

Table S1. The examples of patients for in-context learning; as SOFA score distributed normally, mean and standard deviation (SD) were utilized and as lactic acid level distributed skewed, representative values were selected utilized 1Q, median, and 3Q

Index	Outcome	Sex	Age (yr)	SOFA	Lactic acid
7D_S_1	Survivor	Male	60–69	6	1.5
7D_S_2	Survivor	Male	70–79	9	2.4
7D_S_3	Survivor	Male	≥80	13	4.4
7D_D_1	Death	Female	70–79	9	3.9
7D_D_2	Death	Male	≥80	13	7.1
7D_D_3	Death	Male	50–59	16	11.8
30D_S_1	Survivor	Male	≥80	6	1.5
30D_S_2	Survivor	Female	60–69	9	2.3
30D_S_3	Survivor	Female	70–79	13	4.1
30D_D_1	Death	Male	70–79	8	2.5
30D_D_2	Death	Male	60–69	12	5
30D_D_3	Death	Female	70–79	16	9.9

SOFA: Sequential Organ Failure Assessment.

Table S2. Diagnostic performance of various machine learning models on predicting 7-day in-hospital mortality

	Accuracy	Precision	Recall	F1-score	AUROC
Logistic regression	0.85	0.88	0.97	0.92	0.78
Decision tree	0.79	0.87	0.87	0.87	0.58
Random forest	0.82	0.88	0.91	0.90	0.76
Gradient boosting	0.85	0.88	0.96	0.92	0.79

AUROC: area under the receiver operating characteristic curve.

The bold font indicates the best performance in each test.

Table S3. Diagnostic performance of various machine learning models on predicting 30-day in-hospital mortality

	Accuracy	Precision	Recall	F1-score	AUROC
Logistic regression	0.74	0.75	0.89	0.82	0.75
Decision tree	0.67	0.74	0.78	0.76	0.62
Random forest	0.70	0.75	0.81	0.78	0.71
Gradient boosting	0.74	0.76	0.88	0.82	0.76

AUROC: area under the receiver operating characteristic curve.