

<https://doi.org/10.1038/s41746-024-01215-4>

Deep learning-based prediction of *Clostridioides difficile* infection caused by antibiotics using longitudinal electronic health records

Check for updates

Junmo Kim^{1,7}, Joo Seong Kim^{2,7}, Sae-Hoon Kim³, Sooyoung Yoo⁴, Jun Kyu Lee²✉ & Kwangsoo Kim^{5,6}✉

Clostridioides difficile infection (CDI) is a major cause of antibiotic-associated diarrhea and colitis. It is recognized as one of the most significant hospital-acquired infections. Although CDI can develop severe complications and spores of *Clostridioides difficile* can be transmitted by the fecal-oral route, CDI is occasionally overlooked in clinical settings. Thus, it is necessary to monitor high CDI risk groups, particularly those undergoing antibiotic treatment, to prevent complications and spread. We developed and validated a deep learning-based model to predict the occurrence of CDI within 28 days after starting antibiotic treatment using longitudinal electronic health records. For each patient, timelines of vital signs and laboratory tests with a 35-day monitoring period and a patient information vector consisting of age, sex, comorbidities, and medications were constructed. Our model achieved the prediction performance with an area under the receiver operating characteristic curve of 0.952 (95% CI: 0.932–0.973) in internal validation and 0.972 (95% CI: 0.968–0.975) in external validation. Platelet count and body temperature emerged as the most important features. The risk score, the output value of the model, exhibited a consistent increase in the CDI group, while the risk score in the non-CDI group either maintained its initial value or decreased. Using our CDI prediction model, high-risk patients requiring symptom monitoring can be identified. This could help reduce the underdiagnosis of CDI, thereby decreasing transmission and preventing complications.

Clostridioides difficile (*C. difficile*) is a Gram-positive, anaerobic bacterium capable of forming spores and producing toxins¹. *C. difficile* infection (CDI) is a major cause of antibiotic-associated diarrhea and colitis. It is globally recognized as one of the most significant hospital-acquired infections^{2,3}. In the United States, there were more than 450,000 cases of CDI reported annually from 2011 to 2017. Approximately 29,000 deaths related to CDI were documented^{4,5}. The economic impact of CDI on healthcare costs is substantial in the United States, amounting to 6.3 billion USD. Annual CDI hospital management required nearly 2.4 million days of inpatient stay⁶. Between 2011 and 2013 in Europe, despite variations among countries, CDI occurred at a rate of 7.0 cases per 10,000 patient days⁷. The annual cost of

CDI in Europe was estimated to be €3 billion per year⁸. In Korea, CDI occurred at a rate of 5.06 cases per 100,000 patients, with an economic cost reported to be 15.8 million USD in 2011⁹.

The pathogenesis of CDI is microbial dysbiosis caused by factors such as antibiotics¹⁰. Gut dysbiosis refers to the disruption of the balance of microorganisms in the gastrointestinal tract, leading to the domination and colonization of *C. difficile* in the large intestine¹. The virulence of *C. difficile* is mostly attributed to enzymes and toxins that can induce the breakdown of gut barrier integrity and loss of functionality^{11,12}. *C. difficile* produces two important toxins, Toxin A and Toxin B, in its pathogenesis. Traditionally, Toxin A is known as enterotoxin A and Toxin B is known as cytotoxin B.

¹Interdisciplinary Program in Bioengineering, Seoul National University, Seoul, Republic of Korea. ²Division of Gastroenterology, Department of Internal Medicine, Dongguk University Ilsan Hospital, Dongguk University College of Medicine, Goyang, Republic of Korea. ³Division of Allergy & Clinical Immunology, Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea. ⁴Office of eHealth Research and Businesses, Seoul National University Bundang Hospital, Seongnam, Republic of Korea. ⁵Department of Transdisciplinary Medicine, Institute of Convergence Medicine with Innovative Technology, Seoul National University Hospital, Seoul, Republic of Korea. ⁶Department of Medicine, College of Medicine, Seoul National University, Seoul, Republic of Korea. ⁷These authors contributed equally: Junmo Kim, Joo Seong Kim. ✉e-mail: gerome@dumc.or.kr; kks00716@gmail.com

Risk factors for CDI are mainly known to be antibiotic exposure, old age, and hospitalization. Gastric acid suppression, inflammatory bowel disease, gastrointestinal surgeries, malignancy, transplantations, chronic kidney diseases, and immunosuppressant use are also risk factors for CDI^{1,13}. Severe CDI can lead to the development of pseudomembranous colitis, toxic megacolon, sepsis, and death¹⁴. Spores of *C. difficile* are transmitted by the fecal-oral route. Bacterial infection affecting the colon is spread through both direct and indirect contact. Nevertheless, CDI is occasionally overlooked in clinical settings even when symptoms are present. Thus, it is necessary to monitor high CDI risk groups, particularly those undergoing antibiotic treatment, to prevent complications and spread.

Several studies have employed machine learning algorithms to predict CDI in hospitalized patients. However, none of those studies described CDI prediction performance above an area under the receiver operating characteristic curve (AUROC) of 0.82¹⁵⁻²⁰. Although direct comparisons of these performances are not reasonable due to variations in data and cohort criteria, the predictive accuracy needs to be enhanced. To the best of our knowledge, prior studies have not utilized deep learning techniques, such as recurrent neural network (RNN) and Transformer, which have achieved great performance in several tasks regarding time-series data, to predict CDI caused by antibiotics. Furthermore, prior studies have not elucidated important features for CDI prediction with rational clinical descriptions.

This study aimed to predict the occurrence of CDI within 28 days after starting antibiotic treatment using longitudinal electronic health record (EHR) data, including vital signs, laboratory tests, and patient information such as demographics, comorbidities, and medications. We trained several machine learning and deep learning models to predict CDI and compared the performances of those models. A timeline of vital signs and laboratory tests with a 35-day monitoring period and a patient information vector consisting of age, sex, comorbidities, and medications were constructed for each patient. The data were collected from two locally separated tertiary hospitals. All trained CDI prediction models were externally validated. Important features were deduced from the trained model and CDI risk variation over time was compared between CDI and non-CDI groups.

Results

Study cohorts and dataset construction

This study included 594,759 patients at Seoul National University Hospital (SNUH) between January 2001 and December 2022 and 520,041 patients at Seoul National University Bundang Hospital (SNUBH) between January 2004 and December 2021, all of whom aged 18 years or more and received antibiotics. Following our cohort criteria, 529,049 patients at SNUH and 487,803 patients at SNUBH were excluded. The detailed population flowcharts are presented in Fig. 1. Finally, the numbers of patients in CDI and non-CDI groups were 466 and 65,244 in SNUH and 642 and 31,596 in

SNUBH, respectively. Data from SNUH were used for model development and internal validation, while data from SNUBH were used for external validation.

Table 1 shows the baseline characteristics of the included patients. The baseline characteristics were summarized by the initial monitoring point. The incidence of CDI stood was around 0.71% in SNUH and 2.00% in SNUBH. In both hospitals, the CDI group exhibited a higher average age than the non-CDI group. SNUH had a higher proportion of males in the CDI group, while SNUBH displayed an opposite trend. Both hospitals showed similar trends in vital signs and laboratory tests except for total bilirubin and ALT levels. Regarding comorbidities, the CDI group in both hospitals had a significantly higher prevalence of most diseases than the non-CDI group. Furthermore, the CDI group had a larger number of antibiotics, while antacid usage was more frequent in the non-CDI group of both hospitals.

