

Artificial intelligence-driven cluster analysis for identifying clinical phenotypes in suspected sepsis patients in the emergency department

Received: 21 March 2025

Accepted: 27 March 2026

Published online: 02 April 2026

Cite this article as: Jeong D., Park J.R., Maeng S.J. *et al.* Artificial intelligence-driven cluster analysis for identifying clinical phenotypes in suspected sepsis patients in the emergency department. *BMC Emerg Med* (2026). <https://doi.org/10.1186/s12873-026-01566-z>

Daun Jeong, Jong Rul Park, Seung Jin Maeng, Jung Won Choi, Gun Tak Lee, Sung Yeon Hwang, Chulhong Kim, Jong Eun Park & Tae Gun Shin

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Artificial intelligence-driven cluster analysis for identifying clinical phenotypes in suspected sepsis patients in the emergency department

Daun Jeong^{1,2}, Jong Rul Park³, Seung Jin Maeng⁴, Jung Won Choi⁴, Gun Tak Lee, MD⁴, Sung Yeon Hwang, MD⁴, Chulhong Kim⁵, Jong Eun Park, MD⁴ and Tae Gun Shin, MD⁴

¹Division of Critical Care Medicine, Department of Emergency Medicine, Chung-Ang University Gwangmyeong Hospital, Gwangmyeong-si, Gyeonggi-do, Republic of Korea

²Department of Emergency Medicine, Chung-Ang University College of Medicine, Seoul, Republic of Korea

³Graduate School of Convergence Science and Technology, Major in Medical Science and Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea

⁴Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

⁵Department of Electrical Engineering, Convergence IT Engineering, and Medical Science and Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea

Address for correspondence:

***Correspondence:** Tae Gun Shin, MD, Department of Emergency

Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Republic of Korea. E-mail: tackles@naver.com, Tel.: +82-2-3410-2053; Fax: +82-2-3410-0012

****Co-correspondence:** Jong Eun Park, MD; Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Republic of Korea; E-mail: jongeun.samsung@gmail.com; Tel.: +82-2-3410-2053; Fax: +82-2-3410-0012

ARTICLE IN PRESS

Abstract

Background Identifying high-risk patients in emergency departments (EDs) is crucial due to high mortality rate associated with sepsis and for the timely management. This study aimed to conduct unsupervised consensus clustering of patients with suspected sepsis from a heterogeneous ED population.

Methods This descriptive, retrospective cohort study was conducted between November 2014 and November 2019 on patients with suspected sepsis who visited the ED. Cluster analysis was performed using variables, such as vital signs and clinical laboratory data, to classify patient characteristics. An artificial intelligence model generated a clustering plot based on the distance values of various variables, identifying five distinct clusters, each representing a unique clinical phenotype among the patient subtypes. In addition to analyzing the clinical phenotypes of each cluster, this study examined the prognosis based on antibiotic administration time through subgroup analysis to provide guidance for clinical application.

Results A total of 14,402 patients were included in this study. Cluster analysis combined with an artificial intelligence model identified five distinct clusters among patients with suspected sepsis. The baseline characteristics and clinical outcomes of each of the five clusters are described in detail. Cluster B comprised the largest proportion of patients with septic shock and exhibited the highest percentage of critical care interventions, including vasopressor use, mechanical

ventilation, intensive care unit admission, and 28-d mortality. Cluster E had the lowest rates of these interventions. In multivariable analysis, administration of antibiotics within 3 h significantly decreased 28-d mortality in Cluster A (aOR 0.73; 95% CI, 0.54–0.98, $p = 0.037$) and Cluster E (aOR 0.59; 95% CI, 0.39–0.89, $p = 0.013$).

Conclusion The artificial intelligence model identified five distinct clusters of patients with suspected sepsis who presented to the ED and showed different outcome patterns among the phenotypes.

Keywords Sepsis, Cluster analysis, Artificial intelligence, Emergency Department

Background

Sepsis is defined as a life-threatening organ dysfunction resulting from a dysregulated host response to infection [1]. Pathophysiologically, septic shock is a subtype of sepsis characterized by circulatory, cellular, and metabolic abnormalities. Septic shock is responsible for approximately 40% of hospital mortality [1, 2] with an average 30-d sepsis mortality rate of 24.4% [3].

Given the high mortality rate associated with sepsis, significant efforts have been made to identify high-risk patients [4]. However, most existing models fail to capture the multilevel factors of sepsis that can influence treatment response and are limited by scoring systems [5, 6]. The complex clinical phenotypes of sepsis continue to challenge clinicians due to the heterogeneous circumstances encountered in the emergency department.

The identification of sepsis sub phenotypes has gained attention because of the syndrome's heterogeneous nature [7, 8]. As sepsis is a broad condition with diverse clinical features affecting organ dysfunctions, it is necessary to classify heterogeneous patients into distinctive phenotypes or subsets [8, 9]. These diverse organ dysfunctions result in varying prognoses depending on the phenotypes [10-12]. Given the complex pathophysiology of sepsis, phenotypic classification is crucial to understand the unique features of each cluster.

Although time-critical therapy is recommended for all suspected sepsis, delayed recognition of sepsis due to atypical presentations, combined

cases, hidden infection, or missed case in ED crowding contributes to serious harm and malpractice claims, ranging from approximately 8% to 21% [13-15]. Therefore, decision support tools to prioritize early risk identification, complementing the clinical decision is required.

This study aimed to classify the clinical phenotypes of patients with suspected sepsis presented to the ED, using cluster analysis and artificial intelligence (AI). AI model learn graphical representations from the outputs of the cluster analysis, and classifies distinct phenotypes or subsets. The clinical characteristics of the identified sepsis subtypes are described follows. As an early intervention in sepsis is beneficial, we additionally analyzed prognosis based on antibiotic administration time through subgroup analysis to provide practical guidance for antibiotic administration guidance based on cluster analysis.

Methods

Study design, setting, population, and definition

This single-center, retrospective, descriptive study analyzed patients with suspected sepsis who visited the ED of Samsung Medical Center, a university-affiliated tertiary care referral hospital located in a metropolitan city with an annual visitor count of over 70,000. The study period was from November 2014 to November 2019.

The Third International Consensus Definitions Task Force (Sepsis-3) defines sepsis as a “life-threatening organ dysfunction caused by a dysregulated host response to infection.” Although this definition

captures the overall concept, we adopted the operational criteria used in Sepsis-3 and the accompanying article by Seymour et al. [16]. According to Seymour et al., a 'suspected infection' was identified when both blood cultures were obtained and antibiotics were administered [16]. According to the Sepsis-3 definition, 'sepsis' is characterized by an acute increase in the Sequential Organ Failure Assessment (SOFA) score (delta SOFA) of ≥ 2 points due to infection [16]. The baseline SOFA score was assumed to be zero unless the patient had preexisting organ dysfunction before the onset of infection. The maximum SOFA score (SOFA_max) was defined as the highest SOFA score within 24 h of arrival in the ED. 'Septic shock' was defined as sepsis accompanied by persistent hypotension, requiring vasopressors to maintain a mean arterial pressure (MAP) ≥ 65 mmHg and a serum lactate level > 2 mmol/L, despite adequate volume resuscitation.

Patients aged 18 years and older with suspected sepsis who presented to the ED were included in this study. Suspected sepsis was defined as cases in which blood cultures were obtained, and antibiotics were administered in the ED [16]. The exclusion criteria were as follows: (1) presentation with cardiac arrest; (2) transfer to or from another hospital in the ED; (3) non-infectious conditions such as trauma or bleeding; (4) limitations to invasive care, including prior do-not-resuscitate orders or terminal malignancy; and (5) incomplete electronic medical records. All data were collected retrospectively from the electronic medical records

(EMR) of patients who underwent blood culture and antibiotic administration in the ED between November 2014 and November 2019.

The following demographic and clinical data were collected from the electronic medical records: initial vital signs upon ED arrival, vital signs at the time of shock occurrence within 6 h of arrival in the ED, and demographic characteristics such as age, sex, comorbidities, suspected infectious focus, use of vasopressors, renal replacement therapy, and mechanical ventilation. Mortality outcomes, including the 28-d and in-hospital mortality rates, were tracked using the electronic clinical data of patients with suspected sepsis. Vital signs refers to systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR), respiratory rate (RR), body temperature (BT). The time to antibiotic therapy was calculated by subtracting the ED triage time from the first broad-spectrum drug administration time. Among the suspected infection cohort, we also identified patients who met the new definition of sepsis, defined as a delta SOFA ≥ 2 points, as shown in Table 1.

The study was approved by the Institutional Review Board of the Samsung Medical Center (IRB no.2025-01-082). The requirement for informed consent was waived because this was a retrospective, observational, and anonymous study.

