**Supplementary Material**

**Mortality prediction of patients with sepsis in the emergency department using machine learning models: a retrospective cohort study according to the Sepsis-3 definitions**

**Supplementary Methods**

1. Outlier detection and missing value imputation

2. Preventing underestimation of feature importance

3. Model construction

**Supplementary Tables**

1. Missing rates of the data and types of variables

2. Baseline characteristics of the study population for 7-day mortality

3. Baseline characteristics of the study population for 14-day mortality

4. Cross validation performance for 7-day, 14-day, and 30-day mortalities

**Supplementary Figs**

1. Model calibration in prediction for 7 d mortality

2. Model calibration in prediction for 14 d mortality

3. Partial SHAP dependence plots for the selected values

# Supplementary Methods

# 1. Outlier detection and missing value imputation

We detected outliers using isolation forests [1] and replaced them with the nearest nonoutlier values within the training set. Multivariate imputation by chained equations [2] was performed for imputation in continuous variables, and the imputed values were restricted within the range of the training set. For discrete-value variables, the imputed values were rounded off to the nearest integer.

# 2. Preventing underestimation of feature importance

Hierarchical double clustering was performed in every recursion during recursive feature elimination due to multicollinearity, which caused underestimation of the relative feature importance [3]. Hierarchical clustering on Spearman rank-order correlations of features was repeated, and there were no clusters with a Ward’s linkage of less than one. The importance of each feature was evaluated with a feature set, excluding correlated features belonging to the same cluster.

# 3. Model construction

As a comparative baseline, we built a logistic regression, which is a linear algorithm that uses a logistic function. Four popular supervised learning models were built and compared: support vector machine [4], extreme gradient boosting (XGBoost) [5], light gradient boosting machine (LightGBM) [6], and multilayer perceptron (MLP) [7]. The support vector machine found an optimal hyperplane that divided data points in a multidimensional space. XGBoost and LightGBM were the most promising tree-based gradient boosting algorithms that used level-wise and leaf-wise tree growth, respectively. MLP is a feedforward artificial neural network.

Hyperparameter settings were determined with Bayesian optimization, which built a surrogate model of a posterior distribution and described the objective function [8]. The target metric for the optimal hyperparameters was the area under the receiver operating characteristic curve (AUROC) in cross-validation. MLP was constructed with early stopping, batch normalization [9], and dropout [10] to avoid the overfitting problem, Glorot uniform initializer [11] to initiate the activation function, and the Nesterov Adam optimizer [12] to optimize the weight parameters.

The hyperparameters and its searching space for each model are as follows:

* Logistic regression: L2 regularization tern on weights (0.1 to 10).
* XGBoost: the maximum depth of the tree (5 to 80), a dropout rate which is a fraction of previous trees to drop (0 to 0.25), the minimum loss reduction required to make a further partition on a leaf node of the tree (0 to 10), L2 regularization tern on weights (0 to 5), the minimum sum of instance weight needed in a child (0.5 to 10), and the maximum delta step allowing each leaf output to be (0 to 10).
* Support vector machine: kernel type (polynomial kernel, radial basis function, or sigmoid kernel), L2 regularization tern on weights (0.1 to 10), and tolerance for stopping criterion (0.0001 to 0.01).
* LightGBM: the maximum number of leaves in one tree (30 to 130), the number of boosting iterations (70 to 300), the minimal gain required to perform split a leaf node (0 to 0.2), the minimum sum of instance weight needed in one leaf (0.0001 to 0.01), the minimum number of samples in one leaf (5 to 20), and L2 regularization tern on weights (0 to 0.4).MLP: depth of hidden layers (1 to 4), momentum of batch normalization layers (0.8 to 0.99), a dropout rate (0.5 to 0.7), and L2 regularization tern on weights (0 to 0.4).