CDI prediction performance

The performances of CDI prediction models are presented in Table 2. We fitted tree-based machine learning models, including random forest and gradient boosting machine (GBM)²¹, and RNN and attention-based deep learning models, including simple RNN, long-short term memory (LSTM)²², gated recurrent unit (GRU)²³, Transformer²⁴, and RETAIN²⁵. These models used in internal and external validation were selected after grid search cross-validation in the development process. The results of RNN-based models are presented in Supplementary Fig. 1. The results of tree-based models and attention-based models were omitted due to their poor performances. Consequently, GRU with two layers and 64 nodes was selected as the best model. RNN-based models consistently outperformed both tree-based models and attention-based models. Simple RNN exhibited the best prediction performance in internal validation, with an AUROC of 0.968 (0.957–0.979), while GRU demonstrated the best performance with an AUROC of 0.972 (0.968–0.975) in external validation. For both internal and external validation, GRU showed the highest areas under the precision-recall curve (AUPRC) of 0.250 (0.229–0.270) and 0.535 (0.531–0.539), respectively. We calculated all sensitivities, specificities, precisions, and F1-scores with Youden’s index²⁶. In internal validation, GRU obtained the highest sensitivity while LSTM achieved the highest precision. In external validation, GRU outperformed in all metrics. Meanwhile, since the choice of sensitivity might vary based on specific objectives²⁷, the results with fixed sensitivities of 0.9 and 0.95 are presented in Supplementary Tables 1 and 2. The differences in AUPRC, precision, and F1-score between hospitals were primarily due to the higher CDI incidence observed in the external validation

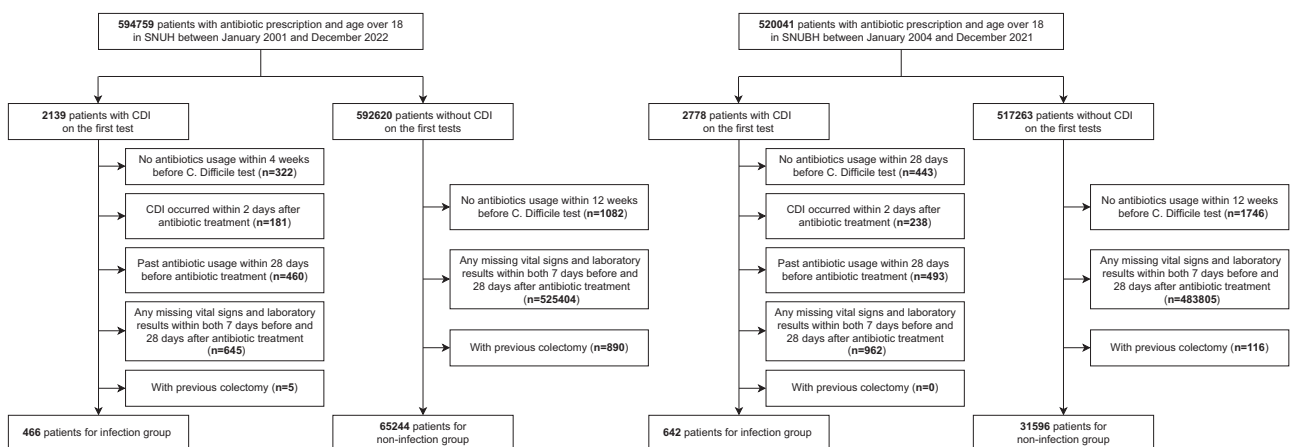


Fig. 1 | Population flowcharts. Data from SNUH were randomly split into development (70% for training and 15% for validation) and internal validation (15%) datasets. Data from SNUBH were used for external validation.

Table 1 | Baseline characteristics of the included patients

	SNUH (n = 65710)			SNUBH (n = 32238)			Between hospitals P-value
	Development and internal validation set			External validation set			
	CDI (n = 466)	Non-CDI (n = 65244)	P-value	CDI (n = 642)	Non-CDI (n = 31596)	P-value	
age	64.54 ± 15.51	57.29 ± 16.72	<0.0001	69.80 ± 16.05	61.76 ± 16.82	<0.0001	<0.0001
male sex	277 (59.44)	35660 (54.66)	0.0399	330 (51.40)	17463 (55.27)	0.0543	0.1382
Vital Signs							
SBP (mmHg)	118.22 ± 18.38	118.03 ± 16.77	0.8255	124.84 ± 22.50	126.14 ± 20.21	0.1464	<0.0001
DBP (mmHg)	70.71 ± 11.60	72.23 ± 11.66	0.0049	70.70 ± 15.03	71.82 ± 13.28	0.0623	0.1382
Heart rate (BPM)	85.69 ± 17.79	79.42 ± 16.07	<0.0001	88.16 ± 18.59	82.46 ± 17.36	<0.0001	<0.0001
Respiratory rate (BPM)	19.92 ± 2.61	19.33 ± 1.70	<0.0001	18.97 ± 3.31	18.81 ± 2.60	0.2046	<0.0001
Body temperature (°C)	36.99 ± 0.66	36.54 ± 0.48	<0.0001	36.94 ± 0.77	36.93 ± 0.71	0.6148	<0.0001
Laboratory tests							
WBC (10 ⁹ /μL)	10.28 ± 5.50	7.62 ± 3.29	<0.0001	10.16 ± 4.99	9.65 ± 4.66	0.0100	<0.0001
Hemoglobin (g/dL)	10.52 ± 1.79	11.68 ± 1.83	<0.0001	11.44 ± 2.37	12.59 ± 2.23	<0.0001	<0.0001
Platelet (10 ⁹ /μL)	242.84 ± 130.51	271.08 ± 118.56	<0.0001	226.93 ± 112.80	232.68 ± 96.03	0.2000	<0.0001
Neutrophil (%)	74.76 ± 16.95	64.22 ± 13.23	<0.0001	74.66 ± 15.90	71.10 ± 16.11	<0.0001	<0.0001
ANC (10 ⁹ /μL)	8.01 ± 4.74	5.12 ± 2.93	<0.0001	8.01 ± 4.74	7.29 ± 4.47	0.0002	<0.0001
Albumin (g/dL)	2.98 ± 0.51	3.56 ± 0.51	<0.0001	3.38 ± 0.67	3.78 ± 0.65	<0.0001	<0.0001
Total protein (g/dL)	5.79 ± 0.85	6.44 ± 0.74	<0.0001	6.42 ± 0.95	6.69 ± 0.87	<0.0001	<0.0001
Total bilirubin (mg/dL)	0.85 ± 0.57	0.76 ± 0.46	0.0014	0.79 ± 0.47	0.81 ± 0.48	0.2237	<0.0001
AST (IU/L)	29.18 ± 17.23	25.89 ± 14.63	<0.0001	33.62 ± 20.03	33.01 ± 22.47	0.4523	<0.0001
ALP (IU/L)	91.71 ± 50.29	78.52 ± 40.78	<0.0001	98.66 ± 47.36	89.50 ± 42.72	<0.0001	<0.0001
ALT (IU/L)	27.56 ± 22.46	26.06 ± 20.43	0.1540	22.01 ± 19.42	26.09 ± 22.54	<0.0001	<0.0001
BUN (mg/dL)	18.30 ± 11.88	14.32 ± 7.42	<0.0001	21.16 ± 11.63	17.54 ± 9.03	<0.0001	<0.0001
Creatinine (mg/dL)	0.92 ± 0.50	0.85 ± 0.32	0.0081	0.97 ± 0.54	0.90 ± 0.40	0.0005	<0.0001
Sodium (mmol/L)	136.89 ± 5.41	138.49 ± 3.59	<0.0001	137.22 ± 5.37	137.84 ± 4.26	0.0034	<0.0001
Potassium (mmol/L)	3.98 ± 0.54	4.18 ± 0.46	<0.0001	4.07 ± 0.58	4.10 ± 0.49	0.2246	0.6764
Chloride (mmol/L)	102.64 ± 6.11	102.62 ± 4.17	0.9257	102.29 ± 6.07	102.38 ± 4.79	0.6919	<0.0001
Total CO ₂ (mmol/L)	24.40 ± 4.20	26.01 ± 3.44	<0.0001	22.91 ± 3.80	23.44 ± 3.25	0.0005	<0.0001
CRP (mg/dL)	6.89 ± 6.02	3.38 ± 4.20	<0.0001	6.95 ± 7.69	5.24 ± 7.40	<0.0001	<0.0001
Comorbidities							
Malignant tumor	194 (41.63)	19301 (29.58)	<0.0001	181 (28.19)	8086 (25.59)	0.1440	<0.0001
Myocardial infarction	11 (2.36)	693 (1.06)	0.0193	22 (3.43)	712 (2.25)	0.0595	<0.0001
Uncomplicated diabetes	46 (9.87)	2089 (3.20)	<0.0001	82 (12.77)	2552 (8.08)	0.0001	<0.0001
Complicated diabetes	14 (3.00)	521 (0.80)	<0.0001	9 (1.40)	247 (0.78)	0.1074	<0.0001
Renal disease	26 (5.58)	571 (0.88)	<0.0001	34 (5.30)	563 (1.78)	<0.0001	0.7613
Heart Failure	23 (4.94)	735 (1.13)	<0.0001	38 (5.92)	983 (3.11)	0.0002	<0.0001
Metastatic carcinoma	30 (6.44)	1419 (2.17)	<0.0001	32 (4.98)	1414 (4.48)	0.5006	<0.0001
Dementia	17 (3.65)	332 (0.51)	<0.0001	58 (9.03)	838 (2.65)	<0.0001	<0.0001
Cerebrovascular disease	70 (15.02)	3729 (5.72)	<0.0001	159 (24.77)	3771 (11.94)	<0.0001	<0.0001
Peripheral vascular disease	19 (4.08)	1113 (1.71)	0.0009	36 (5.61)	1024 (3.24)	0.0023	<0.0001
Pulmonary disease	21 (4.51)	982 (1.51)	<0.0001	52 (8.10)	1514 (4.79)	0.0004	<0.0001
Liver disease	6 (1.29)	102 (0.16)	0.0001	1 (0.16)	46 (0.15)	0.6117	0.5495
Mild liver disease	27 (5.79)	1984 (3.04)	0.0017	22 (3.43)	1182 (3.74)	0.7529	<0.0001
Peptic ulcer disease	14 (3.00)	323 (0.50)	<0.0001	26 (4.05)	570 (1.80)	0.0003	<0.0001
Paraplegia and hemiplegia	11 (2.36)	133 (0.20)	<0.0001	10 (1.56)	138 (0.44)	0.0008	<0.0001
Connective tissue disease	4 (0.86)	601 (0.92)	1.0000	4 (0.62)	236 (0.75)	1.0000	0.0052
HIV	2 (0.43)	55 (0.08)	0.0620	0 (0.00)	5 (0.02)	1.0000	<0.0001
Drug usage							
Number of antibiotics	2.84 ± 1.70	2.31 ± 1.44	<0.0001	2.83 ± 1.57	2.25 ± 1.44	<0.0001	<0.0001
Antacid usage	213 (45.71)	42031 (64.42)	<0.0001	260 (40.50)	19674 (62.27)	<0.0001	<0.0001

