

Development and temporal external validation of a clinical and CT- based model for predicting clinically significant hemopneumothorax after unilateral rib fracture: a retrospective cohort study

Luchao Wang

Huangshan People's Hospital

Zhicheng Xie

Huangshan People's Hospital

Yanwei Zhan

Huangshan People's Hospital

Wei Hong

Huangshan People's Hospital

Weiwei Fang

fangweiwei0302@163.com

Huangshan People's Hospital

Research Article

Keywords: rib fracture, hemopneumothorax, clinical prediction model, temporal external validation, nomogram, calibration

Posted Date: September 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10687011/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

Clinically significant hemopneumothorax (CSHP) is a progression that warrants early recognition after unilateral rib fracture. Monitoring is often guided by initial pneumothorax or hemothorax signs, fracture displacement, and clinical judgment, without a structured risk estimate for this short-term endpoint. Existing rib-injury scores emphasize overall injury burden and cover progression to intervention-requiring hemopneumothorax only to a limited extent.

Methods

This was a single-center retrospective cohort. The development cohort included 484 patients (stratified 7:3 random split by outcome into 338 training and 146 internal test cases), with 115 patients in a temporal external validation set. Predictors were early admission clinical variables and chest CT within 24 hours. On the training set, univariable screening ($P < 0.10$) and multivariable backward elimination ($P < 0.05$) yielded a logistic model and nomogram, compared with SVM, random forest, GBDT, and XGBoost. We reported AUC, Brier score, calibration slope/intercept, Hosmer–Lemeshow test, decision curves, simple-rule comparisons, and feature-contribution analysis.

Results

In total, 599 patients were analyzed. CSHP occurred in 115 of 484 development patients (23.8%) and 30 of 115 external patients (26.1%). The final model included pulmonary contusion, maximum displacement distance, number of displaced fractures, small pneumothorax, main fracture zone, and thoracic vertebral fracture. Logistic AUC was 0.884 (bootstrap 95% CI 0.820–0.936) internally and 0.913 (0.861–0.962) externally; external calibration slope was 0.631, intercept 0.008, and Brier score 0.118. At the training Youden threshold of 0.155 applied externally, sensitivity was 0.933, specificity 0.859, PPV 0.700, and NPV 0.973. External AUC of the full model exceeded the “small pneumothorax” rule by 0.145. Tertiles of predicted probability separated external event rates from 0% in the low stratum to 73.7% in the high stratum.

Conclusions

The model showed good discrimination on temporal external validation and outperformed several simple rules; calibration was limited, so predicted probabilities should not be treated as absolute individual risk. The model is suitable as a reference for early stratification and monitoring planning.

1 Introduction

Rib fractures are common after blunt trauma [1]. Some patients progress from small amounts of gas or fluid at admission to drainage, surgery, or escalation of life support. Existing rib-injury scores capture fracture count, displacement, and flail chest as anatomic burden and help overall severity assessment [2, 3], but seldom answer whether intervention-requiring hemopneumothorax will develop over a short horizon. This study turns early clinical and CT items available at admission into a reproducible risk estimate and evaluates discrimination and calibration in a later-period sample.

Rib fracture burden is a recognized determinant of adverse outcome after blunt chest wall trauma [4]. Existing scoring systems quantify fracture count, displacement, and flail segments, but they were developed to predict pneumonia, respiratory failure, and tracheostomy rather than progression to intervention-requiring hemopneumothorax [2, 3]. Early chest trauma management emphasizes detection of pneumothorax and hemothorax with low-threshold chest drainage [5, 6]; retained hemothorax increases the risk of empyema [7], and pulmonary contusion independently worsens pulmonary outcomes [8]. Although early risk stratification in chest trauma has been investigated [9], a structured, validated risk estimate for short-term clinically significant hemopneumothorax after unilateral rib fracture remains lacking. Reporting follows the TRIPOD statement [10]; discrimination and calibration are assessed with recommended performance measures [11, 12]; machine-learning classifiers, which systematic reviews have not shown to outperform logistic regression for clinical prediction [13], are compared with the bedside logistic nomogram [14].

2 Methods

2.1 Design and participants

This single-center retrospective cohort study was approved by the Ethics Committee of Huangshan People's Hospital, with the requirement for informed consent waived because of the retrospective nature of the study. Eligible participants were adults with unilateral traumatic rib fractures. Patients admitted from January 2021 to December 2024 constituted the development cohort, whereas those admitted from January 2025 to May 2026 constituted the temporal external validation cohort. A total of 599 patients were included in the analysis, comprising 484 patients in the development cohort and 115 patients in the temporal external validation cohort.

2.2 Inclusion and exclusion

Inclusion: age ≥ 18 years; CT-diagnosed unilateral traumatic rib fracture (or contralateral cortical irregularity only); chest CT within 24 hours of admission; complete modeling variables. Exclusion: bilateral fractures; penetrating injury; drainage before admission; massive hemopneumothorax or flail chest requiring emergency management at admission, or immediate ICU mechanical ventilation; poor CT quality; severe chest-wall deformity or lung destruction; prior chest surgery.

2.3 Endpoint

CSHP was defined as any of the following: (1) closed chest drainage (moderate or greater hemothorax: blood volume > 500 ml or CT fluid thickness > 4 cm; moderate or greater pneumothorax: air volume > 20%); (2) progression of hemopneumothorax requiring surgical exploration for hemostasis and fracture fixation; (3) mechanical ventilation or ICU transfer because of hemopneumothorax progression.

2.4 Variables and measurement

Demographics and comorbidities, injury energy, anticoagulation/antiplatelet use, hemoglobin, platelets, INR, AIS chest, ISS [15]; total fracture count, number of displaced fractures, maximum displacement distance, segmental fractures, fracture discontinuity count, main fracture zone (anterior/lateral/posterior coded 0/1/2), posterior rib fractures, laterality; pulmonary contusion, chest-wall hematoma, small hemothorax/pneumothorax, subcutaneous emphysema, mediastinal emphysema, and sternal/thoracic vertebral/clavicular fracture. CT slice thickness was 1.0–1.5 mm with lung and mediastinal window reconstructions. Main fracture zone entered regression as an ordinal code and should be interpreted cautiously.

2.5 Modeling and evaluation

The development cohort was split 7:3 at random, stratified by outcome (random_state = 42). Univariable logistic regression ($P < 0.10$) followed by backward elimination ($P < 0.05$) on the training set produced the final model and a nomogram-style score chart; the same features were used to fit SVM, random forest, GBDT, and XGBoost. Classification thresholds were locked on the training set by the Youden index and applied unchanged to the internal test and temporal external sets. Evaluation included ROC/AUC (bootstrap 95% CI), Brier score, calibration curves with slope/intercept, Hosmer–Lemeshow test, decision curves [16], simple-rule comparisons (small pneumothorax, pulmonary contusion, either, small hemothorax, and multiple cut-points for displaced fracture count and maximum displacement), leave-variable-out sensitivity analyses, tertile stratification of external predicted probabilities, and exploratory subgroups. Analyses were performed in Python (SciPy, statsmodels, scikit-learn, XGBoost, SHAP, and related packages).

The final logistic linear predictor was: $\text{logit}(P) = \beta_0 + \beta_1 \cdot \text{pulmonary contusion} + \beta_2 \cdot \text{maximum displacement} + \beta_3 \cdot \text{displaced fracture count} + \beta_4 \cdot \text{small pneumothorax} + \beta_5 \cdot \text{main fracture zone} + \beta_6 \cdot \text{thoracic vertebral fracture}$; $P = 1/(1 + \exp(-\text{logit}(P)))$. The training intercept was $\beta_0 = -8.770$; remaining coefficients are in Table 2.

3 Results

3.1 Analytic cohorts

Altogether 599 patients were included. Of 484 development patients, 115 had CSHP (23.8%); stratified random split yielded 338 training patients (event rate 23.7%) and 146 internal test patients (event rate 24.0%). The temporal external set included 115 patients with 30 CSHP events (26.1%), slightly higher

than the training rate. The study flow is shown in Fig. 1; between-group distributions of key clinical and CT features in the development cohort are shown in Fig. 2.