Supplementary Tables

Supplementary Table 1. Missing rates of the data and types of variables.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | | Missing rate | Variable type |
| Demographics | | | |
|  | Sex | 0.0 % | Categorical |
|  | Age | 0.0 % | Continuous |
| Vital signs and O2 saturation | | | |
|  | Body temperature | 0.0 % | Continuous |
|  | Systolic blood pressure | 0.0 % | Continuous |
|  | Diastolic blood pressure | 0.1 % | Continuous |
|  | Mean atrial pressure | 0.4 % | Continuous |
|  | Heart rate | 0.0 % | Continuous |
|  | Respiration rate | 0.4 % | Continuous |
|  | SpO2 | 2.5 % | Continuous |
| Comorbidity | | | |
|  | Diabetes Mellitus | 1.4 % | Categorical |
|  | Hypertension | 1.2 % | Categorical |
|  | Malignancy | 1.4 % | Categorical |
|  | Chronic lung disease | 1.4 % | Categorical |
|  | Chronic liver disease | 1.4 % | Categorical |
|  | Chronic kidney disease | 1.4 % | Categorical |
|  | Cardiovascular disease | 1.4 % | Categorical |
|  | Cerebrovascular disease | 1.4 % | Categorical |
|  | Organ transplantation | 1.4 % | Categorical |
|  | AIDS | 1.4 % | Categorical |
|  | Other comorbidities | 1.4 % | Categorical |
| Infection source | | | |
|  | Respiratory | 0.0 % | Categorical |
|  | Genitourinary | 0.0 % | Categorical |
|  | Gastrointestinal | 0.0 % | Categorical |
|  | Bacteremia | 0.0 % | Categorical |
|  | Other Infection sources | 0.0 % | Categorical |
| Laboratory findings | | | |
|  | White blood cell | 0.2 % | Continuous |
|  | Platelet | 0.10% | Continuous |
|  | Glucose | 0.9 % | Continuous |
|  | Bilirubin | 0.2 % | Continuous |
|  | Creatinine | 0.2 % | Continuous |
|  | C-reactive protein | 0.5 % | Continuous |
|  | Procalcitonin | 16.7 % | Continuous |
|  | Initial lactate | 3.1 % | Continuous |
|  | F/U lactate within 12hr | 35.1 % | Continuous |
|  | Lactate clearance | 20.6 % | Continuous |
|  | Blood culture | 0.0 % | Categorical |
|  | Sputum culture | 0.0 % | Categorical |
|  | Urine culture | 0.0 % | Categorical |
| Arterial Blood Gas Analysis | | | |
|  | pH | 1.4 % | Continuous |
|  | PaCO2 | 1.6 % | Continuous |
|  | PaO2 | 1.4 % | Continuous |
|  | HCO3- | 1.6 % | Continuous |
|  | SaO2 | 1.6 % | Continuous |
|  | Septic shock | 0.0 % | Categorical |
|  | Mental status by AVPU | 0.0 % | Categorical |
|  | Glasgow coma scale | 0.1 % | Continuous |
|  | Visit route | 0.0 % | Categorical |
| Treatment | | | |
|  | Time to antibiotics | 0.4 % | Continuous |
|  | Antibiotics administered in previous hospital within 12 h from ED visit | 0.0 % |  |
|  | Antibiotics within 3 h from ED visit | 0.0 % | Categorical |
|  | Source control | 0.0 % | Categorical |
|  | Steroid administration within 12 h | 0.0 % | Categorical |

