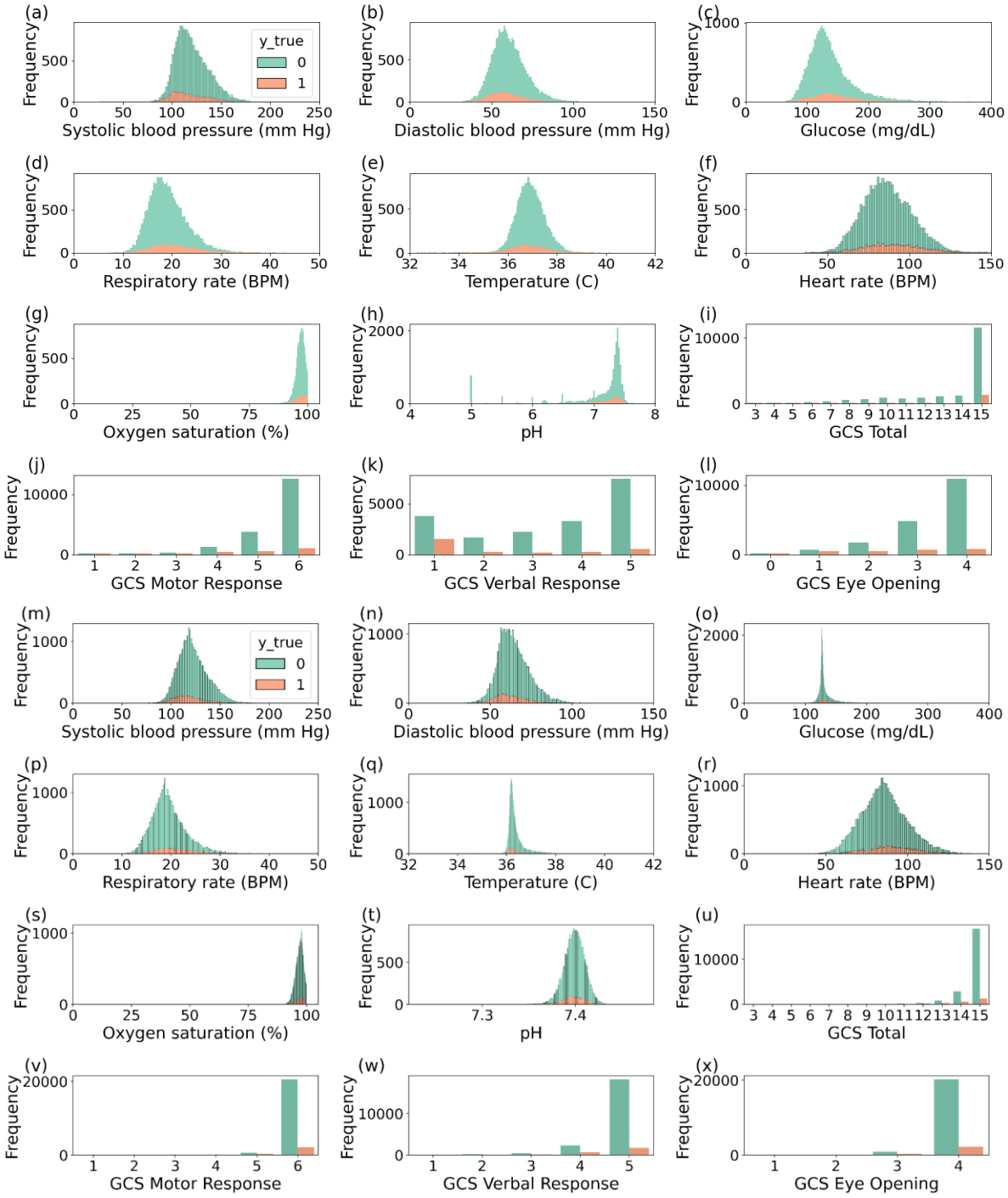
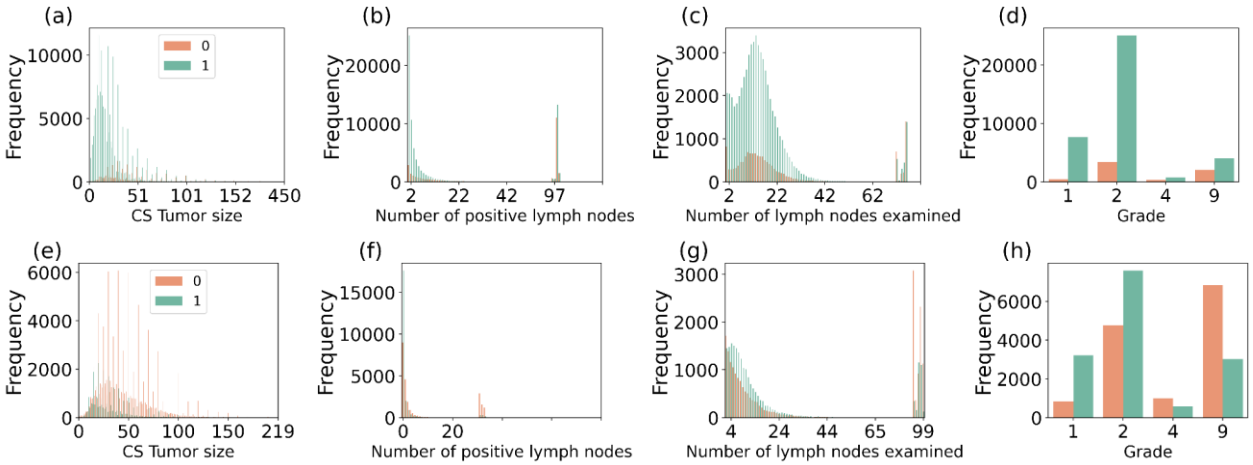
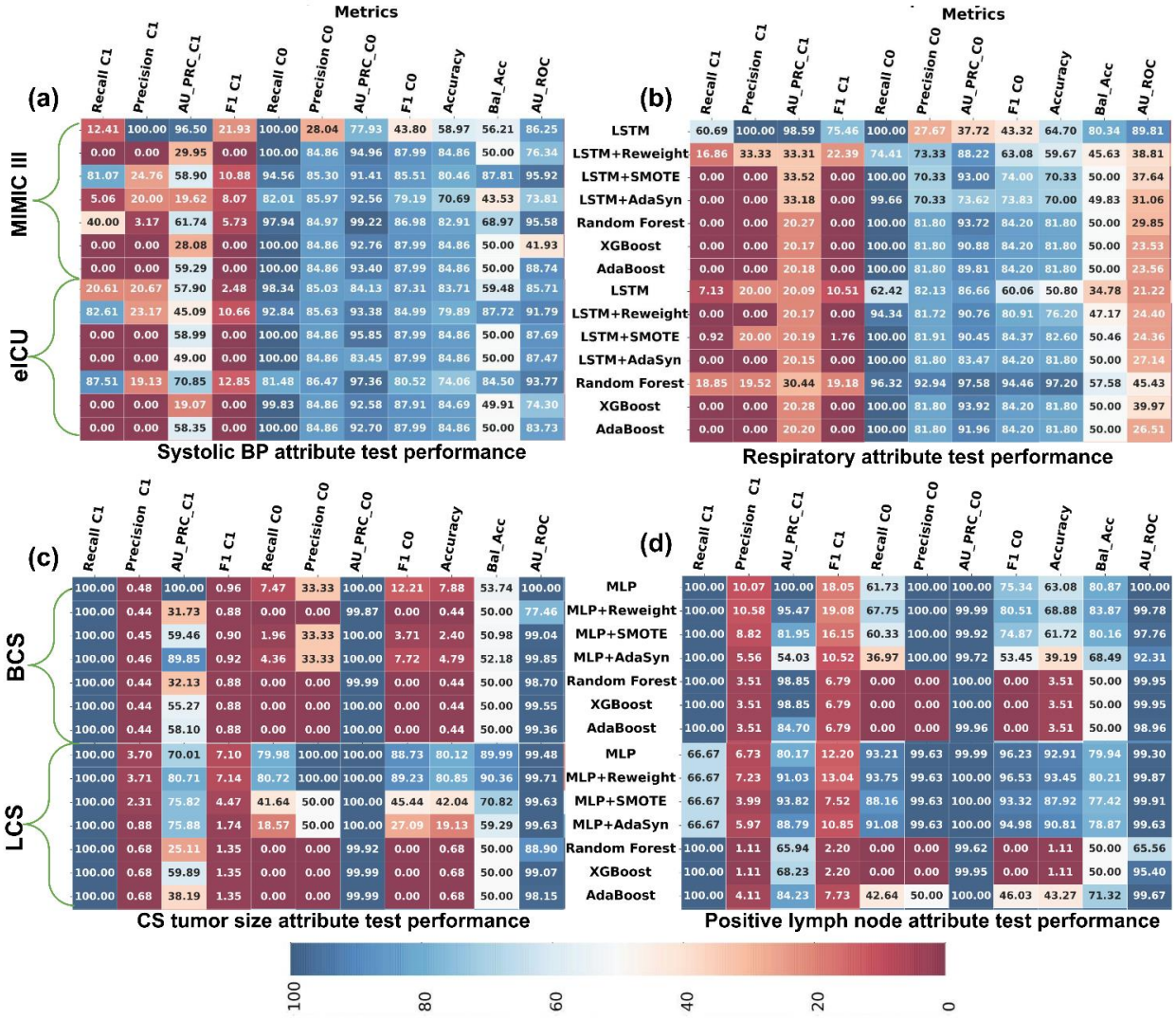
# Supplementary Materials



