

ORIGINAL ARTICLE

A machine learning model for early and accurate prediction of overt disseminated intravascular coagulation before its progression to an overt stage

Yutaka Umemura^{1,2}  | Naoki Okada^{1,3} | Hiroshi Ogura² | Jun Oda² | Satoshi Fujimi¹

¹Division of Trauma and Surgical Critical Care, Osaka General Medical Center, Osaka, Japan

²Department of Traumatology and Acute Critical Medicine, Osaka University Graduate School of Medicine, Osaka, Japan

³Division of Medical Informatics, Kyoto University Graduate School of Informatics, Kyoto, Japan

Correspondence

Yutaka Umemura, Division of Trauma and Surgical Critical Care, Osaka General Medical Center, 3-1-56 Bandai-Higashi, Sumiyoshi, Osaka 558-8558, Japan.
Email: plum00022@gmail.com

Handling Editor: Michael Makris

Abstract

Background: Recent studies suggested an expected survival benefit associated with anticoagulant therapies for sepsis in patients with disseminated intravascular coagulation (DIC). However, anticoagulant therapies for overt DIC are no longer assumed to regulate pathologic progression as overt DIC is a late-phase coagulation disorder. Therefore, methods for early prediction of sepsis-induced DIC before its progression to an overt stage are strongly required.

Objectives: We aimed to develop a prediction model for overt DIC using machine learning.

Methods: This retrospective, observational study included adult septic patients without overt DIC. The objective variable was binary classification of whether patients developed overt DIC based on International Society on Thrombosis and Haemostasis (ISTH) overt DIC criteria. Explanatory variables were the baseline and time series data within 7 days from sepsis diagnosis. Light Gradient Boosted Machine method was used to construct the prediction model. For controls, we assessed sensitivity and specificity of Japanese Association for Acute Medicine DIC criteria and ISTH sepsis-induced coagulopathy criteria for subsequent onset of overt DIC.

Results: Among 912 patients with sepsis, 139 patients developed overt DIC within 7 days from diagnosis of sepsis. Sensitivity, specificity, and area under the receiver operating characteristic curve for predicting onset of overt DIC within 7 days were 84.4%, 87.5%, and 0.867 in the test cohort and 95.0%, 75.9%, and 0.851 in the validation cohort, respectively. Sensitivity and specificity by the diagnostic thresholds were 54.7% and 74.9% for Japanese Association for Acute Medicine DIC criteria and 63.3% and 71.9% for ISTH sepsis-induced coagulopathy criteria, respectively.

Conclusion: Compared with conventional DIC scoring systems, a machine learning model might exhibit higher prediction accuracy.

KEYWORDS

biomarkers, diagnosis, disseminated intravascular coagulation, machine learning, sepsis

Essentials

- Methods to predict sepsis-induced overt disseminated intravascular coagulation (DIC) before its progression to an overt stage are required.
- This was a retrospective, observational study including adult septic patients without overt DIC.
- Machine learning model exhibited higher prediction values compared with conventional screening.
- Parameters of circulatory function were responsible for the sequential progression of overt DIC.

1 | INTRODUCTION

Sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection [1]. Despite the progress made in its medical management over the past few decades, sepsis remains an important global health problem, causing millions of deaths around the world [2,3].

In sepsis, blood coagulation disorders play a key role in the development of microcirculatory dysfunction, multiorgan dysfunction, and subsequent death. Anticoagulant therapies were thus once expected to be beneficial adjunctive therapy against the progression of overwhelming activation of the coagulation system. Although numerous randomized controlled trials so far have evaluated the efficacy of anticoagulant therapies for sepsis over the past few decades, there is still insufficient evidence to support the survival benefit of anticoagulant therapy in sepsis [4–9]. However, most of these trials included septic patients overall, both those with and without disseminated intravascular coagulation (DIC). Recent meta-analyses and several observational studies reported anticoagulant therapy to be associated with lower mortality when it was administered to the population with sepsis-induced DIC [10,11]. These lines of evidence clearly suggested that a survival benefit associated with anticoagulant therapies was expected only in patients with sepsis-induced DIC and not in the overall patient population with sepsis.

Currently, the ISTH overt DIC criteria are the global standard for the diagnosis of DIC and are superior for definitely diagnosing patients as having DIC after excluding other conditions that should be differentiated from DIC. However, as overt DIC is a late-phase coagulation disorder characterized by platelet and fibrinogen consumption, anticoagulant therapies were assumed to no longer regulate the pathologic progression to the advanced stage. Ideally, anticoagulant therapy should be administered to patients with sepsis-induced overt DIC before they develop it. In these contexts, methods for early prediction of sepsis-induced DIC before its progression to an overt stage are strongly required.

Artificial intelligence (AI) technology (especially machine learning technology) is the scientific discipline that uses computer algorithms to help identify patterns of data and make predictions [12]. In the medical field, AI technology supports analysis of large and complex data structures to create prediction models that personalize and improve diagnosis, prognosis, monitoring, and administration of treatments. Prediction algorithms using machine learning technology have been successfully applied in several medical areas, such as coronary heart disease [13], cardiovascular disease [14], end-stage liver

disease [15], and cerebrovascular diseases [16], to evaluate prognosis and risk factors.

Herein, we hypothesized that machine learning technology would enable the establishment of a high-accuracy algorithm to predict overt DIC before its progression to an overt stage based on general clinical data. In this study, we thus aimed to develop a prediction model for overt DIC using machine learning technology and compare its accuracy with that in the conventional scoring systems.

2 | METHODS

2.1 | Design and setting

This was a single-center, retrospective, observational study conducted in an intensive care unit (ICU) of a tertiary care hospital in Japan. We consecutively included adult septic patients without overt DIC at the time of diagnosis of sepsis from January 2014 to January 2021.