Supplementary Table 2. Baseline characteristics of the study population for 7-day mortality.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variables | | All | Survivors on 7-day | Non-survivors on 7-day | *p* value |
| (n = 810) | (n = 649) | (n = 161) |
| Demographics | | | | | |
|  | Female | 337 (41.6 %) | 263 (40.5%) | 74 (46.0%) | 0.244 |
|  | Age, years | 75 (65, 82) | 75 (63, 81) | 79 (69, 84) | < 0.001\*\*\* |
| Vital signs and O2 saturation | | | | | |
|  | Body temperature, °C | 37.1 (36.4, 38.0) | 37.2 (36.5, 38.1) | 36.6 (36.0, 37.5) | < 0.001\*\*\* |
|  | Systolic blood pressure, mmHg | 96 (80, 124.8) | 98 (82, 126) | 87 (72, 110) | < 0.001\*\*\* |
|  | Diastolic blood pressure, mmHg | 60 (50, 73) | 60 (51, 74) | 54 (46, 69) | < 0.001\*\*\* |
|  | Heart rate, bpm | 108 (90, 124) | 108 (90, 124) | 108 (88, 124) | 0.480 |
|  | Respiration rate, /min | 24.5 (5.8) | 24.1 (5.6) | 25.9 (6.4) | < 0.001\*\*\* |
|  | SpO2, % | 92.2 (8.7) | 93.2 (7.4) | 88.5 (12.1) | < 0.001\*\*\* |
| Comorbidity | | | | | |
|  | Diabetes Mellitus | 313 (39.2%) | 245 (37.8%) | 68 (42.2%) | 0.243 |
|  | Hypertension | 423 (52.9%) | 339 (52.2%) | 84 (52.2%) | 0.856 |
|  | Malignancy | 149 (18.6%) | 100 (15.4%) | 49 (30.4%) | < 0.001\*\*\* |
|  | Chronic lung disease | 187 (23.4%) | 146 (22.5%) | 41 (25.5%) | 0.400 |
|  | Chronic liver disease | 46 (5.8%) | 29 (4.5%) | 17 (10.6%) | 0.004\*\* |
|  | Chronic kidney disease | 107 (13.4%) | 91 (14.0%) | 16 (9.9%) | 0.250 |
|  | Cardiovascular disease | 143 (17.9%) | 119 (18.3%) | 24 (14.9%) | 0.426 |
|  | Cerebrovascular disease | 442 (55.3%) | 373 (57.5%) | 69 (42.9%) | 0.003\*\* |
|  | Organ transplantation | 11 (1.4%) | 10 (1.5%) | 1 (0.6%) | 0.620 |
|  | AIDS | 3 (0.4%) | 2 (0.3%) | 1 (0.6%) | 0.900 |
|  | Others | 209 (26.2%) | 172 (26.5%) | 37 (23.0%) | 0.502 |
|  | Unknown | 11 (1.4%) | 6 (0.9%) | 5 (3.1%) | 0.078 |
| Infection source | | | | | |
|  | Respiratory | 534 (65.9%) | 420 (64.7%) | 114 (70.8%) | 0.172 |
|  | Genitourinary | 301 (37.2%) | 252 (38.8%) | 49 (30.4%) | 0.060 |
|  | Gastrointestinal | 91 (11.2%) | 67 (10.3%) | 24 (14.9%) | 0.131 |
|  | Bacteremia | 61 (7.5%) | 46 (7.1%) | 15 (9.3%) | 0.428 |
|  | Others | 49 (6.0%) | 41 (6.3%) | 8 (5.0%) | 0.647 |
| Laboratory findings | | | | | |
|  | White blood cell, 103/μL | 11.9 (7.7, 17.5) | 12.0 (8.1, 17.5) | 11.6 (5.4, 17.8) | 0.048\* |