Primary data analysis

We evaluated the eligibility of 17,733 adult patients who underwent blood culture and antibiotic administration in the ED between November

2014 and November 2019. After excluding patients with cardiac arrest, those transferred to or from another hospital, individuals without an infectious condition, those with limitations in invasive care, and those with inadequate data ($n = 3,331$), 14,402 patients were included in the analysis (Fig.1).

Overall process

This study followed a sequence of steps: (A) unsupervised cluster analysis using machine learning-based clustering was applied using variables such as vital signs and clinical laboratory data to classify patient characteristics; (B) applying artificial intelligence model was applied to imaging data; and (C) analyzing each clinical phenotype of the clustered subtypes was analyzed.

The framework for the overall process analysis is described in Fig. 2.

Detailed steps of computation of the Cluster Index (CI), the unsupervised k-means clustering process, and refinement using a neural network classifier are detailed Supplementary Algorithm S1.

(1) Cohort definition: suspected infection identified by concurrent blood culture and antibiotic administration; exclusions as specified.

(2) Feature preprocessing: 14 features were standardized (z-score); missing values imputed by cohort medians.

(3) Cluster index (CI) computation: a pre-specified linear combination of vitals/labs produced a scalar CI per encounter

(4) Unsupervised k-means clustering: k-means (Euclidean, k-means++ initialization, 100 random restarts, fixed seeds) with k=5 chosen by elbow and silhouette plus clinical interpretability.

(5) Neural Network Refinement model: a neural network trained to predict CI-defined phenotype groups (A-E) from continuous CI-derived image representations for boundary refinement by learning nonlinear patterns.

(6) Clinical analyses: cluster-wise characteristics and outcomes; subgroup analyses for time-to-antibiotics.

Feature Preprocessing and Cluster analysis

This study employed machine learning-based cluster analysis using 14 features: vital signs and clinical laboratory data, including white blood cell count (WBC), lactate (Lac), platelet count (Plt), bilirubin (Bil), C-reactive protein (CRP), creatinine (Cr), absolute neutrophil count (ANC), procalcitonin (PCT), and maximum Sequential Organ Failure Assessment (SOFA_max). The variables incorporated into the distance formulas were selected based on known prognostic factors associated with sepsis. Initial vital signs (vital signs_initial) and laboratory values were used in the data acquisition process, with ANC, Plt, and SOFA_max included as original values because of their suitability for direct use. All continuous variables were standardized using z-score normalization, while categorical variables were one-hot encoded. Missing values were imputed using median values. We then computed a scalar cluster index (CI) as a linear

combination of pre-specified features. The Cluster Index was intentionally designed as a linear combination of multiple standardized clinical variables to reflect aggregate physiological severity. CI coefficients were predefined based on clinical rationale to balance contributions across major physiological domains, rather than estimated from outcome-based regression. Larger coefficients were assigned to variables representing non-redundant organ dysfunction domains to ensure meaningful representation in the composite index. Accordingly, CI stratification does not rely on any single dominant variable but emerges from the combined contribution of multiple components (Supplementary Table S3). K-means clustering was performed using the 14 z-score standardized clinical variables. The Cluster Index (CI) was computed separately and was not included as an input feature for k-means. CI was used solely for visualization and image-based encoding in the neural network refinement model. For clarity, we use CI to denote this score and reserve 'distance' exclusively for the Euclidean distance used by k-means. Vital signs_initial denotes the standardized vector (SBP, DBP, HR, RR, BT) at ED arrival, aggregated as the weighted linear term shown below.

$$\begin{aligned} CI_{\text{initial}} = & \text{vital signs}_{\text{initial}} + WBC \times \frac{1}{4} + \text{Plt} + \text{DBP/SBP} - \text{Bil} \times 5 - \text{sum_of_CRP_Cr} \times \frac{1}{8} \\ & + \text{ANC} - \text{SOFA}_{\text{max}} - \text{PCT} \times \frac{1}{8} \end{aligned}$$

Additionally, the data acquisition process incorporated vital signs at the time of shock onset within 6 h of ED arrival (vital signs_shock) and laboratory values. Numerical values from vital signs and laboratory

values, and the time of shock onset represents baseline of the group, and the baseline offers criteria for clinical decision of emergency patients [17]. The laboratory values used to assess the distance from vital sign shock followed the same trend as those evaluated for the initial vital signs. cluster index in the acquisition environment using shock-related vital signs are presented below.

$$CI_{\text{shock}} = \text{vital signs}_{\text{shock}} + WBC \times \frac{1}{4} + \text{Plt} \times \frac{1}{2} + \text{DBP/SBP} - \text{Bil} \times 10 - \text{sum_of_CRP_Cr} \times \frac{1}{4} \\ + \text{ANC} \times \frac{1}{2} - \text{SOFA_max} - \text{PCT} \times \frac{1}{4}$$

The combined CI for clustering was computed as $CI_{\text{comb}} = CI_{\text{initial}} + CI_{\text{shock}}$ when shock vitals were available, otherwise $CI_{\text{comb}} = CI_{\text{initial}}$. For comparability across encounters, we applied min-max scaling to CI_{comb} within the cohort. The exact coefficients/feature set and an executable pseudo-code are provided in Supplementary Algorithm S1. We assessed the distance values in both the initial data acquisition environment, using initial vital signs, and the second data acquisition environment, based on the vital signs at the time of shock onset. The distances from both environments were normalized to the sum of the two values. A supplementary CI Contribution analysis was conducted to evaluate the relative contribution of individual CI components. Contribution was quantified using standardized magnitude of each weighted CI term. To visualize the clusters, a clustering plot of the distance values of the vital signs and laboratory variables is shown in Fig. 3. Note that CI values are visualized for interpretability and neural network encoding; the distance metric used for k-means clustering was exclusively the Euclidean

distance computed from standardized input variables. When interpreting, one patient may belong to cluster C and the next to cluster A, continuing in this pattern until the last patient. Clusters with vital sign shock data are represented in orange font. The clusters were divided into five groups based on the distance values (long axis). Recent studies have classified patients into three to five clusters as a way to reflect the heterogeneity of sepsis. These phenotypes exhibit significant differences in prognosis, immune responses, and treatment effects, supporting the potential for personalized sepsis care strategies [18, 19]

In this study, k-means clustering was applied using Euclidean distance. The number of clusters ($k = 5$) was selected using the elbow method and the silhouette score, while also reflecting clinical interpretability in accordance with prior phenotype studies. To ensure stability of cluster allocation, k-means++ initialization and 100 random restarts were used with fixed random seeds. Distance-based cluster separation modeling was performed using scikit-learn (version 1.4), enabling full reproducibility of the clustering process.

Applying artificial intelligence model by imaging data

Although k-means deterministically assigns each patient to a single cluster, assignment certainty can be low for borderline cases with similar geometric proximity to multiple centroids. These clinically ambiguous patients are of particular concern in emergency settings because they may still be at high risk of deterioration despite an initial low-risk

numerical label. To support more reliable phenotype assignment in these ambiguous regions, we constructed a continuous Cluster Index (CI) that summarizes multivariate clinical variables. Based on the distribution of CI values, patients were operationally categorized into five ordered phenotype groups (A-E). This design enables refinement of borderline classifications by resolving assignment ambiguity near decision boundaries and enhancing clinical decision support for early escalation. We did not train a de-novo tabular neural classifier as the primary model because the unsupervised k-means step preserves phenotype discovery without outcome labels, and the role of Neural Network is a post-hoc refinement of boundary confidence rather than label creation.

Therefore, we additionally applied an imaging-based neural network as a post hoc refinement layer. Specifically, CI-derived continuous values were encoded as grayscale image representations, and a neural network was trained to predict the CI-defined phenotype group (A-E), which served as operational phenotype labels for refinement. Because the model operates on continuous CI-derived representations rather than discrete cluster identifiers, it can learn nonlinear patterns over the continuous representation space and thereby provide more stable assignments in regions where k-means centroid separation is less distinct. Importantly, this refinement step is intended to improve assignment reliability for borderline cases rather than to redefine the underlying unsupervised phenotype structure or to claim de novo supervised label creation. We did not develop a de novo tabular

supervised classifier as the primary model, because the unsupervised k-means step was retained to preserve phenotype discovery independent of outcome labels, and the role of the neural network was limited to operational refinement of phenotype assignment confidence. A stratified 90/10 split by CI-defined phenotype group (A-E) was used to create a held-out evaluation set; no patient appeared in more than one dataset, preventing data leakage between development and testing.

In this framework, k-means clustering defines the coarse phenotypic structure, while the CI-based neural network provides a nonlinear refinement layer that stabilizes phenotype assignment near decision boundaries without redefining the original unsupervised clusters. The cluster structure remains unchanged, but individual assignments may differ for boundary cases. The imaging encoder preserves subtle continuous differences in CI-related values, enabling the neural network to detect fine-grained changes that may be difficult to capture using hard centroid thresholds alone. Because the refinement model is trained to predict CI-defined phenotypes from continuous CI-derived representations, it provides an operational refinement of assignment reliability near phenotype boundaries rather than merely reproducing k-means centroid-based thresholds. This enhances discrimination by learning nonlinear mappings over continuous CI representations, particularly near regions of overlap. The AI recognized subtle differences in the encoded color corresponding to each distance value. The ability to discriminate between groups can be improved when AI is used to

enhance discrimination by learning nonlinear mappings over continuous CI representations, as it enables the recognition of minimal changes in those values.