All data are shown as mean ± SD and n (%).

P-values were calculated by the Student's t-test for continuous variables and the Fisher's exact test for categorical variables.

SBP systolic blood pressure, DBP diastolic blood pressure, BPM beats (breaths) per minute, WBC white blood count, ANC absolute neutrophil count, AST aspartate transferase, ALP alkaline phosphatase, ALT alanine transaminase, BUN blood urea nitrogen, CRP C-reactive protein, HIV human immunodeficiency virus.

Table 2 | Performance for detecting *Clostridioides difficile* infection

	AUROC	P-value	AUPRC	Sensitivity	Specificity	Precision	F1-score
Internal validation							
Tree-based model							
Random forest	0.834 (0.791–0.877)		0.049 (0.006–0.092)	0.660 (0.651–0.668)	0.875 (0.870–0.881)	0.037 (0.034–0.040)	0.070 (0.065–0.074)
GBM	0.853 (0.813–0.894)	0.1493	0.070 (0.029–0.111)	0.702 (0.694–0.710)	0.868 (0.862–0.874)	0.037 (0.034–0.040)	0.070 (0.066–0.075)
RNN-based model							
Simple RNN	0.968 (0.957–0.979)	<0.0001	0.165 <u>(0.154–0.176)</u>	0.936 <u>(0.932–0.940)</u>	0.877 <u>(0.871–0.883)</u>	0.052 <u>(0.048–0.056)</u>	0.099 <u>(0.094–0.104)</u>
LSTM	0.939 (0.918–0.961)	0.0001	0.118 (0.096–0.140)	0.894 (0.888–0.899)	0.891 (0.885–0.896)	0.056 (0.052–0.060)	0.105 (0.099–0.110)
GRU	0.952 <u>(0.932–0.973)</u>	<0.0001	0.250 (0.229–0.270)	0.936 (0.932–0.940)	0.862 (0.856–0.867)	0.046 (0.043–0.050)	0.088 (0.084–0.093)
Attention-based model							
Transformer	0.871 (0.837–0.904)	0.1053	0.074 (0.040–0.108)	0.755 (0.748–0.763)	0.839 (0.833–0.845)	0.033 (0.030–0.036)	0.063 (0.059–0.067)
RETAIN	0.746 (0.699–0.793)	0.0271	0.025 (0.000–0.072)	0.777 (0.769–0.784)	0.634 (0.626–0.642)	0.015 (0.013–0.017)	0.030 (0.027–0.033)
External validation							
Tree-based model							
Random forest	0.833 (0.818–0.847)		0.086 (0.071–0.100)	0.808 (0.804–0.813)	0.742 (0.738–0.747)	0.060 (0.057–0.063)	0.112 (0.108–0.115)
GBM	0.860 (0.847–0.873)	<0.0001	0.118 (0.105–0.131)	0.773 (0.768–0.777)	0.799 (0.795–0.804)	0.073 (0.070–0.075)	0.133 (0.129–0.136)
RNN-based model							
Simple RNN	0.958 <u>(0.953–0.963)</u>	<0.0001	0.241 (0.237–0.246)	0.935 <u>(0.932–0.937)</u>	0.881 <u>(0.878–0.885)</u>	0.138 <u>(0.134–0.142)</u>	0.241 <u>(0.236–0.245)</u>
LSTM	0.940 (0.934–0.947)	<0.0001	0.271 <u>(0.264–0.277)</u>	0.914 (0.911–0.917)	0.831 (0.827–0.836)	0.099 (0.096–0.103)	0.179 (0.175–0.183)
GRU	0.972 (0.968–0.975)	<0.0001	0.535 (0.531–0.539)	0.938 (0.935–0.940)	0.884 (0.881–0.888)	0.141 (0.138–0.145)	0.246 (0.241–0.251)
Attention-based model							
Transformer	0.871 (0.858–0.883)	<0.0001	0.179 (0.166–0.191)	0.807 (0.803–0.811)	0.786 (0.781–0.790)	0.071 (0.068–0.074)	0.131 (0.127–0.134)
RETAIN	0.755 (0.735–0.774)	<0.0001	0.091 (0.071–0.110)	0.572 (0.566–0.577)	0.803 (0.799–0.808)	0.056 (0.053–0.058)	0.102 (0.098–0.105)

Bold indicates the best and underline indicates the second best.

P-values were calculated using the DeLong method. Sensitivity, specificity, precision, and F1-score were calculated using Youden's index.

AUROC area under the receiver operating characteristic curve, AUPRC area under the precision-recall curve, GBM gradient boosting machine, RNN recurrent neural network, LSTM long short-term memory, GRU gated recurrent unit.

dataset. The receiver operating characteristic (ROC) and precision-recall curves of all experiments are shown in Fig. 2.

Considering that it is practically challenging to collect all features, we validated the GRU-based model with subsets of the features. We categorized the features including vital signs and laboratory tests into four groups: vital signs (SBP, DBP, heart rate, respiratory rate, body temperature), complete blood count (CBC) test (WBC, hemoglobin, platelet, neutrophil, ANC)²⁸, liver function (LF) test (albumin, total protein, total bilirubin, AST, ALP, ALT)^{29,30}, and renal function (RF) test (BUN, creatinine, sodium, potassium, chloride, total CO2)^{31,32}. Those items in each group are usually measured together. CRP was excluded from categories because CRP is an independent test widely used to detect bacterial infection³³. We validated our model with data that utilized only selected feature groups and masked the rest. In addition, considering potential missingness, we randomly masked 20% of the used features and validated the model. When only vital signs and CBC tests were utilized, the performance was slightly dropped, with AUROC decreasing from 0.952 to 0.933 in internal validation and from 0.972 to 0.947 in external validation. Even when the random masking strategy was applied, the AUROC

remained higher than 0.9 in most cases. These results are summarized in Supplementary Table 3.

