



OPEN

Identification of the robust predictor for sepsis based on clustering analysis

Jae Yeon Jang^{1,5}, Gilsung Yoo^{2,5}, Taesic Lee³, Young Uh² & Juwon Kim^{2,4}✉

Sepsis is a life-threatening disorder with high incidence and mortality rate. However, the early detection of sepsis is challenging due to lack of specific marker and various etiology. This study aimed to identify robust risk factors for sepsis via cluster analysis. The integrative task of the automatic platform (i.e., electronic medical record) and the expert domain was performed to compile clinical and medical information for 2,490 sepsis patients and 16,916 health check-up participants. The subjects were categorized into 3 and 4 groups based on seven clinical and laboratory markers (Age, WBC, NLR, Hb, PLT, DNI, and MPXI) by K-means clustering. Logistic regression model was performed for all subjects including healthy control and sepsis patients, and cluster-specific cases, separately, to identify sepsis-related features. White blood cell (WBC), well-known parameter for sepsis, exhibited the insignificant association with the sepsis status in old age clusters (K3C3 and K4C3). Besides, NLR and DNI were the robust predictors in all subjects as well as three or four cluster-specific subjects including K3C3 or K4C3. We implemented the cluster-analysis for real-world hospital data to identify the robust predictors for sepsis, which could contribute to screen likely overlooked and potential sepsis patients (e.g., sepsis patients without WBC count elevation).

Sepsis is a life-threatening disorder caused by overwhelming immune response to infection¹. In 2017, the incidence of sepsis worldwide were estimated 48.9 million cases (95% uncertainty interval [UI] 38.9–62.9), and 11.0 million (95% UI: 11.8–54.5) cases of sepsis-related death has been reported, accounting for about 20% of all global deaths². In Korea, the incidence of sepsis was 265.7 (95% confidence interval [CI], 254.7–277.1) per 100,000 person-years in 2005 and 453.1 (95% CI, 439.0–467.5) in 2012³. Moreover, the 6-month mortality rate among sepsis patients in Korea was 26.5% (95% CI, 24.4% to 28.8%) in 2005 and 30.1% (95% CI, 28.4% to 31.9%) in 2012³.

The diagnostic criteria of Systemic inflammatory response syndrome (SIRS), which is used for identifying sepsis⁴, includes four parameters: presence of fever, tachycardia, tachypnea, and leukocytosis or leukopenia. Albeit the ease of collection of information on those four parameters in clinical setting, these diagnostic markers are non-specific. Furthermore, laboratory markers used for the diagnosis of sepsis have low screening power in specific population group. For example, Seigel et al.⁵ pinpointed that about 52% of patients with bacteremia had a normal range of WBC levels, and 21% of patients diagnosed with severe sepsis or septic shock manifested a normal WBC count at the time of admission to the emergency department. As a consequence, there have been tremendous efforts to find the sensitive and suitable marker that would be used for early diagnosis of sepsis, including C-reactive protein (CRP)^{6–9}, procalcitonin (PCT)^{6,7,10–14} and interleukins^{8,13,15,16}. In the early years, erythrocyte sedimentation rate (ESR) and CRP as markers of inflammation were used in diagnosis of sepsis, however, these markers have limitations due to lack of specificity for sepsis. PCT, a 116-amino-acid peptide, has shown to be strongly associated with systemic bacterial infection and the prognosis of the disease, but the cost for testing is relatively high and not frequently used for screening the disease.

Numerous factors, such as age, gender, and presence of chronic diseases, have been reported to influence sepsis-like immune responses^{17–24} and the infection origins of different organs have resulted in rather heterogeneous clinical outcomes. Previously, Knox et al.²⁵ attempted to identify several sepsis groups susceptible to multiple organ dysfunction syndrome using cluster analysis. Considering that determination of robust or group-specific

¹Division of Hematology-Oncology, Department of Internal Medicine, Yonsei University Wonju College of Medicine, Wonju, South Korea. ²Department of Laboratory Medicine, Yonsei University Wonju College of Medicine, 162 Ilsan-dong, Wonju-city 220-701, Gangwon-do, South Korea. ³Department of Family Medicine, Yonsei University Wonju College of Medicine, Wonju, South Korea. ⁴Center for Precision Medicine and Genomics, Wonju Severance Christian Hospital, Wonju, South Korea. ⁵These authors contributed equally: Jae Yeon Jang and Gilsung Yoo. ✉email: Juwon76@yonsei.ac.kr

biomarkers may be a crucial strategy for early detection of sepsis, the identification of potential predictors for sepsis in heterogeneous subgroups would be meaningful. Therefore, in this study, we conducted the cluster analysis and identified robust or cluster-specific risk factors for sepsis using hematology indices.