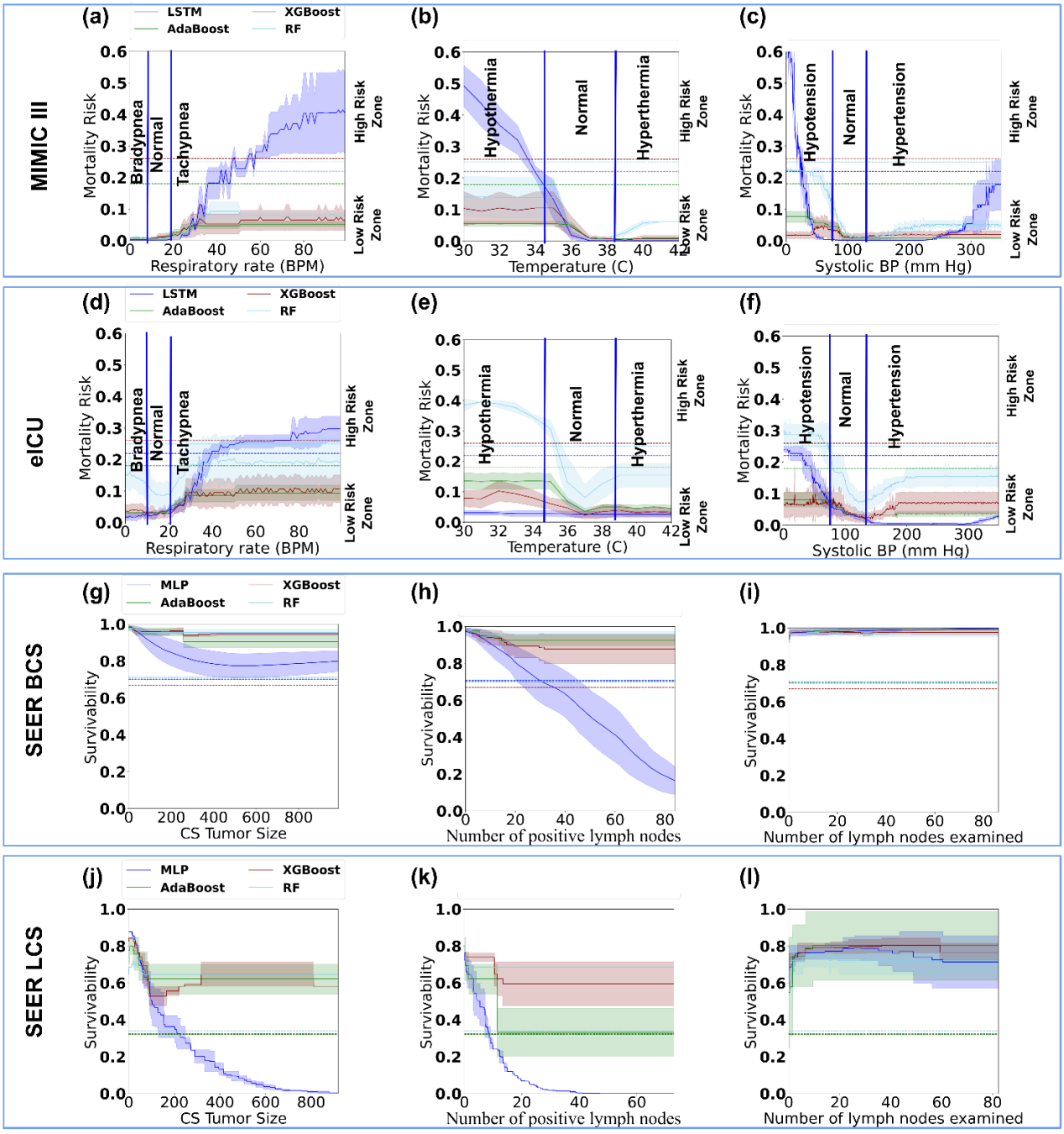

**Supplementary Figure S1. Distributions of different attributes (vitals) of in-hospital mortality risk prediction datasets.** (a)-(l) MIMIC-III and (m)-(x) eICU 48 hours ICU mortality dataset.



