

# Development and Internal Validation of Prediction Models for Early CT- Defined Moderate-to-Large Pleural Effusion After Liver Transplantation: a Retrospective Cohort Study

**Xintao Chen**

State Key Laboratory of Ultrasound in Medicine and Engineering, Chongqing Medical University

**Maoling Qin**

Department of General Surgery, Bishan District Hospital of Traditional Chinese Medicine

**Lunwei Chen**

Department of Hepatobiliary Surgery, The Second Affiliated Hospital of Chongqing Medical University

**Xinhang Yu**

Department of Hepatobiliary Surgery, The Second Affiliated Hospital of Chongqing Medical University

**Wen Luo**

307762@hospital.cqyk.edu.cn

Department of Hepatobiliary Surgery, The Second Affiliated Hospital of Chongqing Medical University

---

## Research Article

**Keywords:** liver transplantation, pleural effusion, computed tomography, machine learning, prediction model, LASSO, internal validation, prognostic nutritional index

**Posted Date:** September 8th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-10627269/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

Pleural effusion is common after liver transplantation (LT), but early moderate-to-large effusions remain difficult to predict. We developed and internally validated models for a study-specific computed tomography (CT)-defined severe postoperative pleural effusion (SPPE) endpoint.

## Methods

We retrospectively studied LT recipients treated from June 2023 to January 2026. Patients with preoperative pleural effusion or unavailable postoperative chest imaging were excluded. SPPE was defined as a new effusion within 5 postoperative days with a maximum CT-estimated pleural-fluid depth  $\geq 4$  cm. The cohort was divided into a training set ( $n = 95$ ) and a held-out validation set ( $n = 32$ ). Four LASSO-selected predictors were used in five models. Conditional out-of-fold and 1000-replicate bootstrap validation kept predictors and hyperparameters fixed.

## Results

Among 127 recipients, 78 (61.4%) met the SPPE definition. Retained predictors were previous abdominal surgery and immediate postoperative prognostic nutritional index (PNI), C-reactive protein, and aspartate aminotransferase measured within 1 hour after LT. Logistic regression achieved the highest held-out AUC (0.658; 95% confidence interval, 0.453–0.864). Bootstrap-corrected AUCs ranged from 0.710 to 0.761. Random forest had the highest corrected AUC (0.761) but the greatest optimism (0.079), whereas logistic regression had a corrected AUC of 0.734 and the lowest optimism (0.026). PNI was the only predictor with a nominally significant post-selection coefficient (odds ratio, 0.868;  $P = 0.010$ ).

## Conclusions

Discrimination was modest, and model ranking depended on the validation method. No algorithm was consistently superior. Prospective external validation is required before clinical use.

## Background

Liver transplantation (LT) is the definitive treatment for selected patients with end-stage liver disease, acute liver failure, and certain unresectable liver malignancies. Despite advances in organ preservation, surgical technique, anesthesia, and postoperative critical care, early respiratory complications remain common and are associated with prolonged mechanical ventilation, longer intensive care unit (ICU) and hospital stays, and worse short-term outcomes [1, 9]. Pleural effusion is among the most frequently reported thoracic abnormalities after LT [1–6]. Its pathogenesis is multifactorial and may involve portal-hypertensive physiology, perioperative fluid and blood product administration, reduced plasma oncotic pressure, diaphragmatic or lymphatic disruption during hepatectomy, postoperative atelectasis, renal or graft dysfunction, and systemic inflammation [1–6, 39].

The clinical spectrum of post-LT pleural effusion is broad. Small collections may resolve without intervention, whereas moderate-to-large or persistent effusions can reduce functional lung volume, worsen dependent

atelectasis and hypoxemia, delay mobilization or liberation from respiratory support, and necessitate thoracentesis or pleural drainage [2–6]. Reported incidence estimates vary widely because studies have used different imaging schedules, observation windows, size classifications, drainage-based definitions, and thresholds for clinical relevance [2–6, 36–40]. Recent studies linking pleural effusion and malnutrition to adverse early outcomes, together with studies of persistent effusions or perioperative pleural drainage, suggest that effusion phenotype and management may have prognostic importance [3, 5, 6].

Previous studies have examined intraoperative determinants of postoperative pulmonary complications, developed a pulmonary complication score, and used machine learning to predict postoperative respiratory support requirements in LT recipients [7, 8, 10]. These efforts support the feasibility of structured perioperative prediction but do not specifically address an early CT-defined moderate-to-large pleural effusion endpoint. This distinction is important because respiratory support requirements and composite pulmonary complications reflect multiple mechanisms, whereas pleural effusion is a discrete radiologic phenotype that may require tailored surveillance and management.

Definitions of severe postoperative pleural effusion remain heterogeneous. Hepatectomy studies have generally classified symptomatic moderate-to-large effusions requiring chest drainage as severe, whereas one nomogram study defined severe effusion as an ultrasound-confirmed moderate-to-large collection with an estimated volume  $\geq 500$  mL [36–38]. In LT recipients, greater pleural-effusion depth has also been associated with the need for drainage and postoperative lung infection [40]. However, no universally accepted CT threshold defines severe pleural effusion after LT. Simple CT measurements can provide reproducible estimates of effusion size, but thresholds developed in other populations are not directly interchangeable with a post-transplant endpoint [12]. Contemporary pleural-disease guidance emphasizes integrating imaging with symptoms, physiology, etiology, and procedural indications rather than relying on size alone [13]. Accordingly, the cutoff used in this study should be regarded as an operational, study-specific radiologic definition rather than a validated universal severity classification.

Machine-learning algorithms can model nonlinear effects and interactions, but greater flexibility does not necessarily improve performance in small tabular clinical datasets. Prediction-model guidance emphasizes prespecification, adequate sample size, appropriate handling of missing data, rigorous internal validation, calibration assessment, and transparent evaluation of clinical utility [22–29]. Comparative evidence also indicates that machine-learning methods often do not outperform well-specified logistic regression when evaluated fairly [26]. These considerations are particularly important in single-center cohorts with limited sample sizes and small validation sets.

We therefore aimed to develop and internally validate models for early CT-defined SPPE after LT using routinely available perioperative and early postoperative variables. We compared multivariable logistic regression with random forest, radial-kernel support vector machine (SVM), extreme gradient boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM). We estimated training performance using conditional out-of-fold predictions; assessed held-out discrimination, overall probabilistic accuracy, graphical calibration, and decision-curve net benefit; complemented the split-sample evaluation with conditional 1000-replicate full-dataset bootstrap validation; and used Shapley additive explanations (SHAP) to characterize random forest predictions. As a secondary objective, we explored whether SPPE was associated with ICU treatment duration, postoperative hospital stay, and day-7 liver biochemical indices among early in-hospital survivors.

## Methods

### Study design, setting, and ethics

This retrospective cohort study was conducted at the Second Affiliated Hospital of Chongqing Medical University. Consecutive adult patients who underwent LT between June 2023 and January 2026 were screened. The report was prepared with reference to the TRIPOD + AI statement for clinical prediction models using regression or machine-learning methods [22]. Because all eligible patients treated during the study period were included, no prospective sample-size calculation was performed. Model complexity and interpretation were constrained by the available sample size [24, 25].

The study was approved by the Ethics Committee of the Second Affiliated Hospital of Chongqing Medical University (Scientific Ethics Pre-Review No. [2025] 361) and was conducted in accordance with the Declaration of Helsinki, national brain-death donation guidelines, and applicable Chinese laws. Organs were procured from controlled brain-dead donors and allocated through the China Organ Transplant Response System. No organs from executed prisoners were used.

### Participants and transplantation procedures

Patients were eligible if they underwent LT during the study period and were older than 18 years. Of 145 potentially eligible recipients, 15 were excluded because preoperative chest CT showed pleural effusion. Three additional recipients were excluded because they died within 3 days after LT before postoperative chest imaging could be obtained. The final analytic cohort comprised 127 recipients. The patient selection and cohort allocation process is shown in Additional file 1.

Transplantations were performed by an experienced liver transplant team. Conventional or piggyback techniques were selected according to recipient anatomy, clinical condition, and surgical judgment. Perioperative care followed institutional protocols.

### Imaging protocol and outcome definition

#### Imaging protocol and outcome definition

All included recipients underwent non-contrast chest CT and contrast-enhanced whole-abdominal CT within 40 hours before LT, followed by postoperative chest CT within 5 days after LT. Patients with any pleural effusion on preoperative chest CT were excluded so that the postoperative endpoint represented a newly developed effusion. Based on size terminology used in previous liver-surgery studies [36–38], the primary endpoint was early CT-defined SPPE. A newly detected postoperative effusion was classified as SPPE when the maximum CT-estimated pleural-fluid depth was  $\geq 4$  cm in either hemithorax. Symptoms, oxygen requirement, thoracentesis, and chest-tube drainage were not required. Thus, SPPE was a study-specific radiologic endpoint rather than a drainage-based clinical complication.

Postoperative CT images were assessed with patients in the supine position. For each examination used for outcome ascertainment, the reader identified the axial image showing the largest pleural-fluid cross-sectional area and measured the maximum fluid depth using electronic calipers in the hospital imaging system. If effusions were bilateral, the greater of the two measurements determined endpoint status. A single reader performed all

measurements to maintain procedural consistency. The 4-cm threshold was an institutional, study-specific criterion and was not assumed to represent a universally accepted CT severity grade [12, 13].

## Candidate predictors and timing

Candidate predictors were extracted from the electronic medical record and included demographic characteristics, smoking and drinking history, comorbidities, previous abdominal surgery, preoperative pulmonary condition, Child–Pugh and Model for End-stage Liver Disease (MELD) scores, preoperative laboratory variables, operative characteristics, transfusion and fluid variables, and immediate postoperative laboratory measurements obtained within 1 hour after LT. Patient and hospitalization identifiers were excluded from modeling. Body mass index (BMI) was dichotomized as  $< 24$  or  $\geq 24$  kg/m<sup>2</sup> according to the prespecified analysis code.

