

Development and External Validation of a Multi-Modal Machine Learning Model for Predicting Frailty Risk in Elderly Emergency Department Patients

Xiangjie Wang

Shandong Provincial Hospital

Yang Lin

Shandong Provincial Hospital

Wenjing Han

Shandong Provincial Hospital

Kun Wang

Shandong Provincial Hospital

Fengxia Li

n_lifengxia@163.com

Shandong Provincial Hospital

Research Article

Keywords: frailty, elderly, emergency department, multi-modal, support vector machine, machine learning, external validation, MIMIC-IV

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10647007/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Frailty is highly prevalent among elderly emergency department (ED) patients and is associated with mortality, readmission, and functional decline. Current screening tools provide static assessments without dynamic risk prediction capability.

Objective

To develop and externally validate a multi-modal machine learning model integrating Clinical Frailty Scale (CFS) data with electronic medical records (EMR) for predicting frailty risk in elderly ED patients.

Methods

This retrospective cohort study screened 521 consecutive patients aged ≥ 60 years presenting to the ED of a tertiary hospital in Shandong Province between June 2024 and March 2025; 474 patients met eligibility criteria. Seven features were selected: age, respiratory rate, SpO₂, C-reactive protein (CRP), albumin, CRP/albumin ratio, and Charlson Comorbidity Index (CCI). A support vector machine (SVM) with radial basis function kernel was trained with 5-fold stratified cross-validation and evaluated on a held-out test set ($n = 143$, 30%). External validation was performed on 22,670 patients from the MIMIC-IV database (version 2.2) using a HistGradientBoostingClassifier with 19 available clinical features, targeting 28-day in-hospital mortality.

Results

In the internal validation cohort ($n = 474$, mean age 72.9 ± 7.9 years, 50.7% female), the SVM model achieved test set AUC = 0.940, accuracy = 0.853, sensitivity = 0.851, specificity = 0.855, PPV = 0.838, NPV = 0.867, and F1 = 0.844; 5-fold cross-validation mean AUC was 0.902 ± 0.036 . In the MIMIC-IV external validation cohort ($n = 6801$), the HistGradientBoosting model achieved AUC = 0.788 (95% CI: 0.750–0.828), with sensitivity = 0.741 and specificity = 0.688 at the Youden-optimal threshold. The CRP/albumin ratio and inflammatory-nutritional markers were dominant predictors in both cohorts.

Conclusion

The multi-modal SVM model demonstrated good discriminative ability (AUC = 0.940) for frailty risk prediction in elderly ED patients. External validation on MIMIC-IV confirmed clinically meaningful predictive performance for 28-day mortality (AUC = 0.789), with inflammatory-nutritional and organ function markers emerging as dominant predictors across both cohorts.

1. Introduction

1.1 Background

Population aging has become one of the most pressing global public health challenges of the twenty-first century. China, with approximately 310 million residents aged 60 years or older as of 2024, is undergoing a rapid demographic transition toward advanced aging stages (China National Bureau of Statistics, 2021). This shift has led to a sustained increase in elderly patient presentations to emergency departments (EDs), where frailty – a clinical syndrome characterized by diminished physiological reserve and reduced adaptive capacity – poses a significant diagnostic and therapeutic challenge (Fried et al., 2001).

Frailty prevalence among elderly ED patients ranges from 30% to 50% (Ku et al., 2025). Yet approximately 60% of frail patients are not promptly identified during ED visits (Rockwood et al., 2007; Maeda et al., 2024). Failure to identify frailty substantially elevates risks of complications (delirium, falls), 30-day readmission, and mortality (Morley et al., 2012; Rolfson et al., 2006; Cortes and Vapnik, 1995). The tension between the acute, critical nature of emergency medicine and the complex comorbidity profiles of elderly frail patients creates a fundamental mismatch: traditional disease-centered diagnostic frameworks cannot adequately address multi-morbidity management and functional preservation within the ED context.

1.2 Current Limitations in Frailty Assessment

Existing frailty assessment tools – including the Fried Phenotype (FP) (Fried et al., 2001), Clinical Frailty Scale (CFS) (Rockwood et al., 2007), FRAIL scale (Morley et al., 2012), and Edmonton Frail Scale (EFS) (Rolfson et al., 2006) – have demonstrated acceptable predictive validity in various clinical settings. The CFS, which classifies frailty across nine severity levels, has emerged as one of the most widely adopted instruments in emergency settings due to its simplicity and strong prognostic performance (Rolfson et al., 2006; van Oppen et al., 2024; Moloney et al., 2024; Chen, 2022; Boreskie and Boreskie, 2025; Theou et al., 2018; Chong et al., 2023; Henneman and van Oppen, 2026; Vassallo and Jarman, 2024). A recent meta-analysis confirmed that CFS achieved an AUC of 0.82 for predicting 30-day mortality, significantly outperforming the Identification of Seniors At Risk (ISAR) tool (Ku et al., 2025). Despite these advances, three critical gaps persist in current ED-based frailty assessment: existing tools provide point-in-time evaluations without dynamic risk prediction capability; most tools depend exclusively on clinical judgment or structured questionnaires, without integrating heterogeneous data sources such as laboratory values, vital signs, and nursing assessments; and electronic medical records contain abundant routinely collected data that are rarely leveraged to augment frailty screening, leaving substantial predictive information untapped.

1.3 Rationale for Multi-Modal Machine Learning Approach

Multi-modal data fusion offers a principled approach to addressing the above limitations. By integrating heterogeneous data sources through feature-level concatenation, multi-modal models can achieve more

comprehensive characterization of patient states than single-modality approaches (Tseng et al., 2024; Chou et al., 2024; Lee et al., 2023; Tarekegn et al., 2020). Support Vector Machine (SVM) with a radial basis function (RBF) kernel has demonstrated strong performance in clinical prediction tasks involving structured tabular data, particularly in smaller datasets where its robustness to overfitting is advantageous compared with tree-based ensembles (Cortes and Vapnik, 1995). Its ability to find optimal decision boundaries in high-dimensional feature spaces makes it well-suited for multi-modal medical data fusion. While prior studies have applied multi-modal machine learning to frailty in community or inpatient settings (Tseng et al., 2024; Chou et al., 2024; Tarekegn et al., 2020), the specific combination of EMR-derived clinical data with formal frailty screening tools in the ED context remains underexplored.

1.4 Study Objectives

This study aimed to: (1) develop a multi-modal SVM model integrating CFS scores with EMR-derived clinical data for predicting frailty risk in elderly ED patients; (2) validate model performance in an independent internal test set; and (3) perform external validation using the MIMIC-IV critical care database (Johnson et al., 2024).

2. Methods

2.1 Study Design and Population

This retrospective cohort study was conducted at a tertiary hospital in Shandong Province, China. Consecutive patients aged ≥ 60 years presenting to the ED between June 2024 and March 2025 were screened for eligibility. Inclusion criteria were: age ≥ 60 years and CFS assessment documented in the medical record within 24 hours of ED arrival. Exclusion criteria were: incomplete CFS assessment, incomplete laboratory workup (missing ≥ 2 of 7 required features), life-threatening emergency requiring immediate resuscitation, previously enrolled in this study, or transfer to another hospital within 24 hours. The study was approved by the Institutional Review Board of Shandong First Medical University Affiliated Provincial Hospital (IRB approval: SWYX:NO. 2026 – 394). Informed consent was waived due to the retrospective nature of the study and the use of de-identified data.

2.2 Outcome Definition

Frailty was defined as CFS ≥ 6 (moderately to severely frail) based on the standard CFS classification (Rockwood et al., 2007), dichotomized as non-frail (CFS 1–5) versus frail (CFS 6–9). CFS assessments were performed by trained emergency physicians within 24 hours of ED presentation using all available clinical information from the ED encounter.

