Received: 7 September 2021 | Revised: 27 September 2021 | Accepted: 28 September 2021

DOI: 10.1002/jcla.24053

RESEARCH ARTICLE

WILEY

Predictive value of red blood cell distribution width in septic shock patients with thrombocytopenia: A retrospective study using machine learning

Jianmin Ling[1] | Tongzhou Liao[2] | Yanqing Wu[3] | Zhaohua Wang[1] | Hai Jin[2] | Feng Lu[2] | Minghao Fang[1]

[1]Department of Emergency and Intensive Care Unit, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Services Computing Technology and System Lab, Cluster and Grid Computing Lab, National Engineering Research Center for Big Data Technology and System, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China

[3]Department of Neurology, Wuhan Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Correspondence

Feng Lu, Services Computing Technology and System Lab, Cluster and Grid Computing Lab, National Engineering Research Center for Big Data Technology and System, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China.
Email: lufeng@hust.edu.cn

Minghao Fang, Department of Emergency and Intensive Care Unit, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China.
Email: fangmh8364@163.com

Abstract

Background: Sepsis-associated thrombocytopenia (SAT) is common in critical patients and results in the elevation of mortality. Red cell distribution width (RDW) can reflect body response to inflammation and oxidative stress. We try to investigate the relationship between the RDW and the prognosis of patients with SAT through machine learning.

Methods: 809 patients were retrospectively analyzed from the Medical Information Mart for Intensive Care III (MIMIC-III) database. The eXtreme Gradient Boosting (XGBoost) and SHapley Additive exPlanations (SHAP) were used to analyze the impact of each feature. Logistic regression analysis, propensity score matching (PSM), receiver-operating characteristics (ROC) curve analysis, and the Kaplan-Meier method were used for data processing.

Results: The patients with thrombocytopenia had higher 28-day mortality (48.2%). Machine learning indicated that RDW was the second most important in predicting 28-day mortality. The RDW was significantly increased in non-survivors by logistic regression and PSM. ROC curve shows that RDW has moderate predictive power for 28-day mortality. The patients with RDW>16.05 exhibited higher mortality through Kaplan-Meier analysis.

Conclusions: Interpretable machine learning can be applied in clinical research. Elevated RDW is not only common in patients with SAT but is also associated with a poor prognosis.

KEYWORDS

inflammation, machine learning, red cell distribution width, septic shock, thrombocytopenia

Jianmin Ling and Tongzhou Liao contributed equally to this work and share first authorship.

1 | INTRODUCTION

According to the Sepsis-3 definition proposed in 2016, sepsis was defined as life-threatening organ dysfunction caused by a dysregulated host response to infection, and septic shock was a subtype of sepsis with circulatory and cellular or metabolic dysfunction.[1] Although sepsis research has made breakthroughs in recent years, it is still a common cause of death in the emergency department or intensive care unit. It had been reported that mortality rates of sepsis can still be as high as 60% in the absence of adequate treatment.[2] In the early stage of sepsis, in order to protect the host from invading pathogens, innate immune cells are activated significantly to clear invading pathogens. If this process is not properly controlled, it will result in exaggerated innate immune and inflammatory responses that could damage organs.[3,4] To better identify patients with multiple organ dysfunction caused by sepsis earlier, the Sequential Organ Failure Assessment (SOFA) score was born.[5] A patient with the SOFA score greater than or equal to two points indicates multiple organ dysfunction and in-hospital mortality greater than 10%.[1,6]

The incidence of thrombocytopenia (platelet counts $<150 \times 10^9$/L) is about 35%–40% in intensive care unit.[7] Sepsis is the most common cause of thrombocytopenia in critical patients, with an incidence of more than 50%.[8] The exact etiology and pathogenesis of sepsis-associated thrombocytopenia (SAT) are still not clear. A traditional opinion of SAT was decreasing production of platelets, called "marrow suppression".[9] However, many scholars believe that thrombopoiesis in marrow may increase rather than suppress during sepsis.[10,11] Increased destruction and assumption due to activation of the immune system and excessive inflammation may be the main cause of SAT.[12-14] Not only the hospital mortality increased, but the duration of ventilation or vasopressor, major bleeding events, renal replacement therapy, and blood transfusion also augmented.[15] Therefore, patients with thrombocytopenia caused by sepsis should be concerned.

Red cell distribution width (RDW) is inexpensive and commonly measured. It is calculated as the size of red blood cells (RBC) divided by the mean corpuscular volume. RDW used to be recognized as an important index in the diagnosis of blood system diseases, such as anemia. Recently, many researches had revealed that increased RDW was associated with poor prognosis and mortality of the cardiovascular disease, cancer, lower respiratory tract disease, critical disease, and even nutritional deficiencies,[16,17] and one of the mechanisms that increased RDW reflects systemic inflammation was widely accepted.[18] Elevated RDW is related to inflammatory status both in acute and chronic conditions, such as COVID-19 infection, chronic kidney disease, type 2 DM, irritable bowel syndrome, rheumatoid arthritis, vertebral disk hernias, and thyroiditis.[19-25]

Artificial intelligence has been widely used in the early warning and mortality prediction of many diseases, including sepsis and septic shock. However, the weakness of machine learning is the lack of explainable ability.[26] To deal with this, we used SHapley Additive exPlanations (SHAP), a model-agnostic machine learning explainable tool, which can analyze the impact of each feature on the classifier and rank them in descending order of the mean marginal contributions to all possible coalitions.