Regarding case mix, the external set had a higher proportion of small pneumothorax than the training set (30.4% vs 24.9%) and a lower proportion of pulmonary contusion (23.5% vs 33.1%); median maximum displacement was 5.6 mm in both. These differences indicate that the external sample was not a simple replica of the development cohort; discrimination and calibration should be interpreted with case mix in mind.

3.2 Baseline characteristics

Baseline characteristics of the development cohort by CSHP status are shown in Table 1. Compared with non-CSHP patients, those with CSHP were older (median 65.0 vs 59.0 years, $P = 0.002$) [17], less often male (27.0% vs 42.3%, $P = 0.005$), and had lower BMI and hemoglobin (BMI 21.5 vs 22.4, $P = 0.002$; hemoglobin 114.0 vs 123.0 g/L, $P < 0.001$). Between-group differences in anticoagulation/antiplatelet use, high-energy injury, emphysema, and prior pulmonary tuberculosis were not statistically significant.

Anatomic burden and CT complications differed more markedly. Patients with CSHP had higher total fracture count, displaced fracture count, maximum displacement, segmental fracture count, posterior rib fracture count, and fracture discontinuity count (median maximum displacement 11.2 vs 4.4 mm; displaced fractures 6.0 vs 3.0; all $P < 0.001$). Pulmonary contusion (77.4% vs 17.3%), small pneumothorax (56.5% vs 16.5%), small hemothorax (92.2% vs 46.3%), subcutaneous emphysema (40.9% vs 10.0%), and clavicular fracture (17.4% vs 4.1%) were more common in the CSHP group (all $P < 0.001$). Thoracic vertebral fracture (11.3% vs 5.4%, $P = 0.048$) and first-rib fracture (13.9% vs 6.8%, $P = 0.027$) were also higher; chest-wall hematoma differed little (9.6% vs 7.6%, $P = 0.628$) and was not retained in the final model. AIS chest and ISS were higher in the CSHP group, but medians were close, suggesting limited ability of global injury scores to separate progression to intervention-requiring hemopneumothorax. Final-model variables followed the selection results.

Table 1
Baseline characteristics of the development cohort by CSHP status

Variable	Non-CSHP (n = 369)	CSHP (n = 115)	P
Sex (male = 1)	156/369 (42.3%)	31/115 (27.0%)	0.005
Age	59.0 (53.0–68.0)	65.0 (55.5–70.5)	0.002
Emphysema	36/369 (9.8%)	18/115 (15.7%)	0.113
Prior pulmonary tuberculosis	11/369 (3.0%)	7/115 (6.1%)	0.210
Anticoagulation/antiplatelet	14/369 (3.8%)	5/115 (4.3%)	1.000
High-energy injury	257/369 (69.6%)	89/115 (77.4%)	0.137
BMI	22.4 (20.3–24.7)	21.5 (19.5–23.2)	0.002
Hemoglobin	123.0 (111.0–134.0)	114.0 (105.0–126.5)	< 0.001
INR	0.9 (0.9–1.0)	0.9 (0.9–1.0)	0.058
Platelets	157.0 (116.0–196.0)	143.0 (105.0–182.5)	0.035
Total fracture count	4.0 (3.0–5.0)	6.0 (5.0–8.0)	< 0.001
Displaced fractures	3.0 (2.0–4.0)	6.0 (4.0–7.0)	< 0.001
Max displacement(mm)	4.4 (3.0–6.7)	11.2 (8.4–13.6)	< 0.001
Segmental fracture count	0.0 (0.0–0.0)	2.0 (0.0–3.0)	< 0.001
Posterior rib fracture count	0.0 (0.0–3.0)	4.0 (2.0–5.0)	< 0.001
Fracture discontinuity count	4.0 (3.0–6.0)	8.0 (6.0–11.0)	< 0.001
Main fracture zone (anterior/lateral/posterior)	1.0 (0.0–2.0)	2.0 (1.0–2.0)	< 0.001
First-rib fracture	25/369 (6.8%)	16/115 (13.9%)	0.027
Chest-wall hematoma	28/369 (7.6%)	11/115 (9.6%)	0.628
Thoracic vertebra fx	20/369 (5.4%)	13/115 (11.3%)	0.048
Sternal fracture	18/369 (4.9%)	10/115 (8.7%)	0.193
Clavicular fracture	15/369 (4.1%)	20/115 (17.4%)	< 0.001
Pulmonary contusion	64/369 (17.3%)	89/115 (77.4%)	< 0.001
Note: Continuous variables are median (IQR); categorical variables are n/N (%). P values are for between-group comparisons.			

Variable	Non-CSHP (n = 369)	CSHP (n = 115)	P
Subcutaneous emphysema	37/369 (10.0%)	47/115 (40.9%)	< 0.001
Small pneumothorax	61/369 (16.5%)	65/115 (56.5%)	< 0.001
Mediastinal emphysema	2/369 (0.5%)	5/115 (4.3%)	0.011
Small hemothorax	171/369 (46.3%)	106/115 (92.2%)	< 0.001
AIS chest	3.0 (2.0–3.0)	3.0 (3.0–3.0)	< 0.001
ISS	9.0 (4.0–9.0)	9.0 (9.0–10.0)	< 0.001
Fracture laterality (right = 1)	171/369 (46.3%)	61/115 (53.0%)	0.250
Note: Continuous variables are median (IQR); categorical variables are n/N (%). P values are for between-group comparisons.			

3.3 Regression results

In training-set univariable logistic regression, burden indicators such as pulmonary contusion, maximum displacement, displaced fracture count, total fractures, and discontinuity count, as well as complication-related items including small pneumothorax, small hemothorax, subcutaneous emphysema, and clavicular fracture, were significantly associated with CSHP (Fig. 3A). After backward elimination, six variables remained: pulmonary contusion, maximum displacement, displaced fracture count, small pneumothorax, main fracture zone, and thoracic vertebral fracture. Multivariable coefficients, ORs, and 95% CIs are in Table 2 and Fig. 3B; the bedside nomogram is shown in Fig. 4.

After adjustment, ORs for pulmonary contusion and small pneumothorax were 5.46 (95% CI 2.13–13.98) and 5.94 (2.25–15.67), indicating that early CT complications remained independent risk factors. Each 1-mm increase in maximum displacement raised the odds of CSHP by about 1.34 (1.20–1.50); each additional displaced fracture raised the odds by about 1.76 (1.36–2.27). Main fracture zone entered as an ordinal anterior/lateral/posterior code (0/1/2); each step toward the posterior rib increased the adjusted OR to 1.98 (1.13–3.48, $P = 0.017$). This coding assumes approximately equal spacing between levels and should be interpreted cautiously. The adjusted OR for thoracic vertebral fracture was 14.23 (2.03–99.70, $P = 0.008$), but the confidence interval was wide: only 22 training patients had thoracic vertebral fracture, of whom 9 had CSHP; among 316 without thoracic vertebral fracture, 71 had CSHP. Sparse crosstabulation indicates unstable effect estimation; the point estimate should not be over-interpreted.

Table 2
Multivariable logistic regression (training set, n = 338)

Variable	β	OR	95% CI lower	95% CI upper	P
Intercept	-8.77	—	—	—	—
Pulmonary contusion	1.697	5.46	2.13	13.98	< 0.001
Max displacement(mm)	0.296	1.34	1.2	1.5	< 0.001
Displaced fractures	0.564	1.76	1.36	2.27	< 0.001
Small pneumothorax	1.781	5.94	2.25	15.67	< 0.001
Main fracture zone (anterior/lateral/posterior)	0.683	1.98	1.13	3.48	0.017
Thoracic vertebra fx	2.655	14.23	2.03	99.7	0.008
Note: Main fracture zone entered the model as an ordinal anterior/lateral/posterior code 0/1/2.					