Preoperative CT examinations were completed within 40 hours before LT. All postoperative blood-test variables used as candidate predictors were obtained from urgent laboratory testing within 1 hour after LT and therefore preceded the postoperative CT outcome window. The prognostic nutritional index (PNI) was calculated as serum albumin (g/L) +  $5 \times$  absolute lymphocyte count ( $10^9$ /L). Although conventionally termed a nutritional index, immediate postoperative PNI was interpreted as a composite of albumin concentration and circulating lymphocyte count rather than a direct measure of nutritional intake.

## Cohort partition and descriptive analysis

Using a fixed random seed and stratification by outcome, the 127 recipients were randomly divided in a 3:1 ratio into a training cohort and a held-out internal validation cohort. Feature screening, imputation, encoding, hyperparameter tuning, and model fitting were performed exclusively in the training cohort. The held-out cohort was reserved for final internal performance assessment and was not used for predictor or hyperparameter selection.

Continuous variables were assessed for normality within outcome groups using the Shapiro–Wilk test. Normally distributed variables were summarized as mean  $\pm$  standard deviation and compared using Welch's two-sample t test. Non-normally distributed variables were summarized as median (25th percentile, 75th percentile) and compared using the Wilcoxon rank-sum test. Categorical variables were summarized as n (%) and compared using Pearson's chi-square test or Fisher's exact test, as appropriate. Screening was based on two-sided unadjusted P values; Benjamini–Hochberg-adjusted P values were calculated descriptively but did not determine variable entry into LASSO.

## Feature screening, LASSO selection, and logistic model

Candidate variables with a univariate P value  $\leq 0.05$  in the training cohort were entered into binomial least absolute shrinkage and selection operator (LASSO) regression [30]. No values were missing in the final analytic dataset. The preprocessing pipeline nonetheless prespecified training-derived median imputation for numeric variables and an explicit “Missing” category for categorical variables to prevent data leakage if missingness occurred. Categorical predictors were dummy encoded, zero-variance columns were removed, and predictors were standardized within the penalized model. Binomial deviance was evaluated using stratified 10-fold cross-validation. Both lambda.min and lambda.1se were displayed; the more parsimonious lambda.1se solution was prespecified for downstream modeling.

The original clinical variables represented by nonzero coefficients at  $\lambda_{1se}$  were entered into an unpenalized multivariable logistic regression model, with complete categorical predictors retained when a dummy-coded level had been selected. A nomogram was generated from the resulting four-predictor model. Odds ratios (ORs), Wald 95% confidence intervals (CIs), and P values were reported descriptively. The fitted logistic equation was also provided to support reproducibility. Because selection and coefficient estimation were performed in the same training data, these inferential estimates are post-selection and should be interpreted as nominal rather than confirmatory.

## Machine learning model development and tuning

Five models were developed using the same  $\lambda_{1se}$ -selected predictor set: logistic regression, random forest, radial-kernel SVM, XGBoost, and LightGBM [31–34]. All preprocessing parameters were estimated exclusively from the training data. Numeric predictors were median imputed, categorical predictors were dummy encoded, validation columns were aligned with the training design matrix, and zero-variance training columns were removed.

Hyperparameters for the four machine-learning algorithms were selected by stratified five-fold cross-validation in the training cohort, with mean area under the receiver operating characteristic curve (AUC) as the tuning criterion. When several configurations performed similarly, a one-standard-error rule favored simpler settings. To reduce overfitting, the candidate grids emphasized shallow trees, low learning rates, subsampling, and regularization. Final models were fitted to the complete training cohort using the selected hyperparameters.

The selected configurations were: random forest, 1200 trees,  $mtry = 1$ , minimum node size = 18, maximum depth = 2, and a 0.65 sampling fraction without replacement; radial SVM, cost = 0.01 and gamma = 0.0625; LightGBM, 2 leaves, maximum depth = 1, learning rate = 0.05, 15 boosting rounds, minimum 15 observations per leaf, feature fraction = 0.75, bagging fraction = 0.70, L1 penalty = 1, L2 penalty = 10, and minimum split gain = 0.20; and XGBoost, maximum depth = 1, learning rate = 0.02, minimum child weight = 5, gamma = 1, 20 boosting rounds, subsample = 0.70, column subsample = 0.75, L1 penalty = 1, and L2 penalty = 10.

## Model performance, bootstrap validation, and interpretability

To reduce resubstitution optimism, training-cohort predictions were generated by stratified five-fold out-of-fold refitting. The LASSO-selected feature set and hyperparameters chosen from the complete training cohort were held fixed during this step; therefore, these estimates were conditional OOF predictions rather than fully nested internal validation estimates. The held-out internal validation cohort was evaluated once using the corresponding final model fitted to the complete training cohort. Discrimination was quantified using the AUC, and overall probabilistic accuracy was summarized using the Brier score. Empirical receiver operating characteristic (ROC) curves were used to calculate AUCs and DeLong 95% CIs; binormal smoothing was applied only for graphical presentation.

Calibration was evaluated graphically by comparing mean predicted probabilities with observed event proportions in rank-based groups: nine groups in the 95-recipient training cohort and four groups in the 32-recipient validation cohort. Because the validation cohort was small, formal calibration intercept and slope estimates were not emphasized. Decision-curve analysis compared model net benefit with treat-all and treat-none strategies across threshold probabilities from 0.01 to 0.80 [29]. These analyses were exploratory because no specific clinical action was prespecified for a given SPPE risk threshold.

To complement the split-sample evaluation, 1000 bootstrap resamples were drawn from the full dataset of 127 recipients. The four selected predictors and tuned hyperparameters from the preceding analysis were held fixed, and each model was refitted in every bootstrap sample. For each resample, optimism was defined as the AUC in the bootstrap sample minus the AUC obtained when the bootstrap-fitted model was applied to the original full dataset. The optimism-corrected AUC was calculated as the apparent full-data AUC minus mean optimism. Corrected AUC distributions, the 2.5th and 97.5th percentiles, and optimism distributions were summarized. Because univariate screening, LASSO selection, and hyperparameter tuning were not repeated within each bootstrap sample, this analysis estimated conditional refitting optimism rather than the optimism of the entire model-development process.

Random forest predictions were interpreted using SHAP values [35]. Global importance was summarized as the mean absolute SHAP value on the predicted-probability scale in the held-out validation cohort, with bootstrap 95% CIs. Beeswarm, heatmap, and individual waterfall plots were generated. SHAP values were estimated using 500 Monte Carlo repetitions with local-accuracy adjustment and the training cohort as the background distribution. These explanations describe fitted-model behavior and do not establish causal effects.

## **Survivor-restricted secondary outcome analysis**

To provide clinical context for the radiologic endpoint, a survivor-restricted secondary analysis excluded seven recipients who died in hospital within 20 postoperative days. This restriction was applied because death can truncate ICU and postoperative hospital length-of-stay outcomes and can preclude or distort assessment of day-7 laboratory values. The resulting cohort comprised 120 recipients: 49 without SPPE and 71 with SPPE.

Secondary outcomes were ICU treatment duration, postoperative hospital stay, and serum alanine aminotransferase (ALT), AST, and total bilirubin on postoperative day 7. Because these variables were non-normally distributed, groups were compared using two-sided Wilcoxon rank-sum tests and summarized as median (interquartile range). Analyses were exploratory and unadjusted, with no correction for multiple comparisons. Because this analysis conditioned on early postoperative survival, it was interpreted separately from the primary prediction analysis and was not used to estimate the association between SPPE and mortality.

## **Statistical software**

Analyses were performed in R using packages including readxl, dplyr, glmnet, rms, pROC, ranger, e1071, xgboost, lightgbm, ggplot2, and fastshap, together with packages for resampling and graphics. A two-sided  $P < 0.05$  was used for descriptive hypothesis testing. No multiplicity correction was used for variable screening.

## **Use of generative artificial intelligence**

ChatGPT (OpenAI) was used solely for English-language polishing and improvement of readability. It was not used for study design, data collection, statistical analysis, interpretation of results, or generation of scientific conclusions. All AI-assisted edits were critically reviewed and approved by the authors, who take full responsibility for the final content of the manuscript.

## **Results**

### **Cohort characteristics and endpoint frequency**

Among 145 adult LT recipients screened, 15 were excluded because of preoperative pleural effusion, and three were excluded because they died within 3 days before postoperative chest imaging could be obtained. The final cohort comprised 127 recipients, of whom 78 (61.4%) met the study-specific CT-defined SPPE endpoint and 49 (38.6%) did not. Stratified random allocation yielded a training cohort of 95 recipients (58 SPPE and 37 non-SPPE) and a held-out internal validation cohort of 32 recipients (20 SPPE and 12 non-SPPE) (Additional file 1).

Demographic, clinical, operative, and laboratory characteristics are presented in Table 1. The relatively high endpoint frequency should be interpreted in the context of scheduled early CT assessment, exclusion of pre-existing effusions, and use of the study-specific 4-cm depth threshold. It should not be interpreted as the incidence of symptomatic or drainage-requiring pleural effusion in broader LT populations.

