were significantly different between any of the groups. An additional analysis was performed using the detection limits as thresholds to create binary variables (Supplementary Fig. S9). Although elevated concentrations of IL-6, IL-15 and OSM were associated with higher 30 day mortality, once accounting for multiple comparisons none of the analytes had a statistically significant odds ratio.

Immune phenotypes in sepsis patients correlate with clinical outcomes. Given the poor predictive performance of routine clinical data and soluble biomarkers we next analysed the cellular compartment by flow cytometry. The first observation was a significant reduction in T cells as percentage of PBMCs in non-survivors compared to survivors at 30 days (Fig. 1A) and 90 days (Fig. 1B), whereas percentages of monocytes and neutrophils amongst white blood cells were not different between the groups. However, and in line with the importance of HLA-DR expression on circulating monocytes as predictor of immunosuppression and poor outcomes in sepsis [65], monocyte HLA-DR expression was lower in non-survivors compared to survivors at both 30 and 90 days (Fig. 2). A trend was also visible for the cell adhesion molecule CD62L, with increased expression on monocytes in non-survivors.

Within the T cell compartment (Supplementary Fig. S10), CD8⁺ T cells tended to be reduced in non-survivors, to a greater extent when predicting mortality within 30 days than 90 days. Notably, T cells, monocytes and neutrophils were comparable between those with and without a microbiologically confirmed infection (Fig. 1C). In contrast, there was a significant increase in the proportion of neutrophils amongst white blood cells and a significant decrease in MAIT and Vδ2⁺ γδ T cells as percentage of T cells in patients with Gram-negative infections, compared to Gram-positive sepsis (Fig. 1D). Of note, expression of the activation marker CD69 (and to a lesser extent, of CD25) on MAIT cells and Vδ2⁺ T cells appeared to differentiate Gram-positive from Gram-negative infections, pointing toward a role of these cells in the immune response to some pathogens but not others (Fig. 3).

A T cell dominated immune signature predicts mortality at 90 days after diagnosis of sepsis.

We next combined all available information to generate predictive models. There was no significant difference between the optimal models chosen for each classifier family when comparing 5-fold CV accuracy and F1 score for the prediction of 30 day mortality (Supplementary Fig. S11). Logistic regression and SVMs showed promise at first, with excellent LOOCV ROC and comparable performance between training and testing data within the model and feature selection process. However, none of the models generalised well when exposed to holdout data (Table 1). Prediction of 90 day mortality was more reliable, with good LOOCV and 5-fold CV performances across all classifiers (Supplementary Fig. S12). While more complex models such as KNN and ensembles of tree-based learners exhibited more over-fitting compared to the simpler logistic regression and linear SVM, the Extra Random Forest model showed superior accuracy, F1 score and AUC scores compared to all other models when tested on holdout data (Table 1), and was therefore chosen for complete case analysis and inspection (Fig. 4). The ROC curve for the

imputed data was comparable to that of the complete case data, and the training LOOCV AUC was almost identical to the complete case AUC. Balanced accuracy and macro F1 score were decreased in complete case analysis compared to the training LOOCV scores but both scores were still greater than 0.7.

The proportion of T cells (as percentage of total PBMCs) was the most noteworthy feature of the Extra Random Forest model. Lower values for T cells influenced a prediction of 90 day mortality, as shown by the gradient for T cells on the beeswarm plot (Fig. 5). Excluding the percentage of T cells, the next most impactful features were elevated blood glucose, higher CXCR3 expression on CD4⁺ T cells, and lower levels of arachidonic acid (a 20-carbon chain polyunsaturated omega-6 fatty acid; C20:4). The percentage of T cells was the dominant factor in the Extra Random Forest model, but where SHAP values were only moderately high, the influence of blood glucose, CD4⁺ T cell CXCR3 expression and arachidonic acid encouraged the prediction of survival. The remaining features in the Extra Random Forest model – magnesium plasma concentration, APACHE II score, and CD25 expression on MAIT cells – appeared important for individual patients, rather than for the wider training cohort.

Neutrophils, CD8⁺ T cells and unconventional T cells form a predictive signature that differentiates Gram-negative and Gram-positive infections. With regard to the top-performing models for predicting Gram-negative cause in sepsis, logistic regression, SVMs and the Extra Random Forest model demonstrated the best LOOCV performance (Supplementary Fig. S13). The logistic regression and linear SVM models performed well on holdout data but the Random Forest model presented the best ROC AUC score overall (Table 1). The Random Forest model selected features of T cells and the neutrophil count, with the proportion of Vδ2⁺ T cells representing the feature with the highest absolute mean SHAP value (Fig. 6). The relationship between the

proportion of $V\delta 2^+$ T cells and their SHAP values was unclear on the beeswarm plot, and was better visualised as a scatterplot (Fig. 7). As the proportion of $V\delta 2^+$ T cells increased, the SHAP value decreased, and thus the impact on the prediction of Gram-positive cause was stronger. Strikingly, there were two Gram-negative cases with high $V\delta 2^+$ $\gamma\delta$ T cell proportions. The model successfully identified the relatively abnormal relationship these outliers had with the proportion of $V\delta 2^+$ T cells, and this was reflected in their low absolute SHAP values. Overall, the model displayed a strong ability to generalise despite such outliers, which highlights the added value of measuring the proportion of $V\delta 2^+$ T cells in the blood of sepsis patients in combination with other immunological features.

The additional features in the Random Forest model included the total neutrophil count, $CD4^+$ T cell memory cluster 2 (a $CD4^+$ T cell cluster characterised by low expression of CD27 and CCR7, moderate expression of CD45RA and high expression of CD57), CD25 expression on $CD8^+$ T cells, a distinct $CD4^+$ $CD8^-$ MAIT cell cluster, and the proportion of T cells (as a percentage of PBMCs). Increased values for all features were associated with higher SHAP values, influencing the model to predict a Gram-negative causative pathogen, with the combination of the chosen features ultimately yielding the correct prediction. In striking contrast to these T cell related features, none of the clinical, soluble protein and lipid parameters determined for each patient were selected in the final Random Forest model, underscoring the overall importance of conventional and unconventional T cell responses for predictive models.

DISCUSSION

We here created supervised machine learning models to predict mortality and underlying cause of infection in patients presenting with acute sepsis. A modelling pipeline was developed that considered the small cohort size, class imbalance and missing data, with a particular focus on interpretability, and employing multiple feature selection algorithms with a diverse choice of hyperparameters. Out of a total of 63 variables derived from routine clinical data, only inspired oxygen (FiO₂) at the time of diagnosis differed between survivors and non-survivors after correcting for multiple comparisons; no other clinically available data were particularly informative, including previously studied biomarkers such as CRP and arterial lactate. This lack of suitable biomarkers amongst routine clinical data highlighted the need for detailed immune phenotyping to yield informative biomarkers of the patient's response to infection.

Investigation of immune cell populations in whole blood confirmed previously well-described observations, such as decreased HLA-DR expression on monocytes [65] and a reduced proportion of circulating T cells [66, 67] amongst non-survivors. Other interesting trends were observed for soluble biomarkers such as cytokines, chemokines and acute phase protein levels in plasma. CXCL10 was decreased in those patients who died within 30 days, increased levels of IL-6 and IL-15 showed a trend towards higher odds of mortality at 30 days, and ferritin levels were higher in Gram-positive infections compared to Gram-negative infections. Ultimately, however, limited data, class imbalance and the detection limits of the assays used made it difficult to reconcile these findings.

A model for predicting 30 day mortality could not be obtained but an Extra Random Forest model for predicting 90 day mortality was identified with a holdout AUC score of 0.85, representing a considerable improvement on the majority of previously reported prognosis biomarkers [5, 14]. The 90 day mortality model showed a diverse selection of input features, including parameters that quantified immune populations, activation profiles of T cells, lipid plasma concentrations and the APACHE II severity score. The diversity of the chosen features highlights the benefits of capturing variables that describe multiple systems and how their combination can contribute to the model performance.

The proportion of T cells (as a percentage of PBMCs) was the main contributing feature to decision-making in the Extra Random Forest model. A comparison of T cells showed a significant difference between survivors and non-survivors at 90 days, in agreement with lymphopenia being a well-documented sign of increased severity and associated with higher mortality [16]. Additional features in the Extra Random Forest model included CXCR3 expression on CD4⁺ T cells, plasma concentrations of arachidonic acid and blood glucose. While the CXCR3 ligand CXCL10 was shown before to correlate with severity in sepsis [69], in the present study concentrations of CXCL10 were lower in non-survivors within 30 days. The relevance of blood glucose levels is supported by the surviving sepsis campaign international guidelines, which recommend tight control of blood glucose levels, with hyperglycaemia associated with increased mortality [20]. In addition, a reduction in arachidonic acid metabolism has been described in sepsis patients compared to healthy controls [70], and eicosanoid lipid mediators that are derived from arachidonic acid have been implicated in the pathogenesis of sepsis [71].