3.4 Performance, calibration, and threshold

On the internal test set, logistic AUC was 0.884 (bootstrap 95% CI 0.820–0.936) and Brier score 0.122; calibration slope was 0.586, intercept 0.017, and Hosmer–Lemeshow P = 4.48e-08. On temporal external validation, AUC was 0.913 (0.861–0.962) and Brier 0.118; calibration slope was 0.631, intercept 0.008, and Hosmer–Lemeshow P = 2.29e-05. The calibration slope substantially below 1 suggests overall overestimation of predicted probabilities or inflated risk separation; the intercept near 0 indicates limited systematic shift. Calibration metrics are summarized in Table 3 and curves in Fig. 5.

Table 3
Calibration-related metrics for the logistic model

Dataset	AUC	Brier	Calibration slope	Calibration intercept
Internal test	0.884	0.122	0.586	0.017
Temporal external	0.913	0.118	0.631	0.008

The classification threshold was locked at 0.155 by the Youden index on the training set and applied unchanged. On the internal test set, sensitivity was 0.743, specificity 0.847, accuracy 0.822, and precision 0.605. On external validation: true positives 28, false positives 12, true negatives 73, false negatives 2; sensitivity 0.933, specificity 0.859, accuracy 0.878, PPV 0.700, NPV 0.973. The low threshold favors fewer missed cases, with more false positives and higher monitoring/follow-up burden

[18]. Apparent training AUC reached 0.968, clearly higher than internal testing, indicating overfitting; reporting and interpretation should prioritize the test set and temporal external validation.

With the same features, SVM, random forest, GBDT, and XGBoost performance on internal and external validation is shown in Table 4 and Fig. 6. Internally, GBDT had the highest AUC (0.890), followed by random forest (0.888), with limited gaps versus logistic regression; at Youden thresholds, tree models had higher specificity and often lower sensitivity (e.g., internal GBDT sensitivity only 0.429). Externally, GBDT AUC was 0.917, XGBoost 0.914, and SVM and logistic both 0.913. Considering discrimination, interpretability, and bedside usability, the primary model remained the logistic nomogram; machine-learning results served mainly as discrimination controls.

Table 4
Multi-model performance (thresholds locked at each model's training Youden threshold)

Model	Dataset	AUC (95% CI)	Sensitivity	Specificity	Accuracy	Brier	HL P
Logistic	Internal test	0.884 (0.820–0.936)	0.743	0.847	0.822	0.122	4.48e-08
SVM	Internal test	0.863 (0.792–0.925)	0.543	0.937	0.842	0.129	1.27e-05
Random Forest	Internal test	0.888 (0.833–0.935)	0.686	0.928	0.870	0.111	0.350
GBDT	Internal test	0.890 (0.826–0.938)	0.429	0.946	0.822	0.137	0.00e+00
XGBoost	Internal test	0.885 (0.826–0.935)	0.543	0.937	0.842	0.124	7.99e-15
Logistic	Temporal external	0.913 (0.861–0.962)	0.933	0.859	0.878	0.118	2.29e-05
SVM	Temporal external	0.913 (0.855–0.965)	0.733	0.906	0.861	0.101	0.127
Random Forest	Temporal external	0.906 (0.855–0.957)	0.867	0.894	0.887	0.109	7.10e-04
GBDT	Temporal external	0.917 (0.870–0.969)	0.700	0.918	0.861	0.118	2.53e-11
XGBoost	Temporal external	0.914 (0.857–0.965)	0.667	0.918	0.852	0.118	1.44e-10
Note: AUC 95% CIs are bootstrap percentile intervals.							

3.5 Rule comparisons and sensitivity analyses

To quantify incremental value over everyday single-cutpoint or single-sign judgments, small pneumothorax, pulmonary contusion, either, small hemothorax, and multiple cut-points for displaced fracture count and maximum displacement were treated as simple rules; AUC, sensitivity, and specificity were computed on the internal test and temporal external sets (Table 5); external AUC comparisons are shown in Fig. 7.

On temporal external validation, full logistic AUC was 0.913. The best simple rule was maximum displacement ≥ 8 mm (AUC 0.819, sensitivity 0.767, specificity 0.871), followed by displaced fractures ≥ 5 (external AUC 0.815 in Table 5) and displaced fractures ≥ 4 (AUC 0.802, sensitivity 0.933, specificity 0.671). “Small pneumothorax alone” had external AUC 0.768 (sensitivity 0.700, specificity 0.835); “pulmonary contusion alone” 0.612; and “pulmonary contusion or small pneumothorax” 0.771. The full model exceeded the small-pneumothorax rule by 0.145, indicating that combining displacement burden with multiple early CT findings separates high- and low-risk patients better than a single sign.

In sensitivity analysis, refitting after removing small pneumothorax yielded external AUC 0.894 (0.019 lower than the full model) and internal AUC 0.863; substantial discrimination remained, suggesting that fracture displacement and contusion information were not dependent on the pneumothorax item. After removing thoracic vertebral fracture, external AUC was 0.913 and internal AUC 0.886, nearly identical to the full model, consistent with sparse events and unstable point estimates for that variable.

Table 5
Comparison of simple clinical rules and the full model

Rule	Dataset	AUC	Sensitivity	Specificity
Small pneumothorax	Internal test	0.649	0.514	0.784
Small pneumothorax	Temporal external	0.768	0.700	0.835
Pulmonary contusion	Internal test	0.729	0.629	0.829
Pulmonary contusion	Temporal external	0.612	0.400	0.824
Pulmonary contusion or small pneumothorax	Internal test	0.718	0.743	0.694
Pulmonary contusion or small pneumothorax	Temporal external	0.771	0.800	0.741
Small hemothorax	Internal test	0.677	0.886	0.468
Small hemothorax	Temporal external	0.704	0.867	0.541
Displaced fractures ≥ 3	Internal test	0.673	0.886	0.459
Displaced fractures ≥ 3	Temporal external	0.724	1.000	0.447
Displaced fractures ≥ 4	Internal test	0.733	0.800	0.667
Displaced fractures ≥ 4	Temporal external	0.802	0.933	0.671
Displaced fractures ≥ 5	Internal test	0.747	0.657	0.838
Displaced fractures ≥ 5	Temporal external	0.815	0.900	0.729
Maximum displacement ≥ 6 mm	Internal test	0.766	0.857	0.676
Maximum displacement ≥ 6 mm	Temporal external	0.769	0.867	0.671
Maximum displacement ≥ 8 mm	Internal test	0.738	0.657	0.820
Maximum displacement ≥ 8 mm	Temporal external	0.819	0.767	0.871
Maximum displacement ≥ 10 mm	Internal test	0.684	0.457	0.910
Maximum displacement ≥ 10 mm	Temporal external	0.676	0.400	0.953
Full logistic model (probability)	Internal test	0.884	—	—

Rule	Dataset	AUC	Sensitivity	Specificity
Full logistic model (probability)	Temporal external	0.913	—	—

3.6 Stratification and exploratory subgroups

External predicted probabilities were divided into tertiles (Table 6, Fig. 8). No CSHP occurred in the low-risk stratum (0/39, observed rate 0.0%); 2 of 38 events occurred in the intermediate stratum (5.3%); and 28 of 38 occurred in the high-risk stratum (73.7%). The stepped rise in event rates supports using model probabilities for early stratification and monitoring intensity; given inadequate calibration slope, stratification should take priority over reading a single probability as absolute risk.

Feature contributions are summarized in Fig. 8. Pulmonary contusion, maximum displacement, displaced fracture count, and small pneumothorax contributed most to predictions, consistent with multivariable directions; main fracture zone and thoracic vertebral fracture ranked lower, the latter still limited by sparse samples. Exploratory subgroup AUCs are in Fig. 9; external subgroup sizes were limited and results are descriptive only, not confirmatory.

Table 6
External predicted-probability tertiles and observed event rates

Risk stratum	Observed rate	n	Events
Low	0.0%	39	0
Intermediate	5.3%	38	2
High	73.7%	38	28

4 Discussion

Placing displacement burden and early CT complications into one checkable set of items still separated high- and low-risk patients in a later-period sample [2, 4, 19]. The full model outperformed “small pneumothorax alone” and several displacement cut-point rules; removing pneumothorax reduced AUC only modestly, indicating that fracture and contusion information were not merely decorative. Pneumothorax remained in the final model, so near-circular concerns cannot be fully dismissed and can only be described as partly diluted. In external strata, the low-risk group had no events and the high-risk group exceeded 70% event rate—an absolute separation that speaks more directly to bedside monitoring allocation than a single AUC figure.