Feature importance analysis

We used Deep SHAP³⁴ to identify important features for deep learning-based CDI prediction. The SHAP values of vital signs, laboratory tests, and patient information in both hospitals exhibited similar patterns, as shown in Fig. 3. This process used GRU as a reference model because it was selected as the best model in the development process. Body temperature and platelet count emerged as the two most influential variables, followed by ANC, BUN, neutrophil percentage, potassium, sodium, and CRP. Malignant tumors showed a relatively high SHAP value in internal validation. However, most comorbidities had a minimal impact on results. Notably, the number of antibiotics used and antacid usage exhibited the highest SHAP values among patient information in both hospitals.

Risk variation over time

To assess temporal differences in risk variation between CDI and non-CDI groups, we calculated continuous risk scores by sequentially entering

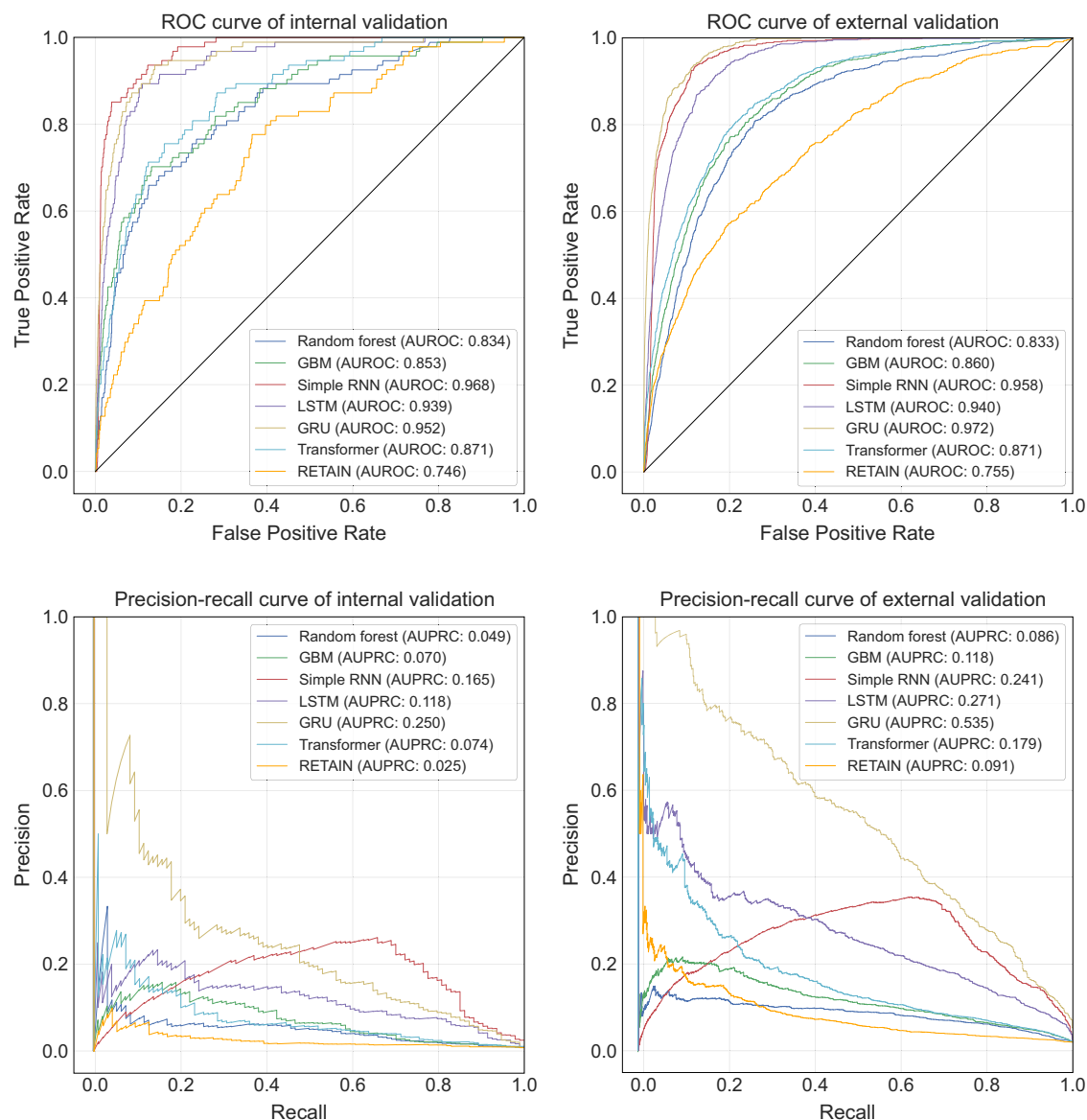


Fig. 2 | Receiver operating characteristic and precision–recall curves of Clostridioides difficile infection prediction models. Curves on the right side (external validation) are smoother than those on the left side (internal validation) because the

amount of data for external validation was much larger than for internal validation. Note that all data from SNUBH were used for external validation, while only split hold-out data from SNUH (15% of the total) were used for internal validation.

timelines ranging from two days to 35 days to the trained model. The risk score was the output value of the model. GRU served as a reference model in this process. Risk score variations over time are shown in Fig. 4. Across the timeline, the risk score in the CDI group exhibited a consistent increase, while the risk score in the non-CDI group either maintained its initial value or decreased. The CDI group had a higher risk score than the non-CDI group initially. This difference in risk score became bigger as time progressed.

Discussion

In this study, we developed and validated several machine learning and deep learning-based CDI prediction models using longitudinal EHR data, including a total of 97,948 patients. For internal and external validation, we used large multicenter datasets from two locally separate tertiary hospitals, SNUH and SNUBH. The model trained with GRU exhibited the best prediction performance with an AUROC of 0.952 for internal validation and 0.972 for external validation. In addition, we identified influential features for CDI prediction through Deep SHAP and assessed temporal differences in risk variation between CDI and non-CDI groups.

The CDI prediction model developed in this study demonstrated superior performance compared to previous studies. Panchavati et al. exhibited CDI detection performance with an AUROC of 0.815 using six hours of inpatient data and XGBoost¹⁵. Oh et al. formulated a CDI prediction model using inpatient data and logistic regression, yielding an AUROC of 0.820¹⁶. Marra et al. performed a cross-sectional study with EHR to predict CDI occurrences three days in advance with an AUROC of 0.604¹⁷. Wiens et al. trained support vector machine and hidden Markov model for CDI prediction, and achieved an AUROC of 0.79 and 0.75, respectively^{19,20}. Regarding CDI caused by antibiotics, Werkhoven et al. performed multiple logistic regression and detected CDI in patients receiving antibiotic therapy with an AUROC of 0.81¹⁸.