Since RDW can reflect systemic inflammation which leads to thrombocytopenia in septic shock patients, in this study, we intended to investigate the association between RDW and the outcomes of septic shock patients with thrombocytopenia extracted from the MIMIC-III database through an interpretable machine learning approach based on artificial intelligence and data analysis methods.

2 | MATERIALS AND METHODS

2.1 | Study design

This is a retrospective cohort study based on the MIMIC-III (version 1.4) database, a free, public ICU database published by the Massachusetts Institute of Technology. The MIMIC-III database contains more than forty thousand patients who were admitted to ICU at Beth Israel Deaconess Medical Center between 2001 and 2012. It records all patients, demographics, vital signs per hour, nursing records, treatment measures, laboratory findings, imaging data, and even in-hospital and out-of-hospital mortality in detail. Because all the private data in this database were anonymized, informed consent and ethical approval were not necessary to sign before the study. One of the authors (YQW) had passed an examination and was approved to access the database (certification number 40107670) and extract the data.

2.2 | Study subjects and setting

Demographics, comorbidities, laboratory findings, sequential organ failure assessment (SOFA) score during ICU admission and primary endpoint, and the 28-day mortality were data we cared about. Structured query language (SQL) was used to extract data from the database. Patients hospitalized in SICU and MICU firstly diagnosed with sepsis, severe sepsis, and septic shock were selected according to ICD code (ICD code: 99591, 99592, and 78552), and septic patients without shock were excluded. Considering other factors that might affect our result, we had determined the following exclusion criteria: 1) hematological malignancies, for example, leukemia, lymphoma, myeloma, etc; 2) immune system diseases, such as, immune thrombocytopenic purpura (ITP), thrombotic thrombocytopenic purpura (TTP), systemic lupus erythematosus (SLE), etc; 3) patients with malignant tumors; 4). pregnant or breastfeeding; and 5) younger than 18 years old or older than 80.

Patients were divided into two groups according to the counts of platelets at the time of admission. Thrombocytopenia was defined as platelet counts less than 150×10^9/L on admission. As for missing data, we handled it according to the missing percentage, such as

fibrinogen (FIB) and D-dimer, which more than 20% were missing, so they were excluded from our study. If continuous variables with missing values are less than 5%, we selected average or median values to replace them.

2.3 | Interpretable machine learning

The eXtreme Gradient Boosting (XGBoost)[27] was used as our machine learning model. We used eight laboratory features, two demographic features, four basic disease features, and these 15 features together, respectively, to construct the predictions models for the 28-day death. The data set was randomly divided into a training data set and the test data set with a 3:1 ratio. The performance of the model was measured by five indicators as follows: area under the curve (AUC), accuracy, precision, recall, and F1-score in the test set.[28] When the AUC and other indicators of the model exceeded 0.8, the model was considered to have met the statistical pattern; when the indicators reached 0.9, the fit was considered quite well.[29] Interpretation of the prediction model was performed using SHAP, which is a value explainable tool based on the tree model, and which can effectively and accurately calculate the contribution of features to the results[26] and reveal the influence of a single factor on the model.

2.4 | Statistical analysis

Continuous variables were expressed as median (IQR) or mean (±SD) and compared with the Mann-Whitney U test or independent sample *t* test. We expressed categorical variables as number (%) and compared data using the chi-square (χ^2) test or Fisher's exact test. Multivariate logistic regression analysis was applied to evaluate the risk factors of non-survivors in patients with thrombocytopenia. Correspondingly, the odds ratio (OR) with 95% confidence interval (CI) of non-survivors was calculated, respectively. All univariates that associated with the outcomes of interest ($p < 0.2$) were selected into our multivariate logistic regression model by stepwise input to assess the risk factors of non-survivors.

Propensity score matching (PSM) was performed to reduce the imbalance between the survivors and non-survivors with thrombocytopenia on admission. The basic characteristics, such as age and gender, comorbidities, laboratory tests except RDW, were matched using a one-to-one nearest neighbor matching algorithm with a caliper width of 0.1. After matching, we used the receiver-operating characteristics (ROC) curve to demonstrate the sensitivity and specificity of RDW, lactate, and SOFA score to predict non-survivors in patients with thrombocytopenia on admission. The Kaplan-Meier method was used with a log-rank test to compare the survival rates between the strata.

All statistical analyses were performed using the Statistical Package for Social Sciences (SPSS) version 26.0 software, and $p < 0.05$ was considered statistically significant.

3 | RESULTS

As shown in Table S1, a total of 809 septic shock patients were enrolled for our analysis according to the inclusion and exclusion criteria (Figure 1). In general, the mean age was 61.2 ± 12.6 years and 444 patients were male (54.9%). As for comorbidities, hypertension (48.0%), diabetes mellitus (38.1%), and heart diseases (55.0%) are more common than stroke (3.2%). Overall, most of the laboratory findings are abnormal, such as RDW, WBC, PT, APTT, and lactate, while the platelet count is normal. Among all septic shock patients, the 28-day mortality is about 44.1%. Compared with patients without thrombocytopenia on admission, those with thrombocytopenia were younger and had a higher proportion of males. All comorbidities, excluding stroke and heart diseases, were not significant between the two groups. RDW levels, PT, APTT, lactate, and SOFA scores were significantly higher in the patients with thrombocytopenia ($p < 0.05$), but Hct, WBC, and platelet counts were lower in this group ($p < 0.05$). Ultimately, the patients with thrombocytopenia had a significantly higher 28-day mortality (48.2%) than the control group (38.5%).