External calibration slope was about 0.63, and Hosmer–Lemeshow results were unfavorable. Adequate discrimination does not imply trustworthy probabilities [20]. A more cautious use is stratification and planning of follow-up and monitoring intensity; absolute probability claims require local recalibration [11,

12]. Tree models had external AUCs close to or slightly above logistic regression but did not stably improve sensitivity [13]; a nomogram remains more practical at the bedside [14]. External AUC slightly higher than internal testing may reflect case-mix differences such as higher external pneumothorax prevalence and event rate, and should not be read as general superiority; multicenter geographic validation is still needed [21]. Occult pneumothorax and small hemothorax are common after blunt trauma and their optimal management is debated [22]; when progression occurs, early thoracoscopic evacuation of retained hemothorax [23] and surgical stabilization of rib fractures may be considered, although the role of fixation outside flail chest remains under investigation [24, 25].

Several limitations should be acknowledged. First, this was a single-center retrospective study, and temporal external validation within the same institution cannot substitute for multicenter geographic external validation [26]. Second, the ordinal coding of the main fracture zone may oversimplify its relationship with the outcome and therefore requires cautious interpretation. Third, the effect estimate for thoracic vertebral fracture was unstable because of the limited number of patients with this finding. Fourth, the use of univariable screening followed by backward variable selection may have increased the risk of overfitting [27]. In addition, the sample size was not determined by an a priori formal calculation [28], and no formal risk-of-bias assessment was performed [30]. Further multicenter studies with larger and geographically independent cohorts, together with more robust validation frameworks [29], are warranted to confirm the generalizability and clinical utility of the model.

5 Conclusions

A model based on pulmonary contusion, maximum displacement, displaced fracture count, small pneumothorax, main fracture zone, and thoracic vertebral fracture showed good discrimination on internal testing and temporal external validation and outperformed several simple rules; calibration was limited. It is suitable as a reference for early stratification, not as an uncalibrated absolute individual-risk calculator.

Declarations

Ethics approval and consent to participate

This retrospective study was approved by the Ethics Committee of Huangshan People's Hospital. The requirement for informed consent was waived by the Ethics Committee because of the retrospective nature of the study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

LW designed the study, collected and analysed the data, and drafted the manuscript. ZX and YZ participated in data collection and validation. WF and WH supervised the study, interpreted the results, and critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Ziegler DW, Agarwal NN. The morbidity and mortality of rib fractures. *J Trauma*. 1994;37(6):975–9. 10.1097/00005373-199412000-00018.
2. Chapman BC, Herbert B, Rodil M, Salotto J, Stovall RT, Biffi W, et al. RibScore: a novel radiographic score based on fracture pattern that predicts pneumonia, respiratory failure, and tracheostomy. *J Trauma Acute Care Surg*. 2016;80(1):95–101. 10.1097/TA.0000000000000867.
3. Ahmad MA, Di Sante E, Giannoudis PV. Assessment of severity of chest trauma: is there an ideal scoring system? *Injury*. 2010;41(10):981–3. 10.1016/j.injury.2010.08.004.
4. Battle CE, Hutchings H, Evans PA. Risk factors that predict mortality in patients with blunt chest wall trauma: a systematic review and meta-analysis. *Injury*. 2012;43(1):8–17. 10.1016/j.injury.2011.01.004.
5. American College of Surgeons Committee on Trauma. ATLS. Advanced Trauma Life Support Student Course Manual. 10th ed. Chicago: American College of Surgeons; 2018.
6. Mowery NT, Gunter OL, Collier BR, Diaz JJ, Haut E, Hildreth A, et al. Practice management guidelines for management of hemothorax and occult pneumothorax. *J Trauma*. 2011;70(2):510–8.

- 10.1097/TA.0b013e31820b5c31.
7. Karmy-Jones R, Holevar M, Sullivan RJ, Fleisig A, Jurkovich GJ. Residual hemothorax after chest tube placement correlates with increased risk of empyema following traumatic injury. *Can Respir J*. 2008;15(5):255–8. 10.1155/2008/918951.
 8. Cohn SM. Pulmonary contusion: review of the clinical entity. *J Trauma*. 1997;42(5):973–9. 10.1097/00005373-199705000-00033.
 9. Huber S, Biberthaler P, Delhey P, Trentzsch H, Winter H, van Griensven M, et al. Predictors of poor outcomes after significant chest trauma in multiply injured patients: a retrospective analysis from the German Trauma Registry (TraumaRegister DGU). *Scand J Trauma Resusc Emerg Med*. 2014;22:52. 10.1186/s13049-014-0052-4.
 10. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55–63. 10.7326/M14-0697.
 11. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol*. 2016;74:167–76. 10.1016/j.jclinepi.2015.12.005.
 12. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–38. 10.1097/EDE.0b013e3181c30fb2.
 13. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22. 10.1016/j.jclinepi.2019.02.004.
 14. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015;16(4):e173–80. 10.1016/S1470-2045(14)71116-7.
 15. Baker SP, O'Neill B, Haddon W Jr, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma*. 1974;14(3):187–96. 10.1097/00005373-197403000-00001.
 16. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26(6):565–74. 10.1177/0272989X06295361.
 17. Bulger EM, Arneson MA, Mock CN, Jurkovich GJ. Rib fractures in the elderly. *J Trauma*. 2000;48(6):1040–7. 10.1097/00005373-200006000-00007.
 18. Deneuville M. Morbidity of percutaneous tube thoracostomy in trauma patients. *Eur J Cardiothorac Surg*. 2002;22(5):673–8. 10.1016/S1010-7940(02)00478-5.
 19. Livingston DH, Shogan B, John P, Lavery RF. CT diagnosis of rib fractures and the prediction of acute respiratory failure. *J Trauma*. 2008;64(4):905–11. 10.1097/TA.0b013e3181668ad7.
 20. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17(1):230. 10.1186/s12916-019-1466-7.

21. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925–31. 10.1093/eurheartj/ehu207.
22. Ball CG, Kirkpatrick AW, Laupland KB, Fox DL, Nicolaou S, Anderson IB, et al. Incidence, risk factors, and outcomes for occult pneumothoraces in victims of major trauma. *J Trauma*. 2005;59(4):917–25. 10.1097/01.ta.0000174663.46453.86.
23. Meyer DM, Jessen ME, Wait MA, Estrera AS. Early evacuation of traumatic retained hemothoraces using thoracoscopy: a prospective, randomized trial. *Ann Thorac Surg*. 1997;64(5):1396–401. 10.1016/S0003-4975(97)00899-0.
24. Pieracci FM, Agarwal S, Doben A, et al. A multicenter, prospective, controlled clinical trial of surgical stabilization of rib fractures in patients with severe, non-flail fracture patterns (Chest Wall Injury Society NONFLAIL). *J Trauma Acute Care Surg*. 2020;88(2):249–57. 10.1097/TA.0000000000002559.
25. Cataneo AJM, Cataneo DC, de Oliveira FH, Arruda KA, El Dib R, de Oliveira Carvalho PE. Surgical versus nonsurgical interventions for flail chest. *Cochrane Database Syst Rev*. 2015;7CD009919. 10.1002/14651858.CD009919.pub2.
26. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ*. 2016;353:i3140. 10.1136/bmj.i3140.
27. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–87. 10.1002/(SICI)1097-0258(19960229)15:4<3C361::AID-SIM168%3E3.0.CO;2-4.
28. Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441. 10.1136/bmj.m441.
29. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605. 10.1136/bmj.b605.
30. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51–8. 10.7326/M18-1376.