Table 1  
Baseline and perioperative characteristics by SPPE status in training and validation cohorts

| Variables                                       | Training cohort               |                               |                               |            | Held-out validation cohort |                          |                  |            |
|-------------------------------------------------|-------------------------------|-------------------------------|-------------------------------|------------|----------------------------|--------------------------|------------------|------------|
|                                                 | Overall<br>(n = 95)           | Non-<br>SPPE (n<br>= 37)      | SPPE (n<br>= 58)              | P<br>value | Overall<br>(n = 32)        | Non-<br>SPPE (n<br>= 12) | SPPE (n<br>= 20) | P<br>value |
| <b>Demographic and clinical characteristics</b> |                               |                               |                               |            |                            |                          |                  |            |
| Sex                                             |                               |                               |                               | 0.677      |                            |                          |                  | 0.204      |
| Female                                          | 21<br>(22.1%)                 | 9<br>(24.3%)                  | 12<br>(20.7%)                 |            | 8<br>(25.0%)               | 1 (8.3%)                 | 7<br>(35.0%)     |            |
| Male                                            | 74<br>(77.9%)                 | 28<br>(75.7%)                 | 46<br>(79.3%)                 |            | 24<br>(75.0%)              | 11<br>(91.7%)            | 13<br>(65.0%)    |            |
| Age, years                                      | 50.92 ±<br>10.39              | 49.78 ±<br>13.65              | 51.64 ±<br>7.68               | 0.455      | 48.62 ±<br>9.62            | 47.83 ±<br>10.70         | 49.10 ±<br>9.18  | 0.736      |
| BMI group                                       |                               |                               |                               | 0.993      |                            |                          |                  | 0.465      |
| < 24 kg/m <sup>2</sup>                          | 59<br>(62.1%)                 | 23<br>(62.2%)                 | 36<br>(62.1%)                 |            | 16<br>(50.0%)              | 5 (41.7%)                | 11<br>(55.0%)    |            |
| ≥ 24 kg/m <sup>2</sup>                          | 36<br>(37.9%)                 | 14<br>(37.8%)                 | 22<br>(37.9%)                 |            | 16<br>(50.0%)              | 7 (58.3%)                | 9<br>(45.0%)     |            |
| Height, cm                                      | 168.00<br>(160.00,<br>170.50) | 167.00<br>(160.00,<br>170.00) | 168.00<br>(161.25,<br>171.00) | 0.391      | 167.16 ±<br>7.95           | 168.92 ±<br>6.67         | 166.10 ±<br>8.62 | 0.310      |
| Weight, kg                                      | 62.64 ±<br>10.07              | 61.00 ±<br>11.35              | 63.69 ±<br>9.11               | 0.229      | 67.78 ±<br>15.03           | 70.08 ±<br>15.44         | 66.40 ±<br>15.00 | 0.516      |
| Smoking<br>history                              |                               |                               |                               | 0.298      |                            |                          |                  | 0.250      |
| No                                              | 58<br>(61.1%)                 | 25<br>(67.6%)                 | 33<br>(56.9%)                 |            | 21<br>(65.6%)              | 6 (50.0%)                | 15<br>(75.0%)    |            |
| Yes                                             | 37<br>(38.9%)                 | 12<br>(32.4%)                 | 25<br>(43.1%)                 |            | 11<br>(34.4%)              | 6 (50.0%)                | 5<br>(25.0%)     |            |
| Drinking<br>history                             |                               |                               |                               | 0.123      |                            |                          |                  | 0.473      |
| No                                              | 63<br>(66.3%)                 | 28<br>(75.7%)                 | 35<br>(60.3%)                 |            | 19<br>(59.4%)              | 6 (50.0%)                | 13<br>(65.0%)    |            |
| Yes                                             | 32<br>(33.7%)                 | 9<br>(24.3%)                  | 23<br>(39.7%)                 |            | 13<br>(40.6%)              | 6 (50.0%)                | 7<br>(35.0%)     |            |
| Diabetes<br>mellitus                            |                               |                               |                               | 0.288      |                            |                          |                  | 1.000      |
| No                                              | 80<br>(84.2%)                 | 33<br>(89.2%)                 | 47<br>(81.0%)                 |            | 27<br>(84.4%)              | 10<br>(83.3%)            | 17<br>(85.0%)    |            |
| Yes                                             | 15<br>(15.8%)                 | 4<br>(10.8%)                  | 11<br>(19.0%)                 |            | 5<br>(15.6%)               | 2 (16.7%)                | 3<br>(15.0%)     |            |

| Variables                                       | Training cohort              |                              |                              |            | Held-out validation cohort    |                               |                              |            |
|-------------------------------------------------|------------------------------|------------------------------|------------------------------|------------|-------------------------------|-------------------------------|------------------------------|------------|
|                                                 | Overall<br>(n = 95)          | Non-<br>SPPE (n<br>= 37)     | SPPE (n<br>= 58)             | P<br>value | Overall<br>(n = 32)           | Non-<br>SPPE (n<br>= 12)      | SPPE (n<br>= 20)             | P<br>value |
| <b>Demographic and clinical characteristics</b> |                              |                              |                              |            |                               |                               |                              |            |
| Previous abdominal surgery                      |                              |                              |                              | 0.021      |                               |                               |                              | 0.248      |
| No                                              | 61<br>(64.2%)                | 29<br>(78.4%)                | 32<br>(55.2%)                |            | 22<br>(68.8%)                 | 10<br>(83.3%)                 | 12<br>(60.0%)                |            |
| Yes                                             | 34<br>(35.8%)                | 8<br>(21.6%)                 | 26<br>(44.8%)                |            | 10<br>(31.2%)                 | 2 (16.7%)                     | 8<br>(40.0%)                 |            |
| Hypertension                                    |                              |                              |                              | 0.566      |                               |                               |                              | 0.053      |
| No                                              | 82<br>(86.3%)                | 31<br>(83.8%)                | 51<br>(87.9%)                |            | 27<br>(84.4%)                 | 8 (66.7%)                     | 19<br>(95.0%)                |            |
| Yes                                             | 13<br>(13.7%)                | 6<br>(16.2%)                 | 7<br>(12.1%)                 |            | 5<br>(15.6%)                  | 4 (33.3%)                     | 1 (5.0%)                     |            |
| Coronary heart disease                          |                              |                              |                              | 0.558      |                               |                               |                              | 1.000      |
| No                                              | 92<br>(96.8%)                | 35<br>(94.6%)                | 57<br>(98.3%)                |            | 32<br>(100.0%)                | 12<br>(100.0%)                | 20<br>(100.0%)               |            |
| Yes                                             | 3 (3.2%)                     | 2 (5.4%)                     | 1 (1.7%)                     |            | 0 (0.0%)                      | 0 (0.0%)                      | 0 (0.0%)                     |            |
| Preoperative pulmonary condition                |                              |                              |                              | 0.134      |                               |                               |                              | 1.000      |
| None                                            | 65<br>(68.4%)                | 29<br>(78.4%)                | 36<br>(62.1%)                |            | 20<br>(62.5%)                 | 8 (66.7%)                     | 12<br>(60.0%)                |            |
| Pneumonia                                       | 21<br>(22.1%)                | 7<br>(18.9%)                 | 14<br>(24.1%)                |            | 11<br>(34.4%)                 | 4 (33.3%)                     | 7<br>(35.0%)                 |            |
| Emphysema                                       | 9 (9.5%)                     | 1 (2.7%)                     | 8<br>(13.8%)                 |            | 1 (3.1%)                      | 0 (0.0%)                      | 1 (5.0%)                     |            |
| Child-Pugh score                                | 9.00<br>(7.00,<br>11.00)     | 9.00<br>(7.00,<br>11.00)     | 9.00<br>(6.25,<br>10.00)     | 0.303      | 9.00 ±<br>2.24                | 8.75 ±<br>2.09                | 9.15 ±<br>2.37               | 0.623      |
| MELD score                                      | 15.00<br>(9.00,<br>23.00)    | 15.00<br>(9.00,<br>26.00)    | 15.00<br>(9.25,<br>20.00)    | 0.258      | 18.47 ±<br>9.08               | 18.33 ±<br>8.78               | 18.55 ±<br>9.48              | 0.948      |
| <b>Preoperative laboratory variables</b>        |                              |                              |                              |            |                               |                               |                              |            |
| Preoperative creatinine, µmol/L                 | 65.20<br>(52.10,<br>77.10)   | 64.10<br>(50.60,<br>76.30)   | 66.10<br>(53.07,<br>77.92)   | 0.617      | 59.10<br>(53.00,<br>70.42)    | 62.40<br>(56.70,<br>70.72)    | 54.60<br>(49.80,<br>70.42)   | 0.321      |
| Preoperative eGFR, mL/min                       | 105.60<br>(89.25,<br>118.20) | 107.00<br>(89.50,<br>119.70) | 103.68<br>(88.27,<br>117.58) | 0.680      | 112.30<br>(102.17,<br>117.95) | 114.45<br>(107.05,<br>116.25) | 110.25<br>(99.22,<br>120.83) | 0.572      |