Materials and methods

Study population. We conducted a cross-sectional study analyzing about 20,000 healthy subjects or patients who visited the Wonju Severance Christian Hospital (WSCH) from January 2017 to December 2020. In detail, subjects were collected using two different sources including Laboratory Medical System and health check-up database from Electronic Medical records (EMR) at WSCH. For screening sepsis patients, three doctors reviewed the ICD-10 code matched to sepsis or SIRS (Table S1). Afterwards, database administrator constructed the dataset including patients diagnosed to sepsis based on the ICD codes from the EMR at WSCH. For healthy control samples, individuals who had undergone health check-up were also collected by the same database administrator. The patients with negative results of blood culture tests and SIRS (ICD 10) were excluded. Then, following subjects were eliminated: those who aged under 19 years or having missing values for laboratory results; who were not diagnosed with sepsis diagnosis at the time of admission; who were diagnosed to rheumatologic diseases, malignancy, and hematologic diseases; and who received immunosuppressant. Finally, a total of 2,490 sepsis patients and 16,916 health check-up participants were enrolled for the analyses of this study. This study was conducted in accordance with the principles of the Declaration of Helsinki and approved by the Institutional Review Board (IRB) of WSCH (No. CR321107). Because this study was performed retrospectively based on existing medical records, the requirement for written consent from the patients was waived, which was approved by IRB (No. CR321107).

Definition of sepsis. We defined sepsis patients as follows: the patients with (1) the ICD-10 code for sepsis; and (2) the ICD-10 code for SIRS and exhibited bacteremia (manually checked by domain expert). Bacteremia was defined using the criteria of the US centers for Disease Control and Prevention/National Healthcare Safety Network surveillance²⁶. The criteria of bacteremia are as follows: (1) the pathogen was identified in one or more blood culture tests; (2) the pathogen was not considered common contaminants.

Measurements. The hematology indices, Delta neutrophil index (DNI) and Myeloperoxidase index (MPXI) were estimated in an automated hematology analyzer ADVIA2120 (Siemens, Tarrytown, NY) following the manufacturer's recommendation. The DNI was calculated using the following formula: $DN (\%) = (\text{the leukocyte subfraction assayed in the MPO channel by cytochemical reaction}) - (\text{the leukocyte subfraction counted in the nuclear lobularity channel by reflected light beam})$ ²⁷. The MPXI was determined with the blood auto-analyzer using 4-chloro-1-naphthol (an MPO substrate in granulocytes), and in these cells black precipitates were formed. As stained white blood cells (WBC) pass through the flow cell, light scatter (y-axis) and absorbance (x-axis) are measured by a tungsten-halogen light source, and the MPXI is defined by the deviation from the mean neutrophil values on the x-axis²⁸.

Cluster analysis. We conducted cluster analysis to identify cluster-specific factors related to sepsis status. Before performing the cluster analysis, we scaled neutrophil-lymphocyte ratio (NLR) and DNI by log-transformation to reduce the effect of outliers (Figure S1). Afterwards, we scaled all continuous variables (V) by using z-normalization as follow.

$$\text{Scaled variable } (V) = \frac{V (\text{an individual subject}) - \text{average of } V (\text{all subjects})}{\text{standard deviation of } V (\text{all subjects})}$$

In the medical field, numerous clustering methods have been implemented to reveal biological and pathophysiological findings. Among them, K-means clustering is widely used in various fields, including clinical medicine and bioinformatics^{29–31}. Ahlqvist et al.²⁹ applied K-means clustering algorithm to patients with newly diagnosed diabetes, identifying five novel clusters that exhibit different associational patterns with the risk of diabetic complication. Kwon et al.³¹ also used the K-means clustering method to yield five sub-groups and their cluster-specific risk factors for sarcopenia. Motivated by these studies, we performed the K-means clustering algorithm to identify the robust or hidden (but potential) risk factors for sepsis.

The determination of features for the cluster analysis is a crucial task. Typically, domain experts or groups have empirically determined the risk factors to be related with the interest disease as the criteria for the cluster analysis^{29,31}. For example, Ahlqvist et al.²⁹ established five clusters based on six clinical variables that are known to be related with diabetes. Similarly, Kwon et al.³¹ made five sub-groups according to nine lifestyle-related features known to be associated with the low muscle mass. Taken together, hematologist and laboratorian had empirically determined seven variables for the cluster analysis based on the literature-based review (Table S2).

Another main task of the K-means clustering is to select k , which means the number of groups to be categorized. Ahlqvist et al.²⁹ determined the optimal k based on silhouette width, besides Kwon et al.³¹ empirically determined 5 as the k . Based on these studies, we empirically determined the number of clusters, and performed the cluster analysis twice according to the change of k from three to four.

We used both healthy participants and sepsis patients for the establishment of clusters. However, due to different characteristics between healthy subjects and sepsis patients, the biased arrangement of them into each cluster was shown. For example, in an experiment (data not shown), a cluster consisted of about 2,000 subjects, of which 1,922 were septic patients. To overcome this problem, we used ComBat algorithm^{32,33} to reduce differential signatures between healthy subjects and septic patients, thereby, yielding more generalized subjects. Afterward,

	Normal group	Sepsis group	P value
n	16,916	2,490	
Age, year	49 ± 0.08	73.3 ± 0.27	< 0.001
Male, n	10,666 (63.1)	1277 (51.3)	< 0.001
WBC count, 10 ⁹ /L	5.5 ± 0.01	13.5 ± 0.18	< 0.001
Neutrophil count, 10 ⁹ /L	3.1 ± 0.01	11.8 ± 0.17	< 0.001
Lymphocyte, 10 ⁹ /L	1.799 ± 0.01	1.001 ± 0.02	< 0.001
NLR	1.8 ± 0.01	19.3 ± 0.43	< 0.001
Hemoglobin, g/dL	14.7 ± 0.01	11.4 ± 0.05	< 0.001
Platelet, 10 ⁹ /L	250 ± 0.44	210.4 ± 2.62	< 0.001
DNI	0.1 ± 0.01	9.1 ± 0.25	< 0.001
MPXI	-0.1 ± 0.03	2.7 ± 0.12	< 0.001

Table 1. General characteristics of study population based on sepsis status. Continuous and categorical variables are described as mean ± standard error and frequency (percent), respectively. Abbreviations: WBC, white blood cell; NLR, neutrophil–lymphocyte ratio; DNI, delta neutrophil index; MPXI, Myeloperoxidase Index; n, number.

we established three to four clusters that had similar characteristics among subjects in the same cluster for the eight variables through the K-means algorithm with setting 100,000 as iteration.