This study followed the principles of the Declaration of Helsinki and the Japanese Clinical Trials Act (Act number 16 of April 14, 2017). The protocol was approved by the Institutional Review Board for Clinical Research of Osaka General Medical Center (IRB number 2022-057). The board waived the requirement for informed consent because of the anonymous nature of the data. Instead, participants had the opportunity to opt out of this research on the medical center's website.

2.2 | Participants

Patients were eligible for this study if they met all of the following criteria: 1) were >20 years old; 2) were diagnosed with sepsis according to the Sepsis-3 criteria [1], ie, had a proven or suspected infection and an acute increase of 2 or more points in the Sequential Organ Failure Assessment (SOFA) score; 3) were admitted to the ICU; and 4) were diagnosed as not having DIC according to the ISTH overt DIC criteria at the time of diagnosis of sepsis [17].

The exclusion criteria were the daily use of warfarin/acetylsalicylic acid or other anticoagulant/thrombolytic therapy before study entry, use of anticoagulant therapies to treat thromboembolic disease within 7 days before the diagnosis of DIC, the limitation of sustained life care, postcardiopulmonary arrest resuscitation status, death within 24 hours after diagnosis of sepsis, decompensated liver

cirrhosis, history of metastatic malignant disease, and chemotherapies and other conditions affecting the activity of the coagulation system. Patients who received low-dose heparin for prophylaxis against deep vein thrombosis or who received intradialytic anticoagulation were not excluded. All patients were typically treated according to the Surviving Sepsis Campaign Guidelines.

2.3 | Data collection

We collected the following baseline and time series data from day 1 to day 7 before the diagnosis of sepsis: age, sex, height, weight, pre-existing comorbidities [18], site of infection, use and amount of catecholamines, use of mechanical ventilation, use of corticosteroids, hourly volume of crystalloids administration, use and amount of albumin, hourly urine output, and laboratory tests.

The severity of illness was evaluated at the time of ICU admission according to the Acute Physiology and Chronic Health Evaluation II and SOFA scores. Severity of coagulopathy was evaluated on the basis of the ISTH sepsis-induced coagulopathy (SIC) criteria [19], the ISTH overt DIC criteria, and the Japanese Association for Acute Medicine (JAAM) DIC criteria [20] at the time of diagnosis of sepsis and once a day from day 1 to day 7 before the diagnosis of overt DIC. For the ISTH SIC and overt DIC criteria, fibrinogen degradation products (FDPs) and D-dimer were chosen as the fibrin-related markers. Thresholds for scoring these markers were decided according to a previous study published by Gando et al. [21] (Supplementary Table S1; no increase, moderate increase, and strong increase were defined by FDP values of <10, 10 to 25, and >25 mg/L and by D-dimer values of <5, 5 to 10, and >10 mg/L, respectively [21]).

2.4 | Endpoint

The prediction endpoint was the onset of overt DIC, defined as the presence of positive ISTH overt DIC criteria from day 2 to day 7 after the diagnosis of sepsis.

2.5 | Dataset and ground truth

Patients who developed overt DIC within 7 days of the diagnosis of sepsis were classified into the positive group; otherwise, they were classified into the negative group. The explanatory variables used were the baseline data and the time series data from day 1 to day 7 in the negative group or those from day 1 to the day before patients fulfilled ISTH overt DIC criteria in the positive group (Supplementary Table S2). Patients in the positive group were randomly assigned to 3 cohorts: the training cohort, validation cohort, or test cohort. Subsequently, equal numbers of patient data were randomly extracted from the negative group and randomly assigned to the 3 cohorts. These exploratory data were preprocessed prior to training.

2.6 | Model

In this study, baseline data obtained at the time of hospital admission and time series data from day 1 to day 7 were used as explanatory variables. Baseline data included age, sex, infection sites, and pre-existing comorbidities, such as hypertension, dementia, congestive heart failure, cerebrovascular disease, chronic pulmonary disease, diabetes, cancer, liver disease, and chronic kidney disease. Time series data included vital signs, laboratory tests, therapeutic interventions (such as the use of catecholamines, mechanical ventilation, and corticosteroids), and volume of crystalloid administration (Supplementary Table S2). Machine learning was performed, setting the objective variable as binary classification of whether patients developed overt DIC at any time point from day 2 to day 7 after the diagnosis of sepsis.

The Light Gradient Boosted Machine (LightGBM) method was used to construct the prediction algorithm in this study (Figure 1). LightGBM is a supervised machine learning technique for decision tree analysis that uses a gradient boosting framework, which enhances predictive accuracy by performing the analysis multiple times to learn the differences between predicted and actual values [22].

2.7 | Training

We set the core parameters of LightGBM as follows: 1) task, train; 2) objective, regression; 3) boosting type, gbdt; 4) number of boosting iterations, 100; 5) learning_rate, 0.1; 6) maximum number of leaves per tree, 31; 7) tree_learner, serial; 8) number of threads, 0; 9) device_type, CPU; and 10) seed, none. Binary error was used as the evaluation index. For the other parameters, the default value was used. The hyperparameter was optimized using Optuna software [23]. In the training process, the early stopping mechanism [22] was used to prevent overfitting. Under these settings, a final model was selected that minimized the binary error. The seed value used for analysis was set to 1 and that of the random number generator was set by `numpy.random.seed(1)`.

2.8 | Statistical analysis

Baseline characteristics were compared between the positive and negative groups by the Mann-Whitney U-test or chi-squared test. Descriptive statistics are summarized as group medians with the first and third quartiles for continuous variables and frequencies with percentages for categorical variables. Standardized mean difference (SMD) was also calculated to examine the balance of patient characteristics between the 2 groups, where >0.1 was generally considered to indicate imbalance.