|  | Platelet, 103/μL | 200 (126, 284) | 201 (131, 284) | 196 (94.8, 274.8) | 0.046\* |
|  | Glucose, mg/dL | 176.8 (141.2) | 177.2 (134.8) | 175.2 (164.5) | 0.870 |
|  | Bilirubin, mg/dL | 1.1 (1.7) | 1.0 (1.5) | 1.5 (2.3) | 0.002\*\* |
|  | Creatinine, mg/dL | 1.8 (1.9) | 1.7 (1.8) | 2.2 (2.0) | 0.009\*\* |
|  | C-reactive protein, mg/L | 12.4 (10.1) | 12.2 (9.9) | 13.2 (10.7) | 0.276 |
|  | Procalcitonin, ng/mL | 13.9 (27.1) | 13.2 (26.4) | 16.7 (29.7) | 0.173 |
|  | Initial lactate, mg/dL | 4.3 (3.7) | 3.7 (3.1) | 6.6 (4.7) | < 0.001\*\*\* |
|  | F/U lactate within 12hr, mg/dL | 3.6 (3.3) | 2.9 (2.6) | 6.0 (4.3) | < 0.001\*\*\* |
|  | Lactate clearance, % | 28.3 (70.8) | 33.3 (54.8) | 9.4 (109.8) | < 0.001\*\*\* |
|  | Arterial Blood Gas Analysis | | | | |
|  | pH | 7.4 (0.1) | 7.4 (0.1) | 7.3 (0.2) | < 0.001\*\*\* |
|  | PaCO2, mmHg | 36.6 (14.2) | 36.4 (12.9) | 37.4 (18.4) | 0.461 |
|  | PaO2, mmHg | 84.9 (53.9) | 86.6 (57.0) | 78.1 (38.8) | 0.073 |
|  | HCO3-, mEq/L | 21.2 (7.5) | 21.8 (6.8) | 18.9 (9.5) | < 0.001\*\*\* |
|  | SaO2, % | 90.9 (10.3) | 92.0 (8.3) | 86.5 (15.0) | < 0.001\*\*\* |
| Clinical severity | | | | | |
|  | Septic shock | 445 (54.9%) | 400 (61.6%) | 45 (28.0%) | < 0.001\*\*\* |
|  | Glasgow coma scale | 10 (8, 13) | 10 (8, 13) | 10 (6, 12) | < 0.001\*\*\* |
|  | SOFA score | 8 (6, 11) | 8 (5, 11) | 11 (8, 13) | < 0.001\*\*\* |
|  | NEWS score | 11 (9, 13) | 11 (9, 12) | 12 (10, 15) | < 0.001\*\*\* |
|  | NEWS2 score | 11 (9, 13) | 11 (9, 13) | 12 (10, 15) | < 0.001\*\*\* |
|  | MEWS score | 6 (5, 8) | 6 (5, 8) | 6 (6, 8) | < 0.001\*\*\* |
|  | qSOFA score of 3 | 204 (25.2%) | 141 (21.7%) | 63 (39.1%) | < 0.001\*\*\* |
| Treatment | | | | | |
|  | Time to antibiotics, min | 96 (46, 162) | 94 (44, 162) | 101 (54, 160) | 0.159 |
|  | Steroid administration within 12hr | 84 (10.4%) | 62 (9.6%) | 22 (13.7%) | 0.165 |
|  | Antibiotics within 3hr | 713 (88.0%) | 573 (88.3%) | 140 (87.0%) | 0.741 |
| Source control | | | | | |
|  | Antibiotics only | 766 (94.6%) | 610 (94.0%) | 156 (96.9%) | 0.207 |
|  | Emergent surgery | 5 (0.6%) | 5 (0.8%) | 0 (0.0%) | 0.579 |
|  | Percutaneous drainage | 23 (2.8%) | 18 (2.8%) | 5 (3.1%) | 0.970 |
|  | Endoscopic intervention | 14 (1.7%) | 14 (2.2%) | 0 (0.0%) | 0.123 |
|  | Removal of infected device | 2 (0.2%) | 2 (0.3%) | 0 (0.0%) | 0.856 |