A compelling finding from this study was the importance attributed to unconventional T cells, even amongst the diversity of available features to select from, which included clinical

variables, protein and lipid mediators as well as the proportions and phenotypes of T cells, monocytes and neutrophils in blood. In predicting 90 day mortality, increased expression of the activation marker CD25 on MAIT cells influenced survival prediction, in line with a recent description of highly activated MAIT cells in clinical and experimental sepsis [72]. The reduction in circulating V δ 2⁺ T cells and MAIT cells in patients with Gram-negative sepsis seen in the present study might reflect recruitment of these cells to sites of infection [73, 74, 75] and agrees with previous observations in septic patients [76, 77, 78] and volunteers challenged with *E. coli* [79]. In contrast, our own previous analysis in a different sepsis cohort showed increased proportions of circulating V γ 9⁺ $\gamma\delta$ T cells in patients infected with HMB-PP⁺ pathogens [80], and others described higher levels of peripheral MAIT cells but not of $\gamma\delta$ T cells in critically ill patients infected with *Streptococcus* spp. (*i.e.* bacteria incapable of producing the corresponding ligands), compared to non-streptococcal causes [81]. Such conflicting findings may be due to key differences in patient characteristics, time points, severity and microbiological definition as well as the methodology used for the flow cytometric analysis. Regardless of these discrepancies, unconventional T cells recently helped differentiate the causative pathogen in patients with acute peritonitis [32] and have been identified as critical players in multi-parameter immune signatures with implications in COVID-19 prognosis [82, 83]. The work presented here provides additional evidence that profiling of V δ 2⁺ T cells and MAIT cells makes valuable contributions to predictive models of acute infection. In support, a recent study described MAIT cell activation (as defined by expression of HLA-DR) combined with the APACHE II score as best indicator of 28 day mortality in sepsis patients [78]. Future studies should include counting beads to be able to accurately report absolute numbers of T cell subsets in blood, which was not possible in the present analysis.

Rigorous validation is a concern in biomarker discovery [84, 85], as it is essential to distinguish between the data used for evaluating a model and the data used for model development, especially the selection of biomarkers to be included. The generation of independent holdout data was the only mechanism to ensure that the ascertained model and feature set combination were not overfitting to the chosen training data, other than the generation of an entirely new patient cohort for validation. The primary limitation in the analysis of cytokines, chemokines and acute phase proteins in plasma were the detection limits, resulting in all but six analytes being below or above the detectable range for 20% or more of the tested samples – highlighting the intrinsic problem of multiplex analyses of biologically distinct mediators. It is also important to note that the selection of features in machine learning models and their associated SHAP values do not imply causation but rather that their combined interaction simply identifies a correlation with the predicted target. To identify the underlying cause, additional experimentation and analysis would be required. Nevertheless, machine learning models can assist in narrowing the list of target variables for investigation in subsequent experiments, thus representing a useful hypothesis-generating exercise [38]. Additionally, the interpretability of machine learning models offers the potential to identify a more general sepsis model by identifying dysregulation patterns across multiple interconnected systems, with the potential to gather insights that move beyond static dogmas of ‘cytokine storm’ and ‘immune paralysis’ [86].

A critical constraint in this study was the patient heterogeneity, reflecting the complex and poorly defined nature of sepsis pathology [87]. Fewer than 70% of the patients enrolled in this study had a microbiologically confirmed infection, a rate comparable to previous observations in sepsis [24]. It was impossible to identify whether this was due to a failure of microbiological culture or the genuine absence of any bacterial infection. Additionally, around 25% of patients

were admitted to the ICU with trauma or following emergency surgery. Although this was included as a categorical variable in the machine learning pipeline to account for a potentially confounding effect, the clinical condition and type of care for such patients would differ from those that had not experienced trauma. There were also insufficient data regarding patient co-morbidity and history of infectious disease before admission to the ICU. Such data form important confounding variables for both prediction of survival and the underlying cause of infection.

There is a growing realisation in the field that the current definition of sepsis is inadequate, which draws focus to the dysregulated host response as the characterising feature of sepsis [1]. In fact, most clinical trials that sought to restore the immune balance have either failed to show benefit or have proven harmful [88]. The complicated patterns of clinical presentation represent a barrier to the advancement of diagnosis and therapy, while the Sepsis-3 definition cannot distinguish the complex heterogeneity observed in the pathophysiology of sepsis [89]. Research into COVID-19, a condition that has many parallels to sepsis, has reported success in uncovering immunological signatures associated with poor outcomes with links back to the underlying biological mechanisms [82]. Thus, focusing on a well-defined pathology within sepsis is likely to yield findings more readily associated with the underlying mechanism driving the immune response. Any future study expanding on the work discussed here should carefully consider the exact recruitment strategy. Reflecting on the success of immunophenotyping of COVID-19, simple strategies could be employed to limit recruitment to those of comparable aetiology, such as culture-positive urosepsis or acute lower respiratory infection. However, this poses the general dilemma of whether to conduct studies in a well-defined but relatively artificial cohort with strict inclusion or exclusion criteria, which will be best suited to define pathophysiological pathways specific for certain subtypes of sepsis, or rather to focus on common mechanisms and biomarkers in an inclusive real-

world scenario as it presents in the clinic. Alternatively, a robust recruitment approach might leverage unsupervised clustering and the identification of endotypes that could be treated as distinct yet overlapping groups [90]. In the future, sepsis is likely to be recognised not as an all-embracing syndrome but rather as a group of related conditions, each characterised by specific cellular alterations and associated biomarkers.

Accepted Manuscript

ACKNOWLEDGEMENTS

We are grateful to all patients and their advocates for participating in this study and to the clinicians and nurses for their cooperation, in particular Jade Cole and Helen Hill. We would also like to thank Mat Clement, Alexander Greenshields-Watson, Ann Kift-Morgan and Kristin Ladell for their help and advice throughout this study.

ETHICAL APPROVAL

Recruitment of sepsis patients was approved by the Health and Care Research Wales Research Ethics Committee under reference 17/WA/0253, protocol number SPON1609-17 and IRAS project ID 231993, and conducted according to the principles expressed in the Declaration of Helsinki.

PATIENT CONSENT

All study participants provided written informed consent for the collection of samples and their subsequent analysis. A waiver of consent system was used when patients were unable to provide prospective informed consent due to the nature of their critical illness or therapeutic sedation at the time of recruitment. In all cases, retrospective informed consent was sought as soon as the patient recovered and regained capacity. In cases where a patient passed away before regaining capacity, the initial consultee's approval would stand.

CONFLICTS OF INTEREST

All authors of the article declare that they have no conflict of interests.

FUNDING

This research was supported by Cardiff University School of Medicine PhD Studentships (R.J.B., L.M.M.), a EU Horizon 2020 Marie Skłodowska-Curie postdoctoral fellowship (L.R.), the European Regional Development Fund via the Welsh Government's Accelerate (M.E.) and Sêr Cymru II programmes (V.O.B., P.G.), a Health and Care Research Wales Clinical Research Time Award (M.P.M.), a Pathway to Portfolio grant (M.P.M.), a Wellcome Trust Institutional Translational Partnership Award (R.J.B, S.M.C., M.P.M., A.A., M.E.), and the Wales Data Nation Accelerator (M.E.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

AUTHOR CONTRIBUTORS

Study design: R.J.B., M.P.M., A.A., M.E. Laboratory experimentation and data acquisition: R.J.B., L.R., L.M.M., S.E.B., D.A.W. Analysis and interpretation of data: R.J.B., S.M.C., B.M., A.A., M.E. Supervision: V.B.O., P.G., M.P.M., A.A., M.E. Access to clinical samples: M.P.M. Drafting of the article: R.J.B., M.E. All authors critically revised the manuscript and approved the final version to be published.

REFERENCES

1. Giamarellos-Bourboulis EJ, Aschenbrenner AC, Bauer M, Bock C, Calandra T, Gatt-Viks I, Kyriazopoulou E, Lupse M, Monneret G, Pickkers P, Scholtze JL, van der Poll T, van de Veerdonk FL, Vlaar APJ, Weis S, Wiersinga WJ, Netea MG. The pathophysiology of sepsis and precision-medicine-based immunotherapy. *Nat Immunol*. 2024 Jan;25(1):19-28. doi: 10.1038/s41590-023-01660-5.
2. Serafim R, Gomes JA, Salluh J, Póvoa P. A Comparison of the Quick-SOFA and Systemic Inflammatory Response Syndrome Criteria for the Diagnosis of Sepsis and Prediction of Mortality: A Systematic Review and Meta-Analysis. *Chest*. 2018 Mar;153(3):646-655. doi: 10.1016/j.chest.2017.12.015.
3. Pellathy TP, Pinsky MR, Hravnak M. Intensive Care Unit Scoring Systems. *Crit Care Nurse*. 2021 Aug 1;41(4):54-64. doi: 10.4037/ccn2021613.
4. van Engelen TSR, Wiersinga WJ, Scicluna BP, van der Poll T. Biomarkers in Sepsis. *Crit Care Clin*. 2018 Jan;34(1):139-152. doi: 10.1016/j.ccc.2017.08.010.
5. Pierrakos C, Velissaris D, Bisdorff M, Marshall JC, Vincent JL. Biomarkers of sepsis: time for a reappraisal. *Crit Care*. 2020 Jun 5;24(1):287. doi: 10.1186/s13054-020-02993-5.
6. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, Shimabukuro D, Chettipally U, Feldman MD, Barton C, Wales DJ, Das R. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine

Learning Approach. JMIR Med Inform. 2016 Sep 30;4(3):e28. doi: 10.2196/medinform.5909.

7. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. Crit Care Med. 2018 Apr;46(4):547-553. doi: 10.1097/CCM.0000000000002936.
8. Delahanty RJ, Alvarez J, Flynn LM, Sherwin RL, Jones SS. Development and Evaluation of a Machine Learning Model for the Early Identification of Patients at Risk for Sepsis. Ann Emerg Med. 2019 Apr;73(4):334-344. doi: 10.1016/j.annemergmed.2018.11.036.
9. Shapiro NI, Trzeciak S, Hollander JE, Birkhahn R, Otero R, Osborn TM, Moretti E, Nguyen HB, Gunnerson KJ, Milzman D, Gaieski DF, Goyal M, Cairns CB, Ngo L, Rivers EP. A prospective, multicenter derivation of a biomarker panel to assess risk of organ dysfunction, shock, and death in emergency department patients with suspected sepsis. Crit Care Med. 2009 Jan;37(1):96-104. doi: 10.1097/CCM.0b013e318192fd9d.
10. Kim H, Hur M, Moon HW, Yun YM, Di Somma S; GREAT Network. Multi-marker approach using procalcitonin, presepsin, galectin-3, and soluble suppression of tumorigenicity 2 for the prediction of mortality in sepsis. Ann Intensive Care. 2017 Dec;7(1):27. doi: 10.1186/s13613-017-0252-y.

11. Zhang Z, Ni H. C-reactive protein as a predictor of mortality in critically ill patients: a meta-analysis and systematic review. *Anaesth Intensive Care*. 2011 Sep;39(5):854-61. doi: 10.1177/0310057X1103900509.
12. Jekarl DW, Lee S, Kim M, Kim Y, Woo SH, Lee WJ. Procalcitonin as a prognostic marker for sepsis based on SEPSIS-3. *J Clin Lab Anal*. 2019 Nov;33(9):e22996. doi: 10.1002/jcla.22996.
13. Arora S, Singh P, Singh PM, Trikha A. Procalcitonin Levels in Survivors and Nonsurvivors of Sepsis: Systematic Review and Meta-Analysis. *Shock*. 2015 Mar;43(3):212-21. doi: 10.1097/SHK.0000000000000305.
14. Kim MH, Choi JH. An Update on Sepsis Biomarkers. *Infect Chemother*. 2020 Mar;52(1):1-18. doi: 10.3947/ic.2020.52.1.1.
15. Song J, Park DW, Moon S, Cho HJ, Park JH, Seok H, Choi WS. Diagnostic and prognostic value of interleukin-6, pentraxin 3, and procalcitonin levels among sepsis and septic shock patients: a prospective controlled study according to the Sepsis-3 definitions. *BMC Infect Dis*. 2019 Nov 12;19(1):968. doi: 10.1186/s12879-019-4618-7.
16. Farkas JD. The complete blood count to diagnose septic shock. *J Thorac Dis*. 2020 Feb;12(Suppl 1):S16-S21. doi: 10.21037/jtd.2019.12.63.
17. Cavaillon JM, Adib-Conquy M. Immune status in sepsis: the bug, the site of infection and the severity can make the difference. *Crit Care*. 2010;14(3):167. doi: 10.1186/cc9046.

18. Das A, Ariyakumar G, Gupta N, Kamdar S, Barugahare A, Deveson-Lucas D, Gee S, Costeloe K, Davey MS, Fleming P, Gibbons DL. Identifying immune signatures of sepsis to increase diagnostic accuracy in very preterm babies. *Nat Commun.* 2024 Jan 9;15(1):388. doi: 10.1038/s41467-023-44387-5.
19. Chen Q, Shi J, Fei A, Wang F, Pan S, Wang W. Neutrophil CD64 expression is a predictor of mortality for patients in the intensive care unit. *Int J Clin Exp Pathol.* 2014 Oct 15;7(11):7806-13.
20. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, Machado FR, McIntyre L, Ostermann M, Prescott HC, Schorr C, Simpson S, Wiersinga WJ, Alshamsi F, Angus DC, Arabi Y, Azevedo L, Beale R, Beilman G, Belley-Cote E, Burry L, Cecconi M, Centofanti J, Coz Yataco A, De Waele J, Dellinger RP, Doi K, Du B, Estenssoro E, Ferrer R, Gomersall C, Hodgson C, Hylander Møller M, Iwashyna T, Jacob S, Kleinpell R, Klompas M, Koh Y, Kumar A, Kwizera A, Lobo S, Masur H, McGloughlin S, Mehta S, Mehta Y, Mer M, Nunnally M, Oczkowski S, Osborn T, Papathanassoglou E, Perner A, Puskarich M, Roberts J, Schweickert W, Seckel M, Sevransky J, Sprung CL, Welte T, Zimmerman J, Levy M. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. *Crit Care Med.* 2021 Nov 1;49(11):e1063-e1143. doi: 10.1097/CCM.0000000000005337.
21. Morgan MP, Szakmany T, Power SG, Olaniyi P, Hall JE, Rowan K, Eberl M. Sepsis Patients with First and Second-Hit Infections Show Different Outcomes Depending

on the Causative Organism. *Front Microbiol.* 2016 Feb 26;7:207. doi: 10.3389/fmicb.2016.00207.

22. Disselkamp M, Coz Yataco AO, Simpson SQ. POINT: Should Broad-Spectrum Antibiotics Be Routinely Administered to All Patients With Sepsis as Soon as Possible? Yes. *Chest.* 2019 Oct;156(4):645-647. doi: 10.1016/j.chest.2019.05.030.
23. Patel JJ, Bergl PA. COUNTERPOINT: Should Broad-Spectrum Antibiotics Be Routinely Administered to All Patients With Sepsis as Soon as Possible? No. *Chest.* 2019 Oct;156(4):647-649. doi: 10.1016/j.chest.2019.05.031.
24. Thorndike J, Kollef MH. Culture-negative sepsis. *Curr Opin Crit Care.* 2020 Oct;26(5):473-477. doi: 10.1097/MCC.0000000000000751.
25. Doernberg SB. Will biomarkers be the answer for antibiotic stewardship? *Lancet Respir Med.* 2020 Feb;8(2):130-132. doi: 10.1016/S2213-2600(19)30406-0.
26. Chun K, Syndergaard C, Damas C, Trubey R, Mukindaraj A, Qian S, Jin X, Breslow S, Niemz A. Sepsis Pathogen Identification. *J Lab Autom.* 2015 Oct;20(5):539-61. doi: 10.1177/2211068214567345.
27. Warhurst G, Dunn G, Chadwick P, Blackwood B, McAuley D, Perkins GD, McMullan R, Gates S, Bentley A, Young D, Carlson GL, Dark P. Rapid detection of health-care-associated bloodstream infection in critical care using multipathogen real-time polymerase chain reaction technology: a diagnostic accuracy study and systematic review. *Health Technol Assess.* 2015 May;19(35):1-142. doi: 10.3310/hta19350.

28. Thomas-Rüddel DO, Poidinger B, Kott M, Weiss M, Reinhart K, Bloos F; MEDUSA study group. Influence of pathogen and focus of infection on procalcitonin values in sepsis patients with bacteremia or candidemia. *Crit Care*. 2018 May 13;22(1):128. doi: 10.1186/s13054-018-2050-9.
29. Masson S, Caironi P, Fanizza C, Thomae R, Bernasconi R, Noto A, Oggioni R, Pasetti GS, Romero M, Tognoni G, Latini R, Gattinoni L. Circulating presepsin (soluble CD14 subtype) as a marker of host response in patients with severe sepsis or septic shock: data from the multicenter, randomized ALBIOS trial. *Intensive Care Med*. 2015 Jan;41(1):12-20. doi: 10.1007/s00134-014-3514-2.
30. Feezor RJ, Oberholzer C, Baker HV, Novick D, Rubinstein M, Moldawer LL, Pribble J, Souza S, Dinarello CA, Ertel W, Oberholzer A. Molecular characterization of the acute inflammatory response to infections with gram-negative versus gram-positive bacteria. *Infect Immun*. 2003 Oct;71(10):5803-13. doi: 10.1128/IAI.71.10.5803-5813.2003.
31. Liuzzi AR, McLaren JE, Price DA, Eberl M. Early innate responses to pathogens: pattern recognition by unconventional human T-cells. *Curr Opin Immunol*. 2015 Oct;36:31-7. doi: 10.1016/j.coi.2015.06.002.
32. Zhang J, Friberg IM, Kift-Morgan A, Parekh G, Morgan MP, Liuzzi AR, Lin CY, Donovan KL, Colmont CS, Morgan PH, Davis P, Weeks I, Fraser DJ, Topley N, Eberl M. Machine-learning algorithms define pathogen-specific local immune fingerprints in peritoneal dialysis patients with bacterial infections. *Kidney Int*. 2017 Jul;92(1):179-191. doi: 10.1016/j.kint.2017.01.017.