To compare the predictive value of the machine learning model with that of the conventional criteria for early detection of coagulopathy, we conducted receiver operating characteristic (ROC) analyses and assessed the sensitivity, specificity, and area under the ROC curve for the JAAM DIC and ISTH SIC criteria at the time of diagnosis of sepsis. Missing values were not imputed in the regression models. All statistical inferences were

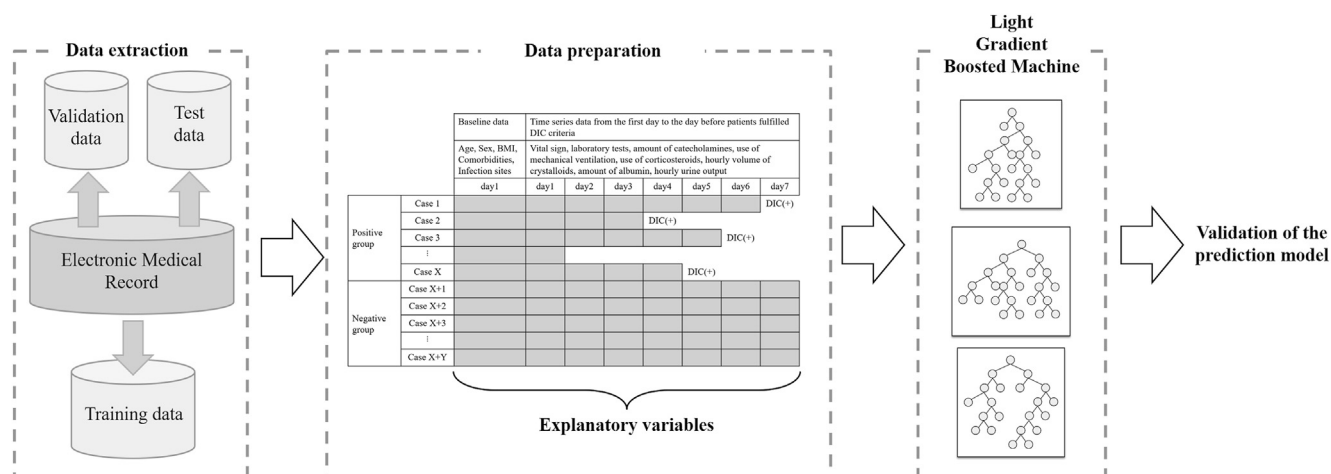


FIGURE 1 Training process. Study patients were randomly divided into the following 3 categories: training cohort, validation cohort, and test cohort. Explanatory variables were the baseline data and the time series data from day 1 to day 7 after the sepsis diagnosis in the negative group or those from day 1 to the day before patients fulfilled International Society on Thrombosis and Haemostasis overt disseminated intravascular coagulation (DIC) criteria in the positive group. The Light Gradient Boosted Machine method was used to construct the prediction model. BMI, body mass index.

2-sided, and a P value of $<.05$ indicated statistical significance. Statistical analyses were conducted using STATA Data Analysis and Statistical Software version 17.0 (StataCorp LLC).

3 | RESULTS

3.1 | Study population

The patient flow diagram is shown in [Figure 2](#). During the study period, 1156 consecutive adult patients were diagnosed as having sepsis based on the Sepsis-3 criteria and admitted to the ICU of Osaka General Medical Center. Among them, 1015 patients were diagnosed as not having overt DIC according to the ISTH overt DIC criteria at the time of diagnosis of sepsis. After excluding 103 patients who met the exclusion criteria, we included 912 patients as the final study cohort. From day 2 to day 7 after the diagnosis of sepsis, 139 patients developed overt DIC and were classified into the positive group, whereas the other 773 patients did not develop overt DIC and were classified into the negative group.

Among the 139 patients in the positive group, we randomly assigned 100 patients to the training cohort, 20 patients to the validation cohort, and 19 patients to the test cohort. Equal numbers of patient data were randomly extracted from the 773 cases in the negative group and assigned to the 3 cohorts.

3.2 | Patient characteristics

Baseline characteristics, vital signs, laboratory tests, and severity scores in the 2 groups are shown in the [Table](#). Patients in the positive group were significantly older and had significantly higher rates of mild-to-moderate liver failure, chronic kidney disease, and chronic heart failure. Distributions of the infectious sites were significantly

different between the groups ($P < .001$). Respiratory infection was the most common infection in the negative group, but the most common infectious site in the positive group was the urinary tract. Many vital signs, including the Glasgow Coma Scale score, systolic blood pressure, diastolic blood pressure, and respiratory rate, were statistically significantly different between the 2 groups.