The results of univariable binary logistic regression showed that age (OR = 1.03, $p < 0.001$), stroke (OR = 1.94, $p = 0.19$), RDW (OR = 1.25, $p < 0.001$), Hct (OR = 1.08, $p = 0.12$), lactate (OR = 1.05, $p < 0.001$), and SOFA score (OR = 1.19, $p < 0.001$) were significantly different between the two groups. All of these univariates were analyzed in the multivariable model. Finally, only stroke (OR = 1.90, $p = 0.2$) was excluded and other variables were risk factors for mortality of patients with septic shock. Detailed data were shown in Table S2.

The SHAP-based interpretation of the XGBoost model is shown in Figure 2. The overall contributions of the top nine features are given, indicating that RDW was the second most important in predicting 28-day death. Higher RDW value leads to a higher risk of death ending for the learning model. Further, SOFA score had the greatest impact on the prediction.

Based on the above results, we found that the 28-day mortality of the patients with thrombocytopenia on admission was significantly higher than others, so we conducted further analysis on this group. Table 1 shows the detailed data. The basic demographics, such as age, gender, and comorbidities, were not significantly different between the two groups ($p > 0.05$). Although RDW increased in both survivors 16.1 ± 2.6) and non-survivors (17.4 ± 2.7), it increased more significantly in the non-survivors ($p < 0.001$). Laboratory findings similar to RDW were PT, APTT, and lactate. Adversely, platelet counts were lower in non-survivors 58(35–91) than survivors 88(48–118). The SOFA score was significantly higher in non-survivors 11(9–14) than survivors 8(5–11).

Again, we estimated the risk factors of non-survivors through binary logistic regression. As shown in Table 2, age (OR = 1.02, $p = 0.03$), RDW (OR = 1.21, $p < 0.001$), lactate (OR = 1.12, $p = 0.01$), and SOFA score (OR = 1.19, $p < 0.001$) were risk factors of non-survivors. After adjusting for confounding factors, the multivariable model obtained the same results.

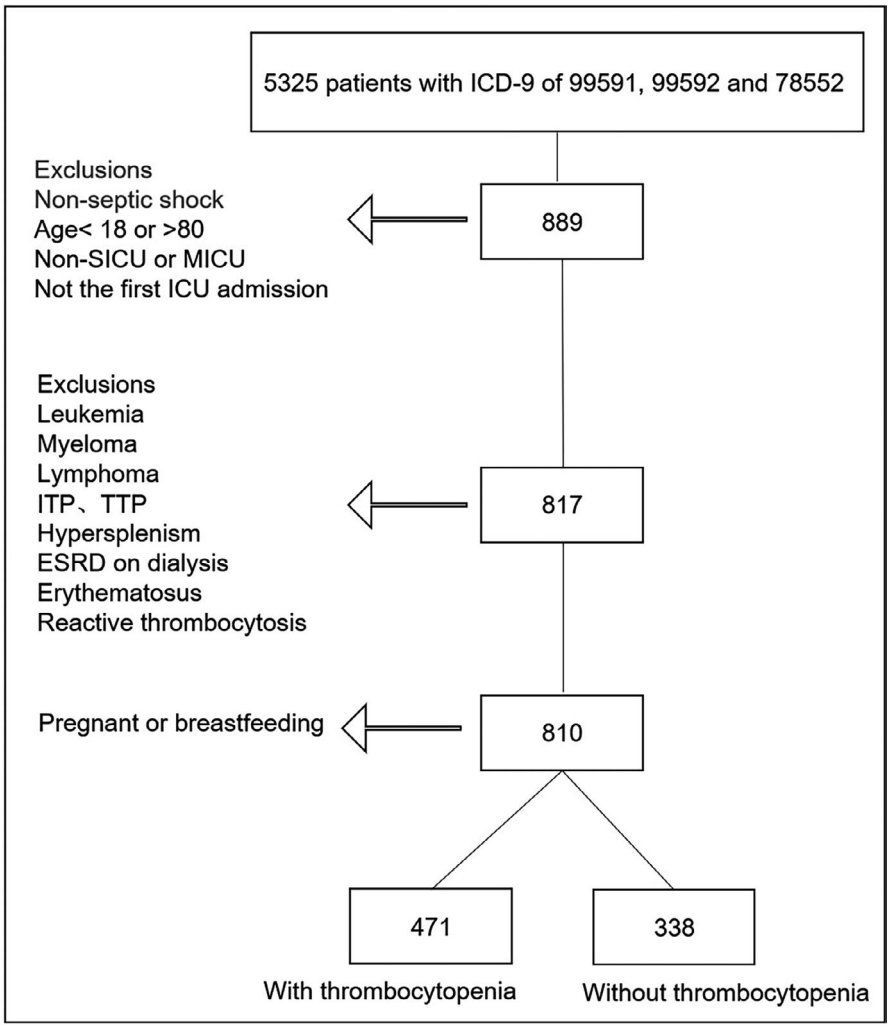

FIGURE 1 Flow Chart Illustrating the Inclusion and Exclusion Criteria of our Study Cohorts