| Variables                                              | Training cohort         |                         |                         |         | Held-out validation cohort |                        |                         |         |
|--------------------------------------------------------|-------------------------|-------------------------|-------------------------|---------|----------------------------|------------------------|-------------------------|---------|
|                                                        | Overall (n = 95)        | Non-SPPE (n = 37)       | SPPE (n = 58)           | P value | Overall (n = 32)           | Non-SPPE (n = 12)      | SPPE (n = 20)           | P value |
| <b>Demographic and clinical characteristics</b>        |                         |                         |                         |         |                            |                        |                         |         |
| Preoperative red blood cell count, 10 <sup>12</sup> /L | 3.51 (2.91, 4.07)       | 3.54 (2.90, 4.05)       | 3.46 (2.93, 4.11)       | 0.828   | 3.43 (2.73, 4.12)          | 3.61 (3.11, 4.10)      | 3.30 (2.68, 4.14)       | 0.627   |
| Preoperative hemoglobin, g/L                           | 107.09 ± 22.40          | 105.68 ± 21.87          | 108.00 ± 22.88          | 0.621   | 105.69 ± 22.95             | 105.33 ± 23.47         | 105.90 ± 23.24          | 0.948   |
| Preoperative platelet count, 10 <sup>9</sup> /L        | 71.00 (42.50, 108.50)   | 65.00 (47.00, 95.00)    | 72.50 (42.25, 126.75)   | 0.374   | 65.00 (52.50, 94.00)       | 57.50 (51.50, 85.25)   | 67.00 (52.50, 94.25)    | 0.447   |
| Preoperative neutrophil count, 10 <sup>9</sup> /L      | 2.69 (1.65, 4.76)       | 2.69 (1.71, 5.28)       | 2.65 (1.65, 4.47)       | 0.807   | 2.62 (1.92, 4.38)          | 2.58 (1.45, 3.16)      | 2.95 (2.00, 6.26)       | 0.179   |
| Preoperative lymphocyte count, 10 <sup>9</sup> /L      | 0.69 (0.44, 1.14)       | 0.64 (0.44, 1.13)       | 0.71 (0.45, 1.14)       | 0.559   | 0.71 ± 0.40                | 0.60 ± 0.18            | 0.77 ± 0.49             | 0.174   |
| Preoperative monocyte count, 10 <sup>9</sup> /L        | 0.33 (0.21, 0.52)       | 0.32 (0.20, 0.49)       | 0.34 (0.24, 0.54)       | 0.577   | 0.38 (0.25, 0.54)          | 0.35 (0.25, 0.46)      | 0.39 (0.27, 0.66)       | 0.533   |
| Preoperative albumin, g/L                              | 34.40 (31.15, 40.55)    | 35.00 (30.70, 40.10)    | 34.15 (31.52, 40.75)    | 0.728   | 35.46 ± 5.42               | 36.42 ± 4.24           | 34.89 ± 6.05            | 0.410   |
| Preoperative C-reactive protein, mg/L                  | 7.37 (5.00, 20.97)      | 8.09 (5.00, 20.75)      | 6.53 (5.00, 20.77)      | 0.446   | 11.93 (5.52, 20.56)        | 12.12 (5.00, 18.37)    | 11.93 (6.15, 22.10)     | 0.570   |
| Preoperative total bilirubin, µmol/L                   | 46.10 (17.95, 155.10)   | 46.10 (17.80, 310.40)   | 45.30 (18.45, 103.80)   | 0.329   | 74.25 (28.23, 285.80)      | 47.00 (39.22, 193.82)  | 84.20 (25.05, 322.70)   | 1.000   |
| Preoperative prothrombin time, s                       | 16.60 (14.35, 21.20)    | 17.20 (14.40, 23.60)    | 16.40 (14.33, 19.82)    | 0.139   | 18.55 (14.90, 22.62)       | 18.40 (16.02, 21.38)   | 18.70 (14.90, 22.62)    | 0.953   |
| Preoperative SII                                       | 320.19 (161.32, 532.61) | 340.36 (206.07, 590.42) | 271.63 (159.88, 502.77) | 0.601   | 399.50 (184.24, 711.48)    | 285.10 (85.55, 526.65) | 495.15 (216.25, 782.38) | 0.192   |
| Preoperative CLR                                       | 14.71 (6.82, 35.12)     | 16.72 (8.08, 43.09)     | 14.02 (6.45, 31.33)     | 0.384   | 17.48 (9.81, 44.22)        | 17.03 (10.79, 29.68)   | 19.84 (9.14, 71.13)     | 0.741   |
| Preoperative NLR                                       | 3.74 (2.29, 6.60)       | 4.32 (2.64, 6.56)       | 3.52 (2.20, 6.62)       | 0.284   | 5.25 (2.60, 9.11)          | 3.87 (2.25, 6.57)      | 6.64 (2.73, 11.12)      | 0.284   |
| Preoperative PLR                                       | 102.78 (65.06, 166.49)  | 96.30 (64.52, 138.39)   | 112.19 (66.01, 173.42)  | 0.604   | 92.58 (72.25, 204.85)      | 105.74 (70.01, 170.74) | 91.24 (78.39, 237.92)   | 0.770   |

| Variables                                                                    | Training cohort                  |                                  |                                  |            | Held-out validation cohort       |                                  |                                  |            |
|------------------------------------------------------------------------------|----------------------------------|----------------------------------|----------------------------------|------------|----------------------------------|----------------------------------|----------------------------------|------------|
|                                                                              | Overall<br>(n = 95)              | Non-<br>SPPE (n<br>= 37)         | SPPE (n<br>= 58)                 | P<br>value | Overall<br>(n = 32)              | Non-<br>SPPE (n<br>= 12)         | SPPE (n<br>= 20)                 | P<br>value |
| <b>Demographic and clinical characteristics</b>                              |                                  |                                  |                                  |            |                                  |                                  |                                  |            |
| Preoperative SIRI                                                            | 1.21<br>(0.58,<br>2.78)          | 1.22<br>(0.58,<br>3.05)          | 1.17<br>(0.59,<br>2.61)          | 0.792      | 2.18<br>(0.72,<br>4.30)          | 1.46<br>(0.61,<br>2.81)          | 2.23<br>(0.96,<br>5.76)          | 0.302      |
| Preoperative PNI                                                             | 39.92 ±<br>7.38                  | 39.31 ±<br>6.89                  | 40.31 ±<br>7.71                  | 0.513      | 38.99 ±<br>5.52                  | 39.42 ±<br>3.78                  | 38.74 ±<br>6.43                  | 0.708      |
| Preoperative CRP-to-albumin ratio                                            | 0.23<br>(0.14,<br>0.58)          | 0.25<br>(0.14,<br>0.69)          | 0.19<br>(0.14,<br>0.56)          | 0.614      | 0.41<br>(0.16,<br>0.55)          | 0.35<br>(0.16,<br>0.49)          | 0.43<br>(0.18,<br>0.57)          | 0.546      |
| Preoperative SIRI-P                                                          | 102.46<br>(41.22,<br>266.19)     | 84.74<br>(46.59,<br>296.44)      | 105.19<br>(29.73,<br>218.35)     | 0.888      | 151.41<br>(35.58,<br>391.88)     | 131.45<br>(27.86,<br>200.57)     | 231.53<br>(64.54,<br>563.45)     | 0.235      |
| <b>Intraoperative characteristics</b>                                        |                                  |                                  |                                  |            |                                  |                                  |                                  |            |
| Surgical technique                                                           | 0.504                            |                                  |                                  |            | 0.626                            |                                  |                                  |            |
| Conventional                                                                 | 80<br>(84.2%)                    | 30<br>(81.1%)                    | 50<br>(86.2%)                    |            | 27<br>(84.4%)                    | 11<br>(91.7%)                    | 16<br>(80.0%)                    |            |
| Piggyback                                                                    | 15<br>(15.8%)                    | 7<br>(18.9%)                     | 8<br>(13.8%)                     |            | 5<br>(15.6%)                     | 1 (8.3%)                         | 4<br>(20.0%)                     |            |
| Operation time, min                                                          | 422.00<br>(390.00,<br>500.00)    | 407.00<br>(380.00,<br>510.00)    | 435.00<br>(397.75,<br>494.25)    | 0.299      | 454.00 ±<br>78.09                | 430.33 ±<br>67.24                | 468.20 ±<br>82.26                | 0.168      |
| Anhepatic phase time, min                                                    | 66.00<br>(54.00,<br>78.00)       | 61.00<br>(54.00,<br>76.00)       | 68.00<br>(55.25,<br>78.00)       | 0.390      | 61.22 ±<br>16.28                 | 62.67 ±<br>9.09                  | 60.35 ±<br>19.56                 | 0.653      |
| Intraoperative fluid volume, mL                                              | 3850.00<br>(3300.00,<br>4700.00) | 3800.00<br>(3250.00,<br>4700.00) | 3875.00<br>(3312.50,<br>4700.00) | 0.703      | 3875.00<br>(3262.50,<br>4400.00) | 3725.00<br>(2893.75,<br>4325.00) | 3875.00<br>(3412.50,<br>4800.00) | 0.459      |
| Intraoperative urine output, mL                                              | 1905.00<br>(1387.50,<br>2725.00) | 1800.00<br>(1375.00,<br>2660.00) | 2050.00<br>(1400.00,<br>2775.00) | 0.485      | 2487.34<br>±<br>1197.31          | 2360.42<br>±<br>1017.29          | 2563.50<br>±<br>1312.84          | 0.629      |
| Blood loss, mL                                                               | 800.00<br>(600.00,<br>1000.00)   | 800.00<br>(600.00,<br>1000.00)   | 800.00<br>(600.00,<br>1000.00)   | 0.801      | 800.00<br>(575.00,<br>1000.00)   | 650.00<br>(500.00,<br>900.00)    | 800.00<br>(600.00,<br>1000.00)   | 0.455      |
| Intraoperative plasma transfusion, U                                         | 1.50<br>(0.00,<br>2.00)          | 1.50<br>(0.00,<br>2.00)          | 1.25<br>(0.00,<br>2.00)          | 0.703      | 1.25<br>(0.00,<br>2.00)          | 0.00<br>(0.00,<br>1.62)          | 2.00<br>(0.00,<br>2.25)          | 0.243      |
| Intraoperative red blood cell transfusion, U                                 | 3.00<br>(0.00,<br>5.00)          | 3.00<br>(0.00,<br>5.00)          | 3.00<br>(0.00,<br>5.00)          | 0.920      | 3.05 ±<br>2.46                   | 2.83 ±<br>2.47                   | 3.17 ±<br>2.51                   | 0.710      |
| <b>Immediate postoperative laboratory variables (within 1 hour after LT)</b> |                                  |                                  |                                  |            |                                  |                                  |                                  |            |