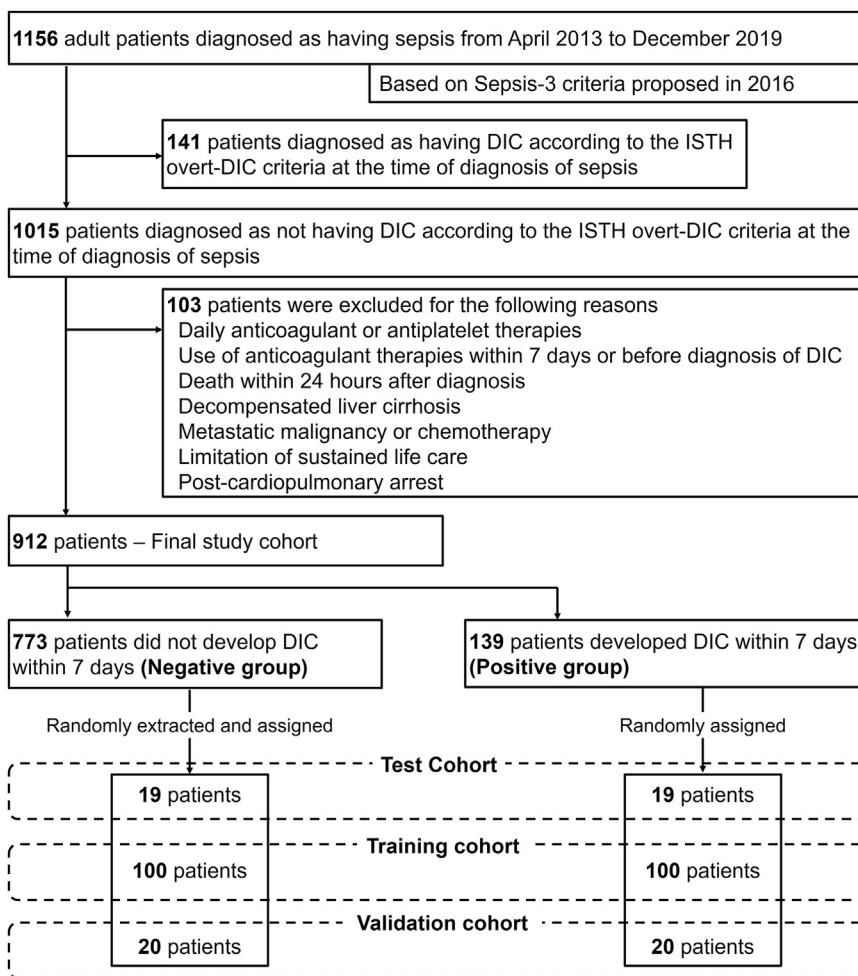
Among the coagulation markers, the positive group had significantly lower platelet counts and fibrinogen levels and higher international normalized ratios, fibrin/FDPs, and D-dimer. Compared with the negative group, the positive group also had significantly higher severity scores and DIC scores (Acute Physiology and Chronic Health Evaluation II score, median 18 vs 23; $P < .001$; SMD = 0.598; SOFA score, median 6 vs 10; $P < .001$; SMD = 0.455; JAAM DIC score, median 2 vs 4; $P < .001$; SMD = 0.288; ISTH overt DIC score, median 2 vs 3; $P < .001$; SMD = 0.176; and ISTH SIC score, median 3 vs 4; $P < .001$; SMD = 0.698). Consequently, the rates of positivity for JAAM DIC and ISTH SIC criteria were both significantly higher in the positive group. In-hospital mortality rates were 12.4% in the negative group and 36.0% in the positive group.

[Supplementary Table S3](#) shows the therapeutic interventions in the DIC-positive and DIC-negative groups. The rates of use of norepinephrine, vasopressin, mechanical ventilation, and renal replacement therapies were significantly higher in the positive group, as was the volume of crystalloids administered within 24 hours. However, the rates of low-dose unfractionated heparin administration for deep vein thrombosis prophylaxis were equal in the 2 groups ($P = .36$).

3.3 | Predictive values in AI model and conventional criteria

[Figure 3](#) presents the ROC curves of the early prediction algorithm constructed by the machine learning method. Sensitivity, specificity,

FIGURE 2 Patient flow diagram. DIC, disseminated intravascular coagulation; ISTH, International Society on Thrombosis and Haemostasis.



and area under the ROC curve for predicting the onset of overt DIC within 7 days after the sepsis diagnosis were 84.4%, 87.5%, and 0.867 in the test cohort and 95.0%, 75.9%, and 0.851 in the validation cohort, respectively. Accuracy, precision, and recall of the model were 0.74, 0.68, and 0.89, respectively. We also conducted ROC analysis for JAAM DIC and ISTH SIC criteria to compare the predictive values for overt DIC. Sensitivity and specificity in the diagnostic thresholds of the criteria (4 points for both criteria) were 54.7% and 74.9% for the JAAM DIC criteria and 63.3% and 71.9% for the ISTH SIC criteria, respectively.

3.4 | Feature importance

Figure 4 shows the “frequencies” and “gains” of covariates as feature importance in the training process. “Frequencies” represent the frequencies that the covariates are referred to when branching the decision tree in the training process, and “gains” represent the decrease in the objective functions from the decision tree after branching based on the covariate. The larger each indicator, the stronger the influence on decision making. The 3 most frequently used covariates for branching were urine output in 24 hours, lactate level, and the

cardiovascular component of the SOFA score (used 7, 5, and 5 times, respectively). Also, these 3 covariates had the highest “gains,” indicating that circulatory dysfunction was responsible for the sequential progression of overt DIC.

4 | DISCUSSION

According to the lines of evidence obtained from the previous randomized controlled trials, meta-analyses, and observational studies, a survival benefit associated with anticoagulant therapies was expected only in patients with sepsis-induced DIC [10,11,24]. However, as overt DIC is a late-phase and uncompensated coagulation disorder, reliable methods to detect patients likely to develop overt DIC at an earlier time before they progress to the overt stage are strongly required.

The present study developed a prediction model for overt DIC using machine learning technology and revealed the following: 1) the machine learning model could predict the progression to overt DIC with higher accuracy than the conventional early diagnostic criteria, and 2) circulation parameters, such as urine output and lactate level, were strongly associated with the sequential progression to overt DIC.

TABLE Baseline characteristics, vital signs, laboratory data, and severity of illness in the 2 groups.