33. Langley RJ, Tsalik EL, van Velkinburgh JC, Glickman SW, Rice BJ, Wang C, Chen B, Carin L, Suarez A, Mohny RP, Freeman DH, Wang M, You J, Wulff J, Thompson JW, Moseley MA, Reisinger S, Edmonds BT, Grinnell B, Nelson DR, Dinwiddie DL, Miller NA, Saunders CJ, Soden SS, Rogers AJ, Gazourian L, Fredenburgh LE, Massaro AF, Baron RM, Choi AM, Corey GR, Ginsburg GS, Cairns CB, Otero RM, Fowler VG Jr, Rivers EP, Woods CW, Kingsmore SF. An integrated clinico-metabolomic model improves prediction of death in sepsis. *Sci Transl Med*. 2013 Jul 24;5(195):195ra95. doi: 10.1126/scitranslmed.3005893.
34. Taneja I, Damhorst GL, Lopez-Espina C, Zhao SD, Zhu R, Khan S, White K, Kumar J, Vincent A, Yeh L, Majdizadeh S, Weir W, Isbell S, Skinner J, Devanand M, Azharuddin S, Meenakshisundaram R, Upadhyay R, Syed A, Bauman T, Devito J, Heinzmann C, Podolej G, Shen L, Timilsina SS, Quinlan L, Manafirasi S, Valera E, Reddy B Jr, Bashir R. Diagnostic and prognostic capabilities of a biomarker and EMR-based machine learning algorithm for sepsis. *Clin Transl Sci*. 2021 Jul;14(4):1578-1589. doi: 10.1111/cts.13030.
35. Chao HY, Wu CC, Singh A, Shedd A, Wolfshohl J, Chou EH, Huang YC, Chen KF. Using Machine Learning to Develop and Validate an In-Hospital Mortality Prediction Model for Patients with Suspected Sepsis. *Biomedicines*. 2022 Mar 29;10(4):802. doi: 10.3390/biomedicines10040802.
36. Smith CL, Dickinson P, Forster T, Craigon M, Ross A, Khondoker MR, France R, Ivens A, Lynn DJ, Orme J, Jackson A, Lacaze P, Flanagan KL, Stenson BJ, Ghazal

- P. Identification of a human neonatal immune-metabolic network associated with bacterial infection. *Nat Commun.* 2014 Aug 14;5:4649. doi: 10.1038/ncomms5649.
37. Ward RA, Aghaeepour N, Bhattacharyya RP, Clish CB, Gaudillière B, Hacohen N, Mansour MK, Mudd PA, Pasupneti S, Presti RM, Rhee EP, Sen P, Spec A, Tam JM, Villani AC, Woolley AE, Hsu JL, Vyas JM. Harnessing the Potential of Multiomics Studies for Precision Medicine in Infectious Disease. *Open Forum Infect Dis.* 2021 Sep 25;8(11):ofab483. doi: 10.1093/ofid/ofab483.
38. Komorowski M, Green A, Tatham KC, Seymour C, Antcliffe D. Sepsis biomarkers and diagnostic tools with a focus on machine learning. *EBioMedicine.* 2022 Dec;86:104394. doi: 10.1016/j.ebiom.2022.104394.
39. Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications. *Comput Biol Med.* 2019 Sep;112:103375. doi: 10.1016/j.combiomed.2019.103375.
40. Wolpert DH, Macready WG. No free lunch theorems for optimization. *IEEE Trans Evol Comput.* 1997; 1(1):67-82. doi: 10.1109/4235.585893.
41. 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.* 2014; 15(90):3133–3181.
42. Wu J, Vodovotz Y, Abdelhamid S, Guyette FX, Yaffe MB, Gruen DS, Cyr A, Okonkwo DO, Kar UK, Krishnamoorthi N, Voinchet RG, Billiar IM, Yazer MH, Namas RA, Daley BJ, Miller RS, Harbrecht BG, Claridge JA, Phelan HA,

- Zuckerbraun BS, Johansson PI, Stensballe J, Morrissey JH, Tracy RP, Wisniewski SR, Neal MD, Sperry JL, Billiar TR; PAMPer study group. Multi-omic analysis in injured humans: Patterns align with outcomes and treatment responses. *Cell Rep Med*. 2021 Dec 21;2(12):100478. doi: 10.1016/j.xcrm.2021.100478.
43. Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, Meyer JG, Quan Q, Muehlbauer LK, Trujillo EA, He Y, Chopra A, Chieng HC, Tiwari A, Judson MA, Paulson B, Brademan DR, Zhu Y, Serrano LR, Linke V, Drake LA, Adam AP, Schwartz BS, Singer HA, Swanson S, Mosher DF, Stewart R, Coon JJ, Jaitovich A. Large-Scale Multi-omic Analysis of COVID-19 Severity. *Cell Syst*. 2021 Jan 20;12(1):23-40.e7. doi: 10.1016/j.cels.2020.10.003.
44. Tomic A, Tomic I, Rosenberg-Hasson Y, Dekker CL, Maecker HT, Davis MM. SIMON, an Automated Machine Learning System, Reveals Immune Signatures of Influenza Vaccine Responses. *J Immunol*. 2019 Aug 1;203(3):749-759. doi: 10.4049/jimmunol.1900033.
45. Fan B, Klatt J, Moor MM, Daniels LA; Swiss Pediatric Sepsis Study; Sanchez-Pinto LN, Agyeman PKA, Schlapbach LJ, Borgwardt KM. Prediction of recovery from multiple organ dysfunction syndrome in pediatric sepsis patients. *Bioinformatics*. 2022 Jun 24;38(Suppl 1):i101-i108. doi: 10.1093/bioinformatics/btac229.
46. Tang G, Luo Y, Lu F, Li W, Liu X, Nan Y, Ren Y, Liao X, Wu S, Jin H, Zomaya AY, Sun Z. Prediction of Sepsis in COVID-19 Using Laboratory Indicators. *Front Cell Infect Microbiol*. 2021 Mar 2;10:586054. doi: 10.3389/fcimb.2020.586054.

47. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, Hotchkiss RS, Levy MM, Marshall JC, Martin GS, Opal SM, Rubenfeld GD, van der Poll T, Vincent JL, Angus DC. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016 Feb 23;315(8):801-10. doi: 10.1001/jama.2016.0287.
48. Raffray L, Burton RJ, Baker SE, Morgan MP, Eberl M. Zoledronate rescues immunosuppressed monocytes in sepsis patients. *Immunology*. 2020 Jan;159(1):88-95. doi: 10.1111/imm.13132.
49. Szakmany T, Lundin RM, Sharif B, Ellis G, Morgan P, Kopczynska M, Dhadda A, Mann C, Donoghue D, Rollason S, Brownlow E, Hill F, Carr G, Turley H, Hassall J, Lloyd J, Davies L, Atkinson M, Jones M, Jones N, Martin R, Ibrahim Y, Hall JE; Welsh Digital Data Collection Platform Collaborators. Sepsis Prevalence and Outcome on the General Wards and Emergency Departments in Wales: Results of a Multi-Centre, Observational, Point Prevalence Study. *PLoS One*. 2016 Dec 1;11(12):e0167230. doi: 10.1371/journal.pone.0167230.
50. Burton RJ, Ahmed R, Cuff SM, Baker S, Artemiou A, Eberl M. CytoPy: An autonomous cytometry analysis framework. *PLoS Comput Biol*. 2021 Jun 8;17(6):e1009071. doi: 10.1371/journal.pcbi.1009071.
51. Burton RJ, Cuff SM, Morgan MP, Artemiou A, Eberl M. GeoWaVe: geometric median clustering with weighted voting for ensemble clustering of cytometry data. *Bioinformatics*. 2023 Jan 1;39(1):btac751. doi: 10.1093/bioinformatics/btac751.

52. Tenny S, Hoffman MR. Odds Ratio. [Updated 2022 May 24]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK431098/>
53. Han J, Lin K, Sequeira C, Borchers CH. An isotope-labeled chemical derivatization method for the quantitation of short-chain fatty acids in human feces by liquid chromatography-tandem mass spectrometry. *Anal Chim Acta*. 2015 Jan 7;854:86-94. doi: 10.1016/j.aca.2014.11.015.
54. Wong KC, Xiang Y, Yin L, So HC. Uncovering Clinical Risk Factors and Predicting Severe COVID-19 Cases Using UK Biobank Data: Machine Learning Approach. *JMIR Public Health Surveill*. 2021 Sep 30;7(9):e29544. doi: 10.2196/29544.
55. Raschka S. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *J Open Source Softw*. 2018; 3(24):638. doi: 10.21105/joss.00638.
56. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. Relief-based feature selection: Introduction and review. *J Biomed Inform*. 2018; 85: 189-203. doi: 10.1016/j.jbi.2018.07.014
57. Zhao Z, Anand R, Wang M. Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE. 2019, pp. 442–452. doi: 10.1109/DSAA.2019.00059.

58. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw.* 2010; 36:1-13. doi: 10.18637/jss.v036.i11.
59. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Machine Learning* 2002; 46:389-422. doi: 10.1023/A:1012487302797
60. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P; SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* 2020 Mar;17(3):261-272. doi: 10.1038/s41592-019-0686-2.
61. Demšar J. Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res.* 2006; 7:1-30.
62. Chen H, Lundberg SM, Lee SI. Explaining a series of models by propagating Shapley values. *Nat Commun.* 2022 Aug 3;13(1):4512. doi: 10.1038/s41467-022-31384-3.
63. Nolt B, Tu F, Wang X, Ha T, Winter R, Williams DL, Li C. Lactate and Immunosuppression in Sepsis. *Shock.* 2018 Feb;49(2):120-125. doi: 10.1097/SHK.0000000000000958.