The values are expressed as n (%), mean (SD), or median (Q1, Q3). \**p* < 0.05, \*\**p* < 0.01, and \*\*\**p* < 0.001. Abbreviations: F/U, follow up; MEWS, modified early warning score; NEWS, national early warning score; NEWS2, national early warning score 2; qSOFA, quick sequential organ failure assessment; SaO2, Arterial oxygen saturation; SOFA, sequential organ failure assessment; SpO2, Saturation of percutaneous oxygen; PaCO2, partial pressure of carbon dioxide; PaO2, partial pressure of oxygen.

Supplementary Table 3. Baseline characteristics of the study population for 14-day mortality.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variables | | All | Survivors on 14-day | Non-survivors on 14-day | *p* value |
| (n = 810) | (n = 597) | (n = 213) |
| Demographics | | | | | |
|  | Female | 337 (41.6 %) | 242 (40.5%) | 95 (44.6%) | 0.341 |
|  | Age, years | 75 (65, 82) | 74 (63, 81) | 79 (69, 84) | < 0.001\*\*\* |
| Vital signs and O2 saturation | | | | | |
|  | Body temperature, °C | 37.1 (36.4, 38.0) | 37.2 (36.5, 38.1) | 36.6 (36.0, 37.5) | < 0.001\*\*\* |
|  | Systolic blood pressure, mmHg | 96 (80, 124.8) | 98 (83, 126) | 92 (74, 116) | < 0.001\*\*\* |
|  | Diastolic blood pressure, mmHg | 60 (50, 73) | 60 (51, 74) | 56 (47, 72) | < 0.001\*\*\* |
|  | Heart rate, bpm | 108 (90, 124) | 108 (90, 124) | 108 (92, 124) | 0.322 |
|  | Respiration rate, /min | 24.5 (5.8) | 24.1 (5.6) | 25.6 (6.2) | 0.001\*\* |
|  | SpO2, % | 92.2 (8.7) | 93.2 (7.3) | 89.5 (11.3) | < 0.001\*\*\* |
| Comorbidity | | | | | |
|  | Diabetes Mellitus | 313 (39.2%) | 219 (36.7%) | 94 (44.1%) | 0.047\* |
|  | Hypertension | 423 (52.9%) | 310 (51.9%) | 113 (53.1%) | 0.684 |
|  | Malignancy | 149 (18.6%) | 86 (14.4%) | 63 (29.6%) | < 0.001\*\*\* |
|  | Chronic lung disease | 187 (23.4%) | 131 (21.9%) | 56 (26.3%) | 0.194 |
|  | Chronic liver disease | 46 (5.8%) | 25 (4.2%) | 21 (9.9%) | 0.003\*\* |
|  | Chronic kidney disease | 107 (13.4%) | 85 (14.2%) | 22 (10.3%) | 0.205 |
|  | Cardiovascular disease | 143 (17.9%) | 110 (18.4%) | 33 (15.5%) | 0.433 |
|  | Cerebrovascular disease | 442 (55.3%) | 349 (58.5%) | 93 (43.7%) | < 0.001\*\*\* |
|  | Organ transplantation | 11 (1.4%) | 10 (1.7%) | 1 (0.5%) | 0.345 |
|  | AIDS | 3 (0.4%) | 2 (0.3%) | 1 (0.5%) | 0.711 |
|  | Others | 209 (26.2%) | 157 (26.3%) | 52 (24.4%) | 0.726 |
|  | Unknown | 11 (1.4%) | 6 (1.0%) | 5 (2.3%) | 0.268 |
| Infection source | | | | | |
|  | Respiratory | 534 (65.9%) | 382 (64.0%) | 152 (71.4%) | 0.062 |
|  | Genitourinary | 301 (37.2%) | 237 (39.7%) | 64 (30.0%) | 0.016\* |
|  | Gastrointestinal | 91 (11.2%) | 58 (9.7%) | 33 (15.5%) | 0.030\* |