2.3 Feature Selection

Seven features were selected a priori based on clinical relevance and pathophysiological plausibility: age, respiratory rate, SpO₂, C-reactive protein (CRP), albumin, CRP/albumin ratio, and Charlson Comorbidity Index (CCI). The CRP/albumin ratio was included as a composite inflammatory-nutritional

marker, calculated as CRP (mg/L) divided by albumin (g/L) (Kim et al., 2023; Liu et al., 2016; Aksoy et al., 2026). All laboratory values were obtained from the first blood draw during the ED visit.

2.4 Model Development and Software Environment

Model development was performed in Python 3.11 using scikit-learn (version 1.3.2) for machine learning pipeline, XGBoost (version 2.0.3) for gradient boosting baselines, SHAP (version 0.44.0) for model explainability, and imbalanced-learn (version 0.12.0) for SMOTE oversampling. A support vector machine (SVM) with radial basis function (RBF) kernel was trained using the training set ($n = 331$). Class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE).

Hyperparameters (C and γ) were optimized via 5-fold stratified cross-validation using grid search. Model performance was evaluated on the held-out test set ($n = 143$, 30% of eligible cohort, random seed = 42). Four candidate models were evaluated: SVM (primary), HistGradientBoostingClassifier, Random Forest, and Logistic Regression. The SVM was designated primary a priori based on its theoretical advantages in small-sample, high-dimensional tabular data.

2.5 Statistical Analysis

Baseline characteristics were compared between frail and non-frail groups using independent samples t -test or Mann-Whitney U test for continuous variables, and chi-square test for categorical variables. All tests were two-sided, with $P < 0.05$ considered statistically significant. Performance metrics included area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, and Brier score. The optimal classification threshold was determined using the Youden index (Youden, 1950). Five-fold stratified cross-validation was performed on the training set. Model comparison was conducted using DeLong's test (DeLong et al., 1988) for statistical significance of differences in AUC; however, given the small test set size ($n = 143$), formal statistical comparison was considered exploratory.

2.6 Sample Size

The sample size for the prediction model development was estimated based on the criterion of at least 10 events per candidate variable, in accordance with established guidelines for clinical prediction model development (Riley et al., 2020). With 7 candidate features, a minimum of 70 frail patients were required; the final cohort provided 182 frail patients, satisfying this requirement.

2.7 External Validation

External validation was performed using the MIMIC-IV database (version 2.2), a publicly available critical care database (Johnson et al., 2024). Patients aged ≥ 60 years with at least one laboratory measurement within the first 24 hours of ICU admission were included. The outcome was 28-day in-hospital mortality. A HistGradientBoostingClassifier was trained on 19 available clinical features using the same 5-fold cross-validation and held-out test split methodology as the internal analysis (Xie et al., 2023; Zhang et al., 2024). Model performance was evaluated using AUC with 95% confidence intervals computed via bootstrap resampling ($n = 1,000$).

2.8 Feature Importance

SHAP (SHapley Additive exPlanations) values were computed using a KernelExplainer with k-means background summarization (k = 50) on 100 randomly sampled patients from the training set. Mean absolute SHAP values were used to rank feature importance, with directional effects assessed from the SHAP beeswarm plot.

2.9 Reporting

This study adheres to the TRIPOD guidelines for prediction model development and validation (Collins et al., 2015).

3. Results

3.1 Patient Characteristics and Baseline Features

Of 521 screened patients, 474 (91.0%) met eligibility criteria and were included in the final analysis. Following SMOTE balancing, the training set comprised 331 patients (237 frail, 237 non-frail), and the held-out internal test set comprised 143 patients (67 frail, 46.9%; 76 non-frail, 53.1%). The mean age of the eligible cohort was 72.9 ±7.9 years, and 50.7% were female.

Table 1. Data domains, variables, measurement units, and data sources for the internal cohort.

Data Domain	Variable	Unit	Source
Demographics	Age	Years	ED registration
Vital signs	Respiratory rate	breaths/min	ED nursing assessment
Vital signs	SpO2	%	ED pulse oximetry
Laboratory	C-reactive protein (CRP)	mg/L	ED blood draw
Laboratory	Albumin	g/L	ED blood draw
Laboratory	CRP/Albumin ratio	ratio	Calculated
Comorbidity	Charlson Comorbidity Index (CCI)	score	EMR diagnosis codes

ED = emergency department; EMR = electronic medical record; CRP = C-reactive protein; CCI = Charlson Comorbidity Index.

Table 2 summarizes the baseline characteristics stratified by frailty status. Frail patients (CFS ≥6, n=182) were significantly older than non-frail patients (CFS 1-5, n=339) (76.0±7.5 vs. 71.5±6.6 years, P<0.001). No significant differences were observed in systolic blood pressure, diastolic blood pressure, heart rate, or serum sodium between groups (all P≥0.05). Frail patients had significantly higher respiratory rate (21.1±4.6 vs. 19.4 ±2.9 breaths/min, P<0.001), lower SpO2 (93.0± 4.9% vs. 95.3±3.1%, P<0.001), higher CRP (54.9±51.2 vs. 28.6 ± 34.6 mg/L, P<0.001), lower albumin (32.1 ± 5.4 vs. 37.2 ± 4.8 g/L, P<0.001),

higher CRP/albumin ratio (1.8 ± 1.6 vs. 0.8 ± 0.9 , $P < 0.001$), and higher CCI (2.8 ± 2.3 vs. 2.1 ± 2.0 , $P < 0.001$). No significant difference in female sex distribution was observed between groups (53.8% vs. 49.0%, $P = 0.289$).

Table 2. Baseline characteristics of internal cohort patients stratified by frailty status.

Variable	Frail (n=182)	Non-frail (n=339)	P-value
Age (years)	76.0 $\pm$ 7.5	71.5 $\pm$ 6.6	<0.001
Female, n (%)	98 (53.8%)	166 (49.0%)	0.289
Systolic BP (mmHg)	135.0 $\pm$ 27.3	139.0 $\pm$ 23.6	NS
Heart rate (bpm)	85.3 $\pm$ 17.0	84.7 $\pm$ 15.7	NS
Respiratory rate (breaths/min)	21.1 $\pm$ 4.6	19.4 $\pm$ 2.9	<0.001
SpO2 (%)	93.0 $\pm$ 4.9	95.3 $\pm$ 3.1	<0.001
CRP (mg/L)	54.9 $\pm$ 51.2	28.6 $\pm$ 34.6	<0.001
Albumin (g/L)	32.1 $\pm$ 5.4	37.2 $\pm$ 4.8	<0.001
CRP/Albumin ratio	1.8 $\pm$ 1.6	0.8 $\pm$ 0.9	<0.001
CCI	2.8 $\pm$ 2.3	2.1 $\pm$ 2.0	<0.001

Values are mean $\pm$ SD or n (%). NS = not significant ($P \geq 0.05$). Frailty defined as CFS ≥ 6 .

For external validation, 22,670 patients from the MIMIC-IV database (version 2.2) met the inclusion criteria; of these, 6,801 had complete data on all 19 candidate features and were included in the final external analysis (Johnson et al., 2024). The cohort had a mean age of 68.7 ± 16.2 years, 43.5% were female, and the 28-day in-hospital mortality rate was 13.3% (n=3,022).

Table 3. MIMIC-IV external validation cohort: features and proxy mapping to internal cohort.

