

Additional file

Interpretable Machine Learning for Sepsis Risk Prediction in Elderly ICU Patients With Heart Failure: Integrating Genetic Evidence and Clinical Data

Table of Contents

- 1. Supplemental Figure S1. Page 2.**
- 2. Supplemental Table S1. Page 3-5.**
- 3. Supplemental Table S2. Page 6.**
- 4. Supplemental Table S3. Page 7.**

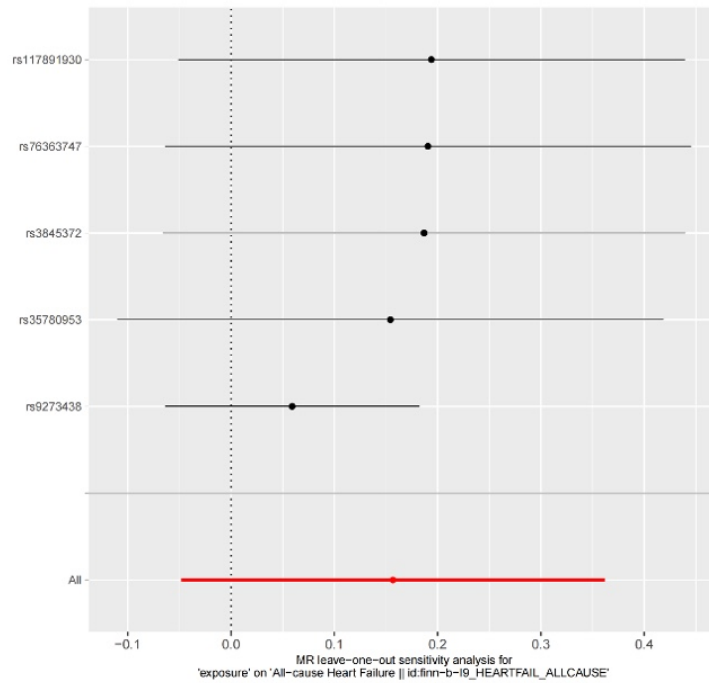
A

Fig S1. Leave-one-out sensitivity analysis of reverse Mendelian randomization (sepsis → all-cause heart failure)

Legend: Leave-one-out sensitivity analysis for reverse Mendelian randomization (MR) analysis (exposure: sepsis; outcome: all-cause heart failure [HF, FinnGen ID: finn-b-I9_HEARTFAIL_ALLCAUSE]) based on FinnGen GWAS data. The x-axis represents the causal effect estimate of each MR analysis. Each dot denotes the inverse variance weighted (IVW) estimate after excluding one individual single nucleotide polymorphism (SNP, labeled with SNP IDs), while the "All" label represents the IVW estimate using all SNPs. This analysis was conducted to evaluate whether individual SNPs unduly influence the overall causal effect estimate, thereby verifying the robustness of reverse MR results.

Table S1. All Candidate Predictor Variables for Sepsis Prediction in Elderly ICU Patients with Heart Failure