Statistical analysis. Student's *t*-test and Chi square test were used to analyze differential characteristics of continuous and categorical variables according to two conditions (i.e., normal and sepsis subjects), respectively. We used a one sample *t*-test to characterize each cluster based on seven variables. We used a logistic regression model to uncover whether a predictor (i.e., independent variable) is related to the sepsis status (i.e., dependent variable: presence vs. absence). In the multivariate model, seven variables (i.e., age, WBC count, NLR, hemoglobin, platelet count, DNI, and MPXI) were used as both independent and confounding variables, and a variable (i.e., sex) was solely used as confounder. We conducted all statistical analyses and illustrations using the R language (version 4.0.1).

Results

Clinical and laboratory characteristics of the study participants. The general characteristics of the enrolled subjects in this study are summarized based on the status of sepsis in Table 1. Subjects with the diagnosis of sepsis were significantly older than those without sepsis. The WBC, neutrophil count, NLR, DNI, and MPXI were higher in patients diagnosed with sepsis while other laboratory indices were all lower in the sepsis group.

Characteristics of each of the three or four clusters. We constructed three or four clusters (C) each for both healthy subjects and sepsis patients using the K-means clustering. Each group (i.e., cluster) included participants with similar characteristics for seven variables. Also, we performed two clustering tasks with grouping all subjects into three or four groups. In case of K-means clustering with setting three as K, 4,941, 6,608, and 7,877 subjects were grouped to three clusters (K3C1, K3C2, and K3C3), respectively. The biochemical characteristics of the three clusters were characterized based on one sample *t*-test (Fig. 1A). For example, the participants in K3C3 were characterized as relatively old age, low levels of WBC counts, NLR, hemoglobin concentration, platelet counts, and high DNI levels (Fig. 1A).

In case of the clustering model with setting four as K, 4,061, 3,541, 6,679, and 5,145 were arranged to K4C1, K4C2, K4C3, and K4C4, respectively. Among them, the notable cluster was the K4C3 which included subjects that are relatively old and have low levels of WBC and platelet counts, and high level of MPXI (Fig. 2A). When setting *p*-value of 0.001 as the cut-off, the statistical significance for the biasness measured via one sample *t*-test of following variables disappeared in K4C3 compared to K3C3: NLR and hemoglobin.

Comparison of general and cluster-specific risk factors. We performed multivariate logistic regression to identify risk factors significantly related to the presence of sepsis for both all and cluster-specific subjects. For three clusters made by the K(=3)-means clustering, age, NLR, hemoglobin, and DNI levels were significantly associated with the sepsis status in both entire and three clusters-specific subjects (Fig. 1B). The subjects in K3C3 had the insignificant association between WBC count and the presence of sepsis, besides the significant relationship between DNI level and the sepsis status. Other laboratory markers significantly related to the sepsis status in K3C3 were NLR, hemoglobin, platelet, DNI, and MPXI (Fig. 1B).

In case of four groups constructed by the K(=4)-means clustering, age, NLR, and DNI levels were the robust factors significantly related with the presence of sepsis in both all and cluster-specific subjects. In K4C3 cluster, WBC count was not a risk factor for the sepsis status whilst NLR, hemoglobin, platelet, DNI, and MPXI levels were significant predictors (Fig. 2B). Taken together, NLR and DNI were the robust predictors in all subjects as well as three or four cluster-specific subjects.