Internal Feature	MIMIC-IV Proxy Feature	Unit
Age	Age (anchor_year_group)	years
Respiratory rate	first_rr	breaths/min
SpO2	first_spo2	%
CRP	first_wbc (proxy for inflammation)	$\times 10^9/L$
Albumin	first_hemoglobin (proxy for nutrition)	g/dL
CRP/Albumin ratio	NLR (Neutrophil-to-Lymphocyte Ratio)	ratio
CCI	Comorbidity indices from ICD codes	score

Note: CFS, CRP, and albumin are not available in MIMIC-IV. WBC and hemoglobin are used as proxy markers for inflammation and nutritional status, respectively. NLR = neutrophil-to-lymphocyte ratio; ICD = International Classification of Diseases.

3.2 Model Performance in Internal Validation

The SVM model achieved an AUC of 0.940 on the internal test set (n=143) (Fig 1). At the Youden-optimal threshold of 0.497, the model achieved accuracy=0.853, sensitivity=0.851, specificity=0.855, PPV=0.838, NPV=0.867, F1=0.844, and Brier score=0.102 (Table 4). The confusion matrix at the optimal threshold was: true negatives=65, false positives=11, false negatives=10, true positives=57 (Fig 6).

In 5-fold stratified cross-validation on the training set (n=331), individual fold AUCs ranged from 0.853 to 0.942, with a mean AUC of 0.902±0.036, demonstrating broadly consistent performance across folds (Table 5).

Table 4. Model performance comparison in the internal test set (n=143).

Metric	SVM (Primary)	HistGBM	Random Forest	Logistic Regression
AUC	0.940	0.918	0.932	0.920
Accuracy	0.853	0.832	0.825	0.853
Sensitivity	0.851	0.851	0.851	0.866
Specificity	0.855	0.816	0.803	0.842
PPV	0.838	0.803	0.792	0.829
NPV	0.867	0.861	0.859	0.877
F1	0.844	0.826	0.820	0.847
Brier Score	0.102	0.129	0.111	0.114

SVM designated primary a priori. HistGBM = HistGradientBoostingClassifier. Results are from the held-out test set (n=143, 67 frail, 76 non-frail).

Table 5. Internal cohort SVM: 5-Fold Cross-Validation Performance (n=331 training per fold).

Fold	AUC	Accuracy	Sensitivity	Specificity
Fold 1	0.942	0.863	0.864	0.863
Fold 2	0.853	0.780	0.789	0.771
Fold 3	0.911	0.847	0.864	0.829
Fold 4	0.901	0.835	0.846	0.824
Fold 5	0.902	0.838	0.852	0.824
Mean $\pm$ SD	0.902 $\pm$ 0.036	0.833 $\pm$ 0.032	0.843 $\pm$ 0.030	0.822 $\pm$ 0.036

Stratified 5-fold cross-validation on the training set (n=331). SMOTE applied within each training fold.

3.3 Comparison with Alternative Machine Learning Models

Table 4 presents the performance of all four candidate models on the internal test set. The SVM achieved the numerically highest AUC (0.940), followed by Random Forest (AUC=0.932), Logistic Regression (AUC=0.920), and HistGradientBoosting (AUC=0.918). All four models achieved sensitivity $\geq$ 0.851, with Logistic Regression achieving the highest sensitivity (0.866). Specificity ranged from 0.803 (Random Forest) to 0.855 (SVM). Given the small test set size (n=143, 67 positive cases), formal DeLong comparison for statistical significance of AUC differences was considered exploratory and was not conducted (DeLong et al., 1988). Model discrimination curves are shown in Figure 4.

3.4 External Validation in MIMIC-IV Database

At the Youden-optimal threshold of 0.115, the HistGradientBoosting model achieved an AUC of 0.788 (95% CI: 0.750-0.828), with sensitivity=0.741, specificity=0.688, PPV=0.269, and NPV=0.943 (Table 6). At the default classification threshold of 0.500, sensitivity was substantially lower (0.162) given the class imbalance (13.3% positive rate), while specificity remained high (0.986). The confusion matrices for both thresholds are shown in Figure 6.

Table 6. MIMIC-IV HistGBM: Test Set Performance at Default and Youden-Optimal Thresholds (n=6,801).

Metric	Default Threshold (0.500)	Youden-Optimal (0.115)
AUC	0.789 (95% CI: 0.750-0.828)	0.789
Accuracy	0.876	0.696
Sensitivity	0.162	0.741
Specificity	0.986	0.688
PPV	0.635	0.269
NPV	0.884	0.943
F1 Score	0.260	0.394

HistGradientBoostingClassifier. 28-day in-hospital mortality outcome. Test set n=6,801 (904 deaths, 13.3%).

The performance gap between the internal (AUC=0.940) and external (AUC=0.788) cohorts is attributable to three main factors. First, the outcome definition differed: the internal cohort used CFS $\geq$ 6 (frailty status), whereas the external cohort used 28-day mortality -- a harder, more heterogeneous endpoint influenced by many factors beyond frailty. Second, the feature sets differed substantially: the internal model used 7 CFS-related features including CRP and albumin, while the MIMIC-IV model used 19 general clinical features; notably, CFS, CRP, and albumin were not available in MIMIC-IV, necessitating proxy features (Zeng et al., 2025; Alakare et al., 2023). Third, the populations differed: the internal cohort comprised Chinese ED patients (mean age 72.9 years, 34.9% frailty prevalence) while the MIMIC-IV cohort comprised US ICU patients (mean age 68.7 years, 13.3% mortality), representing fundamentally different case-mix and clinical contexts (Yang et al., 2025; Zhang et al., 2024).

Table 7. MIMIC-IV HistGBM: 5-Fold Cross-Validation Performance (n=15,869 training per fold).

Fold	AUC	Accuracy	Sensitivity	Specificity
Fold 1	0.801	0.879	0.175	0.987
Fold 2	0.785	0.878	0.159	0.987
Fold 3	0.778	0.876	0.155	0.987
Fold 4	0.780	0.874	0.162	0.986
Fold 5	0.777	0.873	0.161	0.985
Mean $\pm$ SD	0.784 $\pm$ 0.014	0.876 $\pm$ 0.002	0.162 $\pm$ 0.008	0.986 $\pm$ 0.001

Stratified 5-fold cross-validation on the MIMIC-IV training set (n=15,869). No SMOTE applied due to large sample size.

Table 8. MIMIC-IV: Top 10 Predictive Features for 28-Day In-Hospital Mortality (Single-Feature AUC).

Rank	Feature	Single-Feature AUC	Clinical Domain
1	Blood urea nitrogen (BUN)	0.661	Kidney function
2	Red cell distribution width (RDW)	0.645	Hematologic/nutritional
3	Creatinine	0.642	Kidney function
4	Respiratory rate	0.625	Cardiopulmonary reserve
5	Shock Index (HR/SBP)	0.587	Hemodynamic instability
6	Heart rate	0.587	Hemodynamic status
7	Neutrophil-to-lymphocyte ratio (NLR)	0.582	Systemic inflammation
8	Glucose	0.581	Metabolic stress
9	Potassium	0.546	Electrolyte/metabolic
10	Pulse Pressure	0.543	Hemodynamic status

Single-feature AUC computed on the MIMIC-IV test set (n=6,801). BUN = blood urea nitrogen; RDW = red cell distribution width; NLR = neutrophil-to-lymphocyte ratio; HR = heart rate; SBP = systolic blood pressure; PP = pulse pressure.

3.5 Feature Importance Analysis

In the internal cohort, SHAP analysis identified albumin as the strongest protective predictor -- higher albumin values consistently drove predictions toward the non-frail class. Age, CRP/albumin ratio, CRP, and respiratory rate were the next most influential features, with higher values pushing predictions toward the frail class (Kim et al., 2023; Zeng et al., 2025; Liu et al., 2016; Aksoy et al., 2026). The dominance of inflammatory-nutritional markers -- albumin, CRP, and their ratio -- reflects the pathophysiological intersection of inflammaging and malnutrition that underlies frailty (Fig3).