Figures

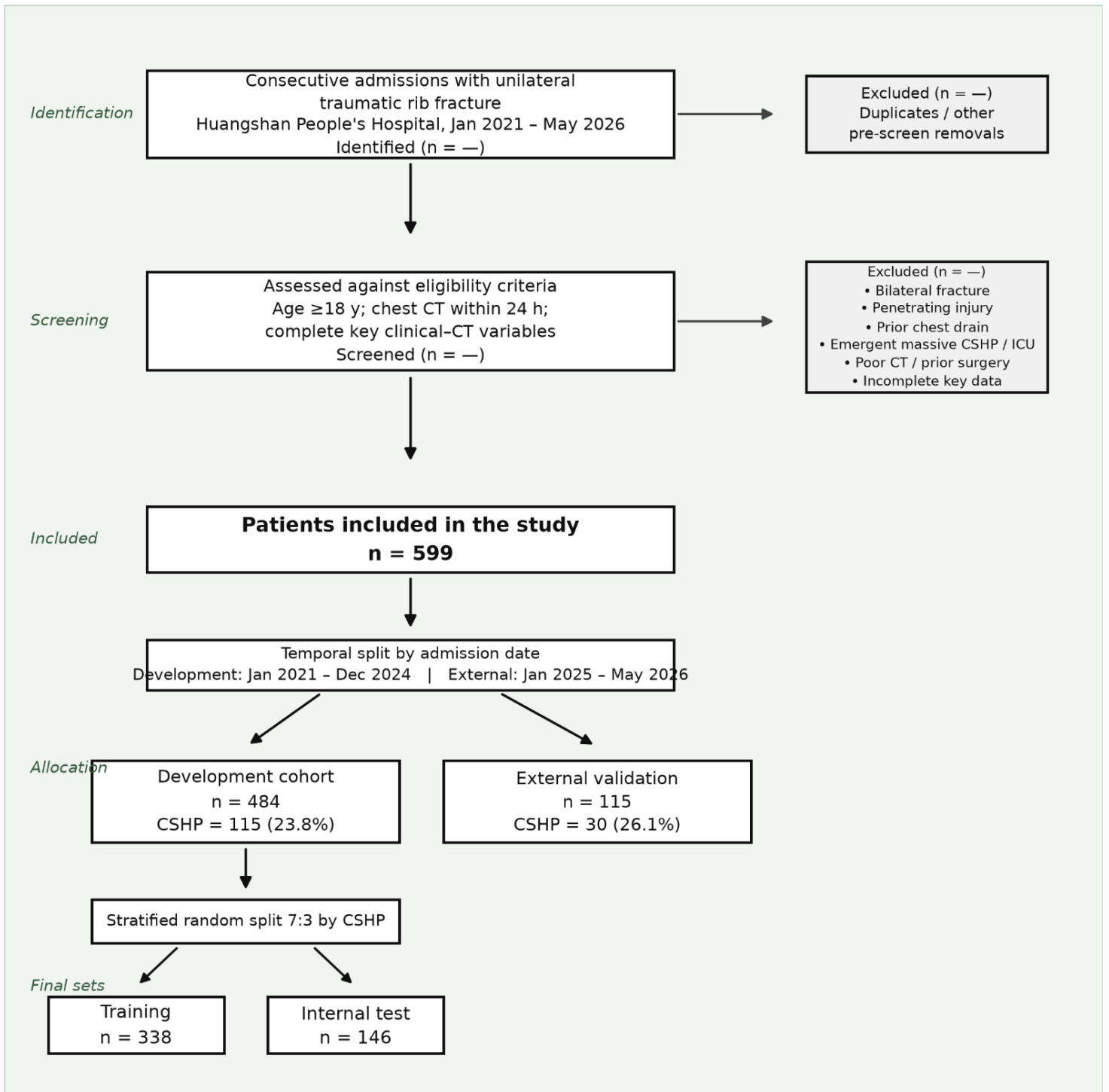


Figure 1

Study flowchart.