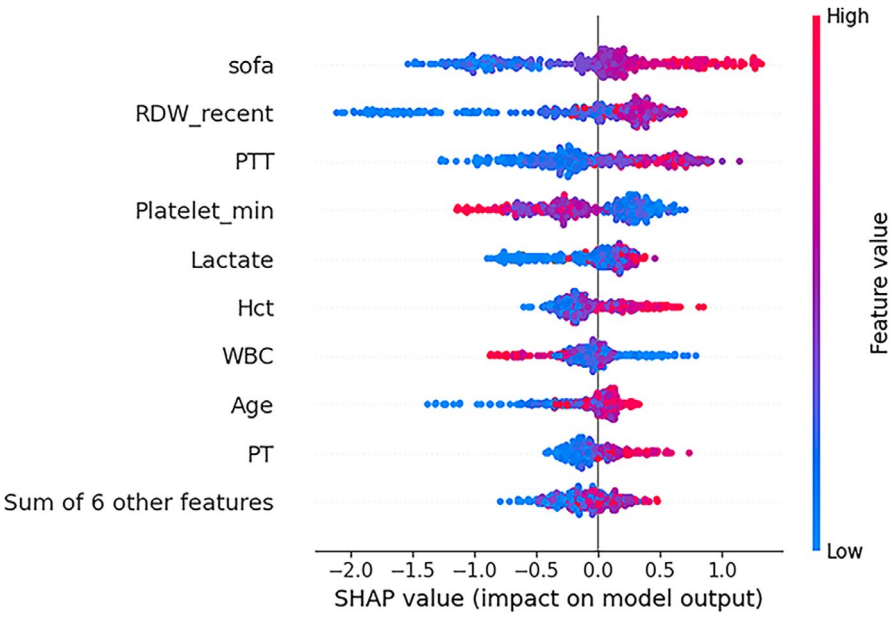

FIGURE 2 SHapley Additive exPlanations values, the SHAP summary plot figure with the top nine features out of the 15 features, and they are sorted in descending order according to their importance for the result. The positive SHAP value indicates the positive contribution is made to the model, while the negative value means the opposite

Propensity score matching (PSM) was frequently conducted by the clinicians when deciding the prediction of death for each septic shock patient with thrombocytopenia on admission (age, gender, comorbidities, RDW, Hb, Hct, WBC, platelet count, PT, APTT, lactate, and SOFA score) to balance patient characteristics and to generate a propensity-matched analysis for the two groups (Table 3). In

TABLE 1 Demographics and baseline characteristics of septic shock patients with thrombocytopenia on admission

	Survivors (*N* = 244)	Non-survivors (*N* = 227)	*p*-value
Characters			
Age, years	59.1 ± 13.2	60.3 ± 11.1	0.27
Male sex, *n* (%)	141 (57.8)	133 (58.6)	0.86
Comorbid diseases			
Hypertension, *n* (%)	117 (48.0)	102 (44.9)	0.51
Diabetes, *n* (%)	82 (33.6)	85 (37.4)	0.38
Stroke, *n* (%)	6 (2.5)	3 (1.3)	0.37
Heart diseases, *n* (%)	127 (52.0)	114 (50.2)	0.69
Laboratory findings			
RDW	16.1 ± 2.6	17.4 ± 2.7	<0.001
Hb, g/dl	10.3 ± 2.0	10.2 ± 2.2	0.94
Hct, (%)	30.9 ± 6.0	30.7 ± 6.7	0.80
WBC, $\times 10^9$/L	10.6 (6.7–17.0)	9.9 (5.4–16.1)	0.41
Platelet count, $\times 10^9$/L	88 (48–118)	58 (35–91)	<0.001
PT, s	16.0 (14.1–19.0)	18.4 (15.6–23.9)	<0.001
APTT, s	36.3 (30.1–44.8)	44.3 (35.0–56.1)	<0.001
Lactate, mmol/L	2.2 (1.5–3.3)	3.2 (2.0–5.7)	<0.001
SOFA score	8 (5–11)	11 (9–14)	<0.001

Abbreviations: APTT, activated partial thromboplastin time; Hb, hemoglobin; Hct, hematocrit; PT, prothrombin time; RDW, red cell distribution width; SOFA, sequential organ failure assessment; WBC, White blood cells.

the matched group, when comparing survivors and non-survivors, we found only RDW (17.2 ± 2.6, $p = 0.001$) was significantly increased in non-survivors.

The ROC curve for RDW, lactate, and SOFA score was associated with mortality of septic shock patients with thrombocytopenia (Figure S1). The area under curve of RDW, lactate, and SOFA score was 0.646, (95% CI: 0.584–0.708); 0.549, (95% CI: 0.485–0.614); 0.563, (95% CI: 0.499–0.627), respectively. RDW was the strongest predictor for mortality of patients with thrombocytopenia on admission. Based on ROC analysis, RDW of 16.05 presented the best sensitivity and specificity in the prediction of mortality (sensitivity 70%, specificity 57%, and area under the curve 0.646, Figure S1). Therefore, we defined the RDW cutoff point as 16.05.