Variable	Overall (n = 5225)	Training cohort (n = 3658)	Internal testing cohort (n = 1567)	P value
General information				
Gender	3046 (58.3%)	2118 (57.9%)	928 (59.2%)	0.392
Age(years)	73.00 (64.00–82.00)	73.00 (64.00–82.00)	73.00 (63.00–82.00)	0.363
Weight(kg)	80.40 (67.10–97.50)	80.50 (67.00–97.50)	80.00 (67.55–97.30)	0.882
Comorbidities				
Hemiplegia	283 (5.4%)	190 (5.2%)	93 (5.9%)	0.309
Cerebrovascular disease	555 (10.6%)	394 (10.8%)	161 (10.3%)	0.628
Myocardial infarction	1066 (20.4%)	742 (20.3%)	324 (20.7%)	0.776
Hypertension	1045 (20.0%)	732 (20.0%)	313 (20.0%)	1
Atrial fibrillation	1814 (34.7%)	1284 (35.1%)	530 (33.8%)	0.391
Cancer	847 (16.2%)	589 (16.1%)	258 (16.5%)	0.776
Liver cirrhosis	424 (8.1%)	302 (8.3%)	122 (7.8%)	0.606
Chronic obstructive pulmonary disease	1255 (24.0%)	881 (24.1%)	374 (23.9%)	0.894
Chronic kidney disease	1860 (35.6%)	1280 (35.0%)	580 (37.0%)	0.172
Diabetes mellitus	2161 (41.4%)	1527 (41.7%)	634 (40.5%)	0.405
Antimicrobial use				
Glycopeptide antibiotics	3917 (75.0%)	2732 (74.7%)	1185 (75.6%)	0.496
Macrolide antibiotics	819 (15.7%)	569 (15.6%)	250 (16.0%)	0.747
Beta-lactam antibiotics	1704 (32.6%)	1208 (33.0%)	496 (31.7%)	0.349
Quinolone antibiotics	1550 (29.7%)	1060 (29.0%)	490 (31.3%)	0.103
Treatment				
Insulin therapy	3393 (64.9%)	2367 (64.7%)	1026 (65.5%)	0.616
Beta-blockers	3827 (73.2%)	2693 (73.6%)	1134 (72.4%)	0.367
ACE inhibitors	1915 (36.7%)	1349 (36.9%)	566 (36.1%)	0.624
Albumin administration	1580 (30.2%)	1107 (30.3%)	473 (30.2%)	0.982
Sedative agents	3553 (68.0%)	2492 (68.1%)	1061 (67.7%)	0.793
Vasopressors	3694 (70.7%)	2595 (70.9%)	1099 (70.1%)	0.58
Furosemide	1080 (20.7%)	749 (20.5%)	331 (21.1%)	0.622
Mechanical ventilation	4623 (88.5%)	3232 (88.4%)	1391 (88.8%)	0.702
CRRT	613 (11.7%)	440 (12.0%)	173 (11.0%)	0.332
ICU physiological score				
SOFA score	6.00 (4.00–9.00)	6.00 (4.00–9.00)	6.00 (4.00–9.00)	0.765

Table S1 (Continued)				
Variable	Overall (n = 5225)	Training cohort (n = 3658)	Internal testing cohort (n = 1567)	P value
SAPS II score	42.00 (34.00–52.00)	42.00 (34.00–52.00)	42.00 (34.00–52.00)	0.487
Laboratory Indicators				
Red blood cell count ($\times 10^{12}/L$)	3.50 (2.90–4.10)	3.50 (2.90–4.10)	3.50 (2.90–4.10)	0.959
White blood cell count ($\times 10^9 /L$)	11.80 (8.40–16.80)	11.90 (8.40–16.90)	11.80 (8.30–16.70)	0.564
Platelet count ($\times 10^9 /L$)	190.00 (133.00–261.00)	191.00 (134.00–263.00)	188.00 (133.00–253.50)	0.101
Calcium(mg/dL)	8.40 (7.90–8.90)	8.40 (7.90–8.90)	8.40 (7.90–8.80)	0.495
Chloride(mmol/L)	102.00 (98.00–107.00)	102.00 (98.00–106.00)	102.00 (98.00–107.00)	0.59
Potassium(mmol/L)	4.30 (3.80–4.80)	4.30 (3.80–4.80)	4.30 (3.80–4.80)	0.9
Sodium(mmol/L)	138.00 (135.00–141.00)	138.00 (135.00–141.00)	138.00 (135.00–141.00)	0.788
Magnesium(mg/dL)	2.00 (1.80–2.30)	2.00 (1.80–2.30)	2.00 (1.80–2.30)	0.209
Glucose(mmol/L)	7.60 (6.10–9.90)	7.60 (6.00–10.10)	7.60 (6.10–9.80)	0.801
Blood urea nitrogen(mg/dL)	29.00 (19.00–48.00)	29.00 (19.00–48.00)	29.00 (19.00–48.00)	0.525
Creatinine(mg/dL)	1.30 (0.90–2.20)	1.30 (0.90–2.20)	1.30 (0.90–2.20)	0.546
Albumin(g/dL)	3.10 (2.60–3.40)	3.10 (2.60–3.40)	3.00 (2.60–3.50)	0.8
Alanine aminotransferase (U/L)	27.00 (15.00–63.00)	27.00 (15.00–62.00)	27.00 (16.00–65.00)	0.472
Aspartate aminotransferase (U/L)	42.00 (24.00–99.00)	42.00 (24.00–100.00)	42.00 (24.00–98.00)	0.897
Alkaline phosphatase(U/L)	87.00 (64.00–127.00)	88.00 (65.00–128.00)	87.00 (63.00–125.00)	0.245
Total bilirubin(mg/dL)	0.70 (0.40–1.30)	0.70 (0.40–1.30)	0.70 (0.40–1.30)	0.285
PaCO ₂ (mmHg)	42.00 (36.00–49.00)	41.00 (36.00–49.00)	42.00 (36.00–49.00)	0.065
PaO ₂ (mmHg)	77.00 (44.00–154.00)	79.00 (45.00–157.00)	74.00 (44.00–151.00)	0.159
pH	7.40 (7.30–7.40)	7.40 (7.30–7.40)	7.40 (7.30–7.40)	0.098
Prothrombin time(s)	15.20 (13.20–19.20)	15.30 (13.20–19.40)	15.10 (13.10–18.80)	0.353
International normalized ratio	1.40 (1.20–1.80)	1.40 (1.20–1.80)	1.40 (1.20–1.70)	0.35
Lactate(mmol/L)	1.80 (1.20–2.80)	1.80 (1.30–2.80)	1.70 (1.20–2.70)	0.329