To implement the artificial intelligence model for classifying patients into groups based on the variable distance values, each distance value in the acquisition environment was normalized. AI model with color input processing layer learns graphical color information to indicate numerical index of inputs and classify images in ultrasound application [20]. The distance values from the initial and second acquisition environments were normalized to the maximum distance value in the initial acquisition environment, which was 1,841. Considering a 480×640 image as the input for the model, five groups based on the distance values of vital signs (with or without shock) were scaled to a numerical range of 0-240 corresponding to 640 vertical axial pixels. The AI model learns images with the scaled numerical range within each clinical phenotype and requires graphically varying inputs to consistently train the AI model at other domains of clinical phenotypes [21]. The second set of distance values derived from the clustering plot was mapped onto 99 of the 640 vertical axial pixels. For the fifth set of distance values, an additional 99 pixels are incorporated into the model input.

$$\text{input}_{480 \times 640} = \left\{ \left(\frac{\text{distance}}{1841} \right) \times 240 \right\} + (\text{group\#} \times 99)$$

Although the clustering plot visualizes each group based on the distance values, fluctuations in these values can occur, with some distant values appearing near the threshold between clusters. Clusters are

equivalent as learning features, and extraction of feature vectors classifies groups with clinical decisions [22]. Clusters gather the vital signs and laboratory variables into groups, and make AI model learn from the numerically grouped mask [23]. The proposed artificial intelligence model enhances the classification by addressing ambiguous values near the threshold value. The neural network classifier comprised two fully connected hidden layers (64 and 32 units, ReLU activation) and a five-class output. Training used the Adam optimizer (learning rate 0.001), batch size 64, and early stopping with 10-fold cross-validation.

Performance metrics included accuracy, macro-F1 score, and inter-cluster separation indices (silhouette score, Davies-Bouldin index). The neural network learns spatial patterns from image-encoded distance values to distinguish each cluster with improved reliability [24, 25].

Sample images representing each cluster of distance values are provided on Fig. 4. From left to right, clusters A to E are displayed. The distance value increased progressively across clusters A-E. Although k-means assigns all patients to a single cluster, the confidence of assignment may be reduced when geometric proximity to multiple centroids is similar.

Therefore, distance values were encoded into two-dimensional color-based images to enable representation learning and refinement of borderline cases. The neural network was trained to predict CI-defined phenotype groups (A-E) from continuous CI-derived image representations, which served as operational labels for refinement, and to capture nonlinear boundaries between clusters. The proposed artificial

intelligence model was trained with using 12,962 images randomly selected for training. The remaining set of 1440 images from the testing set achieving 98% accuracy. Reported accuracy reflects internal consistency of the refinement model and should not be interpreted as ground-truth classification performance in unsupervised phenotyping. The held-out set was stratified by cluster and fixed via random seed (42), 10-fold cross-validation was used for internal model selection prior to final testing. The model consisted of 59 convolution layers and six max pooling layers, with a cross-entropy loss function and Adam optimizer, utilizing a learning rate of 0.0004. The input images, which were normalized using a PyTorch transform class, had a resolution of 256×256 pixels. The model was trained using a batch size of 32 for 10 epochs. After 10 epochs, the training loss decreased, finalizing the training for classifying distant values based on vital signs in patients with and without shock. The plot of the training loss is shown in Fig. 5. All models were developed in Python 3.10 using scikit-learn 1.4 for clustering and PyTorch 2.1 for neural network training. Random seeds were fixed to ensure reproducibility. Full hyperparameter details are provided in Supplementary Table S2 and Supplementary Algorithm S1.

Analyzing each clinical phenotypes of the clustered subtypes

Standard descriptive statistics were applied to all variables including baseline demographics and outcomes. Categorical variables are presented as frequencies with percentages and were compared using the

chi-squared test. Continuous variables were presented as medians with interquartile ranges (IQRs) and compared using the Wilcoxon rank-sum test. Statistical significance was set at $p < 0.05$ (significant). All the analyses were performed using R version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria) and STATA (version 17.0; STATA Corporation, College Station, TX, USA).

Results

Baseline characteristics

Table 1 presents the baseline characteristics of the five clusters, including their vital signs and laboratory test results. Among the 14,402 patients, 7,838 (54.4%) were male and 5,589 (44.2%) had sepsis, defined by an acute increase in SOFA score of 2 or higher (Delta SOFA score ≥ 2). Cluster B included the highest proportion of sepsis patients (71.1%), followed by Clusters A (46.9%) and Cluster C (42.9%). Cluster B exhibited the highest SOFA_max score (5.0, 95% CI: 3.0-8.0), lactate level (2.16, 95% CI: 1.41-3.46), and CRP (8.14, 95% CI: 3.05-16.92) followed by cluster A. Elevated levels of lactate and CRP, along with a higher SOFA score, suggest a more severe form of sepsis and potentially poorer outcomes, as these parameters are valuable tools in assessing the severity, predicting prognosis, and guiding treatment decisions [26]. Therefore, if a patient is assigned to Cluster B or A, the physician may anticipate severe sepsis or a potentially poor outcome.

Additionally, Cluster B had the highest proportion of patients with chronic cardiac disease or metastatic cancer. The number of missing values is detailed in Supplementary Table S1. After excluding inadequate data, including missing data, the analysis of patients with septic shock did not incorporate vasopressor administration time, duration, or missing vital signs_shock data. Since only the vital signs of patients who experienced hypotension within 6 hours from the ED were additionally collected, variables labeled as vital signs_shock existed only in these subgroups.

Table 2 presents the clinical outcomes of the five clusters. The overall proportion of patients with septic shock was 1,118 (7.8%), and the vasopressor requirement was 1695 (11.8%). The overall in-hospital mortality rate was 676 (4.7%) in the entire cohort, ranging from 3.3% to 8.8% across clusters. The 28-d mortality rate of the entire cohort was 989 (6.9%), and 280 (8.4%), 121 (12.3%), 233 (6.3%), 236 (5.7%), and 119 (5.4%) across the clusters, respectively. Cluster B exhibited the poorest outcomes, including the highest rates of septic shock (302; 30.8%), critical care use (described below), in-hospital mortality (86; 8.8%), and 28-d mortality (163; 4.4%). Cluster B also had the highest vasopressor requirement (406; 41.4%), mechanical ventilator use (70; 7.1%), and intensive care unit admission (181; 18.5%), followed by Cluster A. Cluster E had the lowest rates across these metrics. Physicians may predict poor progression if a patient is allocated to Cluster B using

the clustering algorithm employed in this study. It takes at least 2-3 h to obtain blood test results after ED arrival, with additional time required to calculate the SOFA score. Therefore, before the manual calculation of the SOFA score, AI-based clustering results can be used to predict sepsis severity and contribute to a more comprehensive assessment. Notably, Cluster B, which exhibited the highest severity of illness, initiated vasopressors and antibiotics the earliest and maintained their use the longest. The median time to initiate vasopressor therapy was 6.3 h (95% CI: 2.3-33.1), with Cluster B starting vasopressor the earliest time at 3.3 h after ED arrival (95% CI: 1.5-6.3), followed by cluster A at 4.4 h (95% CI: 1.9-13.2). Similarly, Cluster B initiated antibiotics the earliest at 1.5 h after ED arrival (95% CI: 0.5-3.1), followed by cluster A, while cluster B had the longest duration of antibiotics use at 7.3 h (95% CI: 0.4-16.4). Early initiation of antibiotics and vasopressors in Clusters B and A reflected the high severity of these groups, with the fastest physician recognition of severity among all clusters. However, given the guidelines recommending the administration of antibiotics and vasopressors within 1 h of septic shock recognition [27, 28], the median administration time in our study was delayed. Therefore, before the manual calculation of the SOFA score to diagnose sepsis, physicians may intervene promptly using the AI-based clustering method if the patient is allocated to cluster B or A. Thus, the AI-based clustering method may support clinical judgment before the manual calculation of the SOFA score.

In the subgroup analysis, we stratified the time to antibiotic administration using a cut-off value of <3 h and ≥ 3 h to guide timing based on cluster allocation. Multivariable logistic regression revealed that early antibiotics administration within 3 h was significantly associated with decreased 28-d mortality in cluster A (aOR 0.73; 95% CI, 0.54-0.98, $p = 0.037$) and E (aOR 0.59; 95% CI, 0.39-0.89, $p = 0.013$) (Table 3).