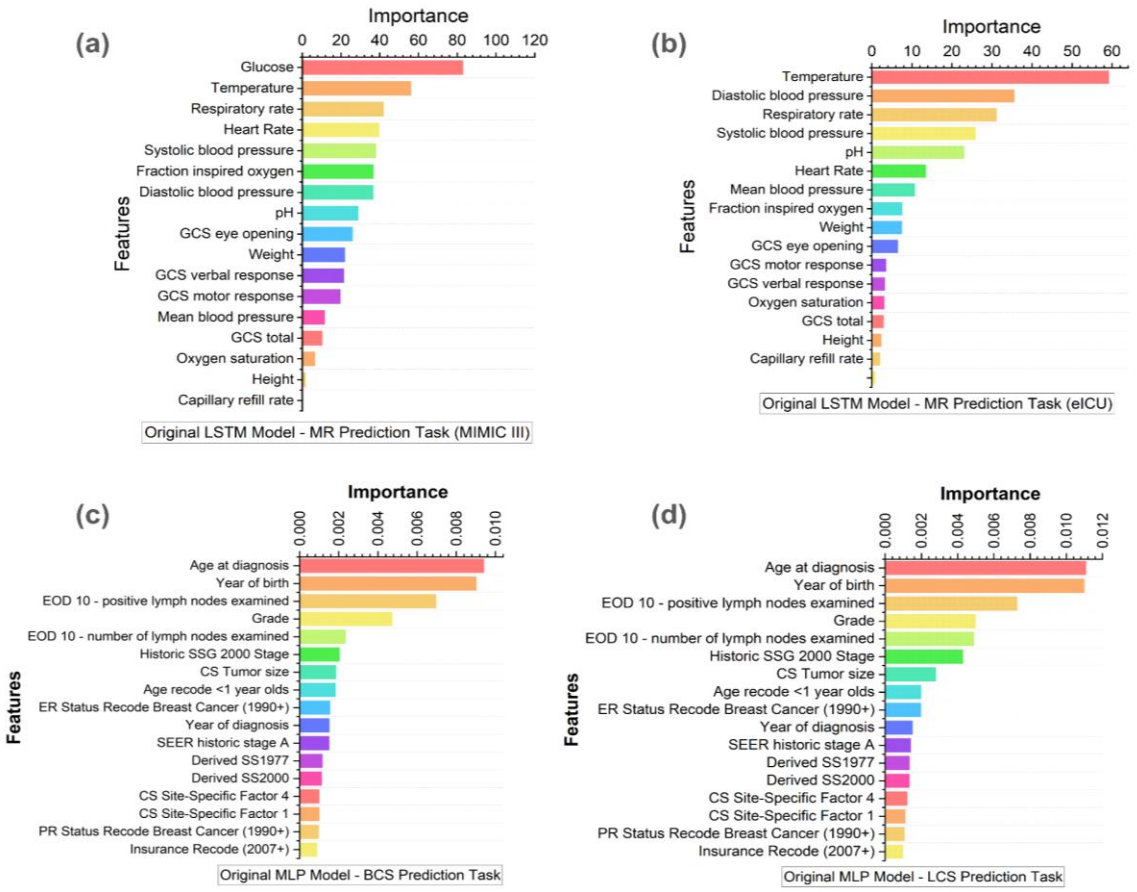

**Supplementary Figure S2: Distributions of the SEER 5-year cancer prediction dataset.** (a)-(d) breast and (e)-(h) lung cancer survivability dataset. Here label 0 represents the death case and 1 is mapped to the survival case. Death cases are the majority in the LCS dataset, whereas survival cases are the majority in the BCS dataset.



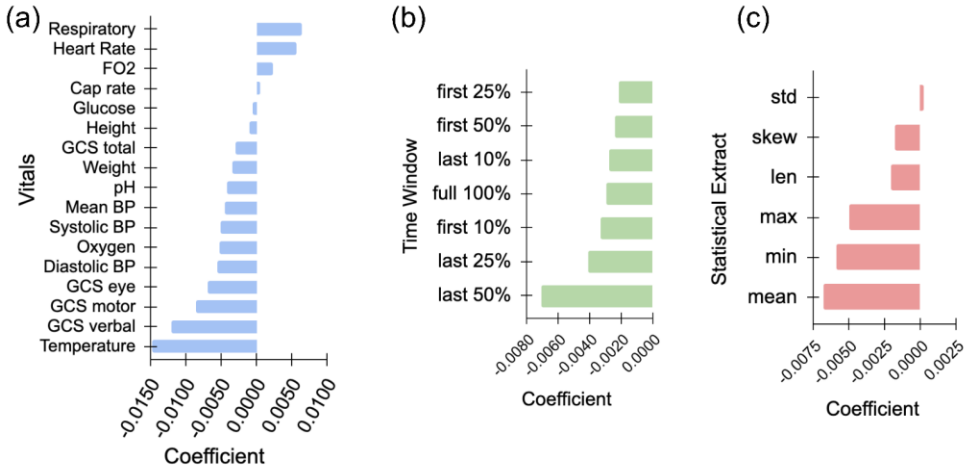

**Supplementary Figure S3. The performance of machine learning models was evaluated on synthesized single-attribute test sets.** C0 and C1 represent class 0 and class 1 respectively. AU_PRC, Bal_Acc, and AU_ROC represent area under Precision-Pecall Curve, Balanced Accuracy, and Area Under Receiver Operating Curve, respectively. The performance on (a) ICU vital systolic blood pressure, (b) ICU vital respiratory rate, (c) cancer attribute CS tumor size, and (d) cancer attribute number of positive lymph nodes. Models were trained on the original datasets, with the top halves of (a) and (b) representing MIMIC III and the bottom halves representing eICU. For (c) and (d), the top halves represent SEER BCS and the bottom halves represent SEER LCS, as indicated on the left. Model names are aligned along the Y-axis. For ICU mortality prediction (a) and (b), Class 1 (C1) corresponds to the death class, whereas for SEER cancer survivability prediction (c) and (d), Class 0 (C0) represents the death class.



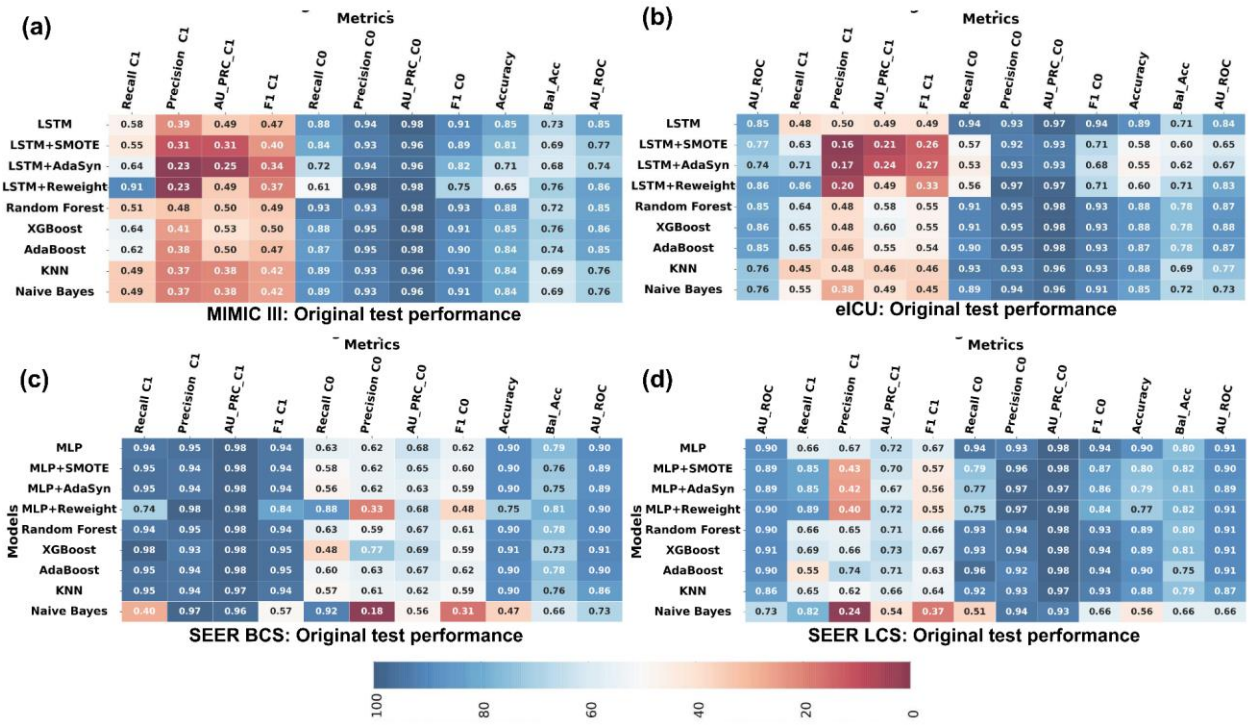

**Supplementary Figure S4**: **Performance comparison between tree-based ensemble methods, AdaBoost, XGBoost, and Random Forest (RF), with LSTM and MLP models under single-attribute varying tests.** (a)-(c) and (d)-(f) Mortality risk prediction results by the models under MIMIC-III and eICU test cases for respiratory rate, temperature, and systolic blood pressure, respectively. (g)-(i) and (j)-(l) 5-year cancer survivability prediction results by the models under SEER BCS and LCS test cases for CS tumor size, the number of positive lymph nodes, and the number of lymph nodes examined, respectively. Horizontal dashed lines represent model-specific thresholds.