64. Dahl RM, Grønlykke L, Haase N, Holst LB, Perner A, Wetterslev J, Rasmussen BS, Meyhoff CS; 6S-Trial and TRISS Trial investigators. Variability in targeted arterial oxygenation levels in patients with severe sepsis or septic shock. *Acta Anaesthesiol Scand*. 2015 Aug;59(7):859-69. doi: 10.1111/aas.12528.
65. Zhuang Y, Peng H, Chen Y, Zhou S, Chen Y. Dynamic monitoring of monocyte HLA-DR expression for the diagnosis, prognosis, and prediction of sepsis. *Front Biosci (Landmark Ed)*. 2017 Mar 1;22(8):1344-1354. doi: 10.2741/4547.
66. Inoue S, Suzuki-Utsunomiya K, Okada Y, Taira T, Iida Y, Miura N, Tsuji T, Yamagiwa T, Morita S, Chiba T, Sato T, Inokuchi S. Reduction of immunocompetent T cells followed by prolonged lymphopenia in severe sepsis in the elderly. *Crit Care Med*. 2013 Mar;41(3):810-9. doi: 10.1097/CCM.0b013e318274645f.
67. Rimmelé T, Payen D, Cantaluppi V, Marshall J, Gomez H, Gomez A, Murray P, Kellum JA; ADQI XIV Workgroup. Immune cell phenotype and function in sepsis. *Shock*. 2016 Mar;45(3):282-91. doi: 10.1097/SHK.0000000000000495.
68. Guo Y, Luan L, Patil NK, Wang J, Bohannon JK, Rabacal W, Fensterheim BA, Hernandez A, Sherwood ER. IL-15 Enables Septic Shock by Maintaining NK Cell Integrity and Function. *J Immunol*. 2017 Feb 1;198(3):1320-1333. doi: 10.4049/jimmunol.1601486.
69. Punyadeera C, Schneider EM, Schaffer D, Hsu HY, Joos TO, Kriebel F, Weiss M, Verhaegh WF. A biomarker panel to discriminate between systemic inflammatory

response syndrome and sepsis and sepsis severity. *J Emerg Trauma Shock*. 2010 Jan;3(1):26-35. doi: 10.4103/0974-2700.58666.

70. Bruegel M, Ludwig U, Kleinhempel A, Petros S, Kortz L, Ceglarek U, Holdt LM, Thiery J, Fiedler GM. Sepsis-associated changes of the arachidonic acid metabolism and their diagnostic potential in septic patients. *Crit Care Med*. 2012 May;40(5):1478-86. doi: 10.1097/CCM.0b013e3182416f05.
71. Amunugama K, Pike DP, Ford DA. The lipid biology of sepsis. *J Lipid Res*. 2021;62:100090. doi: 10.1016/j.jlr.2021.100090.
72. Trivedi S, Labuz D, Anderson CP, Araujo CV, Blair A, Middleton EA, Jensen O, Tran A, Mulvey MA, Campbell RA, Hale JS, Rondina MT, Leung DT. Mucosal-associated invariant T (MAIT) cells mediate protective host responses in sepsis. *Elife*. 2020 Nov 9;9:e55615. doi: 10.7554/eLife.55615.
73. Liuzzi AR, Kift-Morgan A, Lopez-Anton M, Friberg IM, Zhang J, Brook AC, Roberts GW, Donovan KL, Colmont CS, Toleman MA, Bowen T, Johnson DW, Topley N, Moser B, Fraser DJ, Eberl M. Unconventional Human T Cells Accumulate at the Site of Infection in Response to Microbial Ligands and Induce Local Tissue Remodeling. *J Immunol*. 2016 Sep 15;197(6):2195-207. doi: 10.4049/jimmunol.1600990.
74. Chen Z, Wang H, D'Souza C, Sun S, Kostenko L, Eckle SB, Meehan BS, Jackson DC, Strugnell RA, Cao H, Wang N, Fairlie DP, Liu L, Godfrey DI, Rossjohn J, McCluskey J, Corbett AJ. Mucosal-associated invariant T-cell activation and

accumulation after in vivo infection depends on microbial riboflavin synthesis and co-stimulatory signals. *Mucosal Immunol.* 2017 Jan;10(1):58-68. doi: 10.1038/mi.2016.39.

75. Ibidapo-Obe O, Stengel S, Köse-Vogel N, Quickert S, Reuken PA, Busch M, Bauer M, Stallmach A, Bruns T. Mucosal-Associated Invariant T Cells Redistribute to the Peritoneal Cavity During Spontaneous Bacterial Peritonitis and Contribute to Peritoneal Inflammation. *Cell Mol Gastroenterol Hepatol.* 2020;9(4):661-677. doi: 10.1016/j.jcmgh.2020.01.003.
76. Venet F, Bohé J, Debard AL, Bienvenu J, Lepape A, Monneret G. Both percentage of gammadelta T lymphocytes and CD3 expression are reduced during septic shock. *Crit Care Med.* 2005 Dec;33(12):2836-40. doi: 10.1097/01.ccm.0000189745.66585.ae.
77. Choi J, Schmerk CL, Mele TS, Rudak PT, Wardell CM, Deng G, Pavri FR, Kim K, Cepinskas G, He W, Haeryfar SM. Longitudinal analysis of mucosa-associated invariant T cells in sepsis reveals their early numerical decline with prognostic implications and a progressive loss of antimicrobial functions. *Immunol Cell Biol.* 2023 Mar;101(3):249-261. doi: 10.1111/imcb.12619.
78. Tian L, Xu J, Chen C, Lin J, Ju L, Chen L, Zhang Y, Han X, Liu L. HLA-DR+ mucosal-associated invariant T cells predict poor prognosis in patients with sepsis: a prospective observational study. *Scand J Immunol.* 2023 May 10:e13286. doi: 10.1111/sji.13286.

79. Rim S, Sakkestad ST, Zhou F, Gullaksen SE, Skavland J, Chauhan SK, Steinsland H, Hanevik K. Dynamics of circulating lymphocytes responding to human experimental enterotoxigenic *Escherichia coli* infection. *Eur J Immunol*. 2023 Apr 27:e2250254. doi: 10.1002/eji.202250254.
80. Davey MS, Morgan MP, Liuzzi AR, Tyler CJ, Khan MWA, Szakmany T, Hall JE, Moser B, Eberl M. Microbe-specific unconventional T cells induce human neutrophil differentiation into antigen cross-presenting cells. *J Immunol*. 2014 Oct 1;193(7):3704-3716. doi: 10.4049/jimmunol.1401018.
81. Grimaldi D, Le Bourhis L, Sauneuf B, Dechartres A, Rousseau C, Ouaz F, Milder M, Louis D, Chiche JD, Mira JP, Lantz O, Pène F. Specific MAIT cell behaviour among innate-like T lymphocytes in critically ill patients with severe infections. *Intensive Care Med*. 2014 Feb;40(2):192-201. doi: 10.1007/s00134-013-3163-x.
82. Laing AG, Lorenc A, Del Molino Del Barrio I, Das A, Fish M, Monin L, Muñoz-Ruiz M, McKenzie DR, Hayday TS, Francos-Quijorna I, Kamdar S, Joseph M, Davies D, Davis R, Jennings A, Zlatareva I, Vantourout P, Wu Y, Sofra V, Cano F, Greco M, Theodoridis E, Freedman JD, Gee S, Chan JNE, Ryan S, Bugallo-Blanco E, Peterson P, Kisand K, Haljasmägi L, Chadli L, Moingeon P, Martinez L, Merrick B, Bisnauthsing K, Brooks K, Ibrahim MAA, Mason J, Lopez Gomez F, Babalola K, Abdul-Jawad S, Cason J, Mant C, Seow J, Graham C, Doores KJ, Di Rosa F, Edgeworth J, Shankar-Hari M, Hayday AC. A dynamic COVID-19 immune signature includes associations with poor prognosis. *Nat Med*. 2020 Oct;26(10):1623-1635. doi: 10.1038/s41591-020-1038-6.

83. Tian Y, Carpp LN, Miller HER, Zager M, Newell EW, Gottardo R. Single-cell immunology of SARS-CoV-2 infection. *Nat Biotechnol.* 2022 Jan;40(1):30-41. doi: 10.1038/s41587-021-01131-y.
84. Boulesteix AL. Ten simple rules for reducing overoptimistic reporting in methodological computational research. *PLoS Comput Biol.* 2015 Apr 23;11(4):e1004191. doi: 10.1371/journal.pcbi.1004191.
85. Desaire H. How (Not) to Generate a Highly Predictive Biomarker Panel Using Machine Learning. *J Proteome Res.* 2022 Sep 2;21(9):2071-2074. doi: 10.1021/acs.jproteome.2c00117.
86. Ghazal P, Rodrigues PRS, Chakraborty M, Oruganti S, Woolley TE. Challenging molecular dogmas in human sepsis using mathematical reasoning. *EBioMedicine.* 2022 Jun;80:104031. doi: 10.1016/j.ebiom.2022.104031.
87. de Grooth HJ, Postema J, Loer SA, Parienti JJ, Oudemans-van Straaten HM, Girbes AR. Unexplained mortality differences between septic shock trials: a systematic analysis of population characteristics and control-group mortality rates. *Intensive Care Med.* 2018 Mar;44(3):311-322. doi: 10.1007/s00134-018-5134-8.
88. Alcock J. The Emperor Has No Clothes? Searching for Dysregulation in Sepsis. *J Clin Med.* 2018 Aug 29;7(9):247. doi: 10.3390/jcm7090247.
89. van der Poll T, Shankar-Hari M, Wiersinga WJ. The immunology of sepsis. *Immunity.* 2021 Nov 9;54(11):2450-2464. doi: 10.1016/j.immuni.2021.10.012.

90. Seymour CW, Kennedy JN, Wang S, Chang CH, Elliott CF, Xu Z, Berry S, Clermont G, Cooper G, Gomez H, Huang DT, Kellum JA, Mi Q, Opal SM, Talisa V, van der Poll T, Visweswaran S, Vodovotz Y, Weiss JC, Yealy DM, Yende S, Angus DC. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. JAMA. 2019 May 28;321(20):2003-2017. doi: 10.1001/jama.2019.5791.

Accepted Manuscript

Table 1. Holdout performance for the top-performing model selected within each classifier family. Each model is presented as the name of the classifier family, the feature selection method that generated the optimal feature set, and the number of features selected for the top-performing model. The highest ranking metrics are highlighted in bold font. Bootstrapped 95% confidence intervals are shown in square brackets, generated using 100 rounds of resampling.