|  | Bacteremia | 61 (7.5%) | 41 (6.9%) | 20 (9.4%) | 0.295 |
|  | Others | 49 (6.0%) | 38 (6.4%) | 11 (5.2%) | 0.643 |
| Laboratory findings | | | | | |
|  | White blood cell, 103/μL | 11.9 (7.7, 17.5) | 11.9 (8.1, 17.3) | 12.1 (6.3, 18.3) | 0.235 |
|  | Platelet, 103/μL | 200 (126, 284) | 205 (132, 285) | 182 (94.8, 270.2) | 0.005\*\* |
|  | Glucose, mg/dL | 176.8 (141.2) | 174.8 (131.5) | 182.5 (165.4) | 0.497 |
|  | Bilirubin, mg/dL | 1.1 (1.7) | 1.0 (1.4) | 1.4 (2.4) | < 0.001\*\*\* |
|  | Creatinine, mg/dL | 1.8 (1.9) | 1.7 (1.8) | 2.0 (1.9) | 0.042\* |
|  | C-reactive protein, mg/L | 12.4 (10.1) | 12.1 (10.0) | 13.4 (10.3) | 0.123 |
|  | Procalcitonin, ng/mL | 13.9 (27.1) | 13.4 (26.8) | 15.2 (28.2) | 0.461 |
|  | Initial lactate, mg/dL | 4.3 (3.7) | 3.7 (3.2) | 5.9 (4.5) | < 0.001\*\*\* |
|  | F/U lactate within 12hr, mg/dL | 3.6 (3.3) | 2.8 (2.5) | 5.5 (4.1) | < 0.001\*\*\* |
|  | Lactate clearance, % | 28.3 (70.8) | 35.2 (50.9) | 9.9 (104.5) | < 0.001\*\*\* |
|  | Arterial Blood Gas Analysis |  |  |  |  |
|  | pH | 7.4 (0.1) | 7.4 (0.1) | 7.3 (0.1) | < 0.001\*\*\* |
|  | PaCO2, mmHg | 36.6 (14.2) | 36.4 (12.9) | 37.3 (17.2) | 0.445 |
|  | PaO2, mmHg | 84.9 (53.9) | 86.3 (56.5) | 80.8 (45.8) | 0.204 |
|  | HCO3-, mEq/L | 21.2 (7.5) | 21.9 (6.8) | 19.4 (9.0) | < 0.001\*\*\* |
|  | SaO2, % | 90.9 (10.3) | 92.0 (8.4) | 87.8 (13.7) | < 0.001\*\*\* |
| Clinical severity | | | | | |
|  | Septic shock | 445 (54.9%) | 374 (62.6%) | 71 (33.3%) | < 0.001\*\*\* |
|  | Glasgow coma scale | 10 (8, 13) | 10 (8, 13) | 10 (7, 12) | 0.001\*\* |
|  | SOFA score | 8 (6, 11) | 8 (5, 10) | 10 (8, 12) | < 0.001\*\*\* |
|  | NEWS score | 11 (9, 13) | 11 (8, 12) | 12 (10, 14) | < 0.001\*\*\* |
|  | NEWS2 score | 11 (9, 13) | 11 (9, 13) | 12 (10, 14) | < 0.001\*\*\* |
|  | MEWS score | 6 (5, 8) | 6 (5, 8) | 6 (5, 8) | 0.001\*\* |
|  | qSOFA score of 3 | 204 (25.2%) | 128 (21.4%) | 76 (35.7%) | < 0.001\*\*\* |
| Treatment | | | | | |
|  | Time to antibiotics, min | 96 (46, 162) | 96 (45.2, 165.2) | 95 (49, 159) | 0.460 |
|  | Steroid administration within 12hr | 84 (10.4%) | 57 (9.5%) | 27 (12.7%) | 0.248 |
|  | Antibiotics within 3hr | 713 (88.0%) | 526 (88.1%) | 187 (87.8%) | 0.999 |
| Source control | | | | | |
|  | Antibiotics only | 766 (94.6%) | 562 (94.1%) | 204 (95.8%) | 0.466 |
|  | Emergent surgery | 5 (0.6%) | 5 (0.8%) | 0 (0.0%) | 0.406 |
|  | Percutaneous drainage | 23 (2.8%) | 16 (2.7%) | 7 (3.3%) | 0.828 |
|  | Endoscopic intervention | 14 (1.7%) | 12 (2.0%) | 2 (0.9%) | 0.469 |
|  | Removal of infected device | 2 (0.2%) | 2 (0.3%) | 0 (0.0%) | 0.967 |