**Supplementary Figure S5: SHAP-average (of one-hot encoded features) feature importance of machine learning models.** (a) LSTM trained on MIMIC III dataset, (b) LSTM trained on eICU dataset, (c) MLP trained on BCS dataset. (d) MLP trained on LCS dataset.




**Supplementary Figure S6: Logistic regression coefficients (averaged)**. (a) vitals, (b) time-period, and (c) statistical-extract features.




**Supplementary Figure S7: Machine learning model performance on the original test set.** C0 and C1 represent class 0 and class 1 respectively. AU_PRC, Bal_Acc, and AU_ROC represent area under Precision-Pecall Curve, Balanced Accuracy, and Area Under Receiver Operating Curve, respectively. The death class is the minority class in MIMIC III, eICU, and SEER BCS datasets, whereas the majority in the SEER LCS dataset. Death class is represented by 1 in in-hospital mortality risk prediction datasets (MIMIC III and eICU) and 0 in cancer survivability prediction datasets (SEER BCS and LCS).

**Supplementary Notes**

Here, we define the medical terminology used. In-hospital mortality prediction Class 1 refers to the patients who die in the ICU within 48 hours of stay. Breast cancer survivability Class 1 refers to the patients who survive more than 5 years after breast cancer diagnosis. Lung cancer survivability Class 1 refers to the patients who survive more than 5 years after the lung cancer diagnosis. Glasgow Coma Scale (GCS) is a scale for assessing a person's level of consciousness by scoring their eye, verbal, and motor responses. The mortality risk range (threshold, 1.0] is defined as the high-risk zone. The mortality risk range [0, threshold] is defined as the low-risk zone.

Bradypnea means abnormally slow breathing (< 12 BPM). Tachycardia means abnormally fast heart rate (> 100 BPM). Hypothermia means body temperature below normal (< 35°C / 95°F). Hyperthermia means elevated body temperature (> 38.5°C / 101.3°F). Hypoglycemia means low blood glucose levels (< 70 mg/dL). Hyperglycemia means high blood glucose levels (> 180 mg/dL). Hypotension means low blood pressure (< 80/60 mmHg). Hypertension means high blood pressure (> 130/90 mmHg). Hypoxemia means low blood oxygen saturation (< 90% SpO2).

Collaborative Staging (CS) tumor size records the largest dimension (length, width, or height) or the diameter of the primary tumor in millimeters. When describing the extent (size) of the tumor, T0 represents the absence of tumor, T1 means that tumor size is less than 20 mm, T2 20-50 mm, T3 larger than 50 mm, and T4 means any size growing into skin. When describing the spread of tumor to nearby lymph nodes, N0 means 0 positive lymph nodes, N1 1-3, N2 4-9, and N3 means more than 10 positive lymph nodes. ELNs refer to examined number of lymph nodes.

**Supplementary Equations**

Given neuron *i*, its activation value $act(neuron_i(k))$ for input *k*, and a vital zone bounded by $n_1$ and $n_2$ (representing critically low, critically high, or normal vital range), we define the average neural zone activation $NZA_i(n_1, n_2)$ in Equation (1). Given two vital zones $z_1$ and $z_2$, we further define the average neural zone activation difference $\Delta NZA_{avg}\ (z_1, z_2)$ in Equation (2). In Equation (2), $z_{is}$ and $z_{ie}$ are the zone starting and ending values, respectively. $b$ is the number of neurons in the activation layer (16 in our case).

$$NZA_i\ (n_1, n_2)\ =\ \frac{1}{|\ n_1 - n_2|} \sum_{k=n_1}^{n_2} act(neuron_i(k)) \qquad (1)$$

$$\Delta NZA_{avg}\ (z_1, z_2) = \frac{1}{b} \sum_{i=1}^{b} |\ NZA_i(z_{1\,s}, z_{1\,e})\ - NZA_i(z_{2\,s}, z_{2\,e})\ | \qquad (2)$$

**Gradient ascent**

Given a sample $x$, we add the gradient $G_i$ calculated from the trained model with the seed sample $x_i$ to generate a new sample $x_i^{new}$, where $i$ represents the feature to be changed by the gradient ascent process.

$$x_i^{new}\ =\ x_i\ +\ \alpha\ G_i \qquad (3)$$

Where:

$i$ is the selected feature index from [0, 1, …, 17]

$x_i^{new}$ is the sample with updated feature i

$x_i$ is the current feature value

$\alpha$ is the step size or learning rate that controls the magnitude of change in the input feature

$G_i = \frac{\delta y}{\delta x_i}$ is the gradient of the output y with respect to feature $x_i$

**Supplementary Methods**

**Algorithm for Single-Attribute Variation Test Case Generation for Time Series Data**

Input:

- Dataset D (e.g., MIMIC-III or eICU)
- Attribute A∈ $[Diastolic\ bp, Glucose, Respiratory\ rate, \ldots..]$
- Time-series length T = 48 hours
- Seed S

Output: Set of test cases with varied single attribute A

Procedure:

1. Compute minimum $A_{min}$, maximum $A_{max}$, mean $\mu_A$, and standard deviation $\sigma_A$ values of attribute A in dataset D.
2. For each value μ in the range $[A_{min}, A_{max}]$ with step size $\alpha$
   - 2.1. $x_A$: Generate a vector of length T by sampling random values from the normal distribution $N(\mu, \sigma_A^2)$
   - 2.2. $S_\mu$: Replace the feature value of A with the new vector $x_A$ in seed S
   - 2.3. Save the new sample, $S_\mu$ which is a representative of attribute value μ

**Supplementary Table 1:** Best hyperparameters of machine learning models selected using a grid search

| Model | Hyperparameter search set | Best hyperparameters | | | |
|---|---|---|---|---|---|
| | | ***eICU*** | ***MIMIC-III*** | ***SEER-BCS*** | ***SEER-LCS*** |
| **XGBoost** | n_estimators: [50, 100, 150, 200, 300] | 200 | 300 | 300 | 300 |
| | max_depth: [None, 3, 5, 7, 10] | 5 | 3 | 5 | 5 |
| | learning_rate: [0.01, 0.1, 0.2, 0.5, 1] | 0.1 | 0.1 | 0.1 | 0.1 |
| | subsample: [0.1, 0.5, 0.8, 1.0] | 1 | 1 | 0.8 | 0.8 |
| | colsample_bytree: [0.1, 0.5, 0.8, 1.0] | 1 | 1 | 0.8 | 1 |
| **AdaBoost** | n_estimators: [50, 100, 150, 200, 300] | 200 | 200 | 150 | 150 |
| | learning_rate: [0.01, 0.1, 0.2, 0.5, 1] | 0.1 | 0.1 | 1 | 1 |
| **Random Forest** | n_estimators: [50, 100, 150, 200, 300] | 300 | 300 | 300 | 300 |
| | max_depth: [None, 3, 5, 7, 10] | None | None | None | None |
| | min_samples_split: [2, 5, 10] | 10 | 10 | 2 | 10 |
| | min_samples_leaf: [1, 2, 4] | 4 | 4 | 4 | 2 |
| **K Nearest Neighbors** | n_neighbors: [3, 5, 7, 9, 11] | 11 | 11 | 11 | 11 |
| | weights: ['uniform', 'distance'] | distance | distance | distance | distance |
| | metric: ['euclidean', 'manhattan', 'minkowski'] | euclidean | manhattan | euclidean | euclidean |
| | p: [1, 2] | 1 | 1 | 1 | 1 |
| **Naive Bayes** | var_smoothing: [1e-9, 1e-8, 1e-7, 1e-6] | 1e-7 | 1e-6 | 1e-9 | 1e-9 |
| **Transformer** | number of heads: [2, 4, 8] | 4 | 4 | -- | -- |
| | number of transformer blocks: [1, 2, 3] | 2 | 3 | -- | -- |
| | dense units (transformer block): [16, 64, 256] | 16 | 16 | -- | -- |
| | dropout: [0.1, 0.2, 0.3, 0.4, 0.5] | 0.5 | 0.3 | -- | -- |
| | batch normalization: [True, False] | True | True | -- | -- |
| | dense units: [16, 32, 64] | 32 | 64 | -- | -- |