The CDI prediction model has a potential advantage in reducing CDI transmission and preventing complications in clinical settings by reducing underdiagnosis of CDI using patient trajectory. CDI is usually diagnosed by stool examinations such as nucleic acid amplification testing (NAAT), glutamate dehydrogenase (GDH), and enzyme immunoassay (EIA). NAAT and GDH are known to have high sensitivity and yield rapid results. Although EIA exhibits variations in sensitivity, it maintains a high

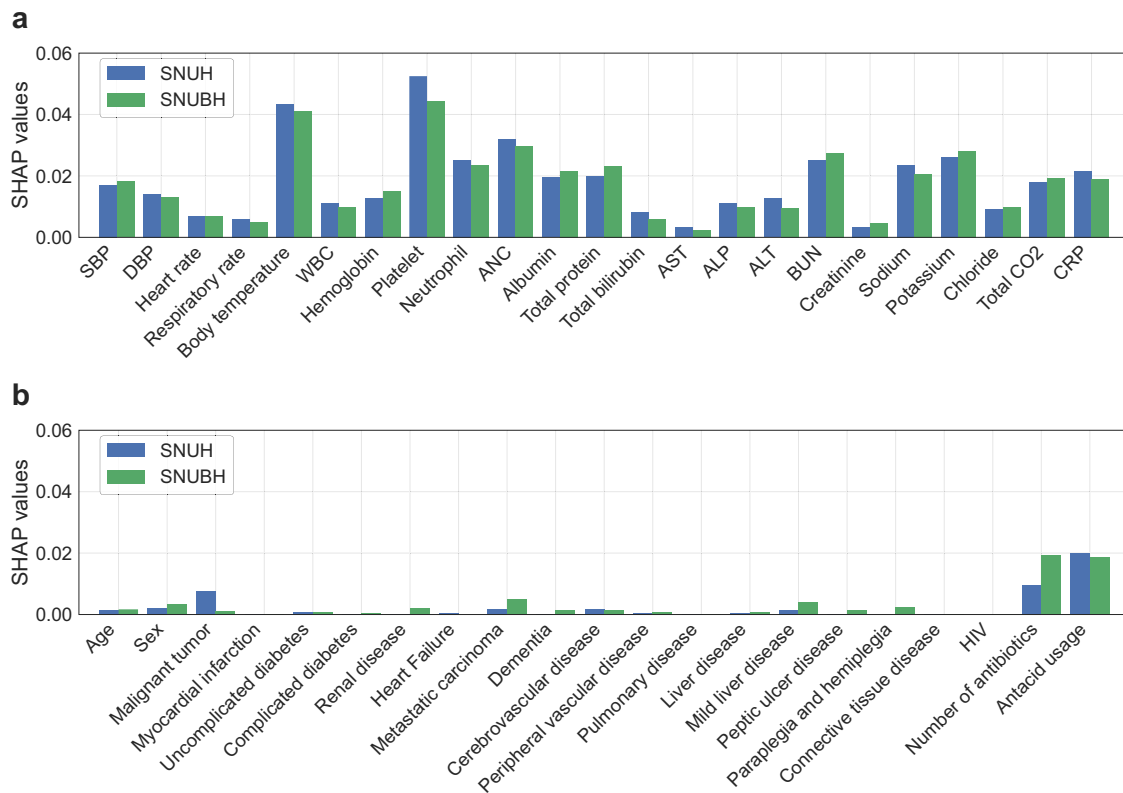


Fig. 3 | SHAP values of vital signs, laboratory tests, and patient information for patients with *Clostridioides difficile* infection. a SHAP values of vital signs and laboratory tests. **b** SHAP values of patient information. We separated the results of timeline data (vital signs and laboratory tests) and patient information data because

those were fed into different layers (RNN or attention layer for timeline and fully connected layer for patient information) and the dimensions of timeline and patient information data were different. Direct comparison of SHAP values of those two data with the same scale was considered to be inappropriate.

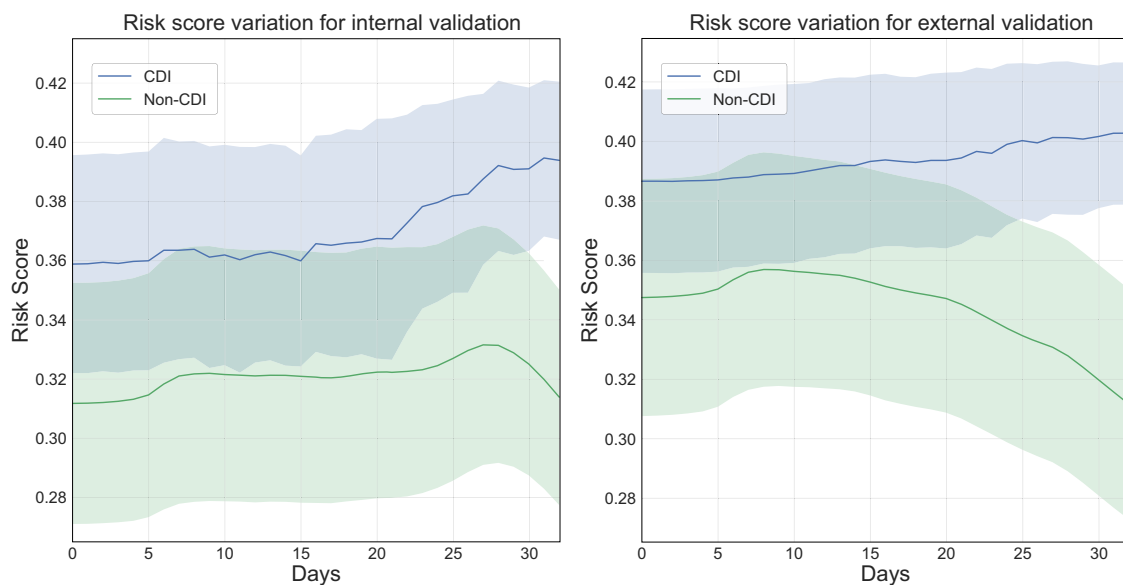


Fig. 4 | Risk score variation over time. The shaded part indicates the standard deviation.

specificity. The diagnostic process for CDI employs a multistep procedure that includes performing EIA if NAAT or GDH is positive. It is typically initiated when new-onset unformed stools occur more than three times within 24 h¹⁴. However, rate of CDI underdiagnosis remains significant despite current diagnostic strategy^{35–37}. Lack of suspicion is one of the important reasons for underdiagnosis³⁸. Improvement in the diagnostic

process through monitoring systems for CDI in high risk patients is necessary. Our prediction model could be utilized to identify patients at risk of developing symptoms. Among the patients who are assigned high risk scores by the prediction model, symptomatic patients should be isolated and undergo stool test, while asymptomatic patients should be closely observed for symptom development. This approach could help reduce the

underdiagnosis of CDI, thereby decreasing transmission and preventing complications. Furthermore, our model showed that the difference in risk score between CDI and non-CDI groups gradually widened over time. This trend implies that clinicians can consider the potential risk of CDI despite the lack of symptoms if the risk score remains high or increases after antibiotic treatment.

Platelet count and body temperature emerged as the two most important features among vital signs and laboratory tests, while the number of antibiotics used and antacid usage stood out as key attributes within patient information. Elevated body temperature is recognized as one of the main symptoms of CDI¹. Platelet level also plays a significant role in CDI, as evidenced by a study indicating associations between abnormal platelet levels and CDI outcomes³⁹. Thrombocytosis is associated with inflammation given that platelets are considered to be acute phase reactants⁴⁰. On the other hand, thrombocytopenia is associated with underlying diseases such as malignancies, hepatic diseases, and immunosuppression, all of which are risk factors for CDI⁴¹. The number of antibiotics used during the observation period correlates with the duration of antibiotic use. In cases where a patient remains unresponsive to antibiotic treatment, there is a suspicion of antibiotic-resistant bacteria, prompting a switch to a broad-spectrum antibiotic. However, the risk of CDI tends to increase as antibiotic treatment continues¹, as the rising trend of risk score was exhibited in the CDI group in our study. It has been reported that gastric acid suppression might have an influence on CDI development^{41,42}. However, there is still controversy as several studies exhibited conflicting results^{43,44}. A notable finding was the significantly lower use of antacids among patients with CDI. The reduced antacid usage in this group might be attributed to considerations of potential drug interactions, as the CDI group presented with a higher prevalence of underlying diseases⁴⁵.

In this study, Transformer outperformed tree-based models in CDI prediction. However, RNN-based models exhibited much better performances. Although Transformers are gaining widespread usages in various fields such as natural language processing (NLP), they were found to be less suitable for training on short numeric timelines (maximum 35 days) in our dataset than RNN-based models. This contrast might be attributed to differing ways in which RNNs and Transformers handle input data: RNN processes data sequentially, while Transformer takes in data all at once with positional encoding and learns relationships between variables with less susceptibility to temporal and sequential dependencies. Several studies dealing with numeric time-series data have employed RNNs for capturing sequential changes^{46–48}. On the other hand, RETAIN, an interpretable attention-based neural network model for temporal EHR data, was initially developed for binary EHR variables. Consequently, it did not perform as well in our dataset, which contained numerous continuous numeric variables in timelines.