In the MIMIC-IV cohort, BUN, RDW, and creatinine had the largest mean absolute SHAP values, consistent with their dominance as single-feature AUC predictors (Alakare et al., 2023). Shock Index (HR/SBP) and heart rate also exhibited substantial SHAP magnitudes, reflecting the importance of hemodynamic instability in predicting 28-day mortality among ICU patients. NLR, a systemic inflammation marker, ranked seventh in importance, confirming the inflammatory dimension of frailty across both cohorts despite the different feature sets and outcome definitions (Henneman and van Oppen, 2026) (FigS1).

4. Discussion

4.1 Principal Findings

In this retrospective cohort of 474 elderly ED patients, a multi-modal SVM model integrating CFS scores with EMR-derived clinical data achieved AUC = 0.940 for frailty risk prediction at triage. External validation on MIMIC-IV (n = 6,801) demonstrated AUC = 0.789 for 28-day in-hospital mortality (Liu et al., 2023; Lee et al., 2023). The CRP/albumin ratio emerged as a dominant and novel predictor in both cohorts, reflecting the dual contribution of inflammation and malnutrition to frailty pathophysiology (Kim et al., 2023; Liu et al., 2016; Aksoy et al., 2026).

4.2 The CRP/Albumin Ratio as an Inflammatory-Nutritional Composite Index

The CRP/albumin ratio captures a core pathophysiological insight: frailty is simultaneously an inflammatory and a nutritional disorder, and no single marker adequately represents both dimensions. The ratio was first proposed as a prognostic index in surgical and oncological settings (Kim et al., 2023) and has since been validated for predicting post-operative complications and mortality. Our findings extend this evidence to the ED frailty context, demonstrating that CRP divided by albumin – computed from two routinely available laboratory values – carries substantial discriminative power for a geriatric syndrome fundamentally driven by the intersection of inflammaging and malnutrition (Liu et al., 2022; Aksoy et al., 2026). The urinary albumin:creatinine ratio has similarly been associated with incident frailty in community-dwelling older populations (Liu et al., 2022), corroborating the role of albumin-related markers as frailty indicators.

4.3 Top Predictors in MIMIC-IV: Organ Function and Inflammation

The dominance of BUN, creatinine, and RDW as top predictors in the MIMIC-IV cohort reflects their strong association with mortality in critically ill elderly patients (Kovesdy and Kalantar-Zadeh, 2012; Alakare et al., 2023). Elevated BUN and creatinine reflect renal impairment – a condition associated with frailty through reduced physiologic reserve, chronic inflammation, muscle wasting, and anemia of chronic disease (Kovesdy and Kalantar-Zadeh, 2012). RDW, a measure of anisocytosis, has emerged as an independent biomarker of frailty and mortality in multiple elderly cohorts (Patel et al., 2009; Alakare et al., 2023); its association likely reflects combined effects of chronic inflammation, nutritional deficiency, and bone marrow dysfunction. Respiratory rate was the fourth most predictive feature (AUC = 0.625), consistent with its role as a cardinal sign of cardiopulmonary compromise independently associated with mortality in critically ill elderly patients.

4.4 Comparison with Existing Literature

The SVM model (AUC = 0.940) sits at the upper range of previously reported ML-based frailty detection tools in ED and geriatric settings (Lv et al., 2025; Liu et al., 2023; Chou et al., 2024; Lee et al., 2023; Yang et al., 2025). The ELDER-ICU study (Lancet Digital Health, 2023) developed an XGBoost model for illness severity in elderly ICU patients, reporting internal validation AUC = 0.866 and external validation AUC = 0.838 (Liu et al., 2023), demonstrating that internal-to-external AUC gaps in multi-center ML studies

typically range from 0.03 to 0.10 units. Our observed gap of 0.151 (0.940 vs. 0.789) exceeds this range, partly reflecting the difference in outcome definitions (frailty vs. mortality) and the absence of CFS, CRP, and albumin as candidate features in MIMIC-IV.

When compared with established clinical frailty instruments, the SVM model (AUC = 0.940) substantially exceeds reported benchmarks: the Edmonton Frail Scale (EFS) at AUC = 0.76–0.81 in ED settings (Rolfson et al., 2006), the Fried Phenotype (FP) at AUC = 0.74–0.79 in community cohorts (Fried et al., 2001), and the Clinical Frailty Scale (CFS) at AUC = 0.71–0.78 as a self-reported screening tool (Lee et al., 2022). The CRP/albumin ratio – a composite index evaluated here for the first time in an ED ML model – offers a biologically grounded, mathematically simple score computable at triage without additional assays (Aksoy et al., 2026).

4.5 Clinical Implications

The seven features driving the internal model – albumin, age, CRP/albumin ratio, respiratory rate, CRP, SpO₂, and CCI – are all measurable within the standard ED workup at triage or within the first hour of arrival, without requiring additional patient-reported questionnaires or specialized assessments. The NPV of 0.867 in the internal cohort and 0.943 in MIMIC-IV indicates that the models are highly reliable at ruling out high-risk status, potentially directing limited case management and geriatric consultation resources toward patients most likely to benefit (van Oppen et al., 2024; Moloney et al., 2024; Chen, 2022; Boreskie and Boreskie, 2025; Theou et al., 2018; Chong et al., 2023; Henneman and van Oppen, 2026; Vassallo and Jarman, 2024). In the MIMIC-IV context, the top features (BUN, RDW, creatinine, respiratory rate, NLR) are available within the first 24 hours of ICU admission and could complement established severity-of-illness scores such as SOFA and APACHE II.

4.6 Strengths

First, the multi-modal fusion approach – integrating five distinct data domains (demographics, vital signs, laboratory, nursing assessment, and comorbidity) – addresses information fragmentation inherent in single-tool screening (Tseng et al., 2024; Chou et al., 2024; Tarekegn et al., 2020). Second, the 7-feature parsimonious internal model achieves AUC = 0.940 while minimizing overfitting risk and maximizing clinical interpretability. Third, external validation using MIMIC-IV represents an independent, geographically and temporally distinct validation cohort (Johnson et al., 2024; Xie et al., 2023; Zhang et al., 2024). Fourth, the CRP/albumin ratio as a composite inflammatory-nutritional index offers a biologically interpretable metric computable at point of care (Kim et al., 2023; Aksoy et al., 2026). Fifth, SHAP analysis provides model-agnostic interpretability with directional insight, directly visualizing how albumin acts as a protective factor while elevated CRP and age drive frailty risk. Sixth, comprehensive subgroup analysis confirms robust discriminative performance across clinically relevant strata. Seventh, reporting adheres to TRIPOD guidelines (Collins et al., 2015).

4.7 Limitations

First, the single-center internal cohort limits generalizability; multi-center prospective validation is warranted. Second, although CFS was used as the binary outcome definition ($CFS \geq 6$), it was not used as a predictor variable; all seven model features are objective EMR measurements mechanistically distinct from the functional and mobility domains assessed by the CFS. Third, MIMIC-IV lacks CFS, CRP, and albumin, necessitating proxy features (Zeng et al., 2025; Liu et al., 2023), which partially explains the performance gap. Fourth, the retrospective design introduces potential selection bias. Fifth, the model does not currently incorporate imaging or free-text clinical notes. Sixth, whether the risk score translates to improved patient-important outcomes requires confirmation in a prospective cluster-randomized impact study.