**Supplementary Table 2:** Training and validation losses of selected in-hospital mortality risk predictor models.

| Prediction task | Model | Trial | Selected Epoch (1-100) | Validation AUPRC (C1) | Training Loss | Validation Loss | Difference (Training loss ~ Validation loss) |
|---|---|---|---|---|---|---|---|
| MIMIC III | *LSTM* | *1* | 39 | 0.568 | 0.273 | 0.278 | 0.005 |
| | | *2* | 29 | 0.564 | 0.279 | 0.278 | 0.001 |
| | | *3* | 33 | 0.566 | 0.274 | 0.279 | 0.005 |
| | *CW-LSTM* | *1* | 37 | 0.556 | 0.270 | 0.284 | 0.014 |
| | | *2* | 24 | 0.556 | 0.280 | 0.286 | 0.006 |
| | | *3* | 20 | 0.561 | 0.283 | 0.281 | 0.002 |
| | *LSTM+ SMOTE* | *1* | 4 | 0.410 | 0.438 | 0.385 | 0.054 |
| | | *2* | 11 | 0.410 | 0.342 | 0.355 | 0.013 |
| | | *3* | 2 | 0.366 | 0.473 | 0.417 | 0.056 |
| | *LSTM+ AdaSyn* | *1* | 18 | 0.337 | 0.318 | 0.411 | 0.093 |
| | | *2* | 6 | 0.413 | 0.395 | 0.353 | 0.042 |
| | | *3* | 33 | 0.408 | 0.305 | 0.343 | 0.038 |
| | *LSTM+ Reweight* | *1* | 63 | 0.568 | 0.436 | 0.429 | 0.007 |
| | | *2* | 58 | 0.559 | 0.428 | 0.431 | 0.003 |
| | | *3* | 66 | 0.561 | 0.429 | 0.424 | 0.005 |
| eICU | *LSTM* | *1* | 81 | 0.476 | 0.277 | 0.280 | 0.003 |
| | | *2* | 87 | 0.510 | 0.274 | 0.283 | 0.009 |
| | | *3* | 86 | 0.492 | 0.279 | 0.276 | 0.003 |
| | *LSTM+ SMOTE* | *1* | 49 | 0.385 | 0.495 | 0.410 | 0.085 |
| | | *2* | 55 | 0.402 | 0.445 | 0.391 | 0.055 |
| | | *3* | 42 | 0.433 | 0.412 | 0.471 | 0.059 |
| | *LSTM+ AdaSyn* | *1* | 39 | 0.320 | 0.410 | 0.419 | 0.009 |
| | | *2* | 38 | 0.376 | 0.454 | 0.499 | 0.045 |
| | | *3* | 35 | 0.336 | 0.439 | 0.468 | 0.029 |
| | *LSTM+ Reweight* | *1* | 99 | 0.473 | 0.532 | 0.526 | 0.006 |
| | | *2* | 90 | 0.495 | 0.512 | 0.539 | 0.026 |
| | | *3* | 89 | 0.454 | 0.494 | 0.502 | 0.008 |

**Supplementary Table 3:** Training and validation losses of selected cancer survivability predictor models.

| Prediction task | Model | Trial | Selected Epoch | Validation AUPRC (minority) | Training Loss | Validation Loss | Difference (Training loss ~ Validation loss) |
|---|---|---|---|---|---|---|---|
| **SEER BCS** | *MLP* | *1* | 3 | 0.673 | 0.231 | 0.230 | 0.002 |
| | | *2* | 3 | 0.672 | 0.232 | 0.229 | 0.003 |
| | | *3* | 4 | 0.675 | 0.230 | 0.228 | 0.002 |
| | *MLP+ SMOTE* | *1* | 4 | 0.654 | 0.323 | 0.340 | 0.017 |
| | | *2* | 2 | 0.650 | 0.350 | 0.340 | 0.010 |
| | | *3* | 5 | 0.647 | 0.310 | 0.338 | 0.028 |
| | *MLP+ AdaSyn* | *1* | 1 | 0.643 | 0.393 | 0.384 | 0.010 |
| | | *2* | 3 | 0.622 | 0.391 | 0.357 | 0.034 |
| | | *3* | 2 | 0.618 | 0.378 | 0.353 | 0.025 |
| | *MLP+ Reweight* | *1* | 7 | 0.677 | 0.389 | 0.352 | 0.037 |
| | | *2* | 5 | 0.666 | 0.391 | 0.354 | 0.037 |
| | | *3* | 5 | 0.672 | 0.389 | 0.361 | 0.028 |
| **SEER LCS** | *MLP* | *1* | 5 | 0.724 | 0.248 | 0.254 | 0.006 |
| | | *2* | 4 | 0.725 | 0.250 | 0.254 | 0.004 |
| | | *3* | 4 | 0.727 | 0.245 | 0.244 | 0.002 |
| | *MLP+ SMOTE* | *1* | 3 | 0.705 | 0.311 | 0.339 | 0.028 |
| | | *2* | 5 | 0.710 | 0.322 | 0.345 | 0.024 |
| | | *3* | 4 | 0.708 | 0.341 | 0.362 | 0.021 |
| | *MLP+ AdaSyn* | *1* | 3 | 0.682 | 0.328 | 0.391 | 0.063 |
| | | *2* | 3 | 0.701 | 0.342 | 0.366 | 0.024 |
| | | *3* | 4 | 0.682 | 0.322 | 0.375 | 0.054 |
| | *MLP+ Reweight* | *1* | 3 | 0.726 | 0.376 | 0.351 | 0.026 |
| | | *2* | 3 | 0.715 | 0.367 | 0.354 | 0.012 |
| | | *3* | 5 | 0.725 | 0.371 | 0.362 | 0.009 |

**Supplementary Table 4**: Number of samples in the training set after resampling

| Dataset | Resampling Method | Total Samples | Class 0 | Class 1 |
|---|---|---|---|---|
| **MIMIC III** | *SMOTE* | 25,222 | 12,694 | 12,528 |
| | *AdaSyn* | 24,872 | 12,694 | 12,178 |
| **eICU** | *SMOTE* | 38,030 | 19,015 | 19,015 |
| | *AdaSyn* | 37,417 | 19,015 | 18,402 |
| **BCS** | *SMOTE* | 347,436 | 173,718 | 173,718 |
| | *AdaSyn* | 344,528 | 170,810 | 173,718 |
| **LCS** | *SMOTE* | 276,218 | 138,109 | 138,109 |
| | *AdaSyn* | 277,264 | 138,109 | 139,155 |

**Supplementary Table 5**: Cost-sensitive learning balanced class weights.

| Dataset | Class 0 | Class 1 |
|---|---|---|
| MIMIC III | 0.578 | 3.694 |
| eICU LSTM | 0.565 | 4.363 |
| LCS MLP | 0.595 | 3.122 |
| BCS MLP | 3.936 | 0.573 |