Patient characteristics	Negative group n = 773	Positive group n = 139	P value	SMD
Age (y)	73 (61-81)	77 (65-83)	.008	1.016277
Sex, male	467 (60.4)	80 (57.6)	.53	0.058171
Body mass index (kg/m ²)	21.1 (18.3-24.4)	21.4 (18.5-24.2)	.84	0.025687
Race, ethnicity			-	-
Asian, Japanese	773 (100)	139 (100)		
Pre-existing comorbidities	-	-	-	-
Hypertension	193 (25)	32 (23)	.62	0.045582
Diabetes mellitus	192 (24.8)	42 (30.2)	.18	0.120616
Cancer	56 (7.2)	11 (7.9)	.78	0.025286
Mild-to-moderate liver failure	21 (2.7)	14 (10.1)	<.001	0.304097
Chronic kidney disease	106 (13.7)	31 (22.3)	.009	0.224945
Hemodialysis	32 (4.1)	11 (7.9)	.05	0.159083
Chronic heart failure	110 (14.2)	33 (23.7)	.005	0.244308
Chronic pulmonary disease	88 (11.4)	13 (9.4)	.48	0.066683
Site of infection	-	-	<.001	-
Respiratory system	267 (34.5)	35 (25.2)	-	0.205623
Abdomen	127 (16.4)	30 (21.6)	-	0.131627
Urinary tract	211 (27.3)	55 (39.6)	-	0.262367
Bone/soft tissue	134 (17.3)	9 (6.5)	-	0.340167
Central nervous system	15 (1.9)	3 (2.2)	-	0.015371
Cardiovascular system	13 (1.7)	3 (2.2)	-	0.034729
Others/unclassifiable/unknown	6 (0.8)	4 (2.9)	-	0.157402
Vital signs	-	-	-	-
Glasgow coma scale	10 (9-14)	10 (6-11)	<.001	1.120492
Systolic blood pressure (mm Hg)	125 (105-147)	113 (87-135)	<.001	1.088008
Diastolic blood pressure (mm Hg)	69 (56-84)	61 (45-82)	.003	0.934345
Respiratory rate (/min)	25 (21-31)	28 (22-33)	.03	0.859915
Heart rate (/min)	103 (88-119)	106 (88-129)	.08	1.01676
Body temperature (°C)	37.2 (36.3-38.2)	37 (36-38.4)	.50	1.329277
Urinary output for 24 h (mL)	1350 (750-2180)	963 (447-1625)	<.001	0.329319
Laboratory tests	-	-	-	-
White blood cell count (10 ³ /μL)	11.7 (7.6-16.8)	8.9 (4.3-14.6)	<.001	0.618032
Platelet count (10 ³ /μL)	202.5 (141.5-281)	139 (99-199)	<.001	0.825432
International normalized ratio	1.2 (1.1-1.3)	1.3 (1.2-1.5)	<.001	0.421353
Fibrin/FDPs (μg/mL)	10.1 (5.5-17.9)	18.75 (9.2-45)	<.001	0.266603
D-dimer (μg/mL)	4.2 (2.1-9.2)	9.25 (4.4-20.2)	<.001	0.291051
Fibrinogen (mg/dL)	470.5 (348-601)	369.5 (263-509)	<.001	0.952803
Sodium (mmol/L)	137 (133-140)	136 (132-140)	.908	1.339317
Potassium (mmol/L)	4.1 (3.7-4.6)	4.2 (3.7-4.8)	.13	1.065221

(Continues)

TABLE (Continued)

Patient characteristics	Negative group n = 773	Positive group n = 139	P value	SMD
Creatinine (mg/dL)	1.16 (0.78-2.04)	1.89 (1.13-3.21)	<.001	0.079825
Bilirubin (mg/dL)	0.8 (0.5-1.2)	0.9 (0.6-1.4)	.01	0.036324
Antithrombin activity (%)	68.2 (56.2-82.7)	56.4 (46.1-73.9)	<.001	1.155063
Lactate (mmol/L)	1.8 (1.1-3.4)	3.7 (2-6.3)	<.001	0.025837
Albumin (g/dL)	2.7 (2.2-3.4)	2.5 (1.9-3.1)	.003	1.057086
C-reactive protein (mg/dL)	14.27 (5.07-24.01)	13.84 (3.71-26.36)	.97	0.236019
Disease severity scores	-	-	-	-
SOFA score	6 (4-8)	10 (8-12)	<.001	0.455536
APACHE II score	18 (13-23)	23 (18-29)	<.001	0.598401
JAAM DIC score	2 (1-4)	4 (2-5)	<.001	0.288529
JAAM DIC-positive	194 (25.1)	76 (54.7)	<.001	0.633665
ISTH overt DIC score	2 (0-3)	3 (2-4)	<.001	0.176317
ISTH SIC score	3 (2-4)	4 (3-4)	<.001	0.69857
ISTH SIC-positive	217 (28.1)	88 (63.3)	<.001	1.678154

Categorical variables are expressed as *n* (%), and continuous variables are presented as medians with first and third quartiles.

APACHE, Acute Physiology and Chronic Health Evaluation; DIC, disseminated intravascular coagulation; FDP, fibrinogen degradation product; ISTH, International Society on Thrombosis and Haemostasis; JAAM, Japanese Association for Acute Medicine; SMD, standardized mean difference; SIC, sepsis-induced coagulopathy; SOFA, Sequential Organ Failure Assessment.

4.1 | Importance of early prediction in sepsis-induced DIC

In a previous multicenter observational study, 13.9% of the septic patients without overt DIC were reported to later develop overt DIC within 3 days from hospital admission [25]. Similarly, in the present study, 15.2% (139 of 912) of patients initially without overt DIC developed overt DIC within 7 days of the sepsis diagnosis, which was equal to the number of patients initially diagnosed with overt DIC. Activation of the coagulation system in sepsis changes dynamically over time, and a delay in intervention is capable of causing progression of illness that is no longer amenable to benefit from anticoagulant therapy [20,26]. Therefore, the possibility of predicting the onset of overt DIC can enable physicians to initiate appropriate, timely interventions against septic coagulopathy before it progresses to the uncompensated stage, thus leading to improvement of clinical outcomes.