A spline curve illustrating the log odds ratio of time to antibiotic administration versus 28-d mortality for each cluster is shown in Fig. 6. Multivariate analysis adjusted for age, SOFA_max, lactate levels, and respiratory infection. The spline curve demonstrated an increasing trend in the 28-d mortality as the time to antibiotic administration increased (Fig. 6).

Discussion

This study employed unsupervised clustering using artificial intelligence on an ED database to categorize patients with suspected infections. The clinical phenotypes of the five identified clusters were correlated with clinical outcomes. In particular, artificial intelligence-based clustering effectively identified high-severity groups characterized by higher rates of septic shock, critical care needs, in-hospital mortality, and 28-d mortality. Additionally, the relationship between antibiotic administration and 28-d mortality varied across clusters, suggesting that the effect of treatment may differ depending on the patient's phenotype.

The heterogeneous nature of sepsis has led to an increased interest in investigating its phenotypes [7, 8, 18, 29]. Recent studies have shown sepsis phenotypes with clinical relevance are often limited to specific populations. For example, some studies have focused on ICU settings [10, 30], while others, such as those by Guilamet et al., identified four clusters within homogenous infectious diseases [31] and Lai et al. conducted clustering to identify subphenotypes in patients with dialysis-requiring sepsis-associated acute kidney injury [32]. However, few studies have explored clinical sepsis phenotypes in heterogeneous populations, which present a major challenge in grouping patients [11]. Additionally, only a limited number of studies have categorized heterogeneous groups based on hemodynamic phenotyping using vital signs [7, 33, 34]. Given that blood pressure is routinely measured in the ED, incorporating systolic blood pressure (SBP) and diastolic blood pressure (DBP) into cluster analysis may offer a more precise and timely identification of phenotypes [10]. This study addressed the dynamic phenotypic changes in patients with septic shock by incorporating both initial vital signs and those recorded at the time of shock. The strengths of this descriptive study include its use of both shock time and baseline vital signs, artificial intelligence-based cluster analysis, a large dataset, and applicability to a diverse ED population, including low-risk patients.

Our result showed that cluster B had the highest rates of metastatic cancer, septic shock, critical care usage, in-hospital mortality, and 28-d mortality. Additionally, we found that early antibiotic administration

within 3 h significantly reduced the 28-d mortality in Clusters A and E. Notably, the cluster with the most severe patients started vasopressors and antibiotics earlier than the other clusters, while also maintaining the longest duration of treatment. Previous studies have highlighted the heterogeneity in the presentation and timing of antibiotic initiation for suspected septic shock [35] and the relationship between the duration of the antibiotic and clinical outcomes remains contentious [31, 36]. However, several studies have demonstrated that the early initiation of antibiotics benefits sepsis patients [37]. Our findings also align with recent research indicating that cancer patients are at a significantly higher risk for sepsis and associated mortality, with hospital mortality being higher in cancer patients than in those without cancer (hazard ratio, 1.25; 95% CI, 1.24 -1.26) [38]. The overall 28-d mortality rate in our cohort was 6.9%, with rates ranging from 5.4% to 12.3% across clusters, consistent with previous reports [39].

Many studies have established scoring systems for risk stratification in the ED [6]. However, traditional scoring systems have limitations when applied to patients who meet the Sepsis-3 criteria (SOFA score ≥ 2). Sepsis-3 diagnostic criteria narrow the sepsis population, potentially leading to false negatives and resulting in delayed or missed diagnoses [40]. In the difficult cases in recognition of sepsis including hidden infection, atypical presentations, combined cases, or missed case in ED crowding, previous studies still emphasizes the importance of sepsis alert systems to prevent the mistakes [13-15, 41].

Therefore, to prevent the mistakes or missed case, clustering analysis combined with AI could be optimized for the sepsis-3 criteria helping to uncover the hidden biological elements of sepsis. Given the complex biology of sepsis, the potential role of metabolic biomarkers, and existing knowledge gaps among healthcare workers, understanding the clinical phenotypes of sepsis through clustering and AI is essential [42, 43]. AI-driven clustering may indicate that multimodal factors are the key predictors of sepsis.

Treatment tailoring has become more prevalent in recent years as the prediction of outcomes using simple scoring systems alone is increasingly limited. The primary goal of identifying clinical phenotypes and host responses related to clinical outcomes is to provide additional guidance for healthcare providers working in diverse and complex environments. Given the high mortality rate associated with sepsis, a personalized approach to treatment is essential, as different phenotypes may respond differently to various treatment strategies [10, 12]. According to Aaron et al., identifying these phenotypes enables personalized care [11]. Understanding these phenotypes using our method would be valuable for identifying sepsis and managing patients in ways that account for individual differences in clinical practice.

Clinically, the result of cluster allocation using an AI-based clustering algorithm can be applied in many ways. The result of phenotyping can be used as the decision support tool to predict sepsis severity, augment recognition and proactive intervention. If the patient is categorized into

Cluster A or B, physicians predict a higher likelihood of poor prognosis, may initiate blood cultures, administer antibiotics, and begin close monitoring even before manual calculation of the SOFA score. Since clusters enable the process evaluation including time to antibiotics subanalyses, physicians can understand the phenotypes of the patient and discuss the disposition of the patient with other department based on the results in Table 2. Also, phenotyping result provides a backup safety net for complex and hidden sepsis, and addresses missed or late identification that contributes to patient harm and medico-legal exposure[13].

The AI-based clustering model presented in this study should be regarded as a supplementary tool for clinical decision-making, rather than a definitive diagnostic guide. Importantly, we did not use supervised outcome labels to drive the initial clustering. This preserves unbiased phenotype discovery while allowing the neural network to function only as a refinement tool to improve cluster boundary reliability. The refinement step is intended to clarify k-means boundary assignments using continuous cluster index information, rather than to claim superior classification accuracy, which is not directly definable in unsupervised phenotype discovery.

This approach, which integrates multiple clinical variables, may assist in identifying complex sepsis phenotypes and supporting timely therapeutic decisions, particularly in cases with heterogeneous clinical features or when physicians face challenges interpreting the trajectory of

sepsis due to disease severity. From a clinical perspective, the primary value of this framework lies not in demonstrating superior classification accuracy, but in providing an early decision-support safety net in complex emergency department environments. Missed or delayed recognition of sepsis due to atypical presentations, hidden infection, or ED crowding remains a preventable harm [13, 44, 45]. By offering phenotype classification before formal SOFA calculation, AI-based clustering may support earlier escalation, closer monitoring, and timely antibiotic administration in patients at risk of deterioration [46, 47]. In such contexts, AI-driven clustering output may serve as an additional reference to enhance physician judgment, including guidance on antibiotic administration timing. Therefore, the clustering results may provide additional guidance for physicians by characterizing the clinical phenotypes of each of the five clusters.

This study has several limitations. First, the data was obtained from a single center, meaning that the groupings may not apply to other hospitals or settings. Second, cluster analysis requires a larger volume of data to capture the underlying risk in the acquisition environment effectively. Third, the black-box nature of AI algorithms limits interpretability, warranting caution in their clinical applications. As emphasized in prior studies on explainable AI in emergency medicine [48], such models should be used in conjunction with established clinical assessments to ensure safe and effective implementation. Fourth, excluding patients with false-positive suspected infection may limit

applicability to real-world clinical environments. Fifth, causal inference cannot be made due to the retrospective observational design.”Finally, after implementing our method in real clinical practice, further feedback through sepsis surveillance will be essential for refinement and validation.

Despite these limitations, this study demonstrated the value of using cluster analysis and artificial intelligence to classify clinical phenotypes among patients with suspected sepsis in the ED. When the model is expanded to a multicenter level with external validation, phenotype identification may be instrumental in risk stratification and guiding future clinical management strategies.

Conclusion

The artificial intelligence model generated five distinct clusters of patients with suspected sepsis who visited the ED. These phenotypes exhibited different clinical characteristics and outcome patterns. While these findings suggest potential value for future risk stratification, prospective external validation is needed to determine how phenotype information may assist clinical decision.

Abbreviations

AI	artificial intelligence
SOFA	Sequential Organ Failure Assessment
SOFA_max	maximum SOFA score
MAP	mean arterial pressure
WBC	white blood cell count
Lac	lactate
Plt	platelet count
Bil	bilirubin
CRP	C-reactive protein
Cr	creatinine
ANC	absolute neutrophil count
PCT	procalcitonin
IQRs	interquartile ranges

Funding

This research was supported by three funders, Grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : RS-2023-KH134607) to Tae Gun Shin.; National Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the by the Ministry of Education (No. 2020R1A6A1A03047902) to Chul Hong Kim.; National Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (RS-2022-NR073656) to Jong Eun

Park. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

JD, JRP, and TGS conceived and designed the study; SJM, JWC, GTL, and SYH supervised data collection, analyzed the data, and interpreted the results; JD, JRP, JEP, SYH, CK, and TGS drafted the manuscript; all authors discussed the results and contributed to the final manuscript.

Clinical trial number: not applicable.

Competing interests

The authors declare no competing interests.