**Supplementary Table 6:** Thresholds of various models. Thresholds are identified through the validation process.

| Models/ Datasets | NN | NN+ SMOTE | NN + AdaSyn | NN+ Reweight | XGBoost | Random Forest | Ada-Boost | KNN | Naive Bayes |
|---|---|---|---|---|---|---|---|---|---|
| **eICU** (NN: LSTM) | 0.24 | 0.11 | 0.14 | 0.24 | 0.22 | 0.24 | 0.24 | 0.21 | 0.17 |
| **MIMIC-III** (NN: LSTM) | 0.22 | 0.19 | 0.18 | 0.22 | 0.26 | 0.25 | 0.18 | 0.21 | 0.06 |
| **SEER-BCS** (NN: MLP) | 0.71 | 0.694 | 0.695 | 0.69 | 0.67 | 0.70 | 0.70 | 0.71 | 0.83 |
| **SEER-LCS** (NN: MLP) | 0.34 | 0.29 | 0.31 | 0.31 | 0.32 | 0.34 | 0.32 | 0.31 | 0.11 |

**Supplementary Table 7:** Mortality risk prediction of two medical experts on attribute-varying test cases. Doctors were given below and told the attribute values may fluctuate.

| Case | Attribute values | Predicted risk (0-1) of attribute-varying test cases | | | |
|---|---|---|---|---|---|
| | | Risk score by doctor 1 | Risk score by doctor 2 | Doctors' avg | Risk by MIMIC III LSTM model |
| **1** | *Systolic blood pressure (mm Hg): 52*<br>*Diastolic blood pressure (mm Hg): 48*<br>*Glucose level (mg/dL): 22*<br>*Respiratory rate (BPM): 37*<br>*Heart rate (BPM): 21*<br>*Oxygen saturation (%): 34* | 0.75 | 0.95 | 0.85 | 0.084 |
| **2** | *Systolic blood pressure (mm Hg): 79*<br>*Diastolic blood pressure (mm Hg): 35*<br>*Glucose level (mg/dL): 47*<br>*Respiratory rate (BPM): 41*<br>*Heart rate (BPM): 20*<br>*Oxygen saturation (%): 43* | 0.80 | 0.90 | 0.85 | 0.137 |
| **3** | *Systolic blood pressure (mm Hg): 61*<br>*Diastolic blood pressure (mm Hg): 32*<br>*Glucose level (mg/dL): 35*<br>*Respiratory rate (BPM): 42*<br>*Heart rate (BPM): 48*<br>*Oxygen saturation (%): 26* | 0.85 | 0.90 | 0.875 | 0.611 |
| **4** | *Systolic blood pressure (mm Hg): 61*<br>*Diastolic blood pressure (mm Hg): 43*<br>*Glucose level (mg/dL): 35*<br>*Respiratory rate (BPM): 42*<br>*Heart rate (BPM): 48*<br>*Oxygen saturation (%): 20* | 0.85 | 0.90 | 0.875 | 0.56 |
| **5** | *Systolic blood pressure (mm Hg): 223*<br>*Diastolic blood pressure (mm Hg): 153*<br>*Glucose level (mg/dL): 253*<br>*Respiratory rate (BPM): 30*<br>*Heart rate (BPM): 117*<br>*Oxygen saturation (%): 68* | 0.50 | 0.50 | 0.50 | 0.117 |
| **6** | *Systolic blood pressure (mm Hg): 295*<br>*Diastolic blood pressure (mm Hg): 153*<br>*Glucose level (mg/dL): 253*<br>*Respiratory rate (BPM): 45*<br>*Heart rate (BPM): 136*<br>*Oxygen saturation (%): 62* | 0.55 | 0.60 | 0.575 | 0.455 |

**Supplementary Table 8:** Deteriorating test cases labeling of mortality risk by 2 medical doctors.

| Attribute name | Case | Predicted risk (0-1) of deteriorating test cases | | | |
|---|---|---|---|---|---|
| | | Medical Doctor 1 | Medical Doctor 2 | Doctors' avg | MIMIC III LSTM model (avg) |
| Respiratory rate | *1* | 0.85 | 0.60 | 0.73 | 0.13 |
| | *2* | 0.90 | 0.80 | 0.85 | 0.19 |
| | *3* | 0.90 | 0.60 | 0.75 | 0.16 |
| Temperature | *1* | 0.90 | 0.40 | 0.65 | 0.18 |
| | *2* | 0.50 | 0.35 | 0.43 | 0.17 |
| | *3* | 0.40 | 0.30 | 0.35 | 0.18 |
| Systolic BP | *1* | 0.95 | 0.90 | 0.93 | 0.13 |
| | *2* | 0.95 | 0.90 | 0.93 | 0.03 |
| | *3* | 0.95 | 0.90 | 0.93 | 0.20 |
| Multi-attribute | *1* | 0.80 | 0.90 | 0.85 | 0.20 |
| | *2* | 0.85 | 0.90 | 0.88 | 0.18 |
| | *3* | 0.80 | 0.90 | 0.85 | 0.19 |

**Supplementary Table 9:** The table shows the gradient-based test case generation steps and prediction. Each step of gradient ascent creates a new test case by changing a single attribute value. Mortality risk (MR) is predicted by three models which are trained on the same MIMIC-III training dataset.

| Attribute | Deteriorating test case | Step factor | # of Steps | Seed attribute value | Final attribute value | MR by LSTM 1 | MR by LSTM 2 | MR by LSTM 3 | MR by LR | MR by CW-LSTM |
|---|---|---|---|---|---|---|---|---|---|---|
| Diastolic BP (mm Hg) | *1* | 0.01 | 16 | 52.94 | 10.77 | 0.04 | 0.22 | 0.22 | 0.03 | 0.05 |
| | *2* | 0.01 | 100 | 55.89 | 1.19 | 0.10 | 0.04 | 0.08 | 0.20 | 0.02 |
| | *3* | 0.01 | 100 | 89.37 | 4.49 | 0.07 | 0.06 | 0.09 | 0.14 | 0.05 |
| Systolic BP (mm Hg) | *1* | 0.2 | 54 | 119.85 | 95.00 | 0.14 | 0.22 | 0.06 | 0.02 | 0.27 |
| | *2* | 0.2 | 35 | 124.70 | 105.35 | 0.07 | 0.11 | 0.22 | 0.13 | 0.19 |
| | *3* | 0.2 | 170 | 125.59 | 67.99 | 0.18 | 0.21 | 0.22 | 0.05 | 0.37 |
| Resp. Rate (BPM) | *1* | 0.01 | 82 | 15.67 | 27.51 | 0.22 | 0.12 | 0.06 | 0.05 | 0.28 |
| | *2* | 0.01 | 138 | 17.49 | 31.14 | 0.17 | 0.22 | 0.20 | 0.45 | 0.22 |
| | *3* | 0.01 | 218 | 20.49 | 42.40 | 0.11 | 0.15 | 0.22 | 0.32 | 0.11 |
| Oxygen Sat. (%) | *1* | 0.01 | 9 | 69.26 | 0.00 | 0.03 | 0.15 | 0.15 | 0.18 | 0.01 |
| | *2* | 0.01 | 21 | 69.83 | 0.00 | 0.03 | 0.01 | 0.03 | 0.67 | 0.01 |
| | *3* | 0.01 | 44 | 80.68 | 0.00 | 0.02 | 0.01 | 0.01 | 0.38 | 0.03 |
| Temp (C) | *1* | 0.001 | 27 | 36.78 | 35.96 | 0.15 | 0.22 | 0.20 | 0.02 | 0.18 |
| | *2* | 0.001 | 38 | 36.62 | 35.64 | 0.15 | 0.14 | 0.23 | 0.14 | 0.10 |
| | *3* | 0.001 | 57 | 36.79 | 35.41 | 0.15 | 0.17 | 0.23 | 0.05 | 0.13 |