Kaplan-Meier survival curves revealed that the patients with RDW less than 16.05 had better prognosis for both the patients with and without thrombocytopenia on admission. (Figure S2).

4 | DISCUSSION

In this retrospective cohort study, we found an association between RDW and clinical outcomes of septic shock patients with thrombocytopenia. Firstly, we compared basic characters between septic shock patients with and without thrombocytopenia on admission, and we found that the mortality of patients with thrombocytopenia on admission was significantly high. Secondly, we sought to investigate the risk factors leading to the death of septic shock patients with thrombocytopenia on admission, and we revealed that age, RDW, lactate, and SOFA score were independent predictors through logistic regression analysis. Besides, the SHAP-based interpretation of the XGBoost model indicated that RDW was the second most important in predicting 28-day death after SOFA. Lastly, after adjustment for covariates by PSM, RDW still showed a good capacity to predict the mortality of patients with thrombocytopenia. In addition, receiver-operating characteristic curve analysis also demonstrated that RDW could predict the mortality of septic shock patients with thrombocytopenia better than others. To our knowledge, this is a unique study focusing on the relationship between RDW and septic shock patients with thrombocytopenia.

Artificial intelligence was widely used in the early warning and mortality prediction of many diseases. Models for predicting mortality in sepsis and septic shock can already achieve high-performance results.[30,31] However, to find the risk factors for the poor prognosis of illness, interpretable machine learning methods are needed, which is the shortcoming of machine learning. SHAP can show the importance of each factor intuitively, which performs better than investigation on the generalizability and interpretability of the proposed model.[32] SHAP has been used in real-time accident detection and feature analysis,[33] but it is rarely used in medical science. In our study, machine learning revealed that RDW was the most important in predicting the 28-day death of patients with SAT except the SOFA score.

Sepsis is now defined as life-threatening organ dysfunction caused by a dysregulated host response to infection, and septic shock occurs when vasopressor is used due to persistent hypotension in the absence of hypovolemia.[1] The initial hyperinflammatory

TABLE 2 Univariable and multivariable analysis to identify the independent predictors of mortality related to septic shock patients with thrombocytopenia on admission

	Univariate analysis		Multivariate analysis	
Variables	**Odds ratio (95% CI)**	***p*-value**	**Odds ratio (95% CI)**	***p*-value**
Age	1.02 (1.00–1.04)	0.03	1.02 (1.00–1.04)	0.03
Sex	1.14 (0.74–1.75)	0.55		
Hypertension	1.31 (0.84–2.04)	0.24		
Diabetes	0.80 (0.50–1.27)	0.35		
Stroke	0.96 (0.22–4.19)	0.96		
RDW	1.21 (1.10–1.33)	<0.001	1.22 (1.11–1.33)	<0.001
Hb, g/dl	0.89 (0.63–1.26)	0.51		
Hct, (%)	1.08 (0.96–1.21)	0.20		
WBC, $\times 10^9$/L	0.99 (0.97–1.01)	0.39		
Platelet count, $\times 10^9$/L	1.02 (1.00–1.04)	0.31		
PT, s	1.14 (0.74–1.75)	0.63		
APTT, s	1.31 (0.84–2.04)	0.49		
Lactate, mmol/L	0.80 (0.50–1.27)	0.01	1.14 (1.05–1.24)	<0.001
SOFA score	0.96 (0.22–4.19)	<0.001	1.21 (1.14–1.28)	<0.001

Abbreviations: APTT, activated partial thromboplastin time; Hb, hemoglobin; Hct, hematocrit; PT, prothrombin time; RDW, red cell distribution width; SOFA, sequential organ failure assessment; WBC, White blood cells.

TABLE 3 Demographics and baseline characteristics of septic shock patients with thrombocytopenia on admission after propensity score matching

	Survivors (*N* = 153)	**Non-survivors (*N* = 153)**	***p*-value**
Characters			
Age, years	60.1 ± 12.5	60.7 ± 10.9	0.65
Male sex, *n* (%)	87 (56.9)	86 (56.2)	0.91
Comorbid diseases			
Hypertension, *n* (%)	75 (49.0)	68 (44.4)	0.42
Diabetes, *n* (%)	52 (34.0)	55 (35.9)	0.72
Stroke, *n* (%)	2 (1.3)	2 (1.3)	1.00
Heart diseases, *n* (%)	83 (54.2)	78 (51.0)	0.57
Laboratory findings			
RDW	16.3 ± 2.6	17.2 ± 2.6	0.001
Hb, g/dl	10.3 ± 2.1	10.5 ± 2.2	0.50
Hct, (%)	30.7 ± 6.1	31.4 ± 6.8	0.37
WBC, $\times 10^9$/L	10.6 (6.5–17.5)	10.7 (5.8–17.9)	0.05
Platelet count, $\times 10^9$/L	69.0 (38.5–107.5)	61.0 (38.0–100.5)	0.42
PT, s	16.4 (14.3–20.0)	17.5 (15.0–22.6)	0.13
APTT, s	38.4 (31.8–46.0)	41.9 (33.9–53.9)	0.14
Lactate, mmol/L	2.5 (1.6–4.2)	2.8 (1.8–4.9)	0.25
SOFA score	9 (7–12)	10 (8–13)	0.20