Regarding cohort definition, we did not include patients with missing values in our dataset, although eliminating patient records with missing features might introduce a potential bias in models because many patients were excluded by the criterion. Considering the potential bias, we conducted an experiment without eliminating any patients with missing features. Instead, we only excluded patients with any missing vital signs. In SNUH, there were 1092 patients in the CDI group and 298,531 in the non-CDI group. In SNUBH, there were 1563 patients in the CDI group and 301,984 in the non-CDI group. We then imputed missing values using multivariate imputation by chained equations (MICE)⁴⁹. In this case, the performance of the random forest in internal validation had an AUROC of 1.0. Respiratory rate and CRP were the two most important variables in the random forest. Even when we trained another random forest with only those two variables and patient information, the AUROC was still 1.0. This implies that the model learned the imputation pattern of missing values reflecting the test pattern, which was relatively easy to learn, rather than the temporal variation of patient trajectory. The GRU-based model also exhibited an AUROC of 1.0 in internal validation. However, a significant decrease in performance was seen in external validation with an AUROC of 0.84. Test items routinely measured may vary between hospitals according to the policy of each hospital or regionality, and the model that learned test patterns did not

perform well in the external cohort. Nevertheless, the original missingness criterion might be difficult to be satisfied in clinical practice. Thus, we validated our model by applying various mitigated missing conditions. Using only vital signs and CBC test, the model achieved an AUROC of 0.933 in internal validation and 0.947 in external validation. Even with a random masking strategy, the model maintained a high performance with an AUROC of 0.929 in internal validation and 0.904 in external validation.

This study has several limitations. First, it was a retrospective study with potential selection bias. However, it was noteworthy that our predictive model maintained good performance in both cohorts, which had different baseline characteristics and the number of events. Second, we monitored only four weeks from the index date. A previous study has reported that the risk of CDI is the highest in the first month after antibiotics use and that it persists until three months⁵⁰. Accordingly, we excluded patients in the non-CDI group who developed CDI within 12 weeks after antibiotics use. However, we monitored them for four weeks, considering the lack of enough vital signs and laboratory test data. Further studies with prolonged monitoring periods are needed. Third, the prevalence of CDI was low in both cohorts. While we used Focal loss⁵¹ to address class imbalance, deep learning models often suffer from those extreme class imbalances. However, the incidence of CDI in the general inpatient population has been reported to be under 2%, and this confirms that our datasets reflected real-world data⁵². In addition, the results of internal and external validation showed similar trends, lending reliability to our findings. Fourth, records of gastrointestinal symptoms were not included in this study. Our model was based on EHR, which included measurements, drugs, and diagnoses records. Symptoms such as diarrhea are usually documented in nursing records. Unfortunately, we could not obtain nursing records owing to internal circumstances. However, if nursing records are accessible and models are trained with symptom history, the appropriate time to use the model can be specified based on symptoms. In addition, prediction performance might be improved with more sophisticated data. Fifth, we did not include patients with missing values in this study. This exclusion was to prevent the model from training test patterns, which might vary between hospitals, rather than the temporal variation of patient trajectory. However, further investigation on handling missing values for longitudinal EHR-based time-series data is required for expanding the study population and generalizing the prediction model. Sixth, we have provided several pieces of clinical evidence that support our feature importance analysis and have conducted a sub-analysis with subsets of the features, but prospective studies utilizing our model are needed to completely evaluate the practical suitability of our model.

In conclusion, we developed a high-performing deep learning-based CDI prediction model from patients with antibiotic treatment. The model was internally and externally validated using data from two locally separate tertiary hospitals. The CDI prediction model can reduce underdiagnosis of CDI and can contribute to the goal of decreasing transmission and preventing complications. This study had limitations in data acquisition, monitoring period, class imbalance, and data missingness. For future works, prospective studies on additional data of gastrointestinal symptoms are needed to specify an appropriate time to use the prediction model, to discover better models with expanded monitoring periods, and to further investigate how to handle missing values.

Methods

Data curation

This study used data from the SNUH between January 2001 and December 2022 and the SNUBH between January 2004 and December 2021. All data were collected from the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)⁵³. OMOP is a public-private partnership established in the United States to inform the appropriate use of observational healthcare databases and study the effects of medical products. OMOP CDM provides standard-based data analysis solutions that support EHR from different sources into a standard data structure, which enables large-scale data analysis. Data from SNUH were randomly split into development (70% for training and 15% for validation) and internal validation (15%) datasets. Data from SNUBH were used for external validation.

The Institutional Review Boards (IRB) at SNUH (IRB No. 2308-101-1459) and SNUBH (IRB No. X-2308-846-906) granted a waiver of approval and informed consent, considering that the data used in this study were de-identified and based on observational electronic medical records from the OMOP CDM. This retrospective, multicenter study was conducted in agreement with the Declaration of Helsinki, the Korean Bioethics and Safety Act (Law No. 16372), and the Human Research Protection Program–Standard Operating Procedure of Seoul National University Hospital.

Cohort definition and main outcomes

Patients with antibiotic prescriptions aged over 18 years were identified and divided into two groups according to whether the first *C. difficile* test was positive or not. We used a *C. difficile* toxin test with EIA for the diagnosis of CDI. Exclusion criteria for the CDI group were as follows: We first excluded patients who had no antibiotic prescriptions within 28 days before *C. difficile* test. Then, patients with CDI that occurred within two days (washout period) after the index date of antibiotics (start of antibiotics) were also excluded, assuming that CDI did not occur due to antibiotics but for other reasons. To remove the potential redundant effect of antibiotics, we excluded the patients with past antibiotic prescriptions within 28 days before the index date. The non-CDI group included patients who had CDI after the first *C. difficile* test and patients who had never experienced CDI. For the former case, those who were diagnosed with CDI within 12 weeks after antibiotic treatment were considered to be at potential risk for CDI, and thus, they were excluded. For both positive and negative groups, patients with any missing vital signs and laboratory tests within both seven days before and 28 days after the index date and patients with previous colectomy procedures were excluded. A brief illustration of the index date definition is shown in Fig. 5.

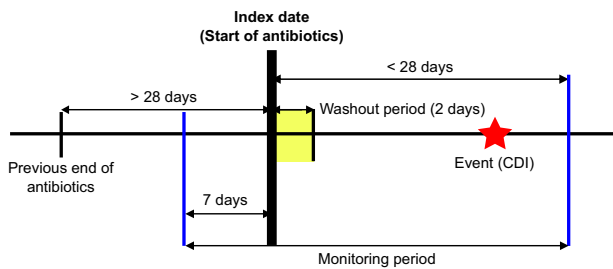


Fig. 5 | Index date definition. We monitored patient records between seven days before and 28 days after the index date, with the maximum length of the monitoring period of 35 days.

Data preprocessing

We used vital signs, laboratory tests, and patient information (including age, sex, comorbidity records, the number of antibiotics used, and antacids usage). OMOP CDM concept IDs of antibiotics, antacids, vital signs, laboratory tests, and colectomy procedures used in this study are shown in Supplementary Tables 4–7. For each patient, we constructed a timeline of vital signs and laboratory tests with patient information vectors. To construct a timeline, we first generated a table with 35 columns, implying the maximum monitoring period. Each column represented a sequential date. The last column was set as the last day. For each vital sign and laboratory item, values were filled in on each date. Blank parts between tests were linearly interpolated. Front and back parts of the timeline without tests were padded with the first and last measured vital signs and laboratory tests, respectively. A patient information vector consisted of age, sex, comorbidity records, the number of antibiotics used, and antacids usage. Age and the number of antibiotics used were numeric, while the rest variables were binary. All numeric variables of patient information, vital signs, and laboratory tests data were standardized before training.