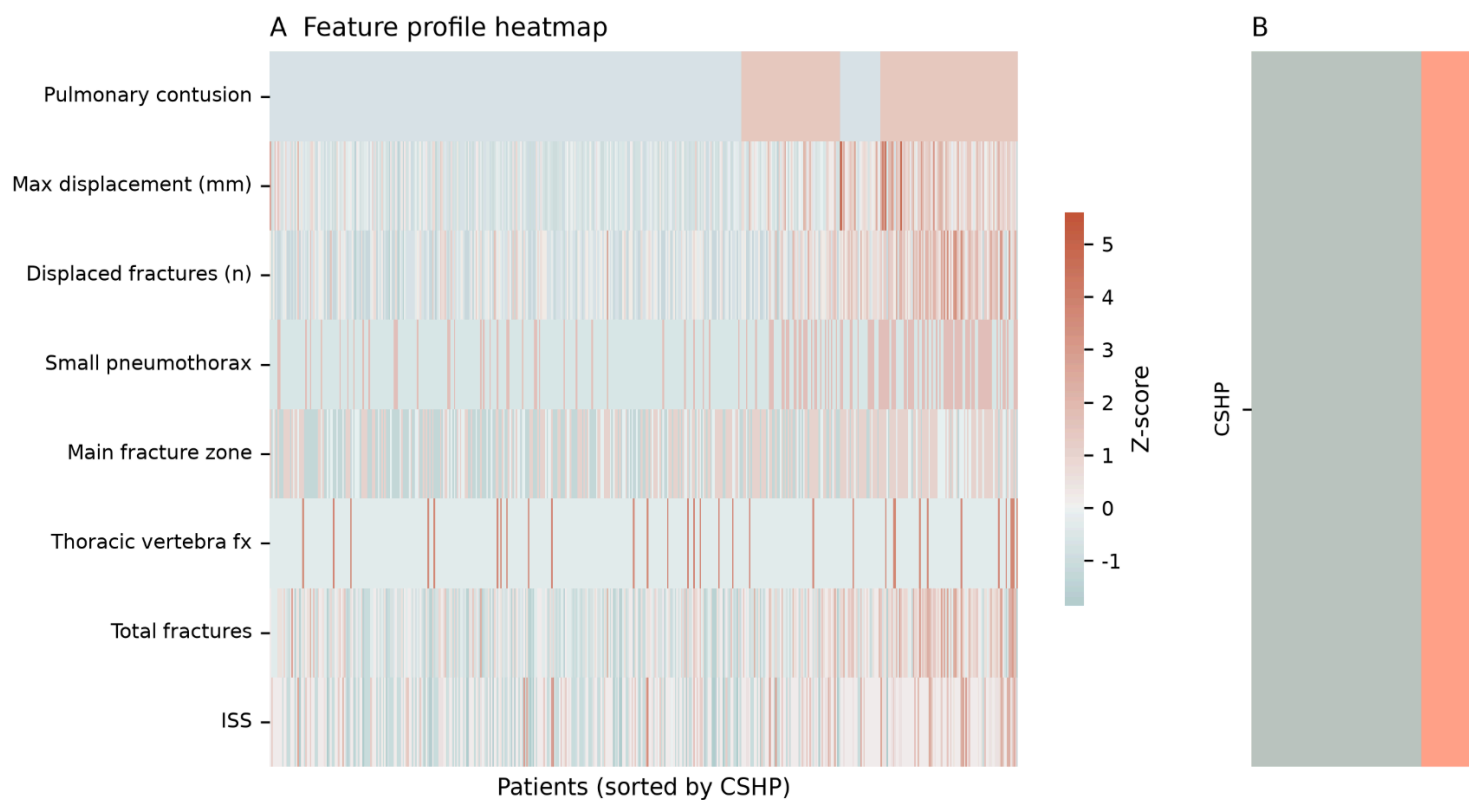


Figure 2

Clinical and CT feature distributions in the development cohort

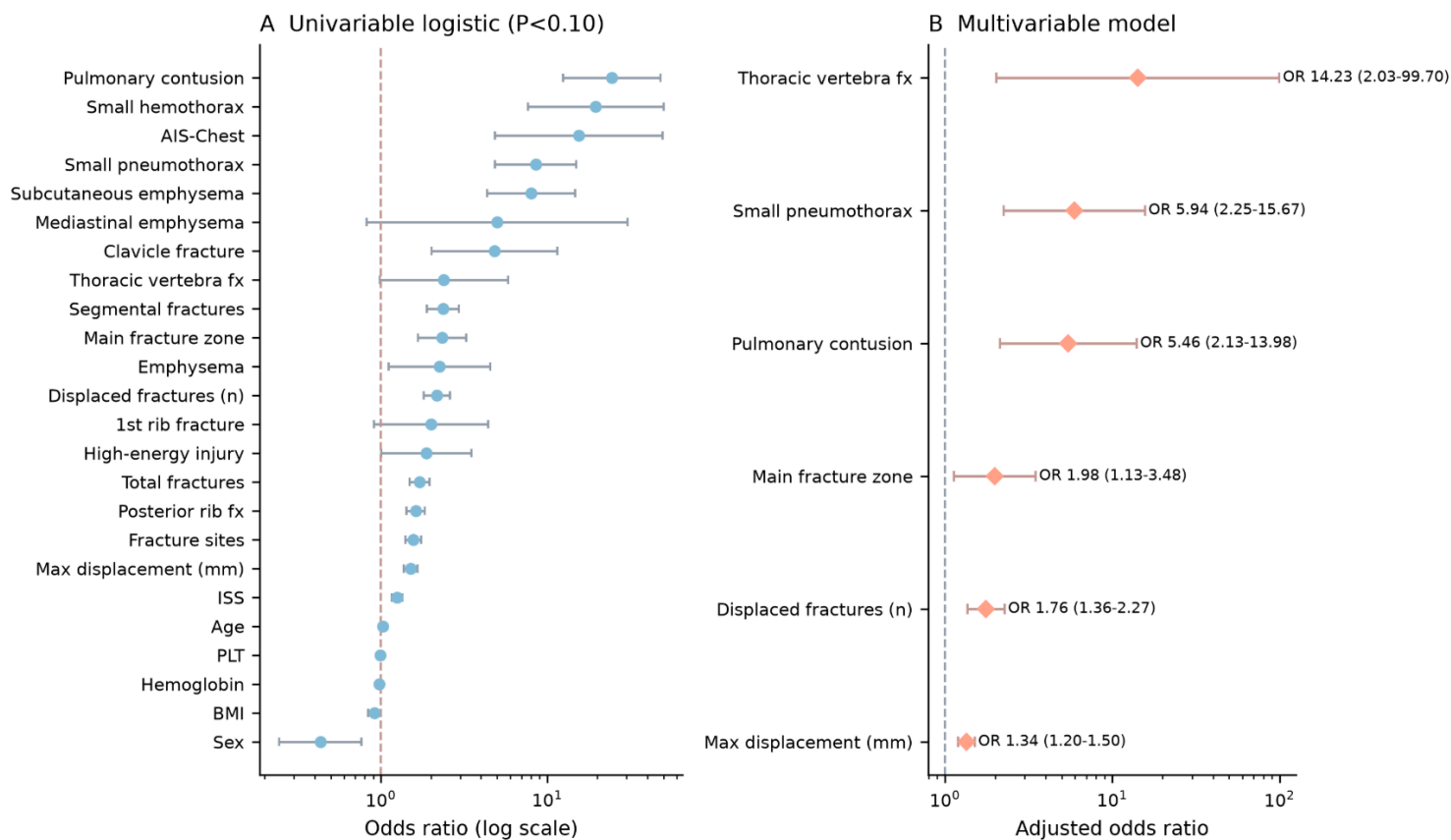


Figure 3

Univariable and multivariable logistic regression forest plots

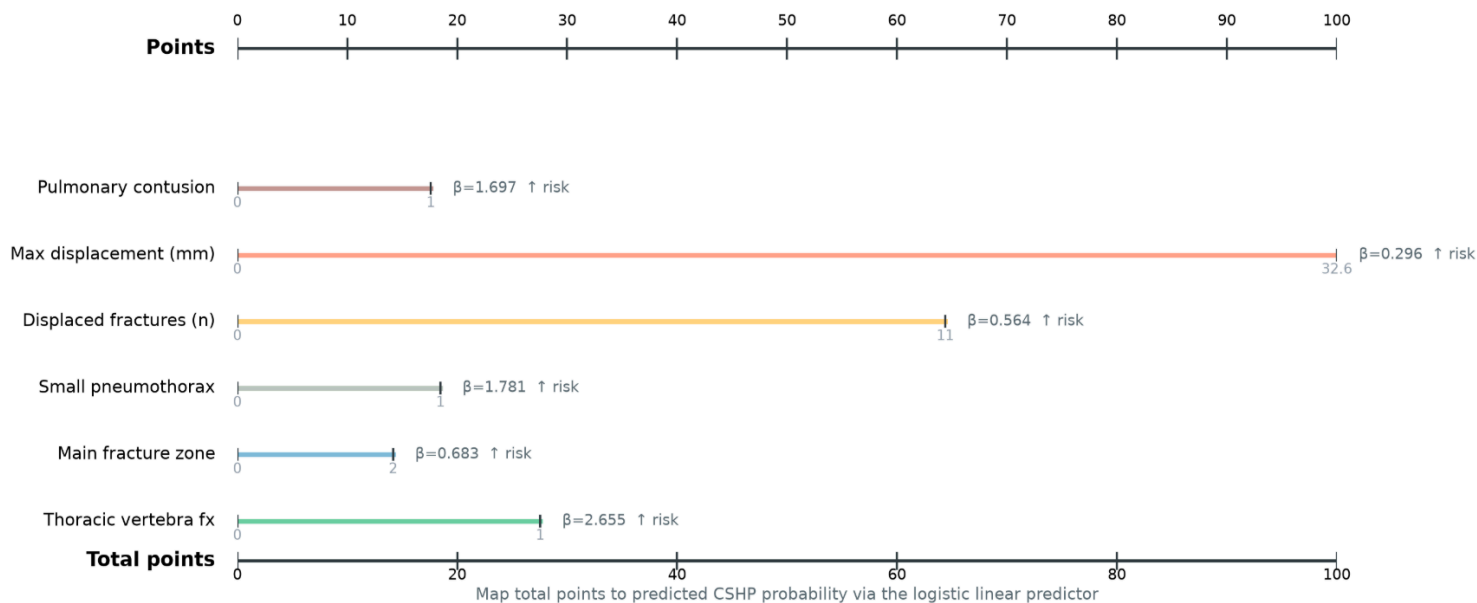


Figure 4