Abbreviations: APTT, activated partial thromboplastin time; Hb, hemoglobin; Hct, hematocrit; PT, prothrombin time; RDW, red cell distribution width; SOFA, sequential organ failure assessment; WBC, White blood cells.

response and hypotension due to maldistribution of blood flow will lead to multiple organ dysfunction, such as, ARDS, AKI, DIC, and, of course, including thrombocytopenia.[34] Although the SOFA score plays an important role in assessing the severity of septic patients, it takes more time because of relatively complicated items. So, we aimed to investigate the value of RDW, which is more convenient and inexpensive in predicting the prognosis of patients with SAT.

The incidence of thrombocytopenia in ICU is about 35%–44%, and sepsis is the most common risk factor for thrombocytopenia.[7,8,35] As reported as SAT, not only the mortality, but also the

length of stay, longer duration of organ support, and major bleeding events were also increased.[15] The mechanisms that contribute to thrombocytopenia are decreasing in the production combines with increasing platelets consumption and destruction.[36] On the one hand, thrombin-mediated platelet activation, such as disseminated intravascular coagulation (DIC), may occur in severe sepsis.[37] On the other hand, acquired hemophagocytic lymphohistiocytosis (HLH) may exist in the circumstance of overproducing cytokines by dysregulated activation and proliferation of lymphocytes.[38] In addition, activated platelets can promote neutrophil recruitment[39] and the formation of neutrophil extracellular traps (NETs) to trap and kill pathogens.[40] Besides, many inflammatory mediators induced in sepsis can directly inhibit or inactivate a disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13 (ADAMTS-13) and results in disseminated platelet-/VWF-rich microthrombi, which used to be called thrombocytopenia-associated multiple organ failure (TAMOF). Multiple organ failure and severe platelet depletion are caused by extensive microthrombus; the same pathological process can also be seen in thrombotic thrombocytopenic purpura (TTP), and hemolytic uremic syndrome (HUS).[41-43] All of these can lead to severe thrombocytopenia. Identifying and treating high-risk septic shock patients with thrombocytopenia earlier may improve prognosis.

Red cell distribution width is calculated as the size of RBC divided by the mean corpuscular volume, and the release of reticulocytes into the circulation may lead to an increase in RDW. In the critical care unit, RDW is very strongly associated with the risk of death and bloodstream infection.[44] As shown in our study, elevated RDW on admission was associated with higher mortality of septic shock patients with thrombocytopenia. In addition, the potency of RDW to predict mortality was almost equal to the classical indicators, such as, lactate and SOFA score. After adjusting confounders (Table 3) by PSM, RDW also elevated significantly in patients with thrombocytopenia on admission, and it showed moderate predictive power for 28-day mortality through ROC analysis. The possible mechanisms are as follows, though the specific one is still unclear. An elevated RDW may indicate an intense inflammation in the body. In outpatients, increasing RDW means a higher level of inflammation markers.[18] Previously, there were researches of a relationship between inflammation and erythropoiesis, such as promotion of red cell apoptosis, myelosuppression of erythroid precursors, reduction of erythropoietin production, reduced bioavailability of iron, and erythropoietin resistance in erythroid precursor cell lines.[45,46] Sepsis can accentuate the suppression of erythrocyte maturation through up-regulating inflammatory cytokines.[47] So, immature RBCs may be released into circulation results in an elevation of RDW. Besides, oxidative stress may be another factor contributing to the RDW-mortality association. High oxidative stress leading to elevated RDW through increasing the destruction of normal RBCs and the release of large premature RBCs into the circulation can be observed in sepsis.[48] Although RDW will increase in the status of inflammation and oxidative stress as mentioned above, it is also elevated in some non-inflammatory diseases, such as acute coronary syndromes among elderly patients, anemia in postmenopausal women, and non-ST-elevation myocardial infarction.[49-51] We infer that only combined with other indicators can take maximum advantage of RDW.

As mentioned above, SAT can be related to the consumption and destruction of platelets because of excessive inflammation, especially in septic shock patients. According to our study, RDW can be used as a reliable and convenient indicator to predict mortality of septic shock patients with thrombocytopenia. However, our study had some limitations. First, it was a single-center retrospective study, and the findings should be further confirmed by multi-center prospective studies. Second, our data were abstracted from a public database, and some variables were unavailable or missing too much. For example, we intended to explore the relationship of RDW, CRP, interleukin, and procalcitonin, which were missed. Lastly, confounders that were not included in our study still existed, although a lot of covariates were well-balanced by PSM. Of course, our machine learning method lacks other verification methods.

5 | CONCLUSIONS

Interpretable machine learning approach can be well applied in clinical research and can reveal risk factors for poor prognosis in critical patients conveniently and accurately. Elevated RDW on admission is common in patients with septic shock, especially in patients complicated with thrombocytopenia. Increased RDW was associated with mortality in septic shock patients with thrombocytopenia. Further studies with detailed data should focus on elucidating mechanisms and how to use this marker to evaluate patients with thrombocytopenia so that appropriate treatment can be implemented early to improve outcomes.