**Supplementary Table 10:** Number of test cases (excluding deteriorating condition) in one test set created from a single seed case in the original MIMIC-III dataset. Five test sets are produced for each attribute or combination of attributes using five different seeds. The deteriorating conditions test is generated by three seeds resulting in 12 test cases.

| Test type | Tested attribute | Number of cases per set | Number of total cases |
|---|---|---|---|
| **GCS test** | *GCS total* | 120 | 600 |
| | *GCS eye-motor* | 24 | 120 |
| | *GCS eye-verbal* | 20 | 100 |
| | *GCS motor-verbal* | 30 | 150 |
| **GCS test total case** | | 194 | 970 |
| **Single attribute critical zone test** | *Diastolic BP* | 250 | 1,250 |
| | *Glucose* | 600 | 3,000 |
| | *Oxygen saturation* | 101 | 505 |
| | *Respiratory rate* | 101 | 505 |
| | *Systolic BP* | 350 | 1,750 |
| | *Temperature* | 13 | 65 |
| **Critical zone test total case** | | 1415 | 7,075 |
| **Double-attribute critical zone test** | *Diastolic BP & Glucose* | 2,500 | 12,500 |
| | *Diastolic BP & Systolic BP* | 3,500 | 17,500 |
| | *Respiratory rate & Heart rate* | 2,500 | 12,500 |
| **Double attribute critical zone test total case** | | 8,500 | 42,500 |
| **Multi-attribute critical zone test (6 attributes)** | *High zone* | 12,695 | 63,475 |
| | *Low zone* | 12,695 | 63,475 |
| **Multi-attribute critical zone test (6 attributes) total case** | | 25,390 | 126,950 |
| **Deteriorating conditions test (gradient ascent method)** | *Single attribute* | 9 | 9 |
| | *Triple-attribute* | 3 | 3 |
| **Deteriorating conditions test total case** | | 12 | 12 |
| **Grand total** | | 35,511 | 177,507 |

**Supplementary Table 11:** Clinical Assessment of Neurological Function: Glasgow Coma Scale (GCS) Scoring System.

| GCS component | Response | GCS score |
|---|---|---|
| **Eye-opening (E)** | *No response* | 1 |
| | *To pain* | 2 |
| | *To verbal* | 3 |
| | *Spontaneous* | 4 |
| **Verbal response (V)** | *No response* | 1 |
| | *Incomprehensible sounds* | 2 |
| | *Inappropriate words* | 3 |
| | *Confused, disoriented* | 4 |
| | *Oriented, converses* | 5 |
| **Motor response (M)** | *No response* | 1 |
| | *Abnormal extension to pain* | 2 |
| | *Abnormal flexion to pain* | 3 |
| | *Withdraws to pain* | 4 |
| | *Localizes pain* | 5 |
| | *Obeys commands* | 6 |

**Supplementary Table 12:** Data distribution of created test set for SEER 5-year breast cancer survivability (BCS) and lung cancer survivability (LCS).

| | | | | | | |
|---|---|---|---|---|---|---|
| **Single attribute test** | **EOD 10 - positive lymph nodes examined continuously** | | | | | |
| | | | Lung cancer set | | Breast cancer set | |
| | *N stage* | Group (count) | # case per set | **# total case** | # case per set | **# total case** |
| | *N0* | 0 | 111 | 333 | 30 | 90 |
| | *N1* | (0, 4) | 333 | 999 | 90 | 270 |
| | *N2* | [4, 10) | 666 | 1,998 | 182 | 546 |
| | *N3* | >10 | 6,978 | 20,934 | 2,260 | 6,780 |
| | **Total** | | **8,088** | **24,264** | **2,562** | **7,686** |
| | **CS Tumor size continuous** | | | | | |
| | | | Lung cancer set | | Breast cancer set | |
| | *T stage* | Group (mm) | # case per set | **# total case** | # case per set | **# total case** |
| | *T0* | 0 | 4 | 12 | 6 | 18 |
| | *T1* | (0, 20] | 57 | 171 | 81 | 243 |
| | *T2* | (20-50] | 91 | 273 | 130 | 390 |
| | *T3* | (50, 988] | 2,637 | 7,911 | 4,080 | 12,240 |
| | **Total** | | **2,789** | **8,367** | **4,297** | **12,891** |
| | **Grades** | | | | | |
| | *Grade stage* | status | Seed Class | | Breast cancer set | |
| | | | | | # case per class | **# total case** |
| | *1* | Well differentiated | Survived (C1) | | 21,723 | 24,875 |
| | | | Death (C0) | | 3,152 | |
| | *2* | Moderately differentiated | Survived (C1) | | 21,723 | 24,875 |
| | | | Death (C0) | | 3,152 | |
| | *3* | Poorly differentiated | Survived (C1) | | 21,723 | 24,875 |
| | | | Death (C0) | | 3,152 | |
| | *4* | Undifferentiated | Survived (C1) | | 21,723 | 24,875 |
| | | | Death (C0) | | 3,152 | |
| | **Total** | | | | | **99,500** |
| **Double attribute test** | **Combinations** | | | | Breast cancer set | |
| | | | | | # case per set | **# total case** |
| | *T-N* | Tumor size and positive lymph node combination | | | 6,177 | 18,531 |
| | *T-ENLs* | Tumor size and number of examined lymph nodes | | | 6,177 | 18,531 |
| | *N-ENLs* | Number of examined lymph nodes and positive lymph node | | | 7,800 | 23,400 |
| | **Total** | | | | | **60,462** |
| **Triple attributes test** | **Three attributes** (Tumor size T4, Number of positive lymph nodes N3, Grade 4) | | | | Breast cancer set | |
| | | | | | # case per class | **# total case** |
| | Seed class survived (C1) | | | | 21,723 | 24,875 |
| | Seed class death (C0) | | | | 3,152 | |
| | **Total** | | | | | **24,875** |

**Supplementary Table 13:** Ranges of important vitals.

| Attribute | Ideal range | Critical risk zone | |
|---|---|---|---|
| | | Low | High |
| Diastolic BP (mm Hg) | 70-80 | 0-60 | 120-250 |
| Glucose (mg/dL) | 70-180 | 0-54 | 250-600 |
| Respiratory Rate (BPM) | 12-18 | 0-10 | 24-100 |
| Temperature (C) | 37 | 30-35 | 39-42 |
| Oxygen Saturation (%) | 95-100 | 0-90 | - |
| Systolic BP (mm Hg) | 100-120 | 0-90 | 180-350 |

**Supplementary Table 14:** Mean standard deviation of the attribute of the original MIMIC-III dataset.

| | All | | Class 0 (Non-death) | | Class 1 (Death) | |
|---|---|---|---|---|---|---|
| | Mean | SD (avg) | Mean | SD (avg) | Mean | SD (avg) |
| **Diastolic bp (mm Hg)** | 61.45 | 15.61 | 61.60 | 14.20 | 60.46 | 24.63 |
| **GCS eye-opening** | 3.22 | 0.59 | 3.34 | 0.58 | 2.47 | 0.63 |
| **GCS motor response** | 5.33 | 0.61 | 5.46 | 0.60 | 4.47 | 0.70 |
| **GCS total** | 11.85 | 1.65 | 12.28 | 1.67 | 9.24 | 1.51 |
| **GCS verbal response** | 3.32 | 0.67 | 3.48 | 0.71 | 2.30 | 0.46 |
| **Glucose (mg/dL)** | 141.54 | 48.34 | 140.72 | 49.67 | 146.81 | 39.78 |
| **Heart Rate (BPM)** | 86.67 | 10.08 | 86.14 | 9.91 | 90.03 | 11.13 |
| **Oxygen saturation (%)** | 97.98 | 11.50 | 98.23 | 12.61 | 96.44 | 4.39 |
| **Respiratory rate (BPM)** | 20.96 | 19.92 | 21.05 | 22.27 | 20.44 | 4.95 |
| **Systolic bp (mm Hg)** | 120.25 | 15.67 | 120.78 | 15.48 | 116.85 | 16.92 |
| **Temperature (C)** | 36.94 | 0.71 | 36.95 | 0.69 | 36.88 | 0.86 |