Table S1: Continuous variables are presented as Mean $\pm$ SD, and

categorical variables as n (%). For comparisons between the training and internal testing cohorts: ① Continuous variables were analyzed using Welch t-test (without assuming equal variances); ② Categorical variables were analyzed using Pearson χ^2 test if all cell counts ≥ 5 , otherwise Fisher's exact test was applied. Statistical significance was set at $P < 0.05$, and $P < 0.001$ was uniformly expressed as such. For variables that were untestable (e.g., variables with all values = 0), the corresponding P-value field was left blank (consistent with CONSORT/STROBE guidelines). n represents the number of patients in each cohort. ICU: Intensive Care Unit; HF: Heart Failure.

Table S2. Performance metrics of all machine learning models in external validation (MIMIC-III)

Model	AUC	Threshold	Accuracy	Sensitivity	Specificity	Precision	F1	PPV	NPV	Kppa
Logistic	0.732	0.68	0.673	0.692	0.664	0.492	0.575	0.488	0.822	0.27
SVM	0.742	0.747	0.699	0.666	0.715	0.523	0.586	0.523	0.827	0.31
GBM	0.684	0.746	0.635	0.638	0.633	0.45	0.528	0.462	0.789	0.24
NeuralNetwork	0.703	0.748	0.685	0.547	0.75	0.507	0.526	0.499	0.778	0.19
KNN	0.606	0.5	0.553	0.751	0.461	0.395	0.518	0.348	0.565	0.01
Xgboost	0.712	0.501	0.632	0.74	0.581	0.454	0.562	0.437	0.794	0.22
Adaboost	0.636	0.5	0.53	0.859	0.375	0.393	0.539	0.392	0.85	0.13
LightGBM	0.667	0.831	0.673	0.517	0.746	0.489	0.503	0.471	0.763	0.17
CatBoost	0.715	0.662	0.658	0.681	0.648	0.476	0.56	0.481	0.812	0.26

Table S2: AUC: Area Under the Receiver Operating Characteristic Curve; PPV: Positive Predictive Value; NPV: Negative Predictive Value; Kappa: Cohen's Kappa coefficient. All performance metrics were calculated based on the external validation cohort from MIMIC-III database. The threshold for each model was determined by maximizing the Youden index (Sensitivity + Specificity - 1).

Table S3

Model Name	Tuning Parameters and Corresponding Values
glm	glm.tune.grid = NULL
svm	sigma = 0.001, C = 0.09
gbm	n.trees = 110, interaction.depth = 5, shrinkage = 0.1, n.minobsinnode = 30
nnet	size = 4, decay = 0.4
xgb	nrounds = 10, max_depth = 4, eta = 0.001, gamma = 10, colsample_bytree = 0.5, min_child_weight = 1, subsample = 0.6
knn	kmax = 12, distance = 1, kernel = "optimal"
ada	mfinal = 2, maxdepth = 2, coeflearn = "Zhu"
lightgbm	objective = "binary", metric = "auc", min_data = 1L, learning_rate = 1.0, num_threads = 2L, force_col_wise = T
catboost	iterations = 100, use_best_model = TRUE, eval_metric = 'AUC', ignored_features = c(4, 9), border_count = 32, depth = 4, learning_rate = 0.03, random_seed = 123