Nomogram-style score based on the final logistic model

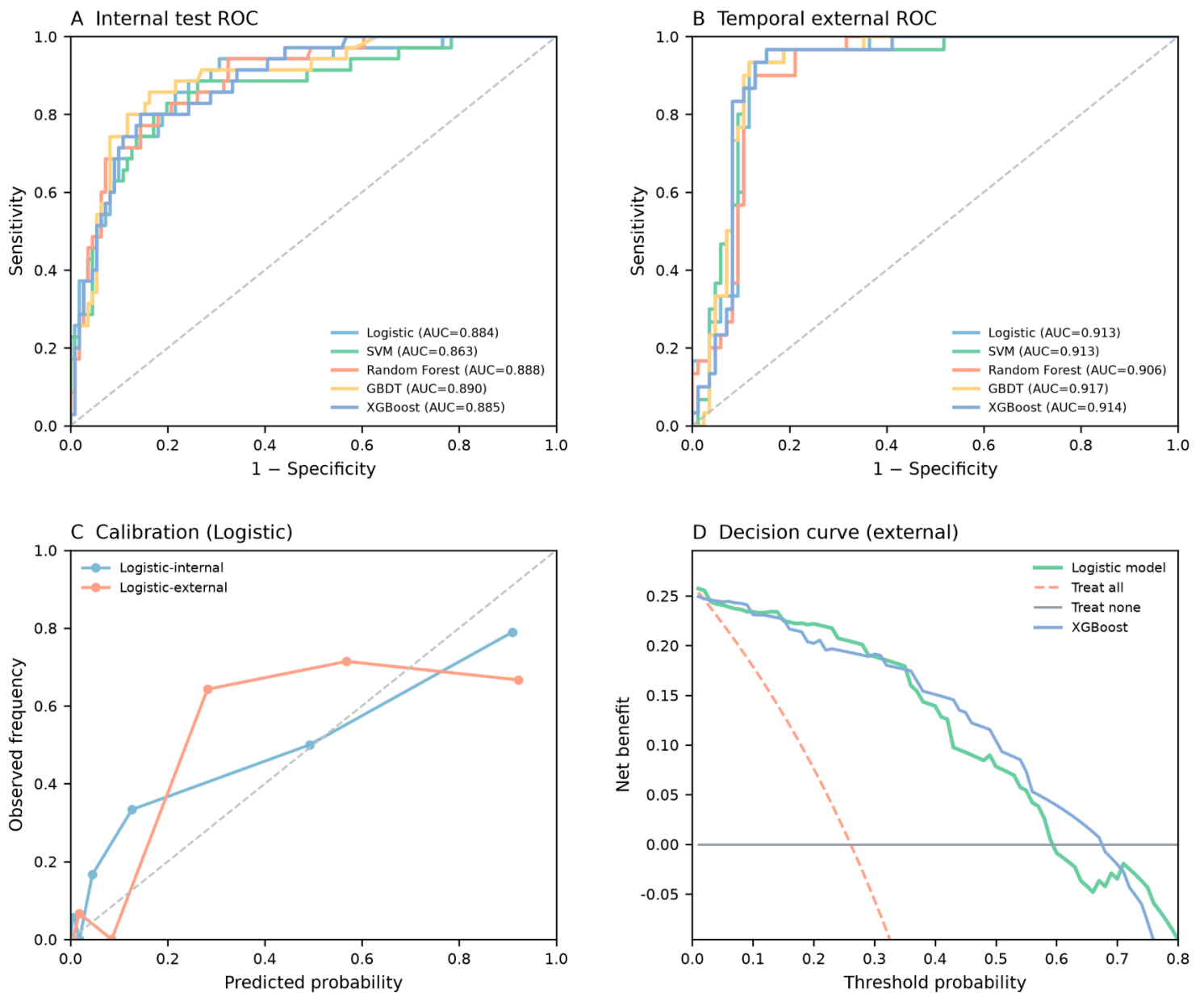


Figure 5

Discrimination, calibration curves, and external decision curve

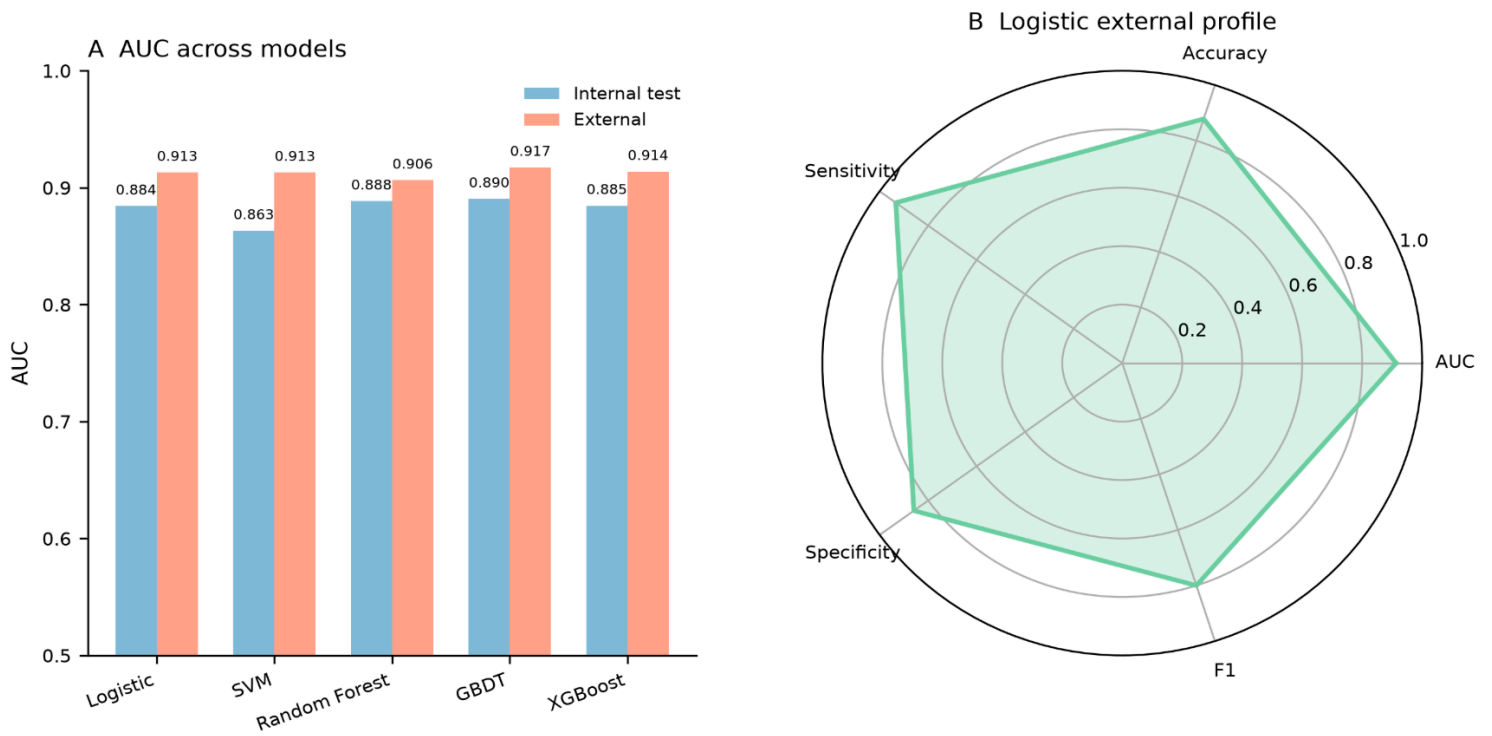


Figure 6

Multi-model AUC comparison and logistic external metric profile

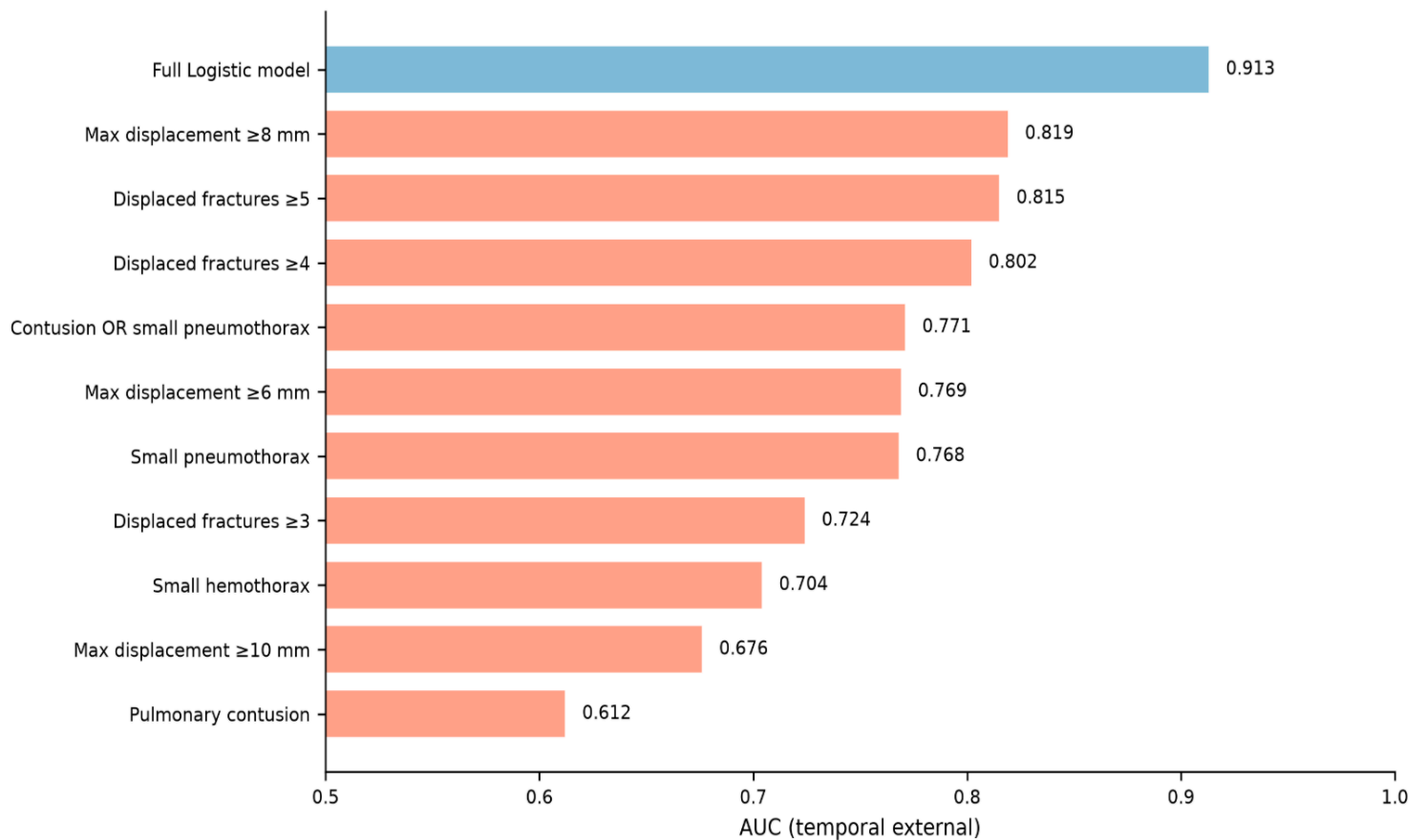