| Variables                                       | Training cohort              |                               |                               |            | Held-out validation cohort    |                                |                               |            |
|-------------------------------------------------|------------------------------|-------------------------------|-------------------------------|------------|-------------------------------|--------------------------------|-------------------------------|------------|
|                                                 | Overall<br>(n = 95)          | Non-<br>SPPE (n<br>= 37)      | SPPE (n<br>= 58)              | P<br>value | Overall<br>(n = 32)           | Non-<br>SPPE (n<br>= 12)       | SPPE (n<br>= 20)              | P<br>value |
| <b>Demographic and clinical characteristics</b> |                              |                               |                               |            |                               |                                |                               |            |
| Activated partial thromboplastin time, s        | 64.40<br>(52.20, 77.75)      | 70.60<br>(53.40, 86.00)       | 62.35<br>(51.57, 74.73)       | 0.255      | 66.20<br>(52.52, 83.90)       | 62.35<br>(51.27, 96.45)        | 66.85<br>(58.00, 78.85)       | 0.683      |
| Prothrombin activity, %                         | 38.99 ± 12.55                | 39.49 ± 10.82                 | 38.67 ± 13.62                 | 0.748      | 35.78 ± 10.64                 | 38.92 ± 11.71                  | 33.90 ± 9.77                  | 0.227      |
| C-reactive protein, mg/L                        | 49.66<br>(21.30, 87.62)      | 74.38<br>(38.76, 104.99)      | 32.31<br>(9.42, 64.31)        | 0.002      | 38.14<br>(13.71, 81.80)       | 25.57<br>(10.22, 79.44)        | 47.19<br>(22.18, 81.80)       | 0.508      |
| Hemoglobin, g/L                                 | 80.62 ± 15.07                | 77.78 ± 15.21                 | 82.43 ± 14.82                 | 0.147      | 83.94 ± 16.81                 | 82.17 ± 16.50                  | 85.00 ± 17.32                 | 0.649      |
| Albumin, g/L                                    | 39.26 ± 4.75                 | 40.94 ± 4.41                  | 38.19 ± 4.68                  | 0.005      | 38.05 ± 5.62                  | 39.82 ± 4.67                   | 36.98 ± 5.98                  | 0.147      |
| Neutrophil count, 10 <sup>9</sup> /L            | 7.85<br>(4.96, 10.07)        | 7.30<br>(4.68, 9.45)          | 8.39<br>(5.37, 10.47)         | 0.207      | 8.29<br>(6.31, 12.29)         | 9.07<br>(7.02, 12.29)          | 7.94<br>(6.10, 11.94)         | 0.521      |
| Lymphocyte count, 10 <sup>9</sup> /L            | 0.29<br>(0.17, 0.45)         | 0.33<br>(0.20, 0.47)          | 0.26<br>(0.16, 0.40)          | 0.290      | 0.36<br>(0.26, 0.45)          | 0.33<br>(0.19, 0.41)           | 0.36<br>(0.31, 0.45)          | 0.572      |
| Monocyte count, 10 <sup>9</sup> /L              | 0.35<br>(0.24, 0.59)         | 0.44<br>(0.25, 0.63)          | 0.32<br>(0.23, 0.54)          | 0.134      | 0.37<br>(0.24, 0.58)          | 0.45<br>(0.25, 0.60)           | 0.33<br>(0.24, 0.54)          | 0.586      |
| Platelet count, 10 <sup>9</sup> /L              | 51.00<br>(34.50, 75.50)      | 45.00<br>(37.00, 64.00)       | 54.00<br>(32.25, 85.50)       | 0.461      | 51.00<br>(31.75, 69.25)       | 52.00<br>(30.50, 59.50)        | 51.00<br>(34.25, 81.50)       | 0.520      |
| D-dimer                                         | 2030.60<br>(954.85, 4229.95) | 1887.80<br>(1101.50, 3489.10) | 2190.10<br>(906.17, 4482.57)  | 0.746      | 2841.65<br>(906.70, 7357.75)  | 6726.10<br>(2218.90, 10000.00) | 2138.00<br>(906.70, 4844.50)  | 0.143      |
| AST, U/L                                        | 1516.00<br>(758.00, 2635.50) | 1012.00<br>(671.00, 1645.00)  | 1849.50<br>(1000.00, 3469.75) | < 0.001    | 2014.50<br>(1119.50, 2769.25) | 1748.00<br>(1119.50, 2309.50)  | 2117.00<br>(1056.75, 3578.75) | 0.755      |
| ALT, U/L                                        | 729.00<br>(402.50, 1351.00)  | 619.00<br>(398.00, 856.00)    | 819.50<br>(492.75, 1753.00)   | 0.019      | 649.00<br>(493.50, 960.50)    | 674.00<br>(507.00, 942.50)     | 646.00<br>(435.25, 1028.00)   | 0.683      |
| Total bilirubin, µmol/L                         | 72.60<br>(40.05, 132.35)     | 79.40<br>(45.80, 213.30)      | 66.05<br>(40.02, 103.55)      | 0.143      | 98.55<br>(62.00, 165.18)      | 80.00<br>(61.62, 170.95)       | 100.95<br>(74.75, 165.18)     | 0.712      |
| eGFR, mL/min                                    | 95.10<br>(64.45, 109.70)     | 106.00<br>(89.80, 111.40)     | 85.35<br>(64.12, 103.65)      | 0.064      | 91.49 ± 35.17                 | 84.47 ± 32.08                  | 95.71 ± 37.04                 | 0.374      |
| SII                                             | 1219.77<br>(660.13,          | 1101.60<br>(493.30,           | 1407.09<br>(771.76,           | 0.110      | 1396.20<br>(850.47,           | 1274.70<br>(848.84,            | 1482.04<br>(863.07,           | 0.683      |

| Variables                                       | Training cohort                |                                |                                |            | Held-out validation cohort    |                                |                               |            |
|-------------------------------------------------|--------------------------------|--------------------------------|--------------------------------|------------|-------------------------------|--------------------------------|-------------------------------|------------|
|                                                 | Overall<br>(n = 95)            | Non-<br>SPPE (n<br>= 37)       | SPPE (n<br>= 58)               | P<br>value | Overall<br>(n = 32)           | Non-<br>SPPE (n<br>= 12)       | SPPE (n<br>= 20)              | P<br>value |
| <b>Demographic and clinical characteristics</b> |                                |                                |                                |            |                               |                                |                               |            |
|                                                 | 2546.40)                       | 2030.74)                       | 2998.39)                       |            | 1877.13)                      | 1652.58)                       | 1899.20)                      |            |
| CLR                                             | 142.20<br>(58.38,<br>300.29)   | 217.24<br>(130.68,<br>351.50)  | 102.42<br>(46.94,<br>255.72)   | 0.013      | 121.11<br>(35.47,<br>298.08)  | 134.58<br>(24.63,<br>366.29)   | 104.59<br>(65.44,<br>251.28)  | 0.922      |
| NLR                                             | 24.63<br>(16.56,<br>37.77)     | 22.95<br>(12.65,<br>28.37)     | 25.61<br>(17.38,<br>44.05)     | 0.097      | 25.25<br>(17.02,<br>36.99)    | 28.45<br>(20.64,<br>36.99)     | 21.73<br>(16.45,<br>35.07)    | 0.471      |
| PLR                                             | 173.08<br>(107.06,<br>297.37)  | 123.08<br>(85.71,<br>325.00)   | 186.25<br>(133.33,<br>292.15)  | 0.163      | 138.92<br>(81.89,<br>235.73)  | 116.39<br>(71.99,<br>284.38)   | 150.56<br>(86.58,<br>225.91)  | 0.892      |
| SIRI                                            | 9.14<br>(4.58,<br>17.49)       | 8.17<br>(4.20,<br>18.34)       | 10.83<br>(4.66,<br>17.35)      | 0.356      | 9.29<br>(5.23,<br>14.67)      | 6.63<br>(3.96,<br>18.46)       | 9.78<br>(6.16,<br>14.18)      | 0.922      |
| PNI                                             | 41.05 ±<br>4.88                | 42.94 ±<br>4.65                | 39.84 ±<br>4.67                | 0.002      | 40.33 ±<br>4.92               | 41.82 ±<br>4.05                | 39.44 ±<br>5.27               | 0.164      |
| CRP-to-<br>albumin ratio                        | 1.25<br>(0.54,<br>2.27)        | 1.90<br>(0.91,<br>2.50)        | 0.85<br>(0.25,<br>1.75)        | 0.006      | 0.94<br>(0.37,<br>2.25)       | 0.62<br>(0.26,<br>2.05)        | 1.22<br>(0.59,<br>2.25)       | 0.425      |
| SIRI-P                                          | 498.41<br>(204.66,<br>1050.97) | 420.86<br>(199.52,<br>1040.51) | 507.22<br>(274.94,<br>1062.36) | 0.694      | 408.65<br>(256.05,<br>931.39) | 374.22<br>(303.22,<br>1070.15) | 414.51<br>(206.24,<br>789.82) | 0.861      |

**Note:** SPPE was defined as a newly developed moderate-to-large pleural effusion detected within 5 postoperative days, operationalized as a maximum CT-estimated pleural-fluid depth  $\geq 4$  cm in either hemithorax. Data are presented as mean  $\pm$  standard deviation, median (25th percentile, 75th percentile), or n (%), as appropriate. P values compare the SPPE and non-SPPE groups within each cohort and are descriptive and unadjusted.

Normally distributed continuous variables were compared using Welch's two-sample t test; non-normally distributed variables were compared using the Wilcoxon rank-sum test. Categorical variables were compared using Pearson's chi-square test or Fisher's exact test, as appropriate.

All postoperative laboratory variables were obtained within 1 hour after liver transplantation.

**Abbreviations:** ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; CLR, C-reactive protein-to-lymphocyte ratio; CRP, C-reactive protein; CT, computed tomography; eGFR, estimated glomerular filtration rate; LT, liver transplantation; MELD, Model for End-stage Liver Disease; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; PNI, prognostic nutritional index; SII, systemic immune-inflammation index; SIRI, systemic inflammation response index; SIRI-P, SIRI-related platelet index; SPPE, severe postoperative pleural effusion.

## Univariate screening

Eight variables met the univariate screening criterion in the training cohort: previous abdominal surgery (P = 0.021), immediate postoperative CRP (P = 0.002), albumin (P = 0.005), AST (P < 0.001), ALT (P = 0.019), CRP-to-lymphocyte ratio (P = 0.013), PNI (P = 0.002), and CRP-to-albumin ratio (P = 0.006). Screening was based on unadjusted P values. In an exploratory Benjamini–Hochberg adjustment, only immediate postoperative AST remained below 0.05 (adjusted P = 0.048), whereas PNI and CRP each had an adjusted P of 0.051.