However, the methods for early and appropriate prediction of sepsis-induced DIC before its progression to an overt stage are not sufficiently established. The ISTH overt DIC criteria are strictly designed to diagnose patients as definitely having overt DIC, but they are not suitable for detecting the early phase of DIC [17]. The JAAM DIC criteria were designed for the early detection of patients with DIC, especially for those with sepsis [20]. However, the JAAM DIC criteria were reported to have low sensitivity for mortality, ie, not all patients with JAAM DIC-positive criteria suffered from

life-threatening coagulation disorders requiring anticoagulant therapies [27]. Similarly, SIC was developed to predict overt DIC with high sensitivity at an earlier time, but it was also reported that only about half of SIC-positive patients subsequently developed overt DIC, and the positive predictive value of SIC might be suboptimal [28,29]. It is thus difficult to construct an early prediction system that achieves both high sensitivity and high specificity with the conventional diagnostic scoring methods. As shown in this study, machine learning methods could be a key technology in constructing a prediction model having clinically reliable accuracy.

Generalization possibility is another key factor in designing an early prediction system. For example, hematological molecular markers, such as thrombin-antithrombin complex, antithrombin III activity, and plasminogen activator inhibitor-1, have been reported to aid in making an earlier and accurate diagnosis of DIC. However, these measurements are costly, still performed only in limited facilities, and are not widely used as a component of DIC diagnostic criteria. The prediction algorithm composed in the present study is based on clinical markers that would generally be obtained in the management of sepsis and thus can be used anywhere by implementing it on electronic medical record systems.

The ultimate goal of diagnosis is not to predict an accurate prognosis but to initiate the best treatment at the best time to improve clinical outcomes. By using machine learning algorithms, it might be possible to accurately detect a progressive coagulopathy before its progression regardless of country, region, or institution in

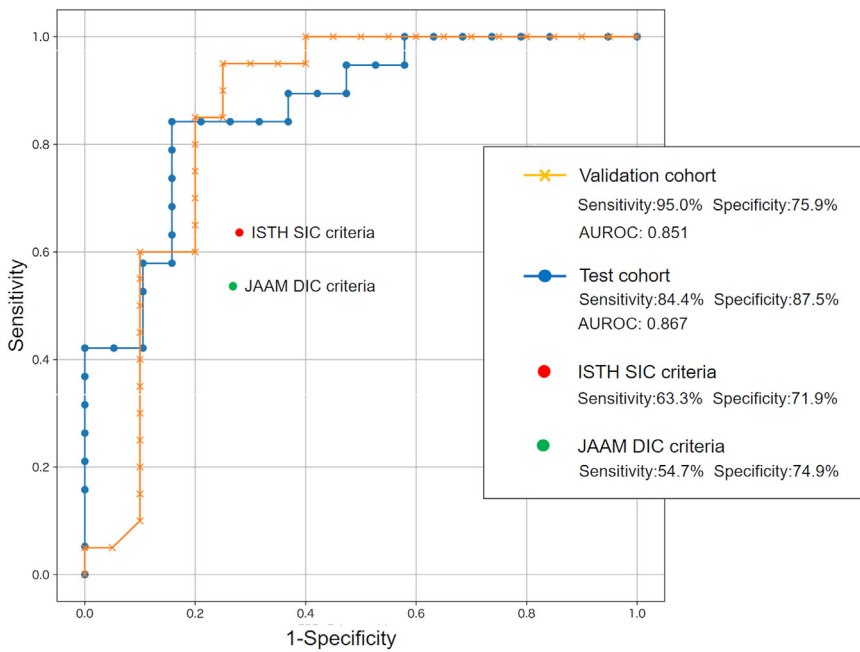


FIGURE 3 Prediction accuracies of the machine learning model. AUROC, area under the receiver operating characteristic curve; DIC, disseminated intravascular coagulation; ISTH, International Society on Thrombosis and Haemostasis; JAAM, Japanese Association for Acute Medicine; SIC, sepsis-induced coagulopathy.

which a patient is located, thus leading to the development of a scientific infrastructure that can verify the true effect of anticoagulant therapy.

4.2 | Enhancement of prediction accuracy

The purpose of this study was to develop a machine learning model that can be widely used in clinical practice settings around the world. It thus required construction of a model with the highest possible accuracy that minimizes the load on the central processing unit rather than a model with the highest accuracy but that heavily loads the central processing unit. As the amount of computation dramatically increases depending on the amount of data in conventional gradient

boosting methods, LightGBM was developed to allow construction of each decision tree at high speed without reducing the accuracy of each decision tree as much as possible. Therefore, LightGBM is one of the best methods for this study purpose as it can handle the time series parameters and construct each decision tree at high speed with as little reduction in the accuracy of each decision tree as possible.

Nevertheless, several future tasks are required to raise the predictive accuracy of machine learning algorithms to a clinically reliable level. First, it will be necessary to modify the learning setting more optimally. In this study, the high recall of 0.89 in the machine learning model represented the detection of patients who subsequently developed overt DIC with minimal omissions. However, one concern is that the prediction could include a substantial number of patients with nonprogressive coagulopathy because model precision was relatively