	Balanced accuracy	Macro F1 score	Macro AUC score
<i>Prediction of 30 day mortality</i>			
Logistic regression-RFE-SVM, top 10	0.58 [0.50–0.67]	0.56 [0.49–0.60]	0.49 [0.44–0.58]
Linear SVM-RFE-SVM, top 10	0.59 [0.5–0.67]	0.57 [0.49–0.60]	0.52 [0.42–0.61]
SVM (cubic polynomial)-RFE-SVM, top 10	0.63 [0.54–0.71]	0.62 [0.53–0.66]	0.59 [0.53–0.64]
KNN-Boruta, top 7	0.63 [0.63–0.67]	0.65 [0.64–0.71]	0.68 [0.64–0.72]
Naive Bayes-RFE-SVM, top 9	0.50 [0.38–0.54]	0.50 [0.38–0.53]	0.52 [0.44–0.57]
Random Forest-Boruta, top 6	0.46 [0.46–0.50]	0.41 [0.40–0.42]	0.59 [0.55–0.64]
Extra Random Forest-Boruta, top 7	0.54 [0.42–0.58]	0.54 [0.40–0.58]	0.73 [0.64–0.83]
XGBoost-Boruta, top 7	0.50 [0.38–0.54]	0.50 [0.38–0.53]	0.48 [0.31–0.56]
<i>Prediction of 90 day mortality</i>			
Logistic regression-RFE-SVM, top 10	0.71 [0.63–0.79]	0.73 [0.64–0.79]	0.69 [0.58–0.78]
Linear SVM-RFE-SVM-top 10	0.71 [0.63–0.79]	0.73 [0.64–0.79]	0.67 [0.58–0.78]
SVM (quartic polynomial)-RFE-SVM, top 8	0.70 [0.63–0.79]	0.72 [0.64–0.79]	0.72 [0.67–0.77]
KNN-RFE-SVM, top 6	0.54 [0.42–0.58]	0.54 [0.40–0.58]	0.54 [0.44–0.64]
Naive Bayes-RFE-SVM, top 10	0.71 [0.63–0.79]	0.73 [0.64–0.79]	0.71 [0.61–0.94]
Random Forest-Boruta, top 3	0.67 [0.58–0.75]	0.67 [0.58–0.72]	0.80 [0.77–0.89]
Extra Random Forest-RFE-SVM, top 7	0.75 [0.67–0.83]	0.79 [0.71–0.88]	0.85 [0.81–0.86]
XGBoost-Boruta, top 6	0.54 [0.42–0.58]	0.54 [0.40–0.58]	0.58 [0.44–0.72]
<i>Prediction of Gram-negative infection</i>			
Logistic regression-RFE-SVM, top 6	0.83 [0.75–1.0]	0.86 [0.80–1.0]	0.76 [0.64–1.0]
Linear SVM-RFE-SVM, top 5	0.83 [0.75–1.0]	0.86 [0.80–1.0]	0.71 [0.57–1.0]
SVM (cubic polynomial)-RFE-SVM, top 5	0.76 [0.68–0.93]	0.76 [0.68–0.86]	0.71 [0.57–0.92]
KNN-RFE-SVM, top 5	0.60 [0.43–0.68]	0.59 [0.4–0.68]	0.81 [0.71–0.93]
Naive Bayes-RFE-SVM, top 6	0.66 [0.50–0.75]	0.67 [0.44–0.80]	0.81 [0.71–1.0]
Random Forest-Boruta, top 6	0.67 [0.50–0.75]	0.68 [0.44–0.80]	0.86 [0.79–0.94]
Extra Random Forest-Boruta, top 6	0.69 [0.61–0.86]	0.67 [0.59–0.75]	0.62 [0.50–0.79]
XGBoost-Boruta, top 5	0.60 [0.43–0.68]	0.60 [0.4–0.68]	0.76 [0.71–0.89]

Figure Legends

Figure 1. Proportion of T cells, monocytes and neutrophils, and conventional and unconventional T cell subsets in patients after sepsis diagnosis. Comparisons shown are between (A) survivors (blue) and non-survivors (red) 30 days after sepsis diagnosis; (B) survivors (blue) and non-survivors (red) 90 days after sepsis diagnosis; (C) those without (blue) and with (red) a microbiologically confirmed infection; and (D) those with a Gram-positive (purple) and Gram-negative (pink) infection, amongst those with a positive bacterial culture. *p* values were generated using two-tailed Mann-Whitney U tests with Bonferroni-Holm corrections for multiple comparisons.

Figure 2. Mean fluorescence intensity (MFI) of HLA-DR, CD86, CD46, CD40 and CD62L on circulating monocytes in sepsis patients. Comparisons between survivors and non-survivors 30 (top) and 90 (bottom) days following a diagnosis of sepsis are shown. *p* values were generated using two-tailed Mann-Whitney U tests with Bonferroni-Holm corrections for multiple comparisons.

Figure 3. Mean fluorescence intensity (MFI) of HLA-DR, CD86, CD46, CD40 and CD62L on MAIT cells, with comparisons between sepsis patients with a Gram-positive versus a Gram-negative infection. *p* values were generated using two-tailed Mann-Whitney U tests with Bonferroni-Holm corrections for multiple comparisons.

Figure 4. Complete case analysis for an Extra Random Forest model tasked with predicting 90 day mortality in sepsis. Performance is documented by a receiver-operating characteristic (ROC) curve (left) and a bar plot (right) showing balanced accuracy, macro F1 score, and ROC area-under-curve (AUC) score. The dotted diagonal line accompanying the ROC curves represents a model with a random performance level.

Figure 5. SHAP (SHapely Additive exPlanations) values for an Extra Random Forest model to predict 90 day mortality. The beeswarm plot (top) shows each observation as a single data point coloured by the value of the feature for that instance, and ranked from the most impactful on the model outcome to the least impactful. The x-axis shows the SHAP value, with a lower value corresponding to an instance having a more significant impact on the negative case for the model (*i.e.* prediction of survival), and a positive value corresponding to having a more significant impact on the positive case for the model (*i.e.* prediction of death). The bar plot on the right-hand side of the beeswarm plot shows the imputation error (with a maximum value of 1) and the percentage of missing values observed in the original data. The heatmap (bottom) shows the SHAP values for each patient. The bar plot on the right-hand y axis shows each feature's mean absolute SHAP value as a measure of a feature's impact on model prediction. The line plot above the heatmap displays each patient's predicted outcome (black line) and the actual outcome (orange line). The dotted line between the possible outcomes is the expected value, equivalent to the observed mortality. Note that predictions reflect performance on the complete training data and do not reflect how the model would perform when exposed to new data.

Figure 6. SHAP (SHapely Additive exPlanations) values for a Random Forest model to predict Gram-negative infection. A lower SHAP value corresponds to an instance having a more significant impact on the negative case for the model (*i.e.* prediction of Gram-positive sepsis), and a positive value corresponds to having a more significant impact on the positive case for the model (*i.e.* prediction of Gram-negative sepsis). The dotted line between the possible outcomes in the heatmap is the expected value, equivalent to the observed incidence of Gram-negative sepsis.

Figure 7. Proportion of Vδ2⁺ T cells plotted against corresponding SHAP (SHapely Additive exPlanations) values that explain the impact on a Random Forest model to predict Gram-negative infection. Each data point represents a unique patient, coloured by the causative pathogen of their acute infection.