Recipients with SPPE had a higher frequency of previous abdominal surgery (44.8% vs 21.6%) and a lower mean immediate postoperative PNI ( $39.84 \pm 4.67$  vs  $42.94 \pm 4.65$ ). Median immediate postoperative AST was higher in the SPPE group (1849.50 vs 1012.00 U/L), whereas immediate postoperative CRP, the CRP-to-lymphocyte ratio, and the CRP-to-albumin ratio were lower. These unadjusted comparisons were used solely for data-driven screening and should not be interpreted as causal associations.

## LASSO selection and nomogram

In 10-fold cross-validation, lambda.min was 0.0236 and retained five nonzero coefficients, whereas lambda.1se was 0.0868 and retained four. The prespecified lambda.1se solution selected previous abdominal surgery and immediate postoperative PNI, CRP, and AST (Fig. 1A and 1B). The resulting four-predictor logistic model was visualized as a nomogram (Fig. 1C).

Multivariable logistic regression results are shown in Table 2. Previous abdominal surgery had an OR of 2.397 (95% CI, 0.818–7.020; P = 0.111). The ORs per one-unit increase were 0.868 for PNI (95% CI, 0.779–0.966; P = 0.010) and 0.991 for CRP (95% CI, 0.980–1.002; P = 0.092). For immediate postoperative AST, the OR was 1.565 per 1000 U/L increase (95% CI, 0.993–2.466; P = 0.053). PNI was the only predictor with a nominal P value < 0.05 in the post-selection refit. The wide CI for previous abdominal surgery and the borderline estimates for CRP and AST indicate limited precision.

Table 2  
Multivariable logistic regression of LASSO-selected predictors in the training cohort

| Variables                                        | $\beta$ | Standard error | Wald z | OR (95% CI)         | P value |
|--------------------------------------------------|---------|----------------|--------|---------------------|---------|
| Previous abdominal surgery (Yes vs No)           | 0.874   | 0.548          | 1.594  | 2.397 (0.818–7.020) | 0.111   |
| Immediate postoperative PNI                      | -0.142  | 0.055          | -2.586 | 0.868 (0.779–0.966) | 0.010   |
| Immediate postoperative C-reactive protein, mg/L | -0.009  | 0.006          | -1.683 | 0.991 (0.980–1.002) | 0.092   |
| Immediate postoperative AST, per 1000 U/L        | 0.448   | 0.232          | 1.931  | 1.565 (0.993–2.466) | 0.053   |

**Note:** The reference category for previous abdominal surgery was No. Odds ratios for PNI and CRP represent the effect per one-unit increase; the immediate postoperative AST odds ratio represents the effect per 1000 U/L increase. Because coefficients were estimated after univariate screening and LASSO selection, confidence intervals and P values are nominal post-selection estimates.

**Abbreviations:** AST, aspartate aminotransferase; CI, confidence interval; CRP, C-reactive protein; LASSO, least absolute shrinkage and selection operator; OR, odds ratio; PNI, prognostic nutritional index.

Using previous abdominal surgery coded as 1 for Yes and 0 for No, the fitted linear predictor was:  $\text{logit}(P[\text{SPPE}]) = 5.841 + 0.874 \times \text{previous abdominal surgery} - 0.142 \times \text{immediate postoperative PNI} - 0.00948 \times \text{immediate postoperative CRP (mg/L)} + 0.000448 \times \text{immediate postoperative AST (U/L)}$ .

## Discrimination, decision curves, calibration, and bootstrap internal validation

Using conditional out-of-fold training predictions, AUCs were 0.776 (95% CI, 0.681–0.872) for logistic regression, 0.763 (0.668–0.857) for random forest, 0.707 (0.600–0.814) for SVM, 0.727 (0.625–0.829) for XGBoost, and 0.656 (0.545–0.768) for LightGBM (Fig. 2A). In the held-out internal validation cohort, the corresponding AUCs were 0.658 (0.453–0.864), 0.642 (0.426–0.858), 0.654 (0.450–0.858), 0.608 (0.401–0.816), and 0.552 (0.333–0.771), respectively (Fig. 2D). No complex algorithm clearly outperformed logistic regression, and the wide, overlapping validation CIs indicated substantial uncertainty in model ranking.

Decision curves suggested that model-based strategies might provide net benefit over treat-all or treat-none approaches across portions of the examined threshold range, but no model showed uniformly superior benefit in the held-out cohort (Fig. 2B and 2E). Calibration plots showed visible deviations and instability, particularly in the 32-patient validation cohort and at probability extremes (Fig. 2C and 2F). These findings do not support claims of robust calibration or immediate threshold-based clinical utility.

Validation-cohort Brier scores were similar across models, ranging from 0.224 to 0.232 (logistic, 0.224; random forest, 0.229; SVM, 0.224; LightGBM, 0.231; XGBoost, 0.232), indicating no clear separation in overall probabilistic accuracy.

In the conditional 1000-replicate full-dataset bootstrap analysis, apparent AUCs were 0.761 for logistic regression, 0.841 for random forest, 0.764 for SVM, 0.765 for LightGBM, and 0.765 for XGBoost. The corresponding optimism-corrected AUCs were 0.734, 0.761, 0.726, 0.710, and 0.711 (Fig. 2G–I). Mean optimism (SD) was 0.026 (0.041), 0.079 (0.028), 0.038 (0.038), 0.055 (0.037), and 0.054 (0.037), respectively. Random forest showed both the highest corrected AUC and the greatest optimism; the corrected distributions overlapped, and model ranking differed from that in the held-out cohort. These results support modest, validation-dependent discrimination but do not establish a consistently superior algorithm.

## Random forest SHAP analysis

In the held-out internal validation cohort, mean absolute SHAP values ranked immediate postoperative PNI first (0.069), followed by AST (0.058), CRP (0.057), and previous abdominal surgery (0.041) (Fig. 3A). The beeswarm plot showed that lower PNI, higher AST, lower CRP, and previous abdominal surgery generally shifted random forest predictions toward higher risk, although the magnitude of the contributions varied across patients (Fig. 3B). The risk-ordered heatmap further illustrated substantial patient-level heterogeneity across the four features (Fig. 3C).

For Patient 103, who met the SPPE endpoint, the waterfall plot showed that immediate postoperative CRP, PNI, previous abdominal surgery, and AST contributed approximately + 0.040, + 0.057, + 0.059, and + 0.060, respectively, shifting the random forest estimate from a background probability of 0.612 to 0.828 (Fig. 3D). This local explanation describes one fitted prediction and does not validate a patient-specific clinical decision.

## Survivor-restricted secondary outcomes

After seven recipients who died in hospital within 20 postoperative days were excluded, 120 recipients remained: 49 without SPPE and 71 with SPPE. ICU treatment duration and postoperative hospital stay did not differ significantly between groups ( $P = 0.526$  and  $P = 0.440$ , respectively). Day-7 AST and total bilirubin were also similar ( $P = 0.343$  and  $P = 0.662$ ), whereas day-7 ALT was nominally higher in the SPPE group ( $P = 0.049$ ) (Fig. 4).

Because five outcomes were examined without multiplicity correction and the ALT  $P$  value was close to 0.05, this isolated result should be regarded as hypothesis-generating. Moreover, the analysis was restricted to early in-hospital survivors and therefore does not characterize the overall association between SPPE and mortality or the complete postoperative course.

## Discussion

### Principal findings

In this single-center study of 127 adult LT recipients, LASSO retained four features for a study-specific early CT-defined moderate-to-large pleural effusion endpoint: previous abdominal surgery and immediate postoperative PNI, CRP, and AST measured within 1 hour after LT. Logistic regression achieved the highest held-out AUC, but discrimination was modest and imprecisely estimated. In the conditional 1000-replicate full-dataset bootstrap analysis, corrected AUCs ranged from 0.710 to 0.761; random forest had the highest corrected AUC but also the greatest optimism, whereas logistic regression showed the smallest optimism. The change in model ranking across validation approaches provided no consistent evidence that a complex algorithm was superior. Overall, the study identifies a measurable but limited early postoperative signal rather than a deployment-ready algorithm.

The decline from moderate conditional OOF training AUCs to lower held-out performance highlights the distinction between model development and transportable prediction. Broad held-out CIs included values close to chance, and the 32-patient validation sample was highly sensitive to a small number of discordant predictions. The OOF estimates reduced direct resubstitution bias but remained conditional on features and hyperparameters selected using the complete training cohort and therefore were not fully nested estimates. Bootstrap correction used all 127 recipients and quantified refitting optimism more efficiently, but it produced a different model ranking and was also conditional on the previously selected predictors and hyperparameters. The held-out and bootstrap findings should therefore be interpreted together as evidence of uncertain, validation-dependent performance.

### Relation to previous post-transplant pulmonary prediction studies

The present endpoint differs from the composite pulmonary complications, pneumonia, or respiratory support outcomes evaluated in previous LT prediction studies. Wang and colleagues developed a pulmonary complication score in 325 recipients using separate development and validation cohorts [8]. More recently, a machine-learning application predicted postoperative respiratory support requirements in 1121 recipients and was evaluated in a larger, temporally separated cohort [10]. These studies addressed clinically important respiratory outcomes, but their composite targets reflect multiple pulmonary processes and are not directly comparable with a CT depth-based pleural effusion endpoint.

Nevertheless, our findings are consistent with the broader prediction-model literature: greater algorithmic complexity does not guarantee better generalization. Systematic comparisons have frequently found no clear

performance advantage of machine learning over logistic regression after accounting for study quality and validation methods [26]. Random forest achieved the highest bootstrap-corrected AUC, but it also had the largest mean optimism, its corrected advantage over logistic regression was modest, and logistic regression ranked first in the held-out cohort. With four selected predictors and only 95 training observations, the nonlinear methods had limited information from which to learn stable interactions. Strong regularization and shallow architectures reduced overfitting but could not create signal absent from the predictors. The differing rankings across validation approaches suggest that any algorithmic advantage was not reproducible in this sample.