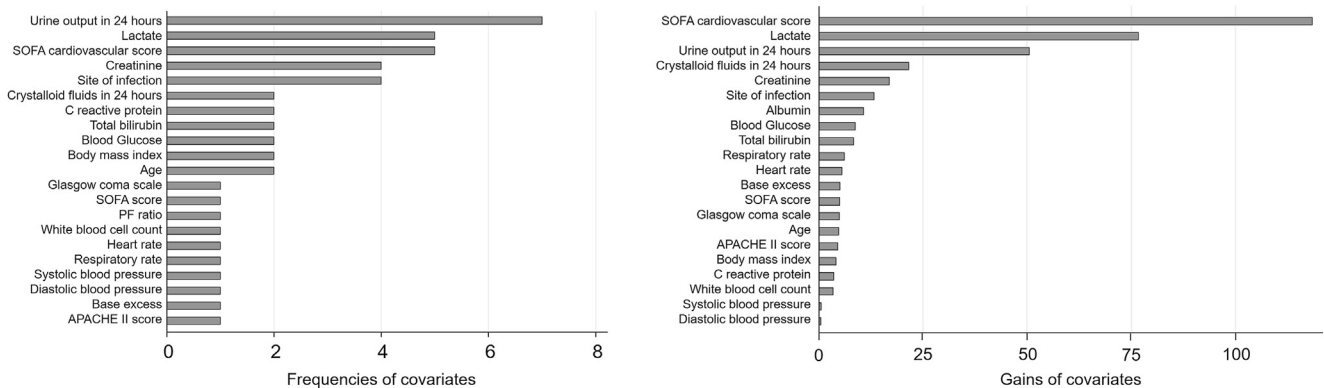


FIGURE 4 Feature importance in the training process. Frequencies represent the frequencies that the covariates are referred to when branching the decision tree in the training process, and gains represent the decrease of the objective functions from the decision tree after branching based on the covariate. APACHE, Acute Physiology and Chronic Health Evaluation; PF, PaO2/FiO2; SOFA, Sequential Organ Failure Assessment.

low (0.68). When actually used in the clinical setting, it will be necessary to increase precision while maintaining recall by adjusting the learning settings.

Second, it will be necessary to optimize the weights of the covariates used to construct the prediction model. The present results suggested that several circulatory-related parameters, such as lactate and urinary output, might especially influence the progression of SIC. By feeding back the results and weighting explanatory variables according to their association with the progression of overt DIC, prediction accuracy would be enhanced when constructing the next model.

Finally, much more patient data will be required to enhance algorithm accuracy. Based on the results of the present study, construction of a clinically reliable model (eg, where both sensitivity and specificity are 95%) will be assumed to require at least 1000 events. As 15% of the septic patients without DIC later developed overt DIC, we need to construct a larger dataset that includes more than 6000 septic patients. Thus, we are currently building a multicenter registry and preparing to perform further clinical research.

4.3 | Limitations

This study has several limitations. First, retrospective datasets are generally less accurate than those in prospective studies. This was a single-center study, and thus, the value of the model could not be externally validated, which could potentially lead to misclassification and overfitting of the model. Due to the low sample size, dividing the study patients to conduct external validation would cause a serious decrease in the quality of learning. Extrapolation of the model needs to be evaluated by a further multicenter study, which is currently being planned. Second, we defined the onset of overt DIC as positivity of ISTH overt DIC criteria; however, onset might be affected not only by SIC but also by other mechanisms of coagulation disorder, such as liver failure, blood loss, or therapeutic drugs. Third, the current version of the prediction model would not be clinically practical for initiating treatment because the timing of DIC onset was broadly defined from day 2 to day 7 after the sepsis diagnosis to assess the progression of coagulopathy over time. Fourth, selection bias may exist due to the study exclusion criteria, which excluded patients who died within 24 hours or received anticoagulant therapies to treat thromboembolic disease other than DIC. Fifth, as the number of cases in the positive group was much lower than that in the negative group, sampling of the cases used for learning potentially caused a bias. Although random sampling was performed with a presetting algorithm, an increase in the number of samples and extrapolation verification are the ultimate solutions to eliminate this bias. In the future, it will thus be necessary to improve the clinical practicality of the model and its prediction accuracy by concomitant use of other learning methods.

5 | CONCLUSION

Compared with the conventional DIC scoring systems, the machine learning model created in this study exhibited higher prediction

accuracy. Machine learning technology will be useful in constructing methods to predict overt DIC at an earlier phase before its progression to the overt stage.

ACKNOWLEDGMENTS

The authors thank all of the nurses and physicians in the participating institutions and all of the patients who contributed to this study.

FUNDING

The authors declare that they have no sources of funding to report.

AUTHOR CONTRIBUTIONS

Y.U. conceived and designed this study; contributed to the acquisition and shaping of the data, analyses, and interpretation of the results; and was responsible for drafting, editing, and submission of the manuscript. N.O. contributed to the study design and acquisition, analysis, and construction of the machine learning models. J.O., H.O., and S.F. contributed to the interpretation of the data and critical appraisal of the manuscript. All of the authors reviewed, discussed, and approved the final manuscript.

RELATIONSHIP DISCLOSURE

There are no competing interests to disclose.

DATA AVAILABILITY

The statistical codes and full dataset are available from the corresponding author.

ORCID

Yutaka Umemura  <https://orcid.org/0000-0001-8357-0313>

REFERENCES

- [1] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. 2016;315:801–10.
- [2] Fleischmann-Struzek C, Mellhammar L, Rose N, Cassini A, Rudd KE, Schlattmann P, et al. Incidence and mortality of hospital- and ICU-treated sepsis: results from an updated and expanded systematic review and meta-analysis. *Intensive Care Med*. 2020;46:1552–62.
- [3] Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *Lancet*. 2020;395:200–11.
- [4] Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, et al. Recombinant human protein C Worldwide Evaluation in Severe Sepsis (PROWESS) study group. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med*. 2001;344:699–709.
- [5] Warren BL, Eid A, Singer P, Pillay SS, Carl P, Novak I, et al. Caring for the critically ill patient. High-dose antithrombin III in severe sepsis: a randomized controlled trial. *JAMA*. 2001;286:1869–78.
- [6] Abraham E, Reinhart K, Svoboda P, Seibert A, Olthoff D, Dal Nogare A, et al. Assessment of the safety of recombinant tissue factor pathway inhibitor in patients with severe sepsis: a multicenter,