Accepted Manuscript

Figure 1

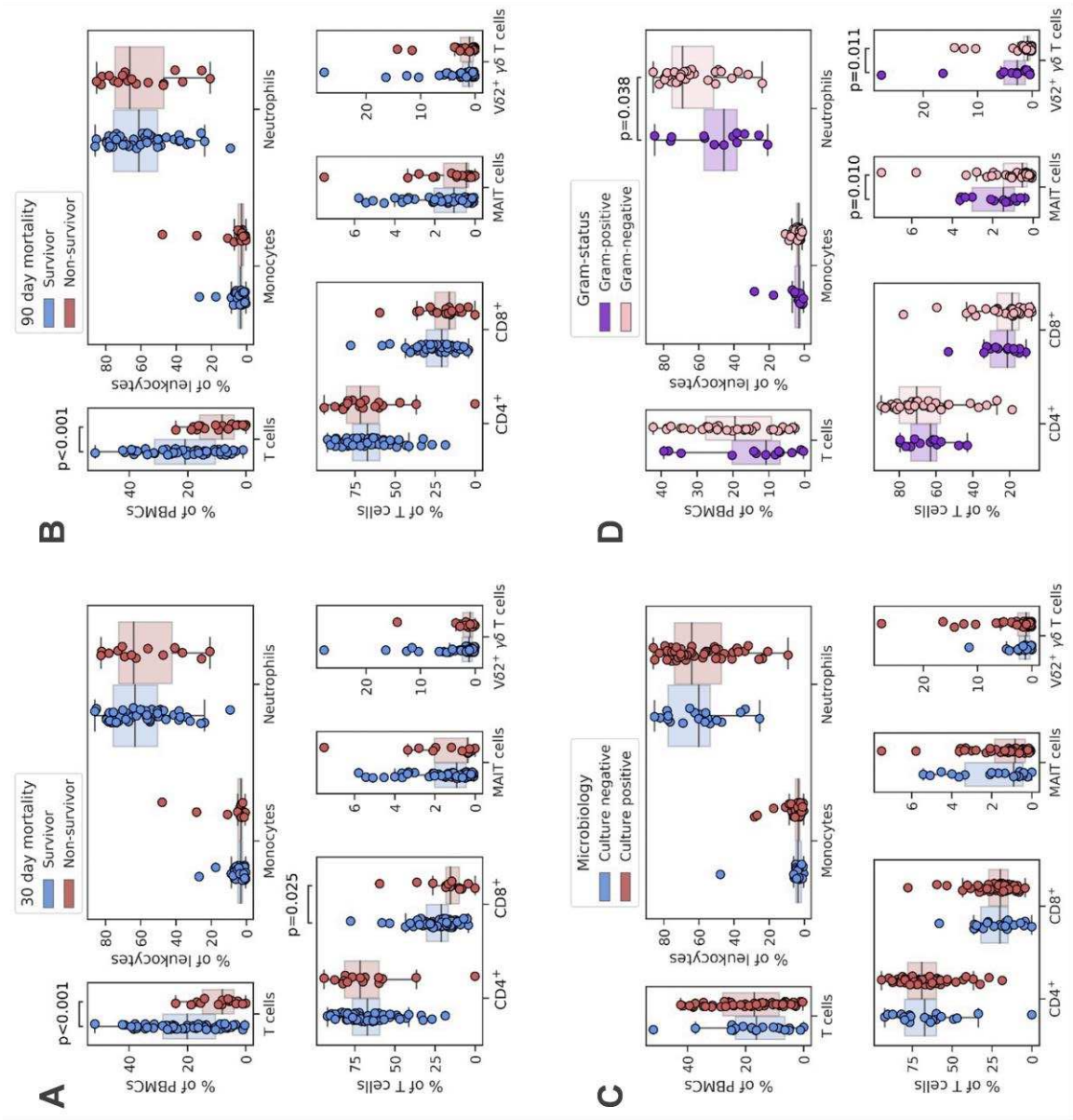


Figure 2

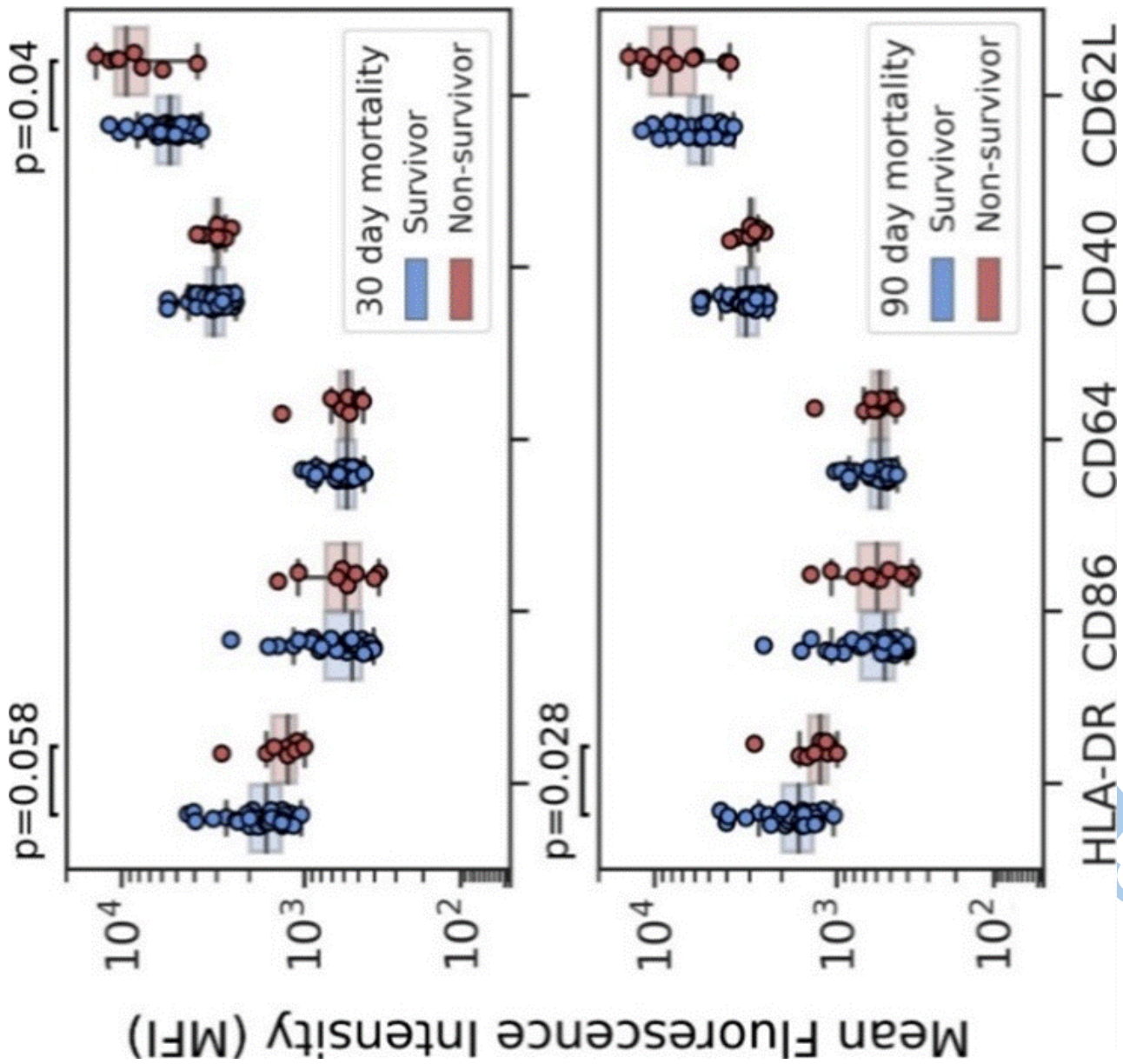


Figure 3

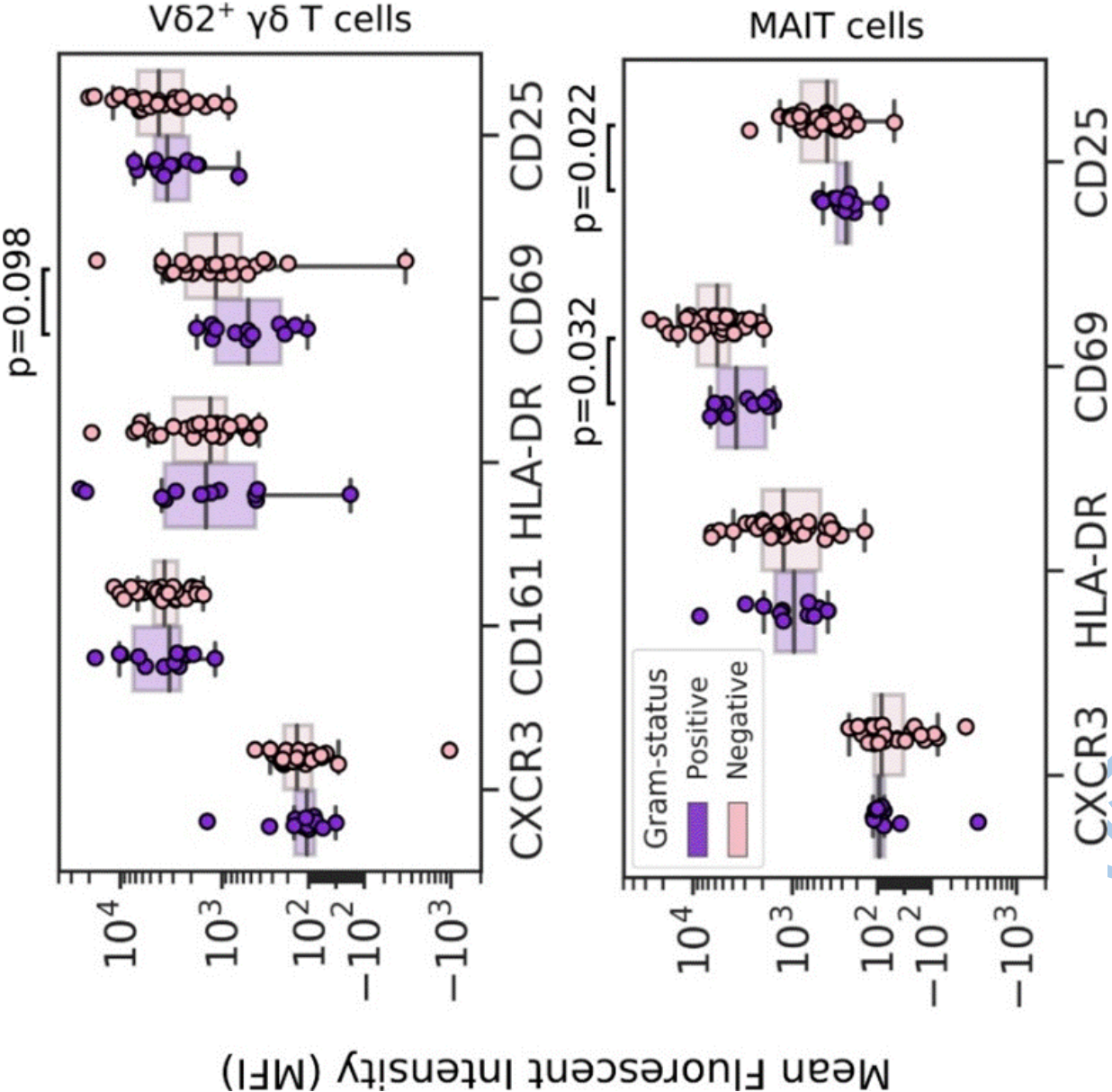