Acknowledgements

Not applicable

Consent for publication

Not applicable

Appendix A.

Supplementary Table 1. Number of missing measurements

Data availability

The datasets used and/or analyzed during the current study are available

from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of the Samsung Medical Center (IRB no.2025-01-082). The requirement for informed consent was waived because this was a retrospective, observational, and anonymous study. All procedures were performed in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

ARTICLE IN PRESS

References

1. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM *et al*: **The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)**. *Jama* 2016, **315**(8):801-810.
2. Shin TG, Ko E, Han SH, Kim T, Choi DH: **Epidemiology of sepsis in emergency departments: insights from the National Emergency Department Information System (NEDIS) database in Korea, 2018-2022**. *Clin Exp Emerg Med* 2025, **12**(3):185-187.
3. Bauer M, Gerlach H, Vogelmann T, Preissing F, Stiefel J, Adam D: **Mortality in sepsis and septic shock in Europe, North America and Australia between 2009 and 2019- results from a systematic review and meta-analysis**. *Crit Care* 2020, **24**(1):239.
4. Rhee C, Filbin MR: **Can Procalcitonin and Other Biomarkers Help Rapidly Identify Sepsis Among Undifferentiated High-Risk Patients in the Emergency Department?** *Crit Care Med* 2024, **52**(6):979-982.
5. Brink A, Alsma J, Verdonschot R, Rood PPM, Zietse R, Lingsma HF, Schuit SCE: **Predicting mortality in patients with suspected sepsis at the Emergency Department; A retrospective cohort study comparing qSOFA, SIRS and National Early Warning Score**. *PLoS One* 2019, **14**(1):e0211133.
6. 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, **153**(3):646-655.
7. Bhavani SV, Semler M, Qian ET, Verhoef PA, Robichaux C, Churpek MM, Coopersmith CM: **Development and validation of novel sepsis subphenotypes using trajectories of vital signs.** *Intensive Care Med* 2022, **48**(11):1582-1592.
8. Maslove DM, Tang B, Shankar-Hari M, Lawler PR, Angus DC, Baillie JK, Baron RM, Bauer M, Buchman TG, Calfee CS *et al*: **Redefining critical illness.** *Nat Med* 2022, **28**(6):1141-1148.
9. Shin TG: **Assessment of organ failure in sepsis patients in the emergency department: clinical evaluation, Sequential Organ Failure Assessment (SOFA) score, and future perspectives.** *Clin Exp Emerg Med* 2024, **11**(4):327-330.
10. Papathanakos G, Andrianopoulos I, Xenikakis M, Papathanasiou A, Koulenti D, Blot S, Koulouras V: **Clinical Sepsis Phenotypes in Critically Ill Patients.** *Microorganisms* 2023, **11**(9):2165.
11. Boussina A, Wardi G, Shashikumar SP, Malhotra A, Zheng K, Nemati S: **Representation Learning and Spectral Clustering for the Development and External Validation of Dynamic Sepsis Phenotypes: Observational Cohort Study.** *J Med Internet Res* 2023, **25**:e45614.

12. Shald EA, Erdman MJ, Ferreira JA: **Impact of Clinical Sepsis Phenotypes on Mortality and Fluid Status in Critically Ill Patients.** *Shock* 2022, **57**(1):57-62.
13. Neilson HK, Fortier JH, Finestone PJ, Ogilby CM, Liu R, Bridges EJ, Garber GE: **Diagnostic Delays in Sepsis: Lessons Learned From a Retrospective Study of Canadian Medico-Legal Claims.** *Crit Care Explor* 2023, **5**(2):e0841.
14. Warmerdam M, Stolwijk F, Boogert A, Sharma M, Tetteroo L, Lucke J, Mooijaart S, Ansems A, Esteve Cuevas L, Rijpsma D *et al*: **Initial disease severity and quality of care of emergency department sepsis patients who are older or younger than 70 years of age.** *PLoS One* 2017, **12**(9):e0185214.
15. Newman-Toker DE, Wang Z, Zhu Y, Nassery N, Saber Tehrani AS, Schaffer AC, Yu-Moe CW, Clemens GD, Fanai M, Siegal D: **Rate of diagnostic errors and serious misdiagnosis-related harms for major vascular events, infections, and cancers: toward a national incidence estimate using the "Big Three".** *Diagnosis (Berl)* 2021, **8**(1):67-84.
16. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, Rubenfeld G, Kahn JM, Shankar-Hari M, Singer M *et al*: **Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3).** *JAMA* 2016, **315**(8):762-774.

17. Park JR, Chung SP, Hwang SY, Shin TG, Park JE: **Myocardial infarction evaluation from stopping time decision toward interoperable algorithmic states in reinforcement learning.** *BMC Med Inform Decis Mak* 2020, **20**(1):99.
18. Seymour CW, Kennedy JN, Wang S, Chang CH, Elliott CF, Xu Z, Berry S, Clermont G, Cooper G, Gomez H *et al*: **Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis.** *Jama* 2019, **321**(20):2003-2017.
19. Sweeney TE, Azad TD, Donato M, Haynes WA, Perumal TM, Henao R, Bermejo-Martin JF, Almansa R, Tamayo E, Howrylak JA *et al*: **Unsupervised Analysis of Transcriptomics in Bacterial Sepsis Across Multiple Datasets Reveals Three Robust Clusters.** *Crit Care Med* 2018, **46**(6):915-925.
20. Misra S, Yoon C, Kim KJ, Managuli R, Barr RG, Baek J, Kim C: **Deep learning-based multimodal fusion network for segmentation and classification of breast cancers using B-mode and elastography ultrasound images.** *Bioeng Transl Med* 2023, **8**(6):e10480.
21. Park E, Misra S, Hwang DG, Yoon C, Ahn J, Kim D, Jang J, Kim C: **Unsupervised inter-domain transformation for virtually stained high-resolution mid-infrared photoacoustic microscopy using explainable deep learning.** *Nat Commun* 2024, **15**(1):10892.

22. Yoon C, Park E, Misra S, Kim JY, Baik JW, Kim KG, Jung CK, Kim C: **Deep learning-based virtual staining, segmentation, and classification in label-free photoacoustic histology of human specimens.** *Light Sci Appl* 2024, **13**(1):226.
23. Cho Y, Misra S, Managuli R, Barr RG, Lee J, Kim C: **Attention-based Fusion Network for Breast Cancer Segmentation and Classification Using Multi-modal Ultrasound Images.** *Ultrasound Med Biol* 2025, **51**(3):568-577.
24. Choi S, Yang J, Lee SY, Kim J, Lee J, Kim WJ, Lee S, Kim C: **Deep Learning Enhances Multiparametric Dynamic Volumetric Photoacoustic Computed Tomography In Vivo (DL-PACT).** *Adv Sci (Weinh)* 2022, **10**(1):e2202089.
25. Jeong H, Lim H, Yoon C, Won J, Lee GY, de la Rosa E, Kirschke JS, Kim B, Kim N, Kim C: **Robust Ensemble of Two Different Multimodal Approaches to Segment 3D Ischemic Stroke Segmentation Using Brain Tumor Representation Among Multiple Center Datasets.** *J Imaging Inform Med* 2024, **37**(5):2375-2389.
26. Turgman O, Schinkel M, Wiersinga WJ: **Host Response Biomarkers for Sepsis in the Emergency Room.** *Crit Care* 2023, **27**(1):97.
27. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, Machado FR, McIntyre L, Ostermann M, Prescott HC *et al*: **Surviving Sepsis Campaign: International Guidelines for**

- Management of Sepsis and Septic Shock 2021.** *Crit Care Med* 2021, **49**(11):e1063-e1143.
28. Park C, Ku NS, Park DW, Park JH, Ha TS, Kim DW, Park SY, Chang Y, Jo KW, Baek MS *et al*: **Early management of adult sepsis and septic shock: Korean clinical practice guidelines.** *Acute Crit Care* 2024, **39**(4):445-472.
 29. Pandey S: **Advances in metabolomics in critically ill patients with sepsis and septic shock.** *Clin Exp Emerg Med* 2025, **12**(1):4-15.
 30. Knox DB, Lanspa MJ, Kuttler KG, Brewer SC, Brown SM: **Phenotypic clusters within sepsis-associated multiple organ dysfunction syndrome.** *Intensive Care Med* 2015, **41**(5):814-822.
 31. Guilamet MCV, Bernauer M, Micek ST, Kollef MH: **Cluster analysis to define distinct clinical phenotypes among septic patients with bloodstream infections.** *Medicine (Baltimore)* 2019, **98**(16):e15276.
 32. Lai CF, Liu JH, Tseng LJ, Tsao CH, Chou NK, Lin SL, Chen YM, Wu VC: **Unsupervised clustering identifies sub-phenotypes and reveals novel outcome predictors in patients with dialysis-requiring sepsis-associated acute kidney injury.** *Ann Med* 2023, **55**(1):2197290.
 33. Zhu JL, Yuan SQ, Huang T, Zhang LM, Xu XM, Yin HY, Wei JR, Lyu J: **Influence of systolic blood pressure trajectory on in-hospital mortality in patients with sepsis.** *BMC Infect Dis* 2023, **23**(1):90.