ACKNOWLEDGEMENTS

There were no sources of support for this project. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. No individual was paid to provide manuscript advice, writing, or editing. Only those authors listed had control over the data, how it was analyzed, interpreted, and presented within the manuscript.

CONFLICT OF INTEREST

All authors have no conflict of interest.

AUTHOR CONTRIBUTIONS

F Lu and MH Fang designed the study and revised the paper; JM Ling and TZ Liao extracted data from database and drafted the manuscript; YQ Wu and ZH Wang analyzed data and created diagrams.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Minghao Fang https://orcid.org/0000-0001-5310-8128

REFERENCES

1. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. 2016;315:801-810.
2. Engel C, Brunkhorst FM, Bone HG, et al. Epidemiology of sepsis in Germany: results from a national prospective multicenter study. *Intensive Care Med*. 2007;33:606-618.
3. Mira JC, Gentile LF, Mathias BJ, et al. Sepsis pathophysiology, chronic critical illness, and persistent inflammation-immunosuppression and catabolism syndrome. *Crit Care Med*. 2017;45:253-262.
4. Kakihana Y, Ito T, Nakahara M, Yamaguchi K, Yasuda T. Sepsis-induced myocardial dysfunction: pathophysiology and management. *J Intensive Care*. 2016;4:22.
5. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the working group on sepsis-related problems of the European society of intensive care medicine. *Intensive Care Med*. 1996;22:707-710.
6. Vincent JL, de Mendonça A, Cantraine F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European society of intensive care medicine. *Crit Care Med*. 1998;26:1793-1800.
7. Vanderschueren S, De Weerdt A, Malbrain M, et al. Thrombocytopenia and prognosis in intensive care. *Crit Care Med*. 2000;28:1871-1876.
8. Mavrommatis AC, Theodoridis T, Orfanidou A, Roussos C, Christopoulou-Kokkinou V, Zakynthinos S. Coagulation system and platelets are fully activated in uncomplicated sepsis. *Crit Care Med*. 2000;28:451-457.
9. de Stoppelaar SF, van 't Veer C, Poll TVD. The role of platelets in sepsis. *Thromb Haemost*. 2014;112:666-677.
10. Segre E, Pigozzi L, Lison D, et al. May thrombopoietin be a useful marker of sepsis severity assessment in patients with SIRS entering the emergency department? *Clin Chem Lab Med*. 2014;52:1479-1483.
11. Zakynthinos SG, Papanikolaou S, Theodoridis T, et al. Sepsis severity is the major determinant of circulating thrombopoietin levels in septic patients. *Crit Care Med*. 2004;32:1004-1010.
12. Thiolliere F, Serre-Sapin AF, Reignier J, et al. Epidemiology and outcome of thrombocytopenic patients in the intensive care unit: results of a prospective multicenter study. *Intensive Care Med*. 2013;39:1460-1468.
13. Eissa DS, El-Farrash RA. New insights into thrombopoiesis in neonatal sepsis. *Platelets*. 2013;24:122-128.
14. Kim CH, Kim SJ, Lee MJ, et al. An increase in mean platelet volume from baseline is associated with mortality in patients with severe sepsis or septic shock. *PLoS One*. 2015;10:e0119437.
15. Menard CE, Kumar A, Houston DS, et al. Evolution and impact of thrombocytopenia in septic shock: a retrospective cohort study. *Crit Care Med*. 2019;47:558-565.
16. Perlstein TS, Weuve J, Pfeffer MA, Beckman JA. Red blood cell distribution width and mortality risk in a community-based prospective cohort. *Arch Intern Med*. 2009;169:588-594.
17. Aktas G, Sit M, Karagoz I, et al. Could red cell distribution width be a marker of thyroid cancer? *J Coll Physicians Surg Pak*. 2017;27:556-558.
18. Lippi G, Targher G, Montagnana M, Salvagno GL, Zoppini G, Guidi GC. Relation between red blood cell distribution width and inflammatory biomarkers in a large cohort of unselected outpatients. *Arch Pathol Lab Med*. 2009;133:628-632.
19. Aktas G. Hematological predictors of novel coronavirus infection. *Rev Assoc Med Bras (1992)*. 2021;67(suppl 1):1-2.
20. Aktas G, Alcelik A, Tekce BK, Tekelioglu V, Sit M, Savli H. Red cell distribution width and mean platelet volume in patients with irritable bowel syndrome. *Prz Gastroenterol*. 2014;9:160-163.
21. Aktas G, Sit M, Dikbas O, et al. Could red cell distribution width be a marker in Hashimoto's thyroiditis? *Exp Clin Endocrinol Diabetes*. 2014;122:572-574.
22. Bilgin S, Aktas G, Zahid Kocak M, et al. Association between novel inflammatory markers derived from hemogram indices and metabolic parameters in type 2 diabetic men. *Aging Male*. 2020;23:923-927.
23. Tekce H, Kin Tekce B, Aktas G, Tanrisev M, Sit M. The evaluation of red cell distribution width in chronic hemodialysis patients. *Int J Nephrol*. 2014;2014:754370.
24. Dagistan Y, Dagistan E, Gezici AR, et al. Could red cell distribution width and mean platelet volume be a predictor for lumbar disc hernias? *Ideggyogyaszati Szemle*. 2016;69:411-414.
25. Cakir L, Aktas G, Mercimek B, Enginyurt O, Mercimek K. Are red cell distribution width and mean platelet volume associated with rheumatoid arthritis? *Biomed Res*. 2016;27:292-294.
26. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2:56-67.
27. Liu C, Cui X, Ackermann TM, Flamini V, Chen W, Castillo AB. Osteoblast-derived paracrine factors regulate angiogenesis in response to mechanical stimulation. *Integr Biol*. 2016;8:785-794.
28. Zhang K, Liu X, Shen J, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell*. 2020;181:1423-1433.e1411.
29. Mao B, Zhang L, Ning P, et al. Preoperative prediction for pathological grade of hepatocellular carcinoma via machine learning-based radiomics. *Eur Radiol*. 2020;30:6924-6932.
30. Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC. Prediction of sepsis patients using machine learning approach: a meta-analysis. *Comput Methods Programs Biomed*. 2019;170:1-9.
31. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med*. 2020;46:383-400.
32. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med*. 2018;46:547-553.
33. Parsa AB, Movahedi A, Taghipour H, Derrible S, Mohammadian AK. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accid Anal Prev*. 2020;136:105405.
34. Oberholzer A, Oberholzer C, Moldawer LL. Sepsis syndromes: understanding the role of innate and acquired immunity. *Shock*. 2001;16:83-96.
35. Strauss R, Wehler M, Mehler K, Kreutzer D, Koebnick C, Hahn EG. Thrombocytopenia in patients in the medical intensive care unit: bleeding prevalence, transfusion requirements, and outcome. *Crit Care Med*. 2002;30:1765-1771.
36. Vardon-Bounes F, Ruiz S, Gratacap MP, Garcia C, Payrastre B, Minville V. Platelets are critical key players in sepsis. *Int J Mol Sci*. 2019;20:3494.
37. Levi M, Ten Cate H. Disseminated intravascular coagulation. *N Engl J Med*. 1999;341:586-592.
38. Halacli B, Unver N, Halacli SO, et al. Investigation of hemophagocytic lymphohistiocytosis in severe sepsis patients. *J Crit Care*. 2016;35:185-190.
39. Grommes J, Alard JE, Drechsler M, et al. Disruption of platelet-derived chemokine heteromers prevents neutrophil extravasation in acute lung injury. *Am J Respir Crit Care Med*. 2012;185:628-636.