The values are expressed as n (%), mean (SD), or median (Q1, Q3). \**p* < 0.05, \*\**p* < 0.01, and \*\*\**p* < 0.001. Abbreviations: F/U, follow up; MEWS, modified early warning score; NEWS, national early warning score; NEWS2, national early warning score 2; qSOFA, quick sequential organ failure assessment; SaO2, Arterial oxygen saturation; SOFA, sequential organ failure assessment; SpO2, Saturation of percutaneous oxygen; PaCO2, partial pressure of carbon dioxide; PaO2, partial pressure of oxygen.

Supplementary Table 4. Cross validation performance for 7-day, 14-day, and 30-day mortalities.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | | AUROC | | | AUPRC | | | Balanced accuracy | | |
| 7-day | 14-day | 30-day | 7-day | 14-day | 30-day | 7-day | 14-day | 30-day |
| Baseline model | | | | | | | | | | |
|  | LogReg | 0.84 (0.06) | 0.80 (0.06) | 0.79 (0.21) | 0.66 (0.09) | 0.64 (0.08) | 0.83 (0.15) | 0.66 (0.12) | 0.65 (0.09) | 0.72 (0.25) |
| Machine learning models | | | | | | | | | | |
|  | XGBoost | 0.90 (0.02) | 0.89 (0.03) | 0.86 (0.08) | 0.74 (0.04) | 0.80 (0.04) | 0.80 (0.10) | 0.76 (0.04) | 0.77 (0.05) | 0.74 (0.11) |
|  | SVM | 0.87 (0.05) | 0.85 (0.05) | 0.86 (0.04) | 0.70 (0.10) | 0.75 (0.10) | 0.80 (0.06) | 0.70 (0.11) | 0.73 (0.10) | 0.76 (0.05) |
|  | LightGBM | 0.93 (0.01) | 0.90 (0.01) | 0.90 (0.01) | 0.81 (0.03) | 0.82 (0.02) | 0.84 (0.02) | 0.78 (0.03) | 0.77 (0.02) | 0.79 (0.03) |
|  | MLP | 0.92 (0.02) | 0.89 (0.02) | 0.90 (0.03) | 0.80 (0.03) | 0.81 (0.04) | 0.84 (0.05) | 0.79 (0.05) | 0.78 (0.04) | 0.78 (0.05) |

The values are aggregated in all folds and expressed as mean (SD). Abbreviations: AUPRC, the area under the precision-recall curve; AUROC, LightGBM, LogReg, logistic regression; MLP multilayer perceptron; SVM, support vector machine; XGBoost, extreme gradient boosting.

****

Supplementary Fig. 1. Model calibration in prediction for 7-day mortality. Orange and blue lines represent calibration curves of raw and calibrated models (isotonic regression), respectively. The legend displays mean (SD) value of brier scores for model prediction, and solid lines and shades represent mean and ± SD of calibration curves.

****

Supplementary Fig. 2. Model calibration in prediction for 14-day mortality. Orange and blue lines represent calibration curves of raw and calibrated models (isotonic regression), respectively. The legend displays mean (SD) value of brier scores for model prediction, and solid lines and shades represent mean and ± SD of calibration curves.

****

Supplementary Fig. 3. Partial SHAP dependence plots for the selected variable.

Supplementary References

[1] Liu FT, Ting KM, Zhou ZH. Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data (TKDD). 2012; 6: 1–39.

[2] Van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. Journal of Statistical Software. 2011; 45: 1–67.

[3] Gregorutti B, Michel B, Saint-Pierre P. Correlation and variable importance in random forests. Statistics and Computing. 2017; 27: 659–678.

[4] Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST). 2011; 2: 1–27.

[5] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016; 785–794.

[6] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, *et al*. Lightgbm: a highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems. 2017; 30: 3146–3154.

[7] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521: 436–444.

[8] Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. ArXiv. 2012: 1206.2944.

[9] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning. 2015; 37: 448–456.

[10] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research. 2014; 15: 1929–1958.

[11] Hanin B, Rolnick D. How to start training: the effect of initialization and architecture. ArXiv. 2018: 1803.01719.

[12] Dozat T. Incorporating nesterov momentum into adam. International Conference on Learning Representations: ICLR 2016-Workshop Track. 2016.