**Supplementary Table 15:** Selected seeds from the MIMIC-III dataset. These five seeds are used to create attribute-based test cases. Mean and SD represent the average and standard deviation values of the particular attribute. Missing values were ignored for calculating the mean and standard deviation.

| **Seeds** | **1** | | **2** | | **3** | | **4** | | **5** | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Attribute** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** |
| **Diastolic BP (mm Hg)** | 55.40 | 8.27 | 55.85 | 7.00 | 55.77 | 4.29 | 66.08 | 5.39 | 66.72 | 10.05 |
| **GCS eye opening** | 4.00 | 0.00 | 4.00 | 0.00 | 4.00 | 0.00 | 4.00 | 0.00 | 4.00 | 0.00 |
| **GCS motor response** | 6.00 | 0.00 | 6.00 | 0.00 | 6.00 | 0.00 | 6.00 | 0.00 | 6.00 | 0.00 |
| **GCS total** | 15.00 | 0.00 | 15.00 | 0.00 | 15.00 | 0.00 | 15.00 | 0.00 | 15.00 | 0.00 |
| **GCS verbal response** | 5.00 | 0.00 | 5.00 | 0.00 | 5.00 | 0.00 | 5.00 | 0.00 | 5.00 | 0.00 |
| **Glucose (mg/dL)** | 118.75 | 13.84 | 153.70 | 22.60 | 97.00 | 8.19 | 126.50 | 7.33 | 109.00 | 21.21 |
| **Heart Rate (BPM)** | 74.27 | 8.29 | 64.40 | 5.72 | 66.90 | 6.22 | 99.34 | 9.89 | 67.67 | 5.84 |
| **Mean BP (mm Hg)** | 73.15 | 5.69 | 84.67 | 7.32 | 72.64 | 4.63 | 84.88 | 4.87 | 84.92 | 10.53 |
| **Oxygen sat. (%)** | 96.71 | 1.68 | 96.81 | 1.13 | 98.51 | 2.01 | 99.06 | 0.91 | 95.10 | 1.79 |
| **Respiratory rate (BPM)** | 15.33 | 3.87 | 16.33 | 3.50 | 17.07 | 2.11 | 15.52 | 3.38 | 14.22 | 2.83 |
| **Systolic BP (mm Hg)** | 108.00 | 8.91 | 116.34 | 12.62 | 106.38 | 6.54 | 122.48 | 5.96 | 121.32 | 14.61 |
| **Temperature (C)** | 37.38 | 0.37 | 36.97 | 0.56 | 37.41 | 0.27 | 37.21 | 0.27 | 37.52 | 0.50 |

**Supplementary Table 16:** Average accuracy of machine learning models under various testing conditions for in-hospital mortality prediction.

| | | Accuracy | | |
|---|---|---|---|---|
| | **Attributes** | **LSTM** | **CW-LSTM** | **LR** |
| **Single attribute critical zone tests** | *Respiratory rate* | 64.67% | 13.00% | 59.00% |
| | *Temperature* | 58.97% | 23.08% | 46.15% |
| | *Diastolic BP* | 23.60% | 23.60% | 23.60% |
| | *Oxygen saturation* | 10.89% | 10.89% | 38.61% |
| | *Systolic BP* | 35.24% | 30.00% | 25.43% |
| | *Glucose* | 33.62% | 33.62% | 33.62% |
| **Average accuracy of single-attribute tests** | | **37.83**% | **22.36**% | **37.74**% |
| **GCS** | | 33.33% | 33.33% | 100% |
| **Multiple attribute critical zone tests** | **Zones** | | | |
| | *High critical zone* | 22.50% | 40.10% | 0.15% |
| | *Low critical zone* | 68.80% | 98.40% | 12.30% |
| **Average accuracy of multi-attribute tests** | | **45.65**% | **69.25**% | **6.23**% |
| **Deteriorating condition tests** | Any critical zones | 83.33% | 66.66% | 33.33% |

**Supplementary Table 17:** Changes in neural zone activation values of the LSTM model. Values in columns 2 and 3 represent the average difference of a neuron's activation score between two different regions (critically low zone, normal zone, and critically high zone).

| **Test Attribute** | $\Delta$□□□**(Low zone, normal zone)** | $\Delta$□□□**(normal zone, high zone)** |
|---|---|---|
| **Temperature** | 0.16 | 0.01 |
| **Glucose** | 0.01 | 0.04 |
| **Diastolic blood pressure** | 0.02 | 0.01 |
| **Respiratory rate** | 0.01 | 0.14 |

**Supplementary Table 18**: Average accuracy of MLP model under different test scenarios for 5-year breast cancer survivability prediction.

| | | | | |
|---|---|---|---|---|
| **Single attribute test** | **EOD 10 - positive lymph nodes examined** | | | |
| | *N stage* | **Group (count)** | | **Accuracy** |
| | *N0* | 0 | | 100% |
| | *N1* | (0, 4) | | 0% |
| | *N2* | [4, 10) | | 0% |
| | *N3* | >10 | | 74.4% |
| | **Average** | | | **43.6%** |
| | **CS Tumor size** | | | |
| | *T stage* | Group (mm) | | **Accuracy** |
| | *T0* | 0 | | 100% |
| | *T1* | (0, 20] | | 0% |
| | *T2* | (20-50] | | 0% |
| | *T3* | (50, 988] | | 0% |
| | **Average** | | | **25%** |
| | **Grades** | | | |
| | *Grade stage* | status | Seed Class | **Accuracy** |
| | *1* | Well-differentiated | Survived (C1) | 96.97% |
| | | | Death (C0) | 48.67% |
| | *2* | Moderately differentiated | Survived (C1) | 4.55% |
| | | | Death (C0) | 57.93% |
| | *3* | Poorly differentiated | Survived (C1) | 7.44% |
| | | | Death (C0) | 65.77% |
| | *4* | Undifferentiated | Survived (C1) | 6.66% |
| | | | Death (C0) | 63.86% |
| | **Average** | | | **43.98%** |
| **Double attribute test** | **Double Combinations** | | | **Accuracy** |
| | T-N: Tumor size and positive lymph node combination | | | 92.97% |
| | T-ENL: Tumor size and number of examined lymph nodes | | | 0% |
| | N-ENL: Number of examined lymph nodes and positive lymph node | | | 19.6% |
| | **Average** | | | **37.52%** |
| **Triple attributes test** | **Three attributes** (Tumor size T4, Number of positive lymph nodes N3, Grade 4) | | | **Accuracy** |
| | Seed class survived (C1) | | | 89.89% |
| | Seed class death (C0) | | | 98.92% |
| | **Average** | | | **94.41%** |

**Supplementary Table 19:** Wasserstein distances between various training and testing datasets.

| Training data | Testing data | Wasserstein distances (average) |
|---|---|---|
| *Original MIMIC-III train set* | *Original MIMIC-III test set* | 12.42 |
| *Original MIMIC-III train set* | *Synthesized MIMIC-III multi-attribute test set* | 33.38 |
| *Original SEER BCS train set* | *Original SEER BCS test set* | 2.1 |
| *Original SEER BCS train set* | *Synthesized SEER BCS triple-attribute test set* | 9.75 |