40. Ma AC, Kubes P. Platelets, neutrophils, and neutrophil extracellular traps (NETs) in sepsis. *J Thromb Haemost*. 2008;6:415-420.
41. Nguyen TC, Cruz MA, Carcillo JA. Thrombocytopenia-associated multiple organ failure and acute kidney injury. *Crit Care Clin*. 2015;31:661-674.
42. Nguyen TC, Carcillo JA. Bench-to-bedside review: thrombocytopenia-associated multiple organ failure–a newly appreciated syndrome in the critically ill. *Crit Care*. 2006;10:235.
43. Fortenberry JD, Nguyen T, Grunwell JR, et al. Therapeutic plasma exchange in children with thrombocytopenia-associated multiple organ failure: the thrombocytopenia-associated multiple organ failure network prospective experience. *Crit Care Med*. 2019;47:e173-e181.
44. Bazick HS, Chang D, Mahadevappa K, Gibbons FK, Christopher KB. Red cell distribution width and all-cause mortality in critically ill patients. *Crit Care Med*. 2011;39:1913-1921.
45. Ghali JK. Anemia and heart failure. *Curr Opin Cardiol*. 2009;24:172-178.
46. Laftah AH, Sharma N, Brookes MJ, et al. Tumour necrosis factor alpha causes hypoferraemia and reduced intestinal iron absorption in mice. *Biochem J*. 2006;397:61-67.
47. Rogiers P, Zhang H, Leeman M, et al. Erythropoietin response is blunted in critically ill patients. *Intensive Care Med*. 1997;23:159-162.
48. Ghaffari S. Oxidative stress in the regulation of normal and neoplastic hematopoiesis. *Antioxid Redox Signal*. 2008;10:1923-1940.
49. Duman TT, Aktas G, Meryem Atak B, Kocak MZ, Kurtkulagi O, Bilgin S. General characteristics of anemia in postmenopausal women and elderly men. *Aging Male*. 2020;23:780-784.
50. Sincer I, Gunes Y, Mansiroglu AK, Aktas G. Differential value of eosinophil count in acute coronary syndrome among elderly patients. *Aging Male*. 2020;23:958-961.
51. Sincer I, Mansiroglu AK, Aktas G, Gunes Y, Kocak MZ. Association between hemogram parameters and coronary collateral development in subjects with non-ST-elevation myocardial infarction. *Rev Assoc Med Bras*. 1992;2020(66):160-165.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Ling J, Liao T, Wu Y, et al. Predictive value of red blood cell distribution width in septic shock patients with thrombocytopenia: A retrospective study using machine learning. *J Clin Lab Anal*. 2021;35:e24053. https://doi.org/10.1002/jcla.24053