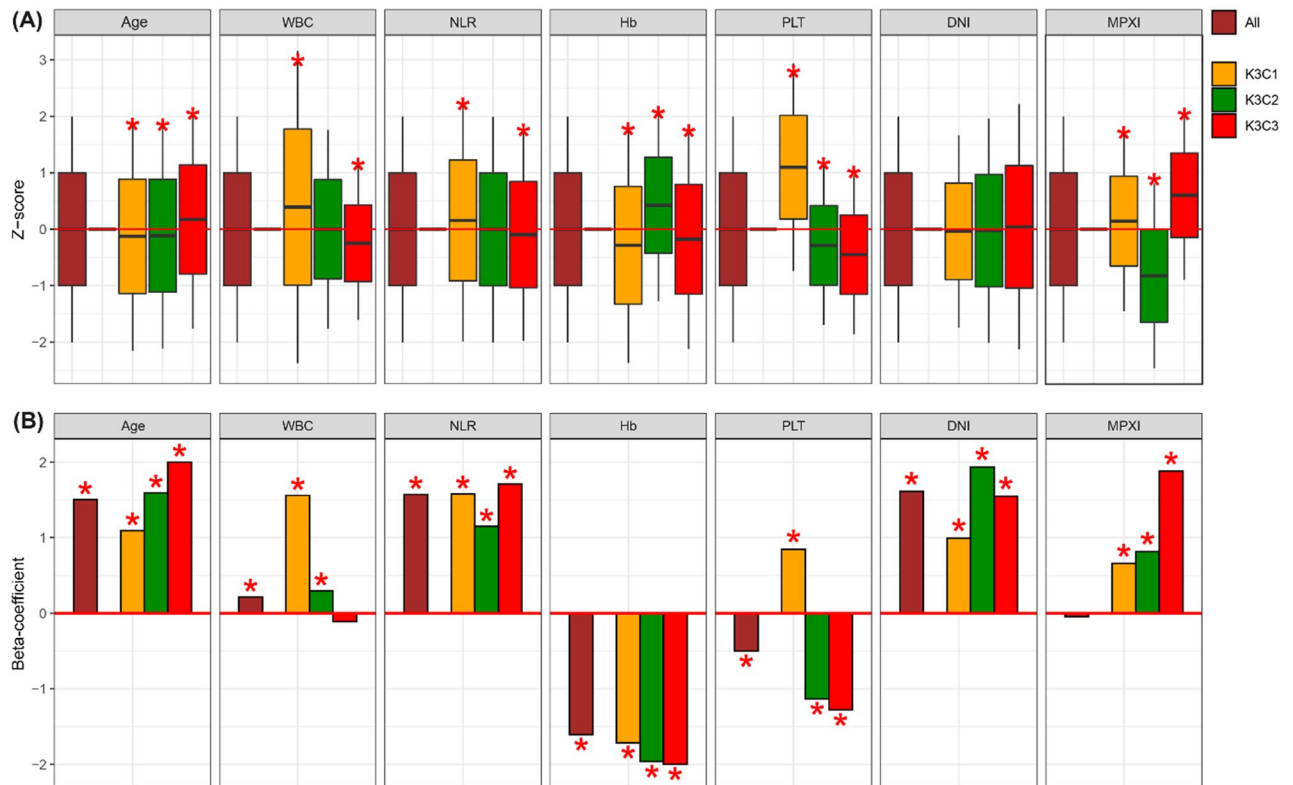


Figure 1. Three subgroups established by K-means clustering. **(A)** Three subgroups were characterized based on seven sepsis-related variables using one sample *t*-test. The brown box plot is about all subjects including healthy control and sepsis patients. Black middle, upper margin, and max lines in each boxplot indicate means, one standard deviation (SD), and 2-SD levels. **(B)** Beta-coefficients were measured from multivariate logistic regression. Brown bar was made by logistic regression for all subjects. Other bars (yellow, green, red) were curated by logistic regression for each cluster-specific subjects. A red asterisk indicates *p*-value less than 0.001.

We compared the distribution of subgroup-specific subjects arranged by K(=3)-means clustering with that by K(=4)-means clustering. 79.9% (6,297/7,877) of K3C3 subjects were included in K4C3, and 85.3% (855/1,002) of sepsis patients in K3C3 were arranged into K4C3 (Fig. 3).

Discussion

We conducted clustering analysis to categorize all subjects (healthy controls and sepsis patients) into three to four groups based on seven clinical and laboratory markers. Then, the associational analysis using multivariate regression model was performed for all subjects and cluster-specific cases. As a result, age, NLR, and DNI were the robust predictors for the sepsis status in all subjects as well as cluster-specific groups.

The WBC levels in elderly subjects (K3C3 and K4C3) exhibited insignificant association with the sepsis status. Aminzadeh et al.³⁴ reported only 60% and 6% sepsis subjects had WBC $\geq 14,000$ and $< 4,000$ WBC counts, respectively. In a study by Caterino et al.³⁵, the predictive performance of WBC was poor for the classification of sepsis (area under the curve: 0.5; 95% CI: 0.3 to 0.7). In elderly patients, as the aging progresses, chronic inflammatory status is pervasive, and cell regeneration ability and phagocytosis are reduced. For these reasons, the response of neutrophils to infection can be delayed^{36–38}. Therefore, normal neutrophil count should not be allowed to exclude patients from being diagnosed with sepsis in elderly population.

Several studies have reported the DNI level as predictors^{39–44} for sepsis and sepsis-related mortality. Seok et al.⁴² reported 73.4% [95% confidence interval (CI): 64.9–80.9] and 97.9% [95% CI: 88.7–99.9] of sensitivity and specificity, respectively, when setting 2.7 of DNI as cut-off, indicating that DNI is a biomarker with high negative predictive value. Park et al.⁴¹ specified subjects diagnosed with severe sepsis or septic shock, and resulted 88.6% and 84.7% of negative predictive value (NPV) and positive predictive value (PPV), respectively, when setting 6.5 of DNI as the cut-off. In addition, several studies have suggested the DNI as prognostic factor for sepsis-related mortality³⁹.

NLR has been considered as an easily obtainable parameter to assess the inflammatory status of patients. It is known to be a useful predictor for the adverse events (e.g., mortality) of patients experiencing diabetes⁴⁵, cancer^{46–48}, and abdomen operation^{49,50}. Specifically, NLR exhibited the significant association with sepsis in both all subjects and cluster groups. NLR have been known to be biomarker for the diagnosis of sepsis⁵¹. Moreover, several expert groups have suggested the NLR as a potential prognostic factor for the sepsis-related complication or mortality^{51–53}.