34. Daulasim A, Vieillard-Baron A, Geri G: **Hemodynamic clinical phenotyping in septic shock.** *Curr Opin Crit Care* 2021, **27**(3):290-297.
35. Prasad PA, Esmaili AM, Oreper S, Beagle AJ, Hubbard C, Raffel KE, Abe-Jones Y, Fang MC, Liu KD, Matthay MA *et al*: **Timing of antibiotic treatment identifies distinct clinical presentations among patients presenting with suspected septic shock.** *J Am Coll Emerg Physicians Open* 2024, **5**(2):e13149.
36. Bisarya R, Song X, Salle J, Liu M, Patel A, Simpson SQ: **Antibiotic Timing and Progression to Septic Shock Among Patients in the ED With Suspected Infection.** *Chest* 2022, **161**(1):112-120.
37. Ruddel H, Thomas-Ruddel DO, Reinhart K, Bach F, Gerlach H, Lindner M, Marshall JC, Simon P, Weiss M, Bloos F *et al*: **Adverse effects of delayed antimicrobial treatment and surgical source control in adults with sepsis: results of a planned secondary analysis of a cluster-randomized controlled trial.** *Crit Care* 2022, **26**(1):51.
38. Liu MA, Bakow BR, Hsu TC, Chen JY, Su KY, Asiedu EK, Hsu WT, Lee CC: **Temporal Trends in Sepsis Incidence and Mortality in Patients With Cancer in the US Population.** *Am J Crit Care* 2021, **30**(4):e71-e79.
39. Freund Y, Lemachatti N, Krastinova E, Van Laer M, Claessens YE, Avondo A, Occelli C, Feral-Pierssens AL, Truchot J, Ortega M *et al*: **Prognostic Accuracy of Sepsis-3 Criteria for In-Hospital**

Mortality Among Patients With Suspected Infection

Presenting to the Emergency Department. *JAMA* 2017, **317**(3):301-308.

40. Fang X, Wang Z, Yang J, Cai H, Yao Z, Li K, Fang Q: **Clinical Evaluation of Sepsis-1 and Sepsis-3 in the ICU.** *Chest* 2018, **153**(5):1169-1176.
41. Kim HJ, Ko RE, Lim SY, Park S, Suh GY, Lee YJ: **Sepsis Alert Systems, Mortality, and Adherence in Emergency Departments: A Systematic Review and Meta-Analysis.** *JAMA Netw Open* 2024, **7**(7):e2422823.
42. She H, Du Y, Du Y, Tan L, Yang S, Luo X, Li Q, Xiang X, Lu H, Hu Y *et al*: **Metabolomics and machine learning approaches for diagnostic and prognostic biomarkers screening in sepsis.** *BMC Anesthesiol* 2023, **23**(1):367.
43. Adegbite BR, Edoa JR, Rylance J, Jacob ST, Kawale P, Adegnika AA, Grobusch MP: **Knowledge of health workers relating to sepsis awareness and management in Lambaréné, Gabon.** *Acta Trop* 2021, **219**:105914.
44. Horberg MA, Nassery N, Rubenstein KB, Certa JM, Shamim EA, Rothman R, Wang Z, Hassoon A, Townsend JL, Galiatsatos P *et al*: **Rate of sepsis hospitalizations after misdiagnosis in adult emergency department patients: a look-forward analysis with administrative claims data using Symptom-Disease Pair**

- Analysis of Diagnostic Error (SPADE) methodology in an integrated health system.** *Diagnosis (Berl)* 2021, **8**(4):479-488.
45. Schols LA, Maranus ME, Rood PPM, Zwaan L: **Diagnostic Discrepancies in the Emergency Department: A Retrospective Study.** *J Patient Saf* 2024, **20**(6):420-425.
46. Khalili M, Enayati M, Patel S, Huschka T, Cabrera D, Parker SJ, Pasupathy K, Mahajan P, Bellolio F: **Identifying diagnostic errors in the emergency department using trigger-based strategies.** *BMJ Open Qual* 2025, **14**(3).
47. Hussain F, Cooper A, Carson-Stevens A, Donaldson L, Hibbert P, Hughes T, Edwards A: **Diagnostic error in the emergency department: learning from national patient safety incident report analysis.** *BMC Emerg Med* 2019, **19**(1):77.
48. Okada Y, Ning Y, Ong MEH: **Explainable artificial intelligence in emergency medicine: an overview.** *Clin Exp Emerg Med* 2023, **10**(4):354-362.

Figure legends

Fig.1 Flowchart.

Fig.2 The framework for the overall process analysis.

Fig. 3 Clustering plot from distance values of vital signs and laboratory values. CI values are continuous; apparent banding may be accentuated by downsampling for print-scale visualization.**Fig. 4** Inputs for proposed artificial intelligence model from distance values.

Fig. 5 Training process of model for classifying distance values of vital signs and laboratory values.

Fig. 6 Spline curve of time to antibiotic administration and 28-d mortality stratified by clusters: (A) all clusters shown in a single graph; (B) each cluster presented in individual graphs.

Table 1 Baseline characteristics of the five clusters

	Overall	A	B	C	D	E	p-value
n	14,402	3,340	980	3,727	4,146	2,209	
Sex	7,838 (54.4)	1,842 (55.1)	511 (52.1)	2,024 (54.3)	2,246 (54.2)	1,215 (55.0)	0.53
Age	63.4 (52.8-73.5)	63.7 (53.3-73.6)	65.8 (56.2-75.1)	63.2 (52.6-73.4)	62.7 (52.1-72.9)	63.5 (52.2-73.4)	<0.001
SBP_0 (mmHg)	119.0 (104.0-136.0)	117.0 (101.0-135.0)	97.0 (85.0-118.0)	120.0 (105.0-137.0)	122.0 (108.0-138.0)	122.0 (109.0-138.0)	<0.001
DBP_0 (mmHg)	71.0 (61.0-81.0)	69.0 (59.0-80.0)	59.0 (52.0-69.0)	72.0 (63.0-82.0)	72.0 (64.0-82.0)	73.0 (64.0-82.0)	<0.001
HR_0 (/min)	100.0 (87.0-115.0)	101.0 (88.0-116.0)	106.0 (90.0-122.0)	100.0 (87.0-115.0)	99.0 (86.0-113.0)	100.0 (87.0-114.0)	<0.001
RR_0 (/min)	18.0 (18.0-20.0)	18.0 (18.0-20.0)	18.0 (18.0-20.0)	18.0 (18.0-20.0)	18.0 (18.0-20.0)	18.0 (18.0-20.0)	<0.001
BT_0	37.6 (36.9-38.3)	37.6 (36.9-38.4)	37.7 (36.8-38.5)	37.6 (36.9-38.3)	37.6 (36.9-38.3)	37.5 (36.9-38.3)	0.31
SaO2	97.0 (95.0-99.0)	97.0 (95.0-99.0)	97.0 (95.0-98.0)	97.0 (95.0-99.0)	98.0 (96.0-99.0)	97.0 (95.0-99.0)	<0.001
DSI_0	1.41 (1.17-1.68)	1.45 (1.20-1.75)	1.73 (1.41-2.09)	1.39 (1.16-1.65)	1.37 (1.15-1.60)	1.37 (1.14-1.61)	<0.001
SSI_0	0.84 (0.68-1.01)	0.86 (0.70-1.04)	1.03 (0.84-1.28)	0.82 (0.67-1.0)	0.81 (0.66-0.96)	0.81 (0.67-0.97)	<0.001
SBP_s (mmHg)	83.0 (78.0-87.0)	83.0 (78.0-87.0)	83.0 (78.0-87.0)	83.0 (78.0-87.0)	85.0 (75.0-87.0)	80.0 (71.50-86.50)	0.89
DBP_s (mmHg)	51.0 (46.0-56.0)	52.0 (46.0-56.0)	51.0 (46.0-56.0)	51.0 (46.0-56.0)	52.0 (46.0-56.0)	48.0 (45.0-52.50)	0.69
HR_s (/min)	101.0 (86.0-119.0)	101.0 (85.25-118.0)	102.0 (86.0-120.0)	101.0 (85.50-118.0)	92.0 (83.0-114.0)	111.0 (97.0-116.0)	0.37
RR_s (/min)	20.0 (18.0-22.0)	20.0 (17.0-22.0)	19.50 (18.0-22.0)	19.0 (18.0-22.0)	20.0 (18.0-23.0)	18.0 (16.0-24.0)	0.87
BT_s	37.40 (36.8-38.2)	37.40 (36.8-38.2)	37.4 (36.7-38.2)	37.4 (36.8-38.2)	37.7 (37.1-38.6)	37.4 (36.7-37.7)	0.33
WBC (10³/μL)	9.19 (5.6-13.2)	9.34 (5.5-13.4)	9.13 (4.6-13.5)	9.15 (5.7-13.3)	9.2 (5.8-13.0)	9.0 (5.5-12.8)	0.22
Platelet (10³/μL)	198.0 (122.0-275.0)	194.0 (119.0-273.50)	169.5 (90.0-258.8)	203.0 (126.0-277.0)	202.0 (126.8-275.0)	202.0 (126.0-280.0)	<0.001
ANC	7.1 (3.84-10.77)	7.3 (3.83-10.97)	7.1 (3.2-11.4)	7.0 (4.0-10.8)	7.0 (4.0-10.6)	6.9 (3.7-10.5)	0.09
Albumin (mg/dL)	3.80 (3.30-4.20)	3.70 (3.20-4.20)	3.40 (2.90-3.90)	3.80 (3.30-4.20)	3.80 (3.30-4.20)	3.80 (3.30-4.20)	<0.001

Bilirubin (mg/dL)	0.70 (0.40-1.20)	0.70 (0.40-1.20)	0.80 (0.50-1.40)	0.70 (0.40-1.20)	0.70 (0.40-1.20)	0.70 (0.40-1.20)	0.002
Creatinine (mg/dL)	0.85 (0.66-1.16)	0.87 (0.67-1.19)	0.95 (0.69-1.52)	0.85 (0.66-1.16)	0.81 (0.65-1.11)	0.84 (0.66-1.10)	<0.001
CRP (mg/dL)	6.66 (2.42-13.55)	7.05 (2.58-13.80)	8.14 (3.05-16.92)	6.66 (2.41-13.42)	6.16 (2.28-12.64)	6.58 (2.26-13.53)	<0.001
Procalcitonin (ng/mL)	0.21 (0.07-1.05)	0.24 (0.07-1.52)	0.77 (0.14-5.83)	0.20 (0.06-0.94)	0.17 (0.06-0.69)	0.19 (0.06-0.67)	<0.001
Lactate (mmol/L)	1.58 (1.16-2.30)	1.64 (1.18-2.44)	2.16 (1.41-3.46)	1.56 (1.14-2.27)	1.51 (1.12-2.11)	1.52 (1.15-2.08)	<0.001
Infection focus							0.35
Respiratory	3,520 (24.4)	834 (25.0)	238 (24.3)	909 (24.4)	985 (23.8)	554 (25.1)	
Gastrointestinal	1,782 (12.4)	446 (13.4)	107 (10.9)	480 (12.9)	502 (12.1)	247 (11.2)	
Urinary	3,762 (26.1)	835 (25.0)	271 (27.7)	981 (26.3)	110 (26.5)	575 (26.0)	
Others	5,338 (37.1)	1,225 (36.7)	364 (37.1)	1,357 (36.4)	1,559 (37.6)	833 (37.7)	
Comorbidities							
Chronic Cardiac disease	1,760 (12.2)	433 (13.0)	142 (14.5)	464 (12.4)	447 (10.8)	274 (12.4)	0.005
Chronic Lung disease	1,151 (8.0)	272 (8.1)	69 (7.0)	292 (7.8)	338 (8.2)	180 (8.1)	0.80
Metastatic Cancer	2,387 (16.6)	606 (18.1)	189 (19.3)	558 (15.0)	630 (15.2)	404 (18.3)	<0.001
Chronic Liver disease	1,105 (7.7)	268 (8.0)	68 (6.9)	276 (7.4)	309 (7.5)	184 (8.3)	0.50
Chronic Renal disease	1,392 (9.7)	319 (9.6)	85 (8.7)	372 (10.0)	407 (9.8)	209 (9.5)	0.77
Delta SOFA score$\geq$2	5,589 (44.2)	1,379 (46.9)	594 (71.1)	1410 (42.9)	1,425 (38.9)	781 (40.7)	<0.001
SOFA_max	2.0 (1.0-4.0)	3.0 (1.0-5.0)	5.0 (3.0-8.0)	2.0 (1.0-4.0)	2.0 (1.0-4.0)	2.0 (1.0-4.0)	<0.001

The data are presented as median (IQR) for continuous variables and as numbers (%) for categorical variables. *IQR* interquartile range, *V/S_0* initial vital signs, *V/S_s* vital signs at the time of shock, *SBP* systolic blood pressure, *DBP* diastolic blood pressure, *HR* heart rate, *RR* respiratory rate, *BT* body temperature, *SOFA* Sequential Organ Failure Assessment, *WBC* white blood cell count, *ANC* absolute neutrophil count, *CRP* C-reactive protein, *SOFA_max* maximal SOFA score within h after ED arrival

Table 2. Clinical outcomes and sepsis interventions of the five clusters

	Overall	A	B	C	D	E	p-value
28-d mortality	989 (6.9)	280 (8.4)	121 (12.3)	233 (6.3)	236 (5.7)	119 (5.4)	<0.001
In-hospital mortality	676 (4.7)	194 (5.8)	86 (8.8)	163 (4.4)	160 (3.9)	73 (3.3)	<0.001
Hospital length of stay	166.0 (94.00-311.00)	167.5 (94.0-311.0)	195.0 (98.0-377.0)	167.0 (93.0-312.0)	163.0 (93.0-304.0)	164.0 (92.0-307.8)	0.001
Intensive Care Unit admission	853 (5.9)	239 (7.2)	181 (18.5)	210 (5.6)	137 (3.3)	86 (3.9)	<0.001
*Septic shock	1,118 (7.8)	371 (11.1)	302 (30.8)	238 (6.4)	141 (3.4)	66 (3.0)	<0.001
Vasopressor use	1,695 (11.8)	539 (16.1)	406 (41.4)	381 (10.2)	245 (5.9)	124 (5.6)	<0.001
Mechanical Ventilator use	357 (2.5)	120 (3.6)	70 (7.1)	81 (2.2)	49 (1.2)	37 (1.7)	<0.001
Time to vasopressor use (h)	6.3 (2.3-33.1)	4.4(1.9-13.2)	3.3 (1.5-6.3)	7.9 (2.6-47.0)	45.9 (11.1-187.9)	43.6 (11.2-254.3)	<0.001
Duration of vasopressor use (h)	20.6 (7.9-43.5)	22.0 (9.0-46.3)	20.5 (9.0-39.6)	18.4 (6.1-38.6)	24.6 (6.4-52.8)	17.7 (6.4-40.3)	0.08
Time to antibiotics use (h)	2.4 (1.0-4.3)	2.3 (0.9-4.1)	1.5 (0.5-3.1)	2.4 (1.0-4.4)	2.5 (1.2-4.6)	2.5 (1.1-4.7)	<0.001

*Patients who met the septic shock criteria according to the new Sepsis-3 definition, characterized by persistent hypotension requiring vasopressors to maintain a mean arterial pressure of ≥ 65 mmHg and a serum lactate level > 2 mmol/L despite adequate volume resuscitation

Table 3. Univariable and multivariable logistic regression analyses of the effect of antibiotic administration within 3 hours on 28-d mortality stratified by clusters

Cluster	OR	95% CI	p-value	aOR ¹	95% CI	p-value
A	1.22	0.94-1.59	0.13	0.73	0.54-0.98	0.037
B	1.47	0.93-2.40	0.11	0.67	0.40-1.16	0.14
C	1.46	1.10-1.94	0.009	0.92	0.67-1.27	0.60
D	1.16	0.89-1.53	0.27	0.88	0.65-1.20	0.42
E	0.84	0.58-1.23	0.37	0.59	0.39-0.89	0.013

OR odds ratio, *CI* confidence interval, *aOR* adjusted odds ratio; ¹Adjusted variables included the respiratory cause of infection, age, initial lactate levels >2 mmol/L, and the SOFA_max score

17733 Assessed for eligibility

3331 excluded

465 Transferred from another hospital

572 Patients who had limitations on invasive care

108 Cardiac arrest

596 Transferred to another hospital during treatment

1149 Non-infection

441 Inadequate data (no vital signs, etc.)

14402 patients with suspected infection were included

2011 Among above suspected infection, patient with
Hypotension within 6 hours from ED arrival

Data Acquisition



Vital signs and Clinical laboratory data at ED arrival (initial)



Additional Vital signs at the time of shock onset within 6 hours from ED arrival (Shock)

Preprocessing and Clustering

Distance Calculation with Normalization and combining of using distance values (*distance_initial* and *distance_shock*)



Distance-based clusters (Unsupervised Machine-learning-based Clustering)