Figure 4

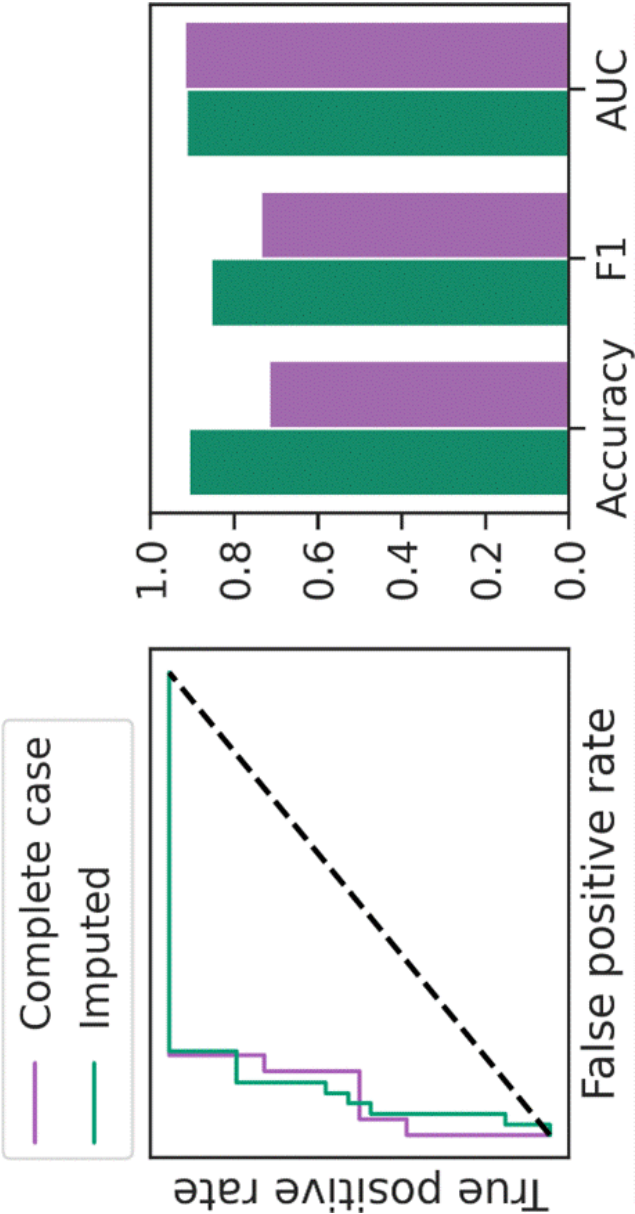


Figure 5

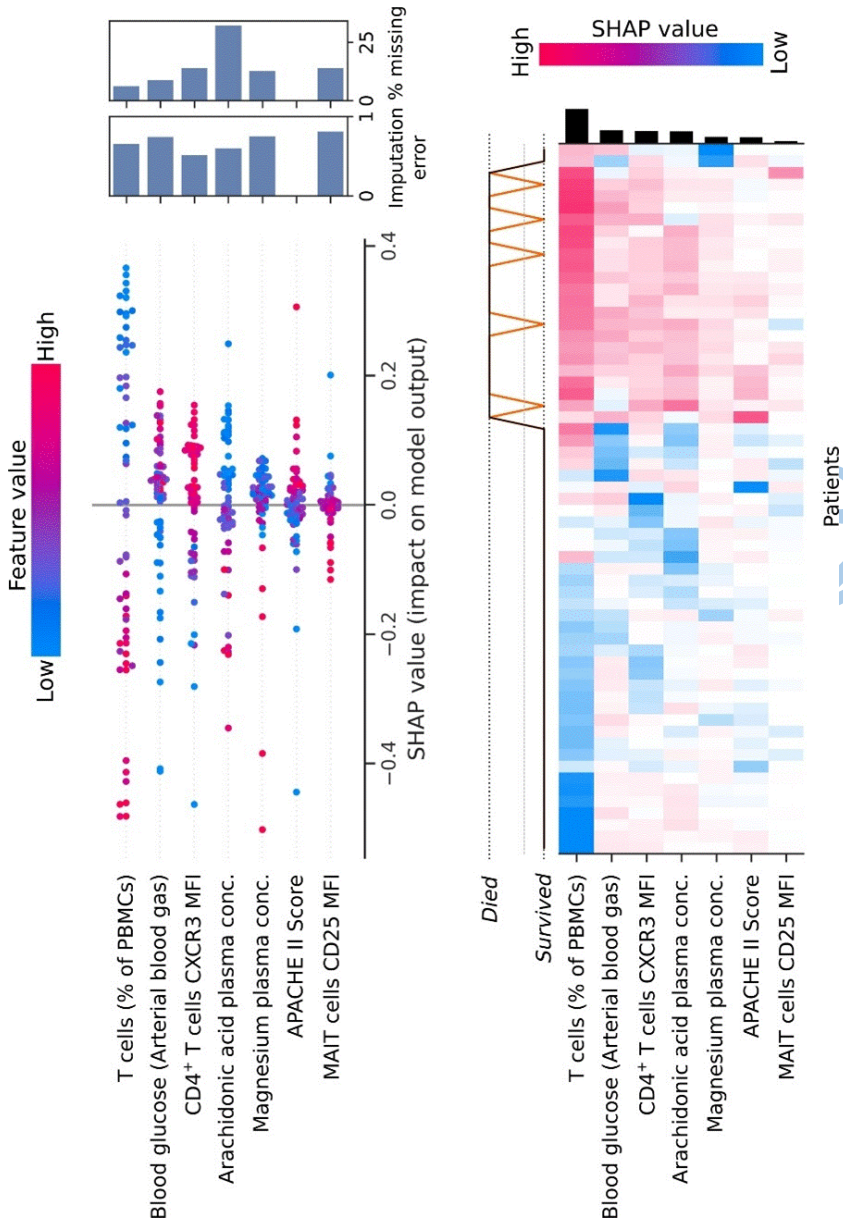


Figure 6

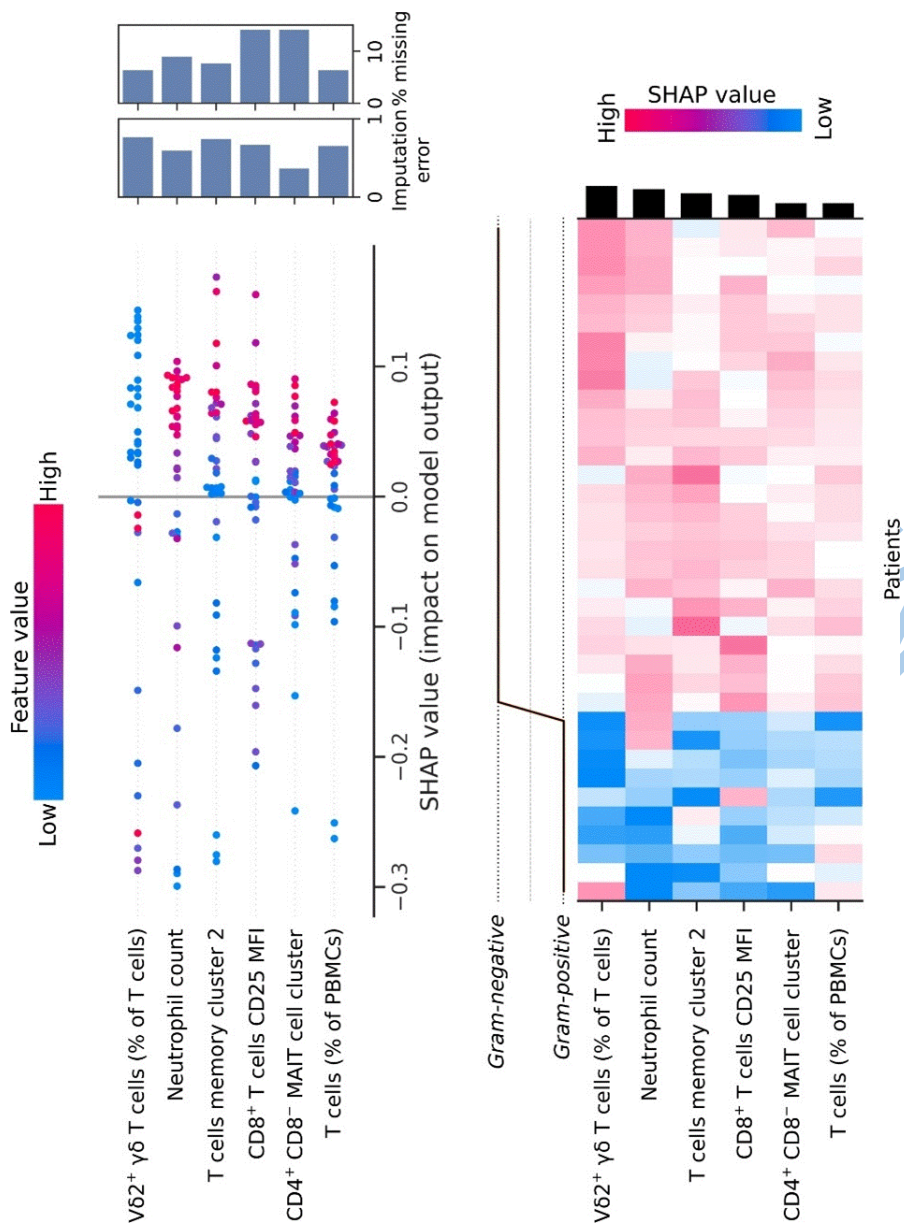


Figure 7