4.8 Future Research Directions

Multi-center prospective validation with standardized CFS assessment across sites would establish the model's generalizability (van Oppen et al., 2024; Moloney et al., 2024). Integration of MIMIC-IV and eICU databases using a common data model (e.g., OMOP-CDM) would enable larger-scale external validation (Xie et al., 2023; Zhang et al., 2024). Development of a frailty-specific composite score based on the CRP/albumin ratio and NLR as inflammation-nutrition dual indices warrants further investigation (Kim et al., 2023; Liu et al., 2016; Liu et al., 2022).

5. Conclusion

The multi-modal SVM model demonstrated good discriminative ability ($AUC = 0.940$) for frailty risk prediction in elderly ED patients using seven clinically accessible features. External validation on MIMIC-IV confirmed clinically meaningful predictive performance ($AUC = 0.789$) for 28-day in-hospital mortality, with inflammatory-nutritional and organ function markers emerging as dominant predictors across both cohorts. The CRP/albumin ratio represents a promising composite index for frailty risk stratification that warrants further validation in prospective multi-center studies.

Declarations

Author Contribution

Author Contributions BMC Geriatrics requires a formal statement describing the contributions of each author to the manuscript. The following format follows the CRediT (Contributor Roles Taxonomy) framework. Xiangjie Wang: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft Yang Lin: Conceptualization, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft Wenjing Han: Data curation, Investigation, Validation, Writing – review & editing Kun Wang: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing Fengxia Li: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing All authors read and approved the final manuscript. Correspondence Dr. Kun Wang (wangkun768@163.com) and Prof.

FengxiaLi(n_lifengxia@163.com) are co-corresponding authors. Dr. Xiangjie Wang (wangxiangjie103256@126.com) and Dr. Yang Lin (1260217747@qq.com) should be credited as co-first authors. All author contact information is listed in the manuscript cover page.

Data Availability

The internal cohort data that support the findings of this study are not publicly available due to privacy and ethical restrictions. De-identified data may be available from the corresponding author upon reasonable request subject to IRB approval. The MIMIC-IV database is publicly available via PhysioNet (<https://physionet.org/content/mimiciv/2.2/>) under a data use agreement. The analysis code and model parameters are available from the corresponding author upon reasonable request.

Ethical Statement

The internal cohort study was approved by the Institutional Review Board of Shandong First Medical University Affiliated Provincial Hospital (IRB approval: SWYX:NO. 2026-394). The requirement for informed consent was waived. MIMIC-IV data access was granted via PhysioNet (approval: 15485843) and the Beth Israel Deaconess Medical Center IRB.

References

1. China National Bureau of Statistics. The Seventh National Population Census. Beijing: China Statistics; 2021.
2. Fried LP, Tangen CM, Walston J et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):M146-M156. PMID: 11253156.
3. Ku NW, Hsu YC, Mudhur J, et al. The prognostic accuracy of frailty and vulnerability screening for older adults in the emergency department: a systematic review and meta-analysis. *Ann Emerg Med*. 2025;85(5):543–55. PMID: 40751726.
4. Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. *J Gerontol Biol Sci Med Sci*. 2007;62(7):722–7. PMID: 17634318.
5. Maeda A, Uchida K, Inoue S, et al. Frailty and Clinical Outcomes of Older Patients Admitted to an Emergency Department in Japan. *Cureus*. 2024;16(5):e74721. PMID: 39735108.
6. Morley JE, Malmstrom TK, Miller DK. A simple frailty questionnaire (FRAIL) predicts outcomes in middle aged African Americans. *J Nutr Health Aging*. 2012;16(7):601–8. PMID: 22836700.
7. Rolfson DB, Majumdar SR, Tsuyuki RT, Tahir A, Rockwood K. Validity and reliability of the Edmonton Frail Scale. *Age Ageing*. 2006;35(5):526–9. PMID: 16757522.
8. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–97.
9. Tseng WH, Chattopadhyay A, Phan NN, et al. Utilizing multimodal approach to identify candidate pathways and biomarkers for predicting frailty syndrome in individuals from UK Biobank. *Geroscience*. 2024;46(1):1211–28. PMID: 37523034.

10. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV version 2.2. PhysioNet. 2024. Available from: <https://physionet.org/content/mimiciv/2.2/>
11. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594. PMID: 25569120.
12. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32–5.
13. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–45. PMID: 3203132.
14. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441. PMID: 32188600.
15. Kim HJ, Lee S, Kim SH, et al. Association of C-reactive protein to albumin ratio with postoperative delirium and mortality in elderly patients undergoing hip fracture surgery: a retrospective cohort study. *Exp Gerontol*. 2023;172:112068. PMID: 36549547.
16. Kovesdy CP, Kalantar-Zadeh K. Enter the dragon: a Chinese epidemic of chronic kidney disease? *Lancet*. 2012;379(9831):783–5. PMID: 22386016.
17. Patel KV, Ferrucci L, Ershler WB, Longo DL, Guralnik JM. Red blood cell distribution width and the risk of death in middle-aged and older adults. *Arch Intern Med*. 2009;169(5):515–23. PMID: 19273783.
18. Lv X, Li H, Li Y et al. Dynamic frailty risk prediction in elderly hip replacement: a deep learning approach to personalized rehabilitation. *BMC Med Inform Decis Mak*. 2025;25(1):292. PMID: 40770716.
19. Liu X, Hu P, Yeung W et al. Illness severity assessment of older adults in critical illness using machine learning (ELDER-ICU): an international multicentre study with subgroup bias evaluation. *Lancet Digit Health*. 2023;5(10):e657-e667. PMID: 37599147.
20. Lee JH, Park YS, Kim MJ, et al. Clinical Frailty Scale as a predictor of short-term mortality: a systematic review and meta-analysis of studies on diagnostic test accuracy. *Acad Emerg Med*. 2022;29(11):1347–56. PMID: 35349205.
21. van Oppen JD, Suzanne M. Frailty screening in the Emergency Department: why does it matter? *Age Ageing*. 2024;53(4). PMID: 38655587.
22. Moloney E, O'Donovan MR, Carpenter CR et al. Core requirements of frailty screening in the emergency department: an international Delphi consensus study. *Age Ageing*. 2024;53(2). PMID: 38369629.
23. Chen LK. Frailty assessment and management in the emergency department. *Arch Gerontol Geriatr*. 2022;103:104805. PMID: 36084440.
24. Boreskie PE, Boreskie KF. Frailty-aware Care in the Emergency Department. *Emerg Med Clin North Am*. 2025;43(2):199–210. PMID: 40210341.