Model development

We trained three kinds of models: tree-based model (including random forest and GBM as a baseline), RNN-based model (including simple RNN, LSTM, and GRU), and attention-based model (including Transformer and RETAIN). As tree-based models were trained with one-dimensional vectors, we concatenated the first and last columns of the timeline and patient information vector of each patient. This implies that tree-based models also used vital signs and laboratory tests before and after the index date as other deep learning models. In the case of RNN and attention-based models, a timeline was fed into the RNN or attention layer and a patient information vector was inputted into the fully connected (dense) layer. The outputs of those layers were then merged, and entered another dense layer to classify CDI or normal cases. A brief illustration of the process of CDI prediction using deep learning models is shown in Fig. 6.

We used grid search cross-validation to find the best-performing model. For tree-based models, the number of trees (from 20 to 200 in 20 intervals), the maximum depth of the tree (one to ten and infinite), and the maximum number of features to consider when looking for the best split (one to ten and the number of features) were used as hyperparameters. For RNN and attention-based models, the number of RNN and attention layers (one to five) and the number of nodes in all hidden layers (32, 64, 128, 256, and 512) were used as hyperparameters. In the case of Transformer, the number of heads for multi-head attention was set to eight. The batch size was set to 256. We used Focal loss⁵¹ to address the extreme class imbalance of

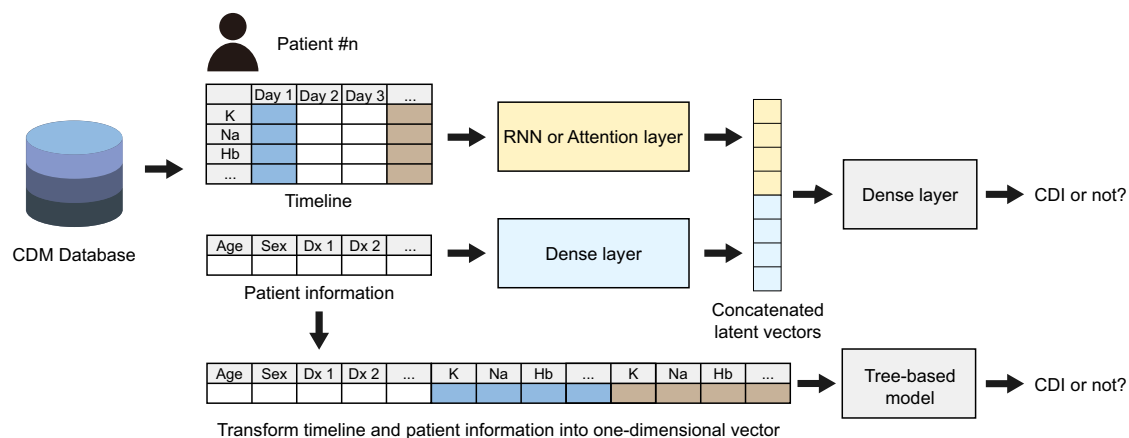


Fig. 6 | The process of Clostridioides difficile infection prediction using tree-based and deep learning models. Timeline variables (vital signs and laboratory tests) were fed into the RNN or attention layer, while patient information variables were fed into the fully connected (dense) layer. The output vectors of those two layers

were concatenated and fed into another fully connected layer to predict CDI. For tree-based models, we transformed timeline and patient information data into a one-dimensional vector by concatenating the first and last columns of the timeline and patient information vector.

the dataset and Adam optimizer⁵⁴ with a learning rate of 0.0001. The model was trained for a maximum of 100 epochs. Early stopping was set with a patience of 20 on the performance measured using the AUROC. All experiments during the development process were performed across five random seeds. The models with the best mean performance were selected and used for internal and external validation. Scikit-learn (version 1.0.2) and Pytorch (version 1.12.0) in Python (version 3.8.10) were used for tree-based and deep learning-based models, respectively.

Identifying important features

To identify important features for deep learning-based CDI prediction, we used Deep SHAP, an enhanced version of the DeepLIFT algorithm⁵⁵. Deep SHAP could compute attribution scores of all nodes and approximate Shapley values implying feature importance scores. Regarding the timeline for each patient, we computed the total absolute SHAP values across all days and then averaged these summed SHAP values across all patients to identify significant vital signs and laboratory items. As for patient information vectors, we similarly averaged the absolute SHAP values across all patients to discern the crucial patient information. SHAP (version 0.42.1) package in Python (version 3.8.10) was used for SHAP value calculation.

Statistical analysis

Characteristics such as age, sex, vital signs, laboratory tests, comorbidities, and drug usage (the number of antibiotics used and antacids usage) between CDI and non-CDI groups and between hospitals were compared by calculating *P*-values using the Student's *t*-test for continuous variables and the Fisher's exact test for categorical variables. To measure and compare the performances of the models, we used AUROC and AUPRC. Confidence intervals (CIs) of AUROC and AUPRC were calculated using DeLong's method⁵⁶, while those of sensitivity, specificity, precision, and F1-score were calculated using Wilson's method⁵⁷. Statistical significance was set at $\alpha = 0.05$. All statistical analyses were performed using scikit-learn (version 1.0.2) in Python (version 3.8.10).

Data availability

The raw data used in this study are not publicly available to preserve participant privacy. The data generated and analyzed during this study are available from the corresponding author upon reasonable request.

Code availability

Codes for training deep learning models and tree-based models, feature importance analysis, and statistical analysis are available at our GitHub repository (https://github.com/kicarussays/drugsafe_EHR).

Received: 2 January 2024; Accepted: 2 August 2024;

Published online: 24 August 2024

References

- Czepiel, J. et al. Clostridium difficile infection: review. *Eur. J. Clin. Microbiol. Infect. Dis.* **38**, 1211–1221 (2019).
- Gorbach, S. L. Antibiotics and Clostridium difficile. *N. Engl. J. Med.* **341**, 1690–1691 (1999).
- Kelly, C. P., Pothoulakis, C. & LaMont, J. T. Clostridium difficile colitis. *N. Engl. J. Med.* **330**, 257–262 (1994).
- Guh, A. Y. et al. Trends in U.S. burden of Clostridioides difficile infection and outcomes. *N. Engl. J. Med.* **382**, 1320–1330 (2020).
- Lessa, F. C. et al. Burden of Clostridium difficile infection in the United States. *N. Engl. J. Med.* **372**, 825–834 (2015).
- Zhang, S. et al. Cost of hospital management of Clostridium difficile infection in United States—a meta-analysis and modelling study. *BMC Infect. Dis.* **16**, 447 (2016).
- Nagy, E. What do we know about the diagnostics, treatment and epidemiology of Clostridioides (Clostridium) difficile infection in Europe? *J. Infect. Chemother.* **24**, 164–170 (2018).
- Jones, A. M., Kuijper, E. J. & Wilcox, M. H. Clostridium difficile: a European perspective. *J. Infect.* **66**, 115–128 (2013).
- Choi, H.-Y. et al. The epidemiology and economic burden of Clostridium difficile Infection in Korea. *BioMed. Res. Int.* **2015**, 510386 (2015).
- Liu, C., Monaghan, T., Yadegar, A., Louie, T. & Kao, D. Insights into the evolving epidemiology of clostridioides difficile infection and treatment: a global perspective. *Antibiotics* **12** <https://doi.org/10.3390/antibiotics12071141> (2023).
- Baktash, A. et al. Mechanistic insights in the success of fecal microbiota transplants for the treatment of Clostridium difficile Infections. *Front. Microbiol.* **9** <https://doi.org/10.3389/fmicb.2018.01242> (2018).
- Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. Clostridium difficile infection. *Nat. Rev. Dis. Prim.* **2**, 16020 (2016).
- Leffler, D. A. & Lamont, J. T. Clostridium difficile infection. *N. Engl. J. Med.* **372**, 1539–1548 (2015).
- McDonald, L. C. et al. Clinical practice guidelines for Clostridium difficile infection in adults and children: 2017 update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clin. Infect. Dis.* **66**, e1–e48 (2018).
- Panchavati, S. et al. A comparative analysis of machine learning approaches to predict C. difficile infection in hospitalized patients. *Am. J. Infect. Control* **50**, 250–257 (2022).
- Oh, J. et al. A generalizable, data-driven approach to predict daily risk of Clostridium difficile infection at two large Academic Health Centers. *Infect. Control Hospital Epidemiol.* **39**, 425–433 (2018).
- Marra, A. R. et al. Modest Clostridioides difficile infection prediction using machine learning models in a tertiary care hospital. *Diagnostic Microbiol. Infect. Dis.* **98**, 115104 (2020).
- van Werkhoven, C. H. et al. Identification of patients at high risk for Clostridium difficile infection: development and validation of a risk prediction model in hospitalized patients treated with antibiotics. *Clin. Microbiol. Infect.* **21**, 786.e781–788 (2015).
- Wiens, J., Horvitz, E. & Gutttag, J. Patient risk stratification for hospital-associated c. diff as a time-series classification task. *Adv. Neural. Inf. Process. Syst.* **25**, 467–475 (2012).
- Wiens, J., Gutttag, J. & Horvitz, E. Learning evolving patient risk processes for c. diff colonization. *ICML Workshop on Machine Learning from Clinical Data.* 79–82 (2012).
- Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
- Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
- Chung, J., Gulcehre, C., Cho, K. & Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv Prepr. arXiv* **1412**, 3555 (2014).
- Vaswani, A. et al. Attention is all you need. *Adv. Neural. Inf. Process. Syst.* **30**, 5998–6008 (2017).
- Choi, E. et al. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. *Adv. Neural. Inf. Process. Syst.* **29**, 3512–3520 (2016).
- Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).
- Abad, C. L. R. & Safdar, N. A review of Clostridioides difficile infection and antibiotic-associated diarrhea. *Gastroenterol. Clin. North Am.* **50**, 323–340 (2021).
- George-Gay, B. & Parker, K. Understanding the complete blood count with differential. *J. PeriAnesthesia Nurs.* **18**, 96–117 (2003).
- Gowda, S. et al. A review on laboratory liver function tests. *Pan Afr. Med. J.* **3**, 17 (2009).
- Hoekstra, L. T. et al. Physiological and biochemical basis of clinical liver function tests: a review. *Ann. Surg.* **257**, 27–36 (2013).
- Bovee, K. C. Renal function and laboratory evaluation. *Toxicol. Pathol.* **14**, 26–36 (1986).
- Frassetto, L. A., Morris, R. C. Jr & Sebastian, A. Effect of age on blood acid-base composition in adult humans: role of age-related renal functional decline. *Am. J. Physiol. Ren. Physiol.* **271**, F1114–F1122 (1996).