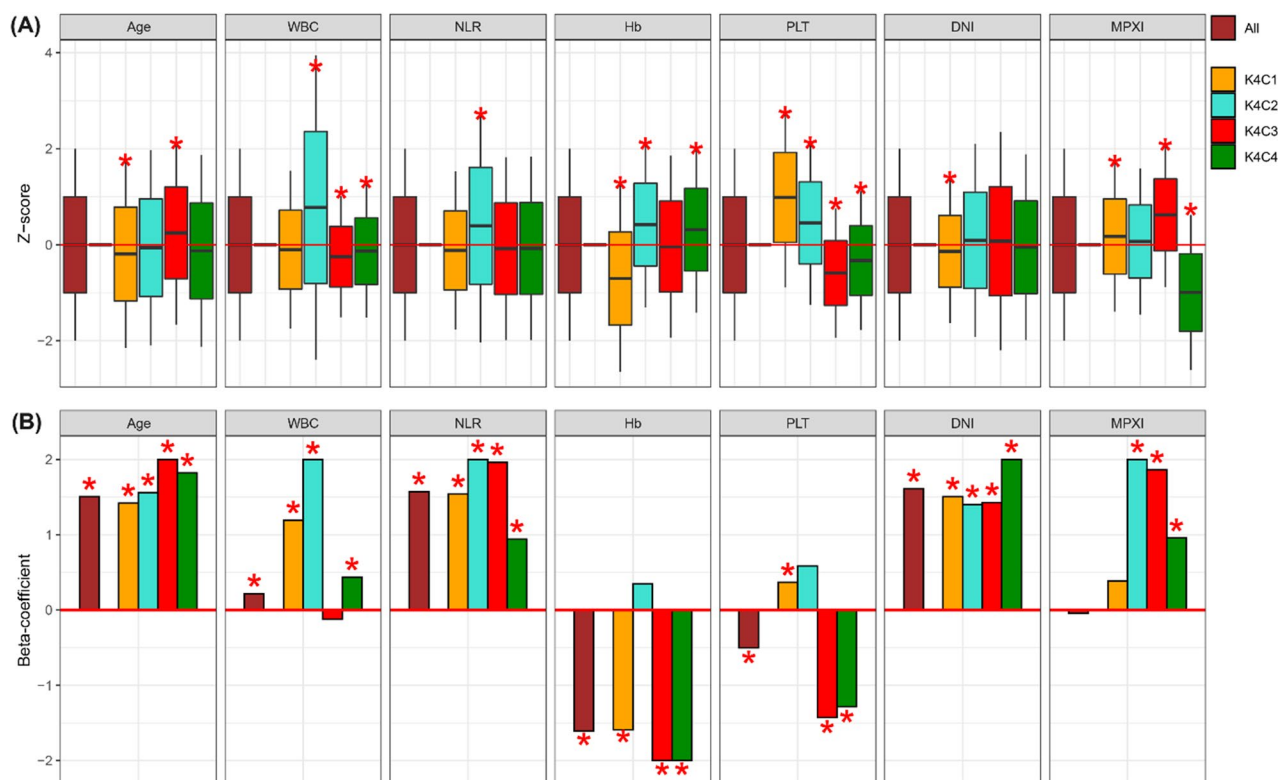


Figure 2. Four subgroups established by K-means clustering. **(A)** Four subgroups were characterized based on seven sepsis-related variables using one sample *t*-test. The brown box plot is about all subjects including healthy control and sepsis patients. Black middle, upper margin, and max lines in each boxplot indicate means, one standard deviation (SD), and 2 SD deviation levels. **(B)** Beta-coefficients were measured from multivariate logistic regression. Brown bar was made by logistic regression for all subjects. Other bars (yellow, light cyan, red, and green) were curated by logistic regression for each cluster-specific subjects. A red asterisk indicates p-value less than 0.001.

(A)	K3C1	K3C2	K3C3	(B)	K3C1	K3C2	K3C3
K4C1	2842 (0.575)	149 (0.023)	1070 (0.136)	K4C1	335/2507 (0.12)	21/128 (0.14)	79/991 (0.07)
K4C2	2064 (0.418)	1135 (0.172)	342 (0.043)	K4C2	283/1781 (0.14)	130/1005 (0.11)	46/296 (0.13)
K4C3	6 (0.001)	376 (0.057)	6297 (0.799)	K4C3	0/6 (0)	46/330 (0.12)	855/5442 (0.14)
K4C4	29 (0.006)	4948 (0.749)	168 (0.021)	K4C4	2/27 (0.07)	671/4277 (0.14)	22/146 (0.13)

Figure 3. Comparison between three and four clusters established by K-means clustering. Colored rectangles indicate cases with prominently high number of common subjects between two subgroups. **(A)** Results are described number and column-based ratio for subjects including healthy control and sepsis patients in common between two subgroups. **(B)** Results are described as number of sepsis patients/healthy cases and ratio [sepsis patients/(sepsis + healthy subjects)].