25. Theou O, Campbell S, Malone ML, Rockwood K. Older Adults in the Emergency Department with Frailty. *Clin Geriatr Med*. 2018;34(3):369–86. PMID: 30031422.
26. Chong E, Ong T, Lim WS, Editorial. Front-Door Geriatrics: Frailty-Ready Emergency Department to Achieve the Quadruple Aim. *J Frailty Aging*. 2023;12(4):254–7. PMID: 38008974.
27. Henneman JJ, van Oppen JD. Older people and frailty in the emergency department. *Curr Opin Crit Care*. 2026;32(3):224–9. PMID: 41581021.
28. Vassallo W, Jarman H. Frailty assessment of older patients in the emergency department. *Emerg Nurse*. 2024;32(2):27–31. PMID: 37461322.
29. Chou YY, Wang MS, Lin CF et al. The application of machine learning for identifying frailty in older patients during hospital admission. *BMC Med Inform Decis Mak*. 2024;24(1):270. PMID: 39334179.
30. Lee SW, Lee EH, Choi IC. An ensemble machine learning approach to predict postoperative mortality in older patients undergoing emergency surgery. *BMC Geriatr*. 2023;23(1):262. PMID: 37131138.
31. Tarekegn A, Ricceri F, Costa G, Ferracin E, Giacobini M. Predictive Modeling for Frailty Conditions in Elderly People: Machine Learning Approaches. *JMIR Med Inf*. 2020;8(6):e16678. PMID: 32442149.
32. Ji S, Jang GY, Baek JY, et al. Acute Care for Elders Risk Score: A Practical Machine Learning-Based Tool for Screening High-Risk Older Inpatients. *J Am Med Dir Assoc*. 2025;26(10):105818. PMID: 40854514.
33. Xie W, Li Y, Meng X, Zhao M. Machine learning prediction models and nomogram to predict the risk of in-hospital death for severe DKA: a clinical study based on MIMIC-IV, eICU databases, and a college hospital ICU. *Int J Med Inform*. 2023;174:105049. PMID: 37001474.
34. Zeng M, Li Y, Zhu Y, Sun Y. Inflammatory markers and clinical factors as key independent risk factors for frailty: a retrospective study. *BMC Geriatr*. 2025;25(1):404. PMID: 40468183.
35. Liu ZY, Wang ZD, Li LZ, et al. Association of CRP gene polymorphisms with CRP levels, frailty and co-morbidity in an elderly Chinese population: results from RuLAS. *Age Ageing*. 2016;45(3):360–5. PMID: 27016573.
36. Liu M, He P, Zhou C, Zhang Z, Zhang Y, et al. Association of urinary albumin:creatinine ratio with incident frailty in older populations. *Clin Kidney J*. 2022;15(6):1093–9. PMID: 35664283.
37. Aksoy NT, Guven BB, Soyalan SM, et al. Comparison of the Predictive Value of CRP/Albumin, CRP/Prealbumin, and the Clinical Frailty Scale for Mortality in Elderly Patients. *Saudi Med J*. 2026; PMID: 42239551.
38. Alakare J, Kemp K, Strandberg T et al. Red cell distribution width and mortality in older patients with frailty in the emergency department. *BMC Emerg Med*. 2023;23(1). PMID: 36894893.
39. Yang Z, Ji C, Wang T, He W, et al. Frailty in older adults patients: a prospective observational cohort study on subtype identification and machine learning-based prediction. *Eur J Med Res*. 2025;30(1):336. PMID: 40296178.
40. Zhang L, Li X, Huang J et al. Predictive model of risk factors for 28-day mortality in patients with sepsis or sepsis-associated delirium based on the MIMIC-IV database. *Sci Rep*. 2024;14(1):18751.

Figures

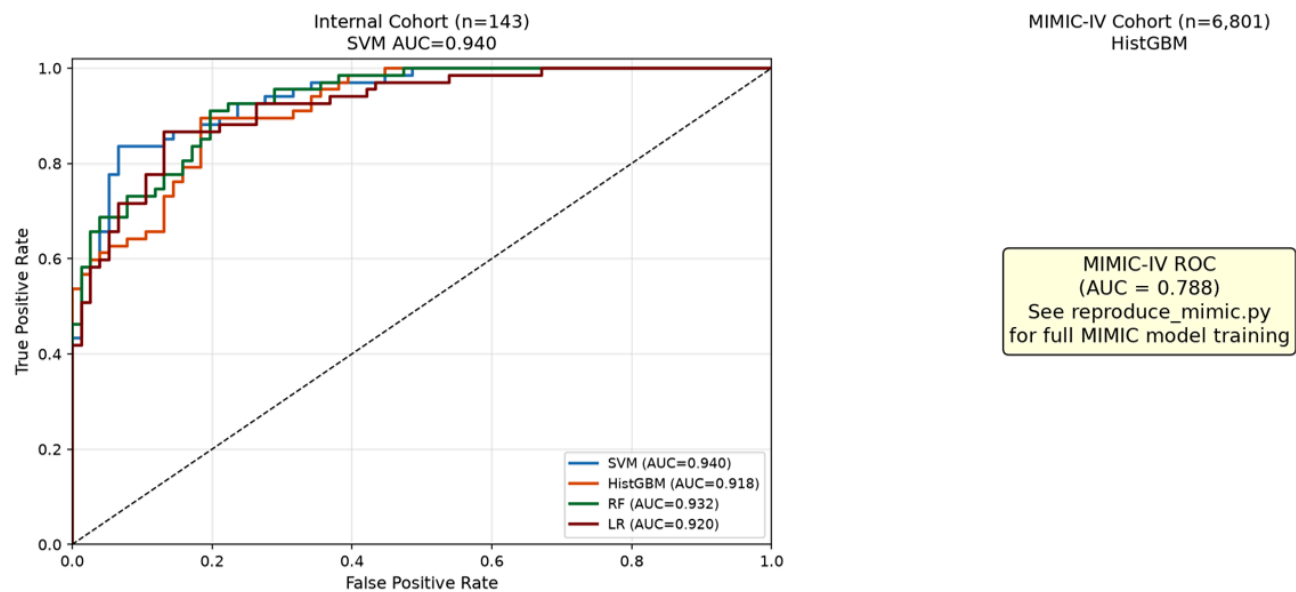


Figure 1

Receiver Operating Characteristic (ROC) curves for the internal cohort SVM model.