- randomized, placebo-controlled, single-blind, dose escalation study. *Crit Care Med.* 2001;29:2081–9.
- [7] Abraham E, Laterre PF, Garg R, Levy H, Talwar D, Trzaskoma BL, et al. Drotrecogin alfa (activated) for adults with severe sepsis and a low risk of death. *N Engl J Med.* 2005;353:1332–41.
- [8] Gando S, Saitoh D, Ishikura H, Ueyama M, Otomo Y, Oda S, et al. A randomized, controlled, multicenter trial of the effects of anti-thrombin on disseminated intravascular coagulation in patients with sepsis. *Crit Care.* 2013;17:R297. <https://doi.org/10.1186/cc13163>
- [9] Vincent JL, Ramesh MK, Ernest D, LaRosa SP, Pacht J, Aikawa N, et al. A randomized, double-blind, placebo-controlled, phase 2b study to evaluate the safety and efficacy of recombinant human soluble thrombomodulin, ART-123, in patients with sepsis and suspected disseminated intravascular coagulation. *Crit Care Med.* 2013;41:2069–79.
- [10] Umemura Y, Yamakawa K, Ogura H, Yuhara H, Fujimi S. Efficacy and safety of anticoagulant therapy in three specific populations with sepsis: a meta-analysis of randomized controlled trials. *J Thromb Haemost.* 2016;14:518–30.
- [11] Yamakawa K, Umemura Y, Hayakawa M, Kudo D, Sanui M, Takahashi H, et al. Benefit profile of anticoagulant therapy in sepsis: a nationwide multicentre registry in Japan. *Crit Care.* 2016;20:229.
- [12] Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet.* 2019;393:1577–9.
- [13] Forrest IS, Petrazzini BO, Duffy Á, Park JK, Marquez-Luna C, Jordan DM, et al. Machine learning-based marker for coronary artery disease: derivation and validation in two longitudinal cohorts. *Lancet.* 2023;401:215–25.
- [14] Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol.* 2021;18:465–78.
- [15] Yu YD, Lee KS, Man Kim J, Ryu JH, Lee JG, Lee KW, et al. Artificial intelligence for predicting survival following deceased donor liver transplantation: retrospective multi-center study. *Int J Surg.* 2022;105:106838.
- [16] Patel UK, Anwar A, Saleem S, Malik P, Rasul B, Patel K, et al. Artificial intelligence as an emerging technology in the current care of neurological disorders. *J Neurol.* 2021;268:1623–42.
- [17] Taylor FB Jr, Toh CH, Hoots WK, Wada H, Levi M. Scientific Subcommittee on Disseminated Intravascular Coagulation (DIC) of the International Society on Thrombosis and Haemostasis (ISTH). Towards definition, clinical and laboratory criteria, and a scoring system for disseminated intravascular coagulation. *Thromb Haemost.* 2001;86:1327–30.
- [18] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40:373–83.
- [19] Iba T, Nisio MD, Levy JH, Kitamura N, Thachil J. New criteria for sepsis-induced coagulopathy (SIC) following the revised sepsis definition: a retrospective analysis of a nationwide survey. *BMJ Open.* 2017;7:e017046. <https://doi.org/10.1136/bmjopen-2017-017046>
- [20] Gando S, Iba T, Eguchi Y, Ohtomo Y, Okamoto K, Koseki K, et al. A multicenter, prospective validation of disseminated intravascular coagulation diagnostic criteria for critically ill patients: comparing current criteria. *Crit Care Med.* 2006;34:625–31.
- [21] Gando S, Saitoh D, Ogura H, Fujishima S, Mayumi T, Araki T, et al. A multicenter, prospective validation study of the Japanese Association for Acute Medicine disseminated intravascular coagulation scoring system in patients with severe sepsis. *Crit Care.* 2013;17:R111. <https://doi.org/10.1186/cc12783>
- [22] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc; 2017:3149–57.
- [23] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: Association for Computing Machinery; 2019:2623–31.
- [24] Kienast J, Juers M, Wiedermann CJ, Hoffmann JN, Ostermann H, Strauss R, et al. Treatment effects of high-dose antithrombin without concomitant heparin in patients with severe sepsis with or without disseminated intravascular coagulation. *J Thromb Haemost.* 2006;4:90–7.
- [25] Umemura Y, Yamakawa K, Hayakawa M, Hamasaki T, Fujimi S, Japan Septic Disseminated Intravascular Coagulation (J-Septic DIC) study group. Screening itself for disseminated intravascular coagulation may reduce mortality in sepsis: a nationwide multicenter registry in Japan. *Thromb Res.* 2018;161:60–6.
- [26] Yamakawa K, Umemura Y, Muraio S, Hayakawa M, Fujimi S. Optimal timing and early intervention with anticoagulant therapy for sepsis-induced disseminated intravascular coagulation. *Clin Appl Thromb Hemost.* 2019;25:1076029619835055.
- [27] Helms J, Severac F, Merdji H, Clere-Jehl R, François B, Mercier E, et al. Performances of disseminated intravascular coagulation scoring systems in septic shock patients. *Ann Intensive Care.* 2020;10:92.
- [28] Yamakawa K, Yoshimura J, Ito T, Hayakawa M, Hamasaki T, Fujimi S. External validation of the two newly proposed criteria for assessing coagulopathy in sepsis. *Thromb Haemost.* 2019;119:203–12.
- [29] Iba T, Arakawa M, Di Nisio M, Gando S, Anan H, Sato K, et al. Newly proposed sepsis-induced coagulopathy precedes International Society on Thrombosis and Haemostasis overt-disseminated intravascular coagulation and predicts high mortality. *J Intensive Care Med.* 2020;35:643–9.

SUPPLEMENTARY MATERIAL

The online version contains supplementary material available at <https://doi.org/10.1016/j.rpth.2024.102519>