33. Nabulsi, M., Hani, A. & Karam, M. Impact of C-reactive protein test results on evidence-based decision-making in cases of bacterial infection. *BMC Pediatr.* **12**, 140 (2012).
34. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural. Inf. Process. Syst.* **30**, 4765–4774 (2017).
35. Davies, K. A. et al. Underdiagnosis of Clostridium difficile across Europe: the European, multicentre, prospective, biannual, point-prevalence study of Clostridium difficile infection in hospitalised patients with diarrhoea (EUCLID). *Lancet Infect. Dis.* **14**, 1208–1219 (2014).
36. Alcalá, L. et al. Impact of clinical awareness and diagnostic tests on the underdiagnosis of Clostridium difficile infection. *Eur. J. Clin. Microbiol. Infect. Dis.* **34**, 1515–1525 (2015).
37. Alcalá, L. et al. The undiagnosed cases of Clostridium difficile infection in a whole nation: where is the problem? *Clin. Microbiol. Infect.* **18**, E204–E213 (2012).
38. Cataldo, M. A. et al. Hospitalized patients with diarrhea: rate of Clostridioides difficile infection underdiagnosis and drivers of clinical suspicion. *Anaerobe* **70**, 102380 (2021).
39. Buchrits, S. et al. The importance of abnormal platelet count in patients with Clostridioides difficile Infection. *J. Clin. Med.* **10** <https://doi.org/10.3390/jcm10132957> (2021).
40. Thomas, M. R. & Storey, R. F. The role of platelets in inflammation. *Thromb. Haemost.* **114**, 449–458 (2015).
41. Trifan, A. et al. Proton pump inhibitors therapy and risk of Clostridium difficile infection: systematic review and meta-analysis. *World J. Gastroenterol.* **23**, 6500–6515 (2017).
42. Arriola, V. et al. Assessing the risk of hospital-acquired Clostridium difficile infection with proton pump inhibitor use: a meta-analysis. *Infect. Control Hosp. Epidemiol.* **37**, 1408–1417 (2016).
43. Beaulieu, M., Williamson, D., Pichette, G. & Lachaine, J. Risk of Clostridium difficile-associated disease among patients receiving proton-pump inhibitors in a Quebec medical intensive care unit. *Infect. Control Hosp. Epidemiol.* **28**, 1305–1307 (2007).
44. Campbell, K. A., Phillips, M. S., Stachel, A., Bosco, J. A. 3rd & Mehta, S. A. Incidence and risk factors for hospital-acquired Clostridium difficile infection among inpatients in an orthopaedic tertiary care hospital. *J. Hosp. Infect.* **83**, 146–149 (2013).
45. Wedemeyer, R. S. & Blume, H. Pharmacokinetic drug interaction profiles of proton pump inhibitors: an update. *Drug Saf.* **37**, 201–211 (2014).
46. Li, H. et al. Spatial dynamic graph convolutional network for traffic flow forecasting. *Appl. Intell.* **53**, 14986–14998 (2023).
47. Choya, T. et al. CrowdFlowTransformer: Capturing Spatio-Temporal Dependence for Forecasting Human Mobility. *IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. 496–501 (2023).
48. Lim, H., Park, J.-S. & Han, Y. Rare Sound Event Detection Using 1D Convolutional Recurrent Neural Networks. *DCASE*. 80–84 (2017).
49. Azur, M. J., Stuart, E. A., Frangakis, C. & Leaf, P. J. Multiple imputation by chained equations: what is it and how does it work? *Int. J. Methods Psychiatr. Res.* **20**, 40–49 (2011).
50. Hensgens, M. P., Goorhuis, A., Dekkers, O. M. & Kuijper, E. J. Time interval of increased risk for Clostridium difficile infection after exposure to antibiotics. *J. Antimicrob. Chemother.* **67**, 742–748 (2012).
51. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*. 2980–2988.
52. Marra, A. R. et al. Incidence and outcomes associated with Clostridium difficile infections: a systematic review and meta-analysis. *JAMA Netw. Open* **3**, e1917597 (2020).
53. Hripscak, G. et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud. Health Technol. Inf.* **216**, 574–578 (2015).
54. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. *arXiv Prepr. arXiv* **1412**, 6980 (2014).
55. Shrikumar, A., Greenside, P. & Kundaje, A. Learning important features through propagating activation differences. *Proceedings of the 34th International Conference on Machine Learning Vol. 70* (eds Precup Doina & Teh Yee Whye) 3145–3153 (PMLR, Proceedings of Machine Learning Research, 2017).
56. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).
57. Wilson, E. B. Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc.* **22**, 209–212 (1927).

Acknowledgements

This research was supported by a grant from Korea Institute of Drug Safety and Risk Management in 2023. J.K.L. received funding from Korea Institute of Drug Safety and Risk Management (No. 22233018800). J.K. was supported by a fellowship program of the AI Institute at Seoul National University (AIIS). The funders played no role in the study design, data collection, data analysis, data interpretation, or writing of this manuscript.

Author contributions

J.S.K. and J.K.L. contributed to the conceptualization and design of this study. J.K. handled and mainly analyzed the research data. All authors interpreted the results. J.K. constructed machine learning and deep learning models and conducted statistical analysis. J.K. and J.S.K. wrote the original draft of the paper. S.H.K. and S.Y. revised the paper. J.S.K., S.H.K. and J.K.L. reviewed clinical evidence of this study. K.S.K. and S.Y. provided the data. J.K. and J.S.K. verified the quality of the data. J.K., K.S.K. and S.Y. had full access to all raw data. All authors had the final responsibility to submit for publication.

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/s41746-024-01215-4>.

Correspondence and requests for materials should be addressed to Jun Kyu Lee or Kwangsoo Kim.

Reprints and permissions information is available at <http://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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2024