Figure 7

Temporal external validation: AUC of the full model versus simple clinical rules

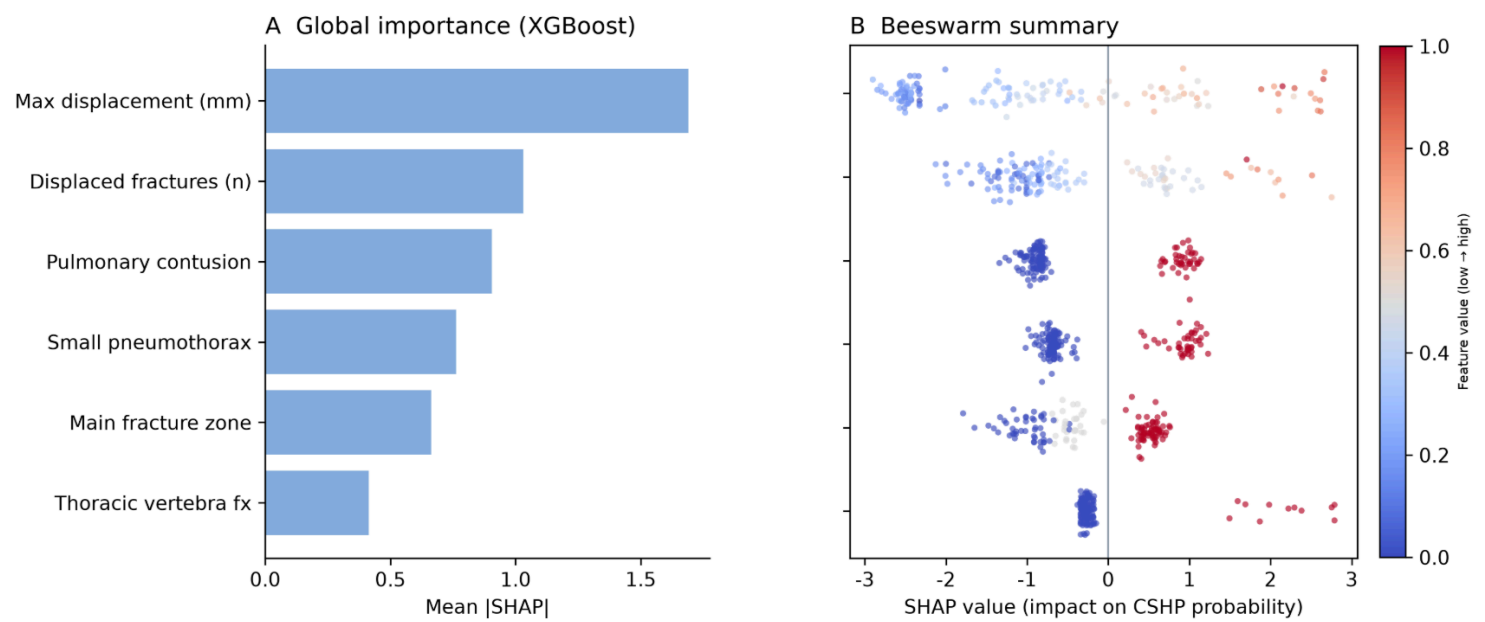


Figure 8

Summary of model feature contributions

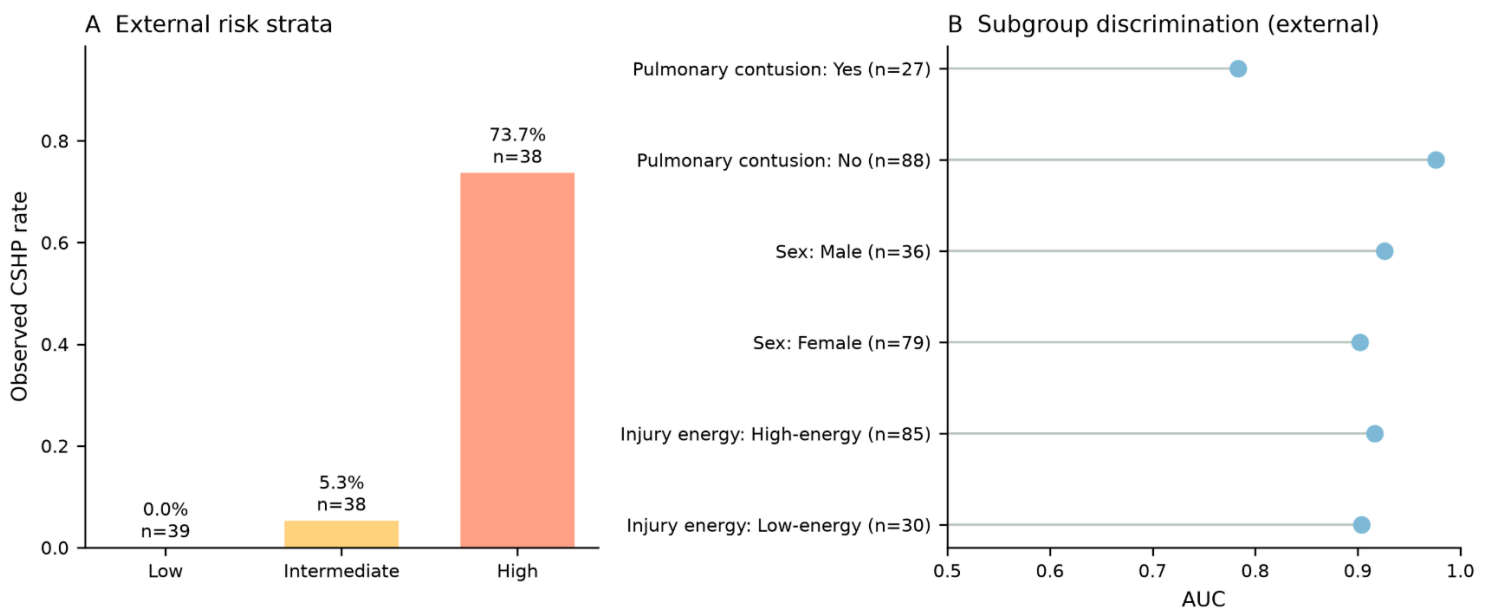


Figure 9

External risk stratification and exploratory subgroup performance