## Interpretation of the selected features

Immediate postoperative PNI was the most consistent selected feature, with the only nominally significant coefficient in the logistic refit and the largest mean absolute SHAP value in the random forest. Nutritional impairment, sarcopenia, and reduced muscle strength are increasingly recognized as determinants of adverse outcomes and resource use before and after LT [5, 14–18]. However, PNI measured within 1 hour after transplantation requires cautious interpretation. At this time point, albumin is influenced by pre-existing hepatic and nutritional reserve, crystalloid administration, transfusion, exogenous albumin, blood loss, redistribution, and capillary leak. Lymphocyte count may reflect chronic immunonutritional status, operative stress, corticosteroid exposure, and rapid leukocyte redistribution. Immediate postoperative PNI should therefore be viewed as a composite of oncotic, inflammatory, immune, and dilutional status rather than a pure nutritional measure.

A lower immediate postoperative PNI may identify reduced plasma oncotic reserve and stress-associated lymphopenia, both of which could favor extravascular fluid accumulation or indicate greater physiologic vulnerability. This interpretation is compatible with prior observations linking malnutrition and pleural effusion to adverse early LT outcomes [5], but the present study cannot determine whether PNI is causal, reflects treatment intensity, or summarizes unmeasured illness severity. Future studies should examine albumin and lymphocyte count separately, account for albumin administration, transfusion, and cumulative fluid balance, and compare preoperative, immediate postoperative, and longitudinal PNI values.

Immediate postoperative AST was retained by LASSO, was borderline in the logistic refit, and ranked second by mean absolute SHAP value. Early AST elevation after graft reperfusion may reflect hepatocellular injury, ischemia-reperfusion stress, graft quality, hemodynamic instability, or extrahepatic tissue injury [19, 20]. However, a measurement within 1 hour is not equivalent to established early allograft dysfunction criteria, which incorporate later aminotransferase, bilirubin, and coagulation measurements [19]. To improve clinical interpretability, the logistic result is reported per 1000 U/L increase. Future models should evaluate nonlinear transformations rather than assume linearity across the observed range.

The CRP findings were counterintuitive: CRP was lower in the SPPE group in the univariate analysis and had an inverse, nonsignificant coefficient after adjustment. CRP does not rise instantaneously in response to an inflammatory insult at the end of surgery; transplant and postoperative biomarker studies indicate that changes evolve over subsequent hours to days [16, 21]. A value obtained within 1 hour after LT may therefore reflect preoperative inflammation, recipient hepatic synthetic function, donor-graft behavior, hemodilution, transfusion, and perioperative treatment more than a new postoperative inflammatory response. The inverse association should not be interpreted as protective and is more plausibly timing-dependent, confounded, or unstable. Serial CRP trajectories may be more biologically informative.

Previous abdominal surgery may act as a proxy for adhesions, altered anatomy, prolonged dissection, diaphragmatic manipulation, bleeding, transfusion, or greater operative complexity. These mechanisms could contribute to pleural fluid formation through inflammatory, lymphatic, and fluid-balance pathways. However, the estimate was imprecise (OR, 2.397; 95% CI, 0.818–7.020), and information on adhesion grade, reoperation type, and operative complexity was unavailable. Previous abdominal surgery should therefore be regarded as a hypothesis-generating marker rather than an established independent risk factor.

## Clinical and radiologic interpretation of the endpoint

The 61.4% endpoint frequency exceeds that in many clinically defined or drainage-based reports and must be interpreted within the study design. Previous hepatectomy studies generally required a symptomatic moderate-to-large effusion and chest drainage, or used an ultrasound-estimated volume  $\geq 500$  mL, to define severe effusion [36–38]. By contrast, our endpoint classified a new moderate-to-large effusion as SPPE solely on the basis of a maximum CT-estimated depth  $\geq 4$  cm within 5 days; symptoms and intervention were not required. The observed proportion therefore represents the frequency of this specific imaging phenotype in the analyzed cohort, not the universal incidence of clinically severe or drainage-requiring pleural effusion after LT.

In the supine position, pleural fluid redistributes posteriorly, and a single axial depth is only a surrogate for three-dimensional volume. Although simple CT measurements can improve size estimation [12], the 4-cm cutoff has not been validated against respiratory mechanics, drainage volume, or patient-centered outcomes in LT recipients. Contemporary pleural guidance does not define severity using one CT depth alone [13]. A standardized single-reader protocol improved within-study consistency but precluded assessment of interobserver reliability. Future studies should prespecify slice selection and measurement orientation, use at least two blinded readers, quantify intra- and interobserver agreement, and relate CT measurements to ultrasound estimates, oxygenation, ventilation, drainage, and recovery.

The model is not intended to determine whether thoracentesis or chest-tube drainage is required. Such decisions depend on symptoms, gas exchange, effusion laterality and progression, suspected infection or bleeding, mechanical ventilation, coagulation status, and procedural risk. Following validation, a model of this type might support intensified surveillance or earlier bedside lung ultrasonography rather than automatic intervention. Point-of-care thoracic ultrasonography is well suited to serial assessment because it can evaluate pleural fluid, lung aeration, and congestion without repeated transport or radiation [11].

## Association with the early postoperative course among survivors

The survivor-restricted analysis provides a clinical counterpoint to the radiologic prediction findings. Among the 120 recipients remaining after the exclusion of seven early in-hospital deaths, SPPE was not associated with longer ICU treatment or postoperative hospitalization, and day-7 AST and total bilirubin did not differ significantly. These results suggest that, among early survivors, a moderate-to-large CT-detected effusion does not necessarily translate into prolonged resource use or generalized graft biochemical dysfunction. The endpoint may encompass a broad spectrum, from clinically tolerated fluid accumulation to effusions requiring active management, whereas ICU and discharge timing are influenced by numerous graft, infectious, renal, surgical, and social factors.

The nominally higher day-7 ALT in the SPPE group may reflect ongoing hepatocellular stress, shared perioperative injury, or greater overall physiologic burden. However, ALT was one of five unadjusted outcomes, its P value was 0.049, and neither day-7 AST nor total bilirubin differed. The result therefore lacks consistency across liver

biochemical markers and may represent chance, residual confounding, or selection effects rather than a direct biological relationship.

Interpretation is further constrained by conditioning on survival. The secondary cohort retained 71 of 78 SPPE recipients and all 49 non-SPPE recipients, indicating that the seven excluded early in-hospital deaths occurred in the SPPE group. This pattern is clinically important, but the secondary analysis was neither designed nor powered to evaluate mortality. Excluding deaths can bias length-of-stay and laboratory recovery comparisons toward survivors with more favorable courses. Accordingly, the absence of longer ICU or hospital stay among survivors should not be interpreted as evidence that SPPE is benign or unrelated to adverse outcomes in the full cohort. Future studies should evaluate mortality, discharge, pleural intervention, and respiratory support using prespecified time-to-event or competing-risk methods.

## **Model performance, calibration, and clinical utility**

Bootstrap correction clarified the extent to which apparent full-data discrimination reflected optimism. Random forest decreased from an apparent AUC of 0.841 to a corrected AUC of 0.761, whereas logistic regression decreased from 0.761 to 0.734. This pattern is consistent with greater model flexibility producing larger apparent gains but also greater overfitting. However, the bootstrap procedure did not repeat the entire feature-selection and tuning pipeline, so the corrected values should not be interpreted as unbiased estimates of external performance or as proof that random forest is preferable.

Discrimination alone is insufficient for clinical implementation. A model may rank patients reasonably well while systematically overestimating or underestimating absolute risk [27, 28]. In the validation cohort, grouped calibration curves were unstable, and calibration intercept and slope estimates were not emphasized because the sample was too small for precise estimation. The high endpoint prevalence also means that calibration and decision thresholds may change substantially across centers with different CT schedules, surgical practices, fluid-management strategies, or endpoint definitions.

Decision-curve analysis explored whether risk-based strategies might provide net benefit compared with treating all or none [29]. However, decision curves require a clinically meaningful action at each threshold. In this study, no downstream action was prespecified; possible responses ranged from ultrasound surveillance and respiratory physiotherapy to repeat CT, whereas invasive drainage carries a different benefit–harm balance. The curves should therefore be interpreted as methodological exploration rather than evidence that the model should direct a specific intervention.

SHAP analysis improved transparency by showing how the random forest used each feature at the global and individual levels [35]. It also illustrated heterogeneous and potentially bidirectional contributions that a single OR cannot fully describe. Nevertheless, SHAP values inherit the uncertainty, confounding, and predictor correlations of the fitted model. Rankings may be unstable in a 32-patient validation sample, and local explanations can appear precise even when predicted probabilities are poorly calibrated. Model explanation is not causal explanation and should not be used to infer that modifying a feature will alter risk.

## **Strengths and limitations**

This study has several strengths. Preoperative pleural effusion was excluded using CT performed within 40 hours before surgery, and outcome measurement followed a standardized supine CT procedure with one reader and electronic calipers. The analysis compared regression and multiple machine-learning approaches using a common

feature set, applied conservative regularization, used conditional out-of-fold rather than direct resubstitution predictions for training-cohort performance plots, reserved a held-out cohort for final assessment, and added 1000-replicate full-dataset bootstrap correction to quantify refitting optimism. Discrimination, Brier score, graphical calibration, and net benefit were evaluated, and global and local SHAP analyses improved model transparency.

The limitations are substantial. First, this retrospective single-center study included only 127 recipients, 78 events, and 32 held-out validation cases. Modern sample-size principles emphasize the number of candidate parameters, outcome prevalence, anticipated model fit, and required precision rather than a simple events-per-variable rule [24, 25]. The available sample limited precision, assessment of nonlinearity, and reliable model comparison. A random split also uses data inefficiently in a small cohort. The OOF estimates were not nested because the feature set and hyperparameters were chosen using the complete training cohort. The added bootstrap analysis used the full dataset to estimate refitting optimism, but feature screening, LASSO selection, and hyperparameter tuning were fixed rather than repeated within each replicate. Consequently, neither resampling analysis captures the full uncertainty of the development pipeline. Temporal and geographic validation remain necessary to assess transportability.