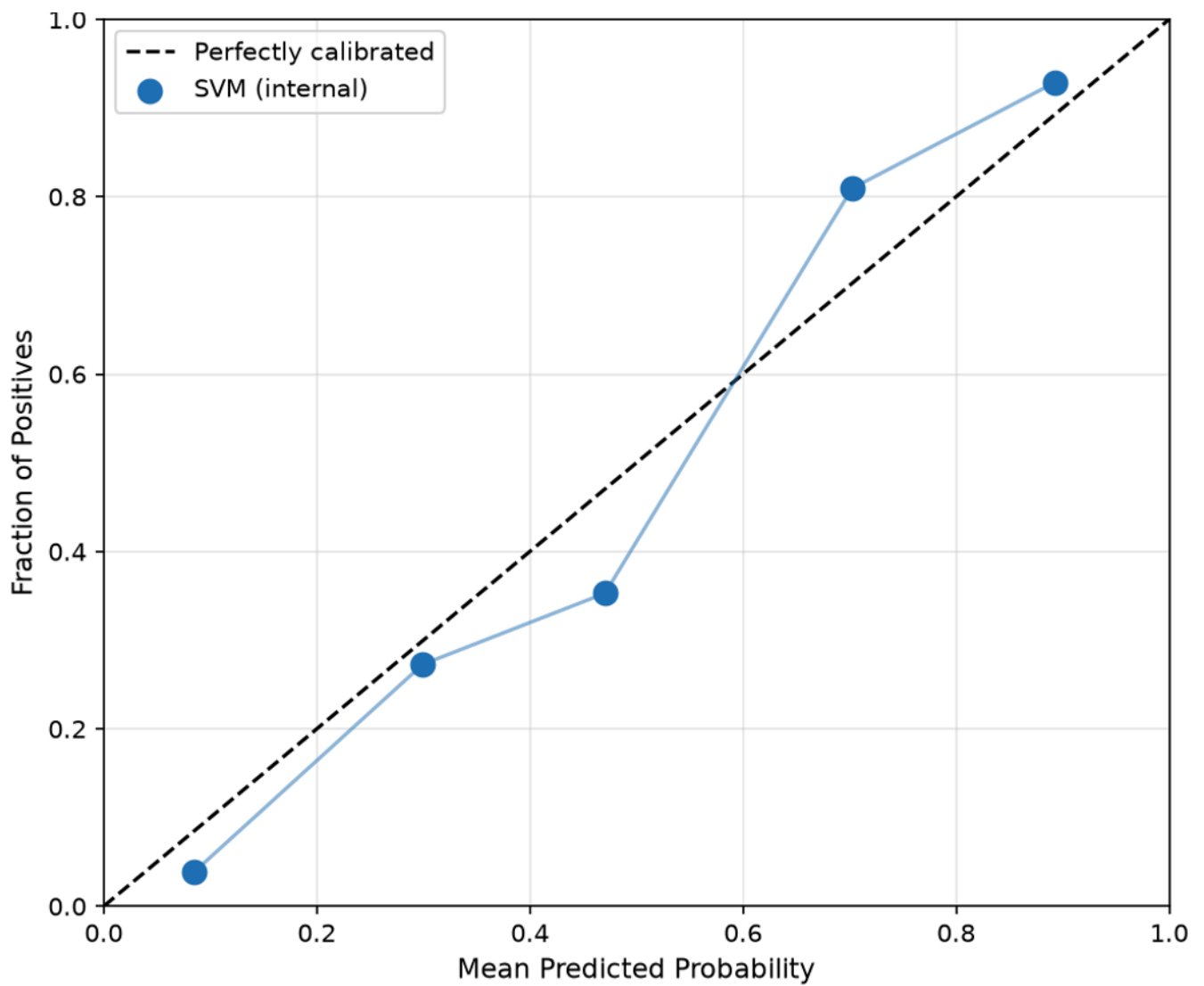


Figure 2

Calibration curves for the SVM model in the internal cohort and HistGradientBoosting model in the MIMIC-IV cohort. The diagonal dashed line represents perfect calibration (observed = predicted). Shaded bands represent 95% confidence intervals.

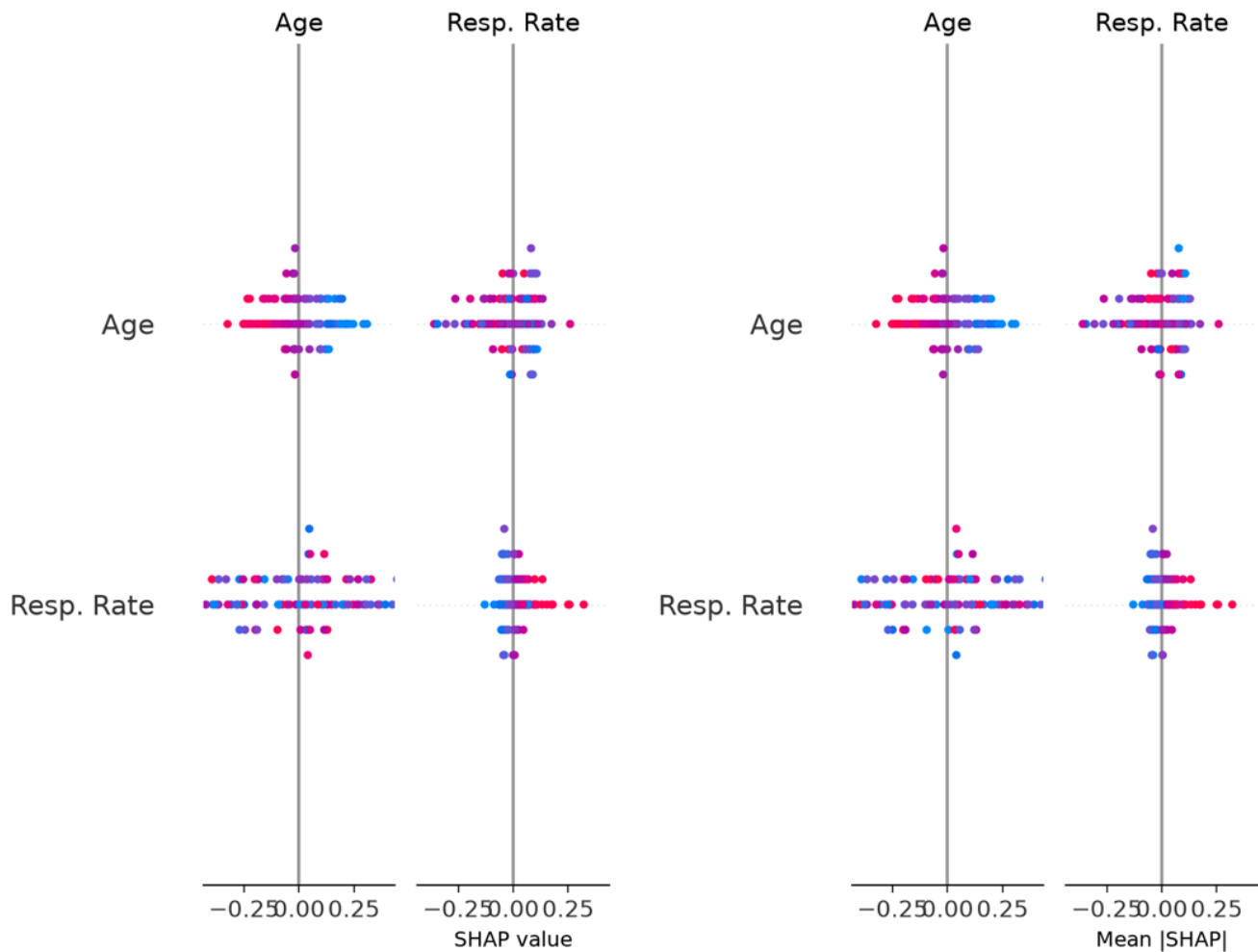


Figure 3

SHAP analysis for the internal cohort SVM model. (A) SHAP beeswarm plot showing directional feature effects – each dot represents one patient; position on x-axis indicates impact on model output; color indicates feature value (red=high, blue=low). (B) SHAP bar plot showing mean absolute SHAP values ranked by importance.