Clustering is an establishment task to arrange heterogeneous samples or variables to several groups, in which each group includes homogenous subjects or variables. Many methods, including K-means clustering, auto-encoder³⁰, and NMF, have been used to construct groups, clusters, or modules. Moon and Lee developed the integrative non-linear representation method (Joint deep semi-NMF, JDSNMF) based on deep learning and modified NMF considering both intensity and direction of feature for the clustering task⁵⁴. We implemented a widely used and simple method, K-means clustering to establish several subgroups. Future study implementing integrative algorithm^{30,54} for clustering is needed to establish the sophisticated module.

Our study is a multidiscipline work including the database professional, laboratorian, computer analyst, and domain expert (i.e., Hemato-oncologist). Data analyzed in this study is automatically curated from real-world hospital database by database professional. Afterward, laboratorian and domain expert empirically determined

candidate variables that have potential relationship with the sepsis status. Crucial limitation of the present study is the determination of sepsis status based on ICD codes and blood culture results retrospectively. Therefore, it is crucial to establish the prediction model using robust predictors for sepsis status and apply this in real-world setting^{55,56}.

In conclusion, this study revealed that the DNI and NLR are promising diagnostic marker for sepsis in specific group that exhibited insignificant association between WBC count and sepsis status. These findings could contribute to screen sepsis patients without leukocytosis in the emergency care units.