Second, predictors were screened using univariate P values before LASSO. This two-stage data-driven approach may discard jointly informative predictors, increase selection instability, and understate uncertainty. The subsequent logistic-regression P values and CIs are nominal because they do not account for the selection process. Future analyses should prespecify clinically plausible predictors, use subject-matter knowledge to limit dimensionality, model nonlinear effects appropriately, and apply penalization without preliminary significance filtering.

Third, the outcome used a study-specific classification in which a moderate-to-large effusion was operationalized as a maximum CT depth  $\geq 4$  cm. This threshold has not been externally validated and the endpoint did not require symptoms, oxygen supplementation, reintubation, ventilation, pneumonia, thoracentesis, chest drainage, or mortality. Atelectasis could coexist, but its severity was not included in the endpoint.

Fourth, three recipients who died within 3 days without postoperative imaging were excluded from the prediction cohort. Because death prevented endpoint verification, this exclusion may have introduced survivor and verification bias. Separately, seven recipients with ascertained pleural-effusion status who died within 20 days were excluded from the secondary outcome analysis. Conditioning on survival may introduce survivor selection or collider bias and may attenuate associations with ICU duration, hospital stay, or day-7 laboratory outcomes. These survivor-restricted findings cannot be interpreted as evidence that SPPE is unrelated to mortality or overall postoperative prognosis.

Fifth, the survivor-restricted analysis was exploratory, unadjusted, and involved five comparisons without multiplicity correction. The day-7 ALT result ( $P = 0.049$ ) may represent a chance finding, and effect estimates with CIs or multivariable adjustment were unavailable. Sixth, although one reader reduced procedural variability, reader blinding and repeatability were not assessed, and the single-center imaging workflow may not be reproducible elsewhere. Seventh, donor and graft characteristics, ventilation parameters, immunosuppression, albumin administration, and longitudinal biomarker trajectories were not incorporated into the final model. Eighth, calibration was assessed graphically in only nine training and four validation groups; calibration-in-the-large, calibration slope, optimism-corrected Brier score, bootstrap-corrected calibration measures, and uncertainty intervals for net benefit were not reported. Finally, SHAP analysis focused on the random forest even though model

ranking differed across held-out and bootstrap validation; it should therefore be viewed as complementary rather than evidence of clinical superiority.

## Implications for future research

The next step should be prospective, preferably multicenter, data collection using standardized postoperative imaging or ultrasonography schedules, standardized immediate postoperative biomarker collection, and clinically anchored endpoint definitions. The 4-cm CT depth threshold should be validated against estimated effusion volume, respiratory physiology, drainage requirements, and patient-centered outcomes. Multiple blinded readers should assess measurement reproducibility. Predictor collection should include baseline nutrition and frailty, perioperative fluid and blood product balance, albumin administration, donor and graft factors, ventilation variables, renal function, and serial laboratory measurements. Future studies should ensure that every candidate predictor is measured before the intended prediction time and should prespecify mortality and recovery outcomes.

Model development should use a prespecified predictor set, contemporary sample-size planning, multiple imputation when missingness is material, and flexible but penalized functional forms. Internal validation should repeat the entire modeling pipeline, including predictor selection and hyperparameter tuning, within bootstrap or repeated resampling. External evaluation should report AUC, calibration-in-the-large, calibration slope, Brier score, clinically interpretable thresholds, and decision-curve net benefit with uncertainty. Recalibration or model updating should be anticipated when endpoint prevalence or clinical practice differs. A dynamic model incorporating biomarker trajectories and bedside lung ultrasound may ultimately be more useful than a single static prediction immediately after surgery.

## Conclusions

In this exploratory single-center cohort, previous abdominal surgery and immediate postoperative PNI, CRP, and AST measured within 1 hour after LT formed a parsimonious feature set for a study-specific early CT-defined moderate-to-large pleural effusion endpoint. Logistic regression achieved the highest held-out AUC, whereas random forest achieved the highest optimism-corrected AUC in the conditional full-dataset bootstrap analysis. The differing rankings and overlapping uncertainty indicated that no model was consistently superior. Discrimination was modest, calibration remained uncertain, and neither the OOF nor the bootstrap procedure repeated the full selection and tuning pipeline. Prospective multicenter validation using standardized imaging, clinically anchored endpoints, and rigorous calibration assessment is required before clinical implementation.

## Abbreviations

ALT, alanine aminotransferase; AST, aspartate aminotransferase; AUC, area under the receiver operating characteristic curve; BMI, body mass index; CI, confidence interval; CLR, C-reactive protein-to-lymphocyte ratio; CRP, C-reactive protein; CT, computed tomography; eGFR, estimated glomerular filtration rate; ICU, intensive care unit; LASSO, least absolute shrinkage and selection operator; LightGBM, Light Gradient Boosting Machine; LT, liver transplantation; MELD, Model for End-stage Liver Disease; NLR, neutrophil-to-lymphocyte ratio; OOF, out-of-fold; OR, odds ratio; PLR, platelet-to-lymphocyte ratio; PNI, prognostic nutritional index; RF, random forest; ROC, receiver operating characteristic; SD, standard deviation; SHAP, Shapley additive explanations; SII, systemic immune-inflammation index; SIRI, systemic inflammation response index; SIRI-P, SIRI-related platelet index; SPPE, severe postoperative pleural effusion; SVM, support vector machine; XGBoost, extreme gradient boosting.

## Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Second Affiliated Hospital of Chongqing Medical University (Scientific Ethics Pre-Review No. [2025] 361) and was conducted in accordance with the Declaration of Helsinki, national brain-death donation guidelines, and applicable Chinese laws. Because this retrospective study used anonymized data, the requirement for written informed consent was waived by the ethics committee. Organs were procured from controlled brain-dead donors and allocated through the China Organ Transplant Response System. No organs from executed prisoners were used.

Clinical trial number

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are not publicly available because of institutional and ethical restrictions but are available from the corresponding author on reasonable request, subject to applicable institutional and ethical requirements.

Competing interests

The authors declare that they have no competing interests.

Funding

No specific funding was received for this work.

Authors' contributions

XTC, MQL, LWC, and XHY contributed equally to the conception and design of the study, data acquisition and curation, interpretation of the results, and drafting or critical revision of the manuscript. XTC performed the statistical analyses and prepared the visualizations. WL supervised the study, provided clinical and methodological guidance, and critically revised the manuscript. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work.

Acknowledgements

Not applicable.

## References

1. Feltracco P, Barbieri S, Galligioni H, et al. Early respiratory complications after liver transplantation. *World J Gastroenterol.* 2013;19(48):9271-9281. doi:10.3748/wjg.v19.i48.9271.

2. Lui JK, Spaho L, Hakimian S, et al. Pleural effusions following liver transplantation: a single-center experience. *J Intensive Care Med*. 2021;36(8):862-872. doi:10.1177/0885066620932448.
3. Wiering L, Makiola E, Kuehn S, et al. Perioperative pleural drainage in liver transplantation: a retrospective analysis from a high-volume liver transplant center. *Ann Transplant*. 2020;25:e918456. doi:10.12659/AOT.918456.
4. Agrafiotis AC, Moris D, Vernadakis S. Surgical chest complications after liver transplantation. *World J Transplant*. 2022;12(11):359-364. doi:10.5500/wjt.v12.i11.359.
5. Clouse JW, Mangus RS, Vega CA, et al. Pleural effusion and malnutrition are associated with worse early outcomes after liver transplant. *Am Surg*. 2023;89(12):5881-5890. doi:10.1177/00031348221126962.
6. Patel BH, et al. Implications of pleural fluid composition in persistent pleural effusion following orthotopic liver transplant. *Med Sci (Basel)*. 2023;11(1):24. doi:10.3390/medsci11010024.
7. Oh EJ, et al. Intraoperative factors modifying the risk of postoperative pulmonary complications after living donor liver transplantation. *Transplantation*. 2023;107(8):1748-1755. doi:10.1097/TP.0000000000004544.
8. Wang Y, Ning Z, Yang L, Wang T. Construction and validation of a pulmonary complication score for patients after liver transplantation. *Clin Transplant*. 2023;37(3):e14872. doi:10.1111/ctr.14872.
9. Panackel C, et al. Pulmonary assessment of the liver transplant recipient. *J Clin Exp Hepatol*. 2023;13(5):895-911. doi:10.1016/j.jceh.2023.04.003.
10. Wang Y, Zhou Y, Gong Y, Ding Z, Yang L, Wang T. A machine learning-driven app for predicting the need for post-operative respiratory support in liver transplant recipients. *Int J Med Inform*. 2026;209:106263. doi:10.1016/j.ijmedinf.2026.106263.
11. Kajal K, Premkumar M, Chaluvashetty SB, Bhujade H, Kulkarni AV. Point-of-care thoracic ultrasonography in patients with cirrhosis and liver failure. *Cureus*. 2021;13(6):e15559. doi:10.7759/cureus.15559.
12. Moy MP, Levsky JM, Berko NS, Godelman A, Jain VR, Haramati LB. A new, simple method for estimating pleural effusion size on CT scans. *Chest*. 2013;143(4):1054-1059. doi:10.1378/chest.12-1292.
13. Roberts ME, Rahman NM, Maskell NA, et al. British Thoracic Society Guideline for pleural disease. *Thorax*. 2023;78(11):1143-1156. doi:10.1136/thorax-2023-220304.
14. Bischoff SC, Bernal W, Dasarathy S, et al. ESPEN practical guideline: clinical nutrition in liver disease. *Clin Nutr*. 2020;39(12):3533-3562. doi:10.1016/j.clnu.2020.09.001.
15. Chapman B, Goh SK, Parker F, et al. Malnutrition and low muscle strength are independent predictors of