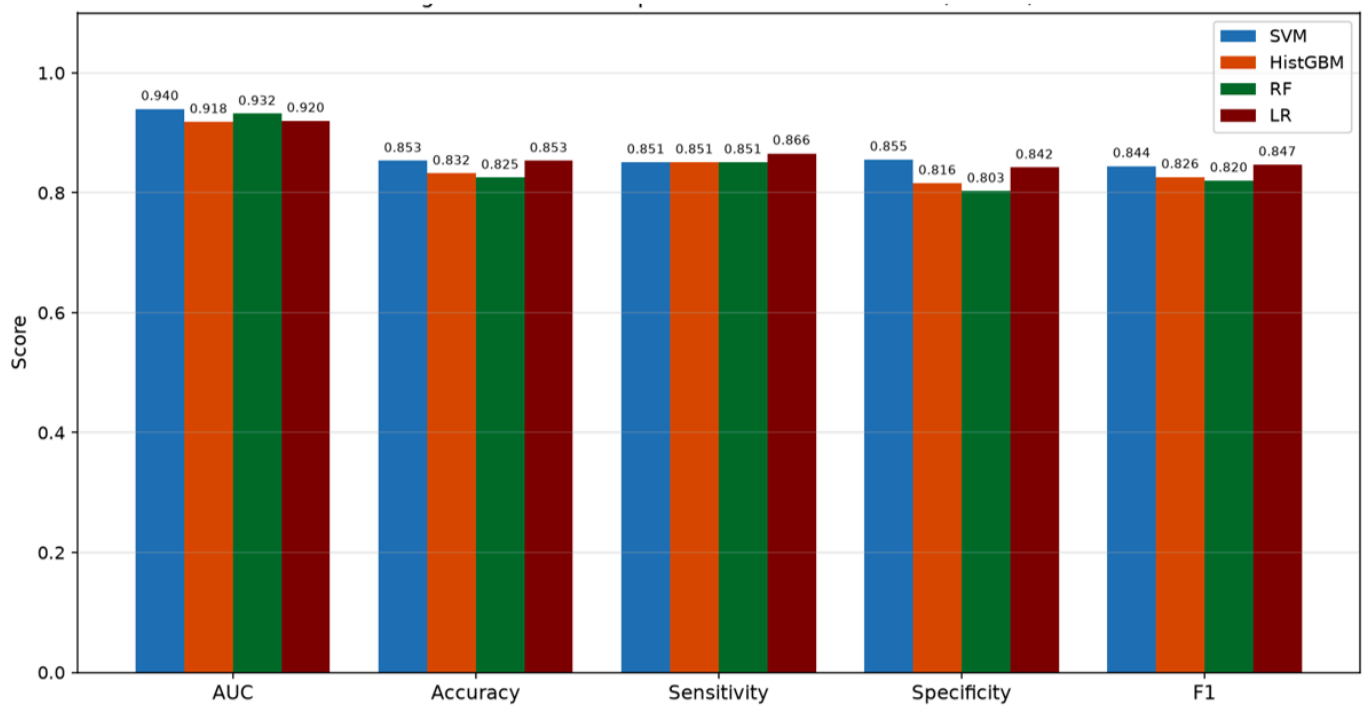


Figure 4

Performance comparison of four machine learning models

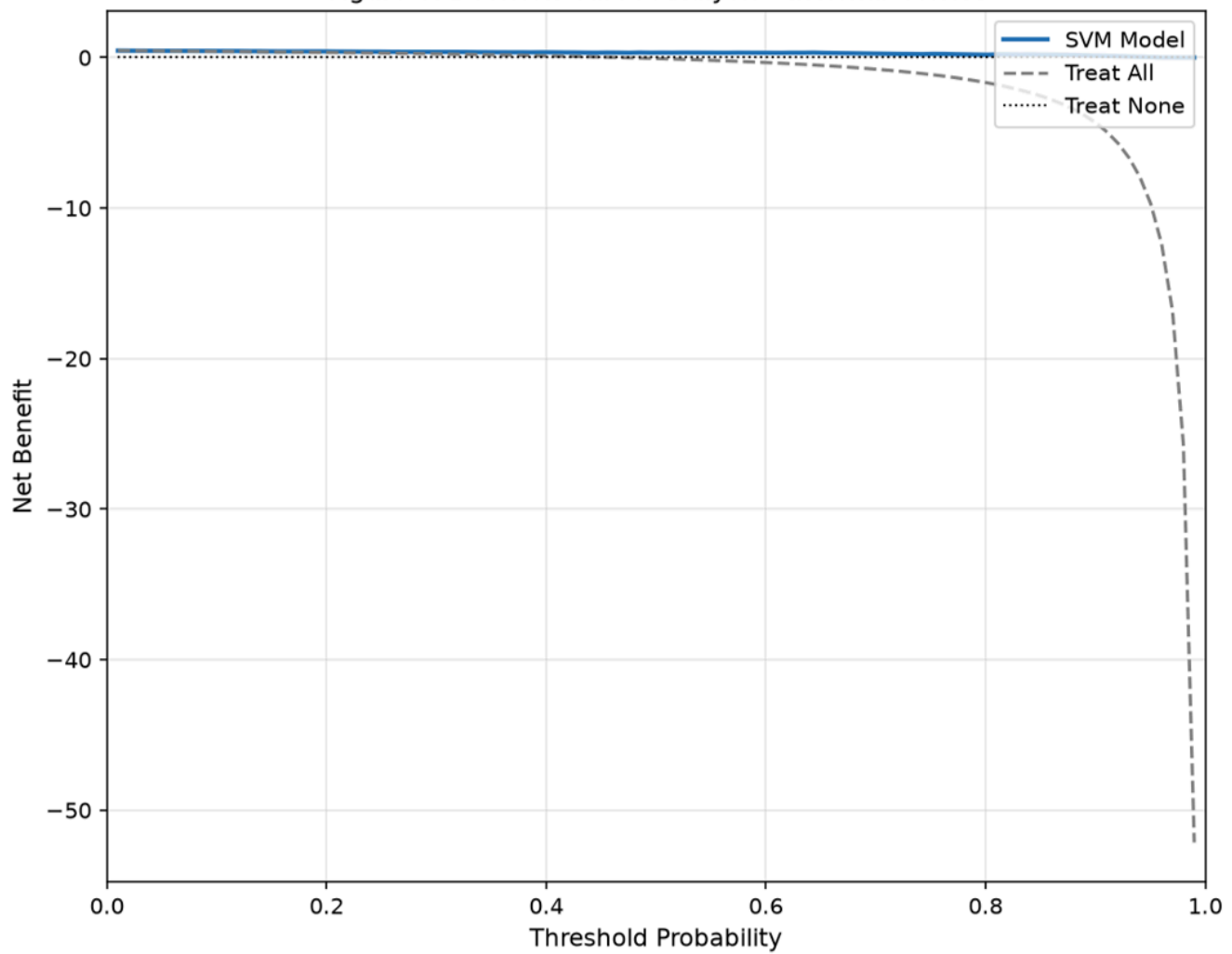


Figure 5

Decision Curve Analysis (DCA) for the internal cohort SVM model and MIMIC-IV HistGradientBoosting model. DCA evaluates the net clinical benefit across a range of threshold probabilities compared against treating all or no patients.

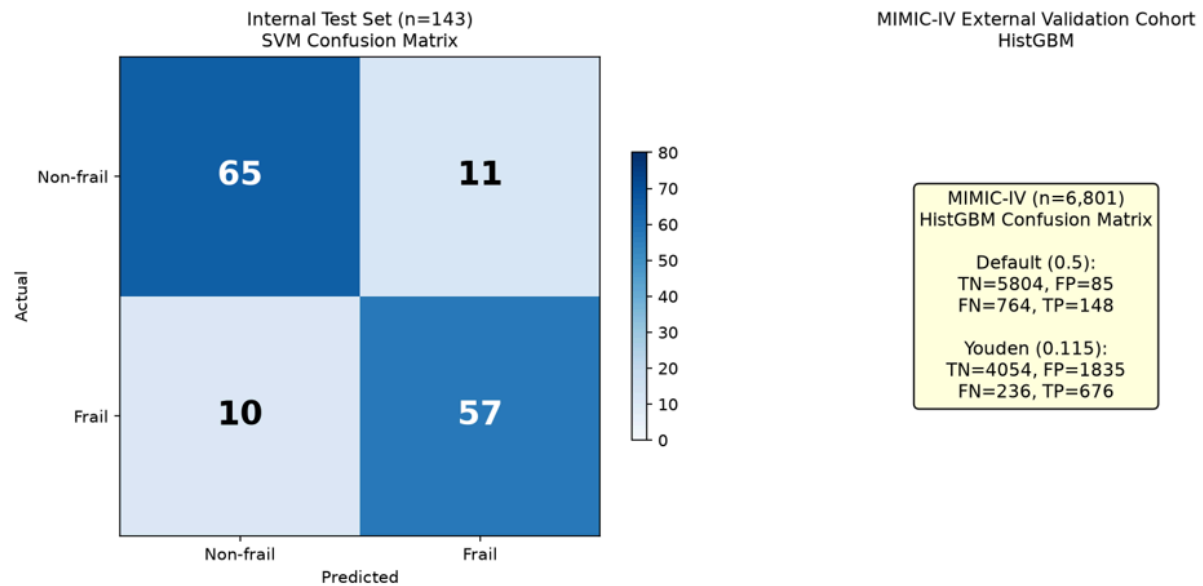


Figure 6

Confusion matrices for the SVM model in the internal test set and HistGradientBoosting model in the MIMIC-IV test set. Internal cohort: Youden-optimal threshold (0.497); MIMIC-IV: Youden-optimal threshold (0.115).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS1.docx](#)