Applying AI model

Artificial Intelligence model with encoding of distance values as color images

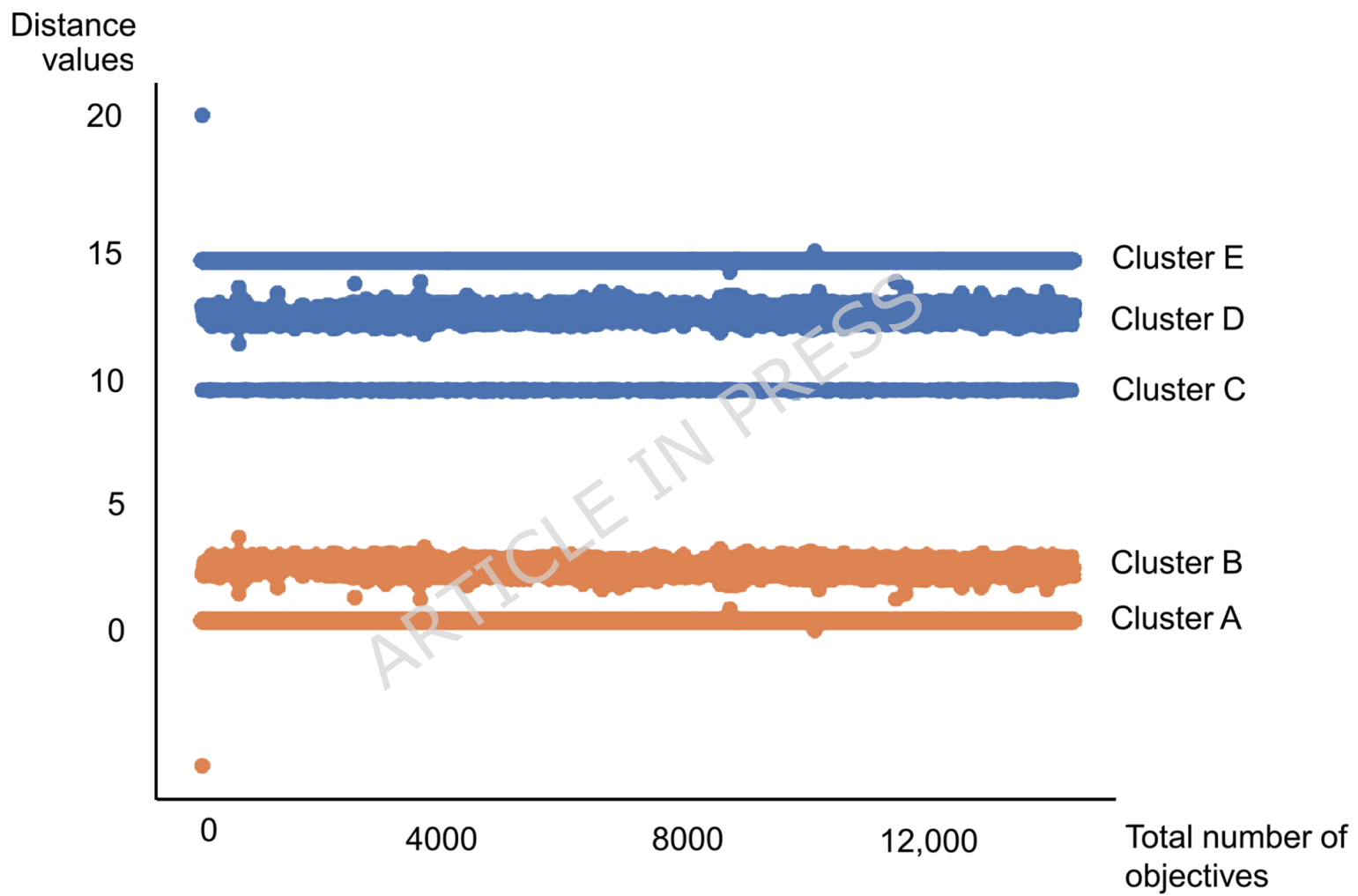
CNN model trained with 14,402 images



Phenotype analysis

Clinical outcomes and sepsis interventions of the 5 clusters





Distance
values



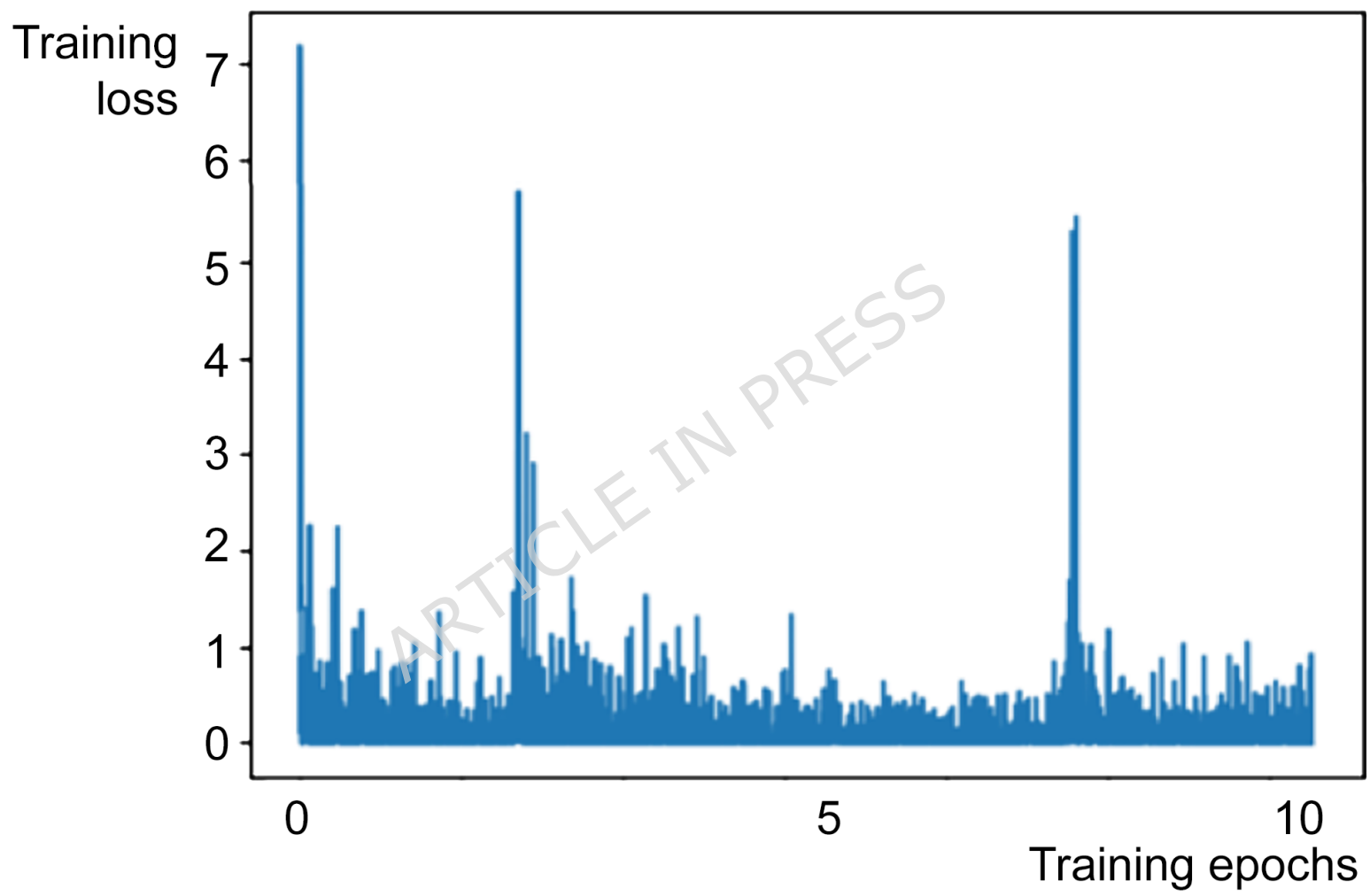
Cluster A

Cluster B

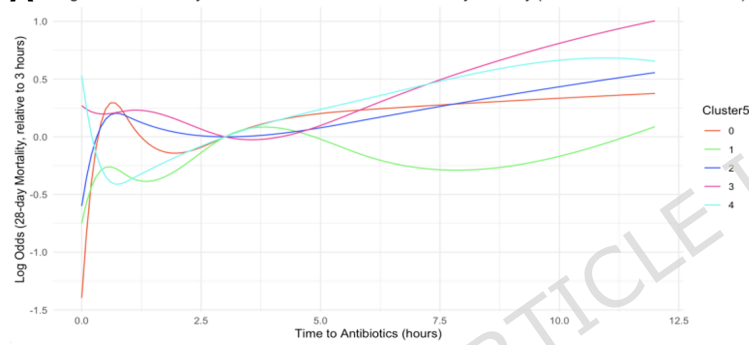
Cluster C

Cluster D

Cluster E



A Log Odds Stratified by Cluster5: Time to Antibiotics vs. 28-day Mortality (Reference time set to 3 hours)



B Log Odds Stratified by Cluster5: Time to Antibiotics vs. 28-day Mortality (Reference time set to 3 hours)