Received: 22 November 2021; Accepted: 24 January 2022

Published online: 11 February 2022

References

- Singer, M. *et al.* The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* **315**, 801–810. <https://doi.org/10.1001/jama.2016.0287> (2016).
- Rudd, K. E. *et al.* Global, regional, and national sepsis incidence and mortality, 1990–2017: Analysis for the Global Burden of Disease Study. *Lancet* **395**, 200–211. [https://doi.org/10.1016/s0140-6736\(19\)32989-7](https://doi.org/10.1016/s0140-6736(19)32989-7) (2020).
- Kim, J., Kim, K., Lee, H. & Ahn, S. Epidemiology of sepsis in Korea: a population-based study of incidence, mortality, cost and risk factors for death in sepsis. *Clin. Exp. Emerg. Med.* **6**, 49–63. <https://doi.org/10.15441/ceem.18.007> (2019).
- Bone, R. C. *et al.* Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* **101**, 1644–1655. <https://doi.org/10.1378/chest.101.6.1644> (1992).
- Seigel, T. A. *et al.* Inadequacy of temperature and white blood cell count in predicting bacteremia in patients with suspected infection. *J. Emerg. Med.* **42**, 254–259. <https://doi.org/10.1016/j.jemermed.2010.05.038> (2012).
- Brunkhorst, F. M., Al-Nawas, B., Krummenauer, F., Forycki, Z. F. & Shah, P. M. Procalcitonin, C-reactive protein and APACHE II score for risk evaluation in patients with severe pneumonia. *Clin. Microbiol. Infect.* **8**, 93–100. <https://doi.org/10.1046/j.1469-0691.2002.00349.x> (2002).
- Simon, L., Gauvin, F., Amre, D. K., Saint-Louis, P. & Lacroix, J. Serum procalcitonin and C-reactive protein levels as markers of bacterial infection: A systematic review and meta-analysis. *Clin. Infect. Dis.* **39**, 206–217. <https://doi.org/10.1086/421997> (2004).
- Gabay, C. & Kushner, I. Acute-phase proteins and other systemic responses to inflammation. *N. Engl. J. Med.* **340**, 448–454. <https://doi.org/10.1056/nejm199902113400607> (1999).
- Póvoa, P. *et al.* C-reactive protein as a marker of infection in critically ill patients. *Clin. Microbiol. Infect.* **11**, 101–108. <https://doi.org/10.1111/j.1469-0691.2004.01044.x> (2005).
- Tang, B. M., Eslick, G. D., Craig, J. C. & McLean, A. S. Accuracy of procalcitonin for sepsis diagnosis in critically ill patients: Systematic review and meta-analysis. *Lancet Infect. Dis.* **7**, 210–217. [https://doi.org/10.1016/s1473-3099\(07\)70052-x](https://doi.org/10.1016/s1473-3099(07)70052-x) (2007).
- Nakamura, A. *et al.* Efficacy of procalcitonin in the early diagnosis of bacterial infections in a critical care unit. *Shock* **31**, 586–591. <https://doi.org/10.1097/SHK.0b013e31819716fa> (2009).
- Assicot, M. *et al.* High serum procalcitonin concentrations in patients with sepsis and infection. *Lancet* **341**, 515–518. [https://doi.org/10.1016/0140-6736\(93\)90277-n](https://doi.org/10.1016/0140-6736(93)90277-n) (1993).
- Pettilä, V., Hynninen, M., Takkunen, O., Kuusela, P. & Valtonen, M. Predictive value of procalcitonin and interleukin 6 in critically ill patients with suspected sepsis. *Intensive Care Med.* **28**, 1220–1225. <https://doi.org/10.1007/s00134-002-1416-1> (2002).
- Karzai, W., Oberhoffer, M., Meier-Hellmann, A. & Reinhart, K. Procalcitonin—A new indicator of the systemic response to severe infections. *Infection* **25**, 329–334. <https://doi.org/10.1007/bf01740811> (1997).
- Sfeir, T., Saha, D. C., Astiz, M. & Rackow, E. C. Role of interleukin-10 in monocyte hypo-responsiveness associated with septic shock. *Crit. Care Med.* **29**, 129–133. <https://doi.org/10.1097/00003246-200101000-00026> (2001).
- Cannon, J. G. *et al.* Circulating interleukin-1 and tumor necrosis factor in septic shock and experimental endotoxin fever. *J. Infect. Dis.* **161**, 79–84. <https://doi.org/10.1093/infdis/161.1.79> (1990).
- Brodin, P. *et al.* Variation in the human immune system is largely driven by non-heritable influences. *Cell* **160**, 37–47. <https://doi.org/10.1016/j.cell.2014.12.020> (2015).
- Brodin, P. & Davis, M. M. Human immune system variation. *Nat. Rev. Immunol.* **17**, 21–29. <https://doi.org/10.1038/nri.2016.125> (2017).
- Kurts, C., Panzer, U., Anders, H. J. & Rees, A. J. The immune system and kidney disease: Basic concepts and clinical implications. *Nat. Rev. Immunol.* **13**, 738–753. <https://doi.org/10.1038/nri3523> (2013).
- Dahmer, M. K., Cornell, T. & Quasney, M. W. Genetic and epigenetic factors in the regulation of the immune response. *Curr. Opin. Pediatr.* **28**, 281–286. <https://doi.org/10.1097/mop.0000000000000356> (2016).
- Zimmermann, P. & Curtis, N. Factors that influence the immune response to vaccination. *Clin. Microbiol. Rev.* <https://doi.org/10.1128/cmr.00084-18> (2019).
- Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638. <https://doi.org/10.1038/nri.2016.90> (2016).
- Gavazzi, G. & Krause, K. H. Ageing and infection. *Lancet Infect. Dis.* **2**, 659–666. [https://doi.org/10.1016/s1473-3099\(02\)00437-1](https://doi.org/10.1016/s1473-3099(02)00437-1) (2002).
- Ongrádi, J. & Kövesdi, V. Factors that may impact on immunosenescence: an appraisal. *Immun. Ageing* **7**, 7. <https://doi.org/10.1186/1742-4933-7-7> (2010).
- Knox, D. B., Lanspa, M. J., Kuttler, K. G., Brewer, S. C. & Brown, S. M. Phenotypic clusters within sepsis-associated multiple organ dysfunction syndrome. *Intensive Care Med.* **41**, 814–822. <https://doi.org/10.1007/s00134-015-3764-7> (2015).
- Horan, T. C., Andrus, M. & Dudeck, M. A. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. *Am. J. Infect. Control.* **36**, 309–332. <https://doi.org/10.1016/j.ajic.2008.03.002> (2008).
- Nahm, C. H., Choi, J. W. & Lee, J. Delta neutrophil index in automated immature granulocyte counts for assessing disease severity of patients with sepsis. *Ann. Clin. Lab. Sci.* **38**, 241–246 (2008).
- Yonezawa, K. *et al.* Association between the neutrophil myeloperoxidase index and subsets of bacterial infections. *Int. J. Lab. Hematol.* **32**, 598–605. <https://doi.org/10.1111/j.1751-553X.2010.01227.x> (2010).
- Ahlqvist, E. *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: A data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol.* **6**, 361–369. [https://doi.org/10.1016/s2213-8587\(18\)30051-2](https://doi.org/10.1016/s2213-8587(18)30051-2) (2018).
- Baek, B. & Lee, H. Prediction of survival and recurrence in patients with pancreatic cancer by integrating multi-omics data. *Sci. Rep.* **10**, 18951. <https://doi.org/10.1038/s41598-020-76025-1> (2020).
- Kwon, Y. J., Kim, H. S., Jung, D. H. & Kim, J. K. Cluster analysis of nutritional factors associated with low muscle mass index in middle-aged and older adults. *Clin. Nutr.* **39**, 3369–3376. <https://doi.org/10.1016/j.clnu.2020.02.024> (2020).

32. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
33. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
34. Aminzadeh, Z. & Parsa, E. Relationship between age and peripheral white blood cell count in patients with sepsis. *Int. J. Prev. Med.* **2**, 238–242 (2011).
35. Caterino, J. M., Scheatzle, M. D., Forbes, M. L. & D'Antonio, J. A. Bacteremic elder emergency department patients: Procalcitonin and white count. *Acad. Emerg. Med.* **11**, 393–396. <https://doi.org/10.1197/j.aem.2003.10.027> (2004).
36. Butcher, S., Chahel, H. & Lord, J. M. Review article: Ageing and the neutrophil: no appetite for killing?. *Immunology* **100**, 411–416. <https://doi.org/10.1046/j.1365-2567.2000.00079.x> (2000).
37. Weiskopf, D., Weinberger, B. & Grubeck-Loebenstien, B. The aging of the immune system. *Transpl. Int.* **22**, 1041–1050. <https://doi.org/10.1111/j.1432-2277.2009.00927.x> (2009).
38. Wenisch, C., Patruta, S., Daxböck, F., Krause, R. & Hörl, W. Effect of age on human neutrophil function. *J. Leukoc. Biol.* **67**, 40–45. <https://doi.org/10.1002/jlb.67.1.40> (2000).
39. Ahn, C. *et al.* The delta neutrophil index (DNI) as a prognostic marker for mortality in adults with sepsis: A systematic review and meta-analysis. *Sci. Rep.* **8**, 6621. <https://doi.org/10.1038/s41598-018-24211-7> (2018).
40. Park, S. Y., Lee, J. S., Oh, J. & Park, J. Y. Delta neutrophil index as a predictive and prognostic factor for Candidemia patients: A matched case-control study. *BMC Infect. Dis.* **20**, 396. <https://doi.org/10.1186/s12879-020-05117-0> (2020).
41. Park, B. H. *et al.* Delta neutrophil index as an early marker of disease severity in critically ill patients with sepsis. *BMC Infect. Dis.* **11**, 299. <https://doi.org/10.1186/1471-2334-11-299> (2011).
42. Seok, Y. *et al.* Delta neutrophil index: A promising diagnostic and prognostic marker for sepsis. *Shock* **37**, 242–246. <https://doi.org/10.1097/SHK.0b013e3182454acf> (2012).
43. Kim, H. *et al.* Usefulness of the delta neutrophil index as a promising prognostic marker of acute cholangitis in emergency departments. *Shock* **47**, 303–312. <https://doi.org/10.1097/shk.0000000000000722> (2017).
44. Celik, I. H. *et al.* The value of delta neutrophil index in neonatal sepsis diagnosis, follow-up and mortality prediction. *Early Hum. Dev.* **131**, 6–9. <https://doi.org/10.1016/j.earlhumdev.2019.02.003> (2019).
45. Azab, B., Chainani, V., Shah, N. & McGinn, J. T. Neutrophil-lymphocyte ratio as a predictor of major adverse cardiac events among diabetic population: A 4-year follow-up study. *Angiology* **64**, 456–465. <https://doi.org/10.1177/0003319712455216> (2013).
46. Hung, H. Y. *et al.* Effect of preoperative neutrophil-lymphocyte ratio on the surgical outcomes of stage II colon cancer patients who do not receive adjuvant chemotherapy. *Int. J. Colorectal Dis.* **26**, 1059–1065. <https://doi.org/10.1007/s00384-011-1192-x> (2011).
47. Tomita, M., Shimizu, T., Ayabe, T., Yonei, A. & Onitsuka, T. Preoperative neutrophil to lymphocyte ratio as a prognostic predictor after curative resection for non-small cell lung cancer. *Anticancer Res.* **31**, 2995–2998 (2011).
48. Sharaiha, R. Z. *et al.* Elevated preoperative neutrophil:lymphocyte ratio as a predictor of postoperative disease recurrence in esophageal cancer. *Ann. Surg. Oncol.* **18**, 3362–3369. <https://doi.org/10.1245/s10434-011-1754-8> (2011).
49. Kahramanca, S. *et al.* Neutrophil-to-lymphocyte ratio as a predictor of acute appendicitis. *Ulus Travma Acil. Cerrahi Derg.* **20**, 19–22. <https://doi.org/10.5505/tjtes.2014.20688> (2014).
50. Ishizuka, M., Shimizu, T. & Kubota, K. Neutrophil-to-lymphocyte ratio has a close association with gangrenous appendicitis in patients undergoing appendectomy. *Int. Surg.* **97**, 299–304. <https://doi.org/10.9738/cc161.1> (2012).
51. de Jager, C. P. *et al.* Lymphocytopenia and neutrophil-lymphocyte count ratio predict bacteremia better than conventional infection markers in an emergency care unit. *Crit Care* **14**, R192. <https://doi.org/10.1186/cc9309> (2010).
52. Liu, X. *et al.* Prognostic significance of neutrophil-to-lymphocyte ratio in patients with sepsis: A prospective observational study. *Med. Inflamm.* **2016**, 8191254. <https://doi.org/10.1155/2016/8191254> (2016).
53. Huang, Z., Fu, Z., Huang, W. & Huang, K. Prognostic value of neutrophil-to-lymphocyte ratio in sepsis: A meta-analysis. *Am. J. Emerg. Med.* **38**, 641–647. <https://doi.org/10.1016/j.ajem.2019.10.023> (2020).
54. Moon, S. & Lee, H. JDSNMF: Joint deep semi-non-negative matrix factorization for learning integrative representation of molecular signals in Alzheimer's disease. *J. Pers. Med.* **11**, 686. <https://doi.org/10.3390/jpm11080686> (2021).
55. Lee, T., Kim, J., Uh, Y. & Lee, H. Deep neural network for estimating low density lipoprotein cholesterol. *Clin. Chim. Acta* **489**, 35–40. <https://doi.org/10.1016/j.cca.2018.11.022> (2019).
56. Hwang, S. *et al.* A deep neural network for estimating low-density lipoprotein cholesterol from electronic health records: Real-time routine clinical application. *JMIR Med. Inform.* **9**, e29331. <https://doi.org/10.2196/29331> (2021).

Author contributions

J.K.: Conceptualization, investigation, analysis and interpretation; J.-Y.J. and G.Y.: writing-original draft preparation; J.K.: writing-review and editing, supervision; J.K. and Y.U.: Data collection and statistical analyses; T.-S.L. and G.Y.: Organization and design of materials; J.-Y.J.: data review and analysis. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-06310-8>.

Correspondence and requests for materials should be addressed to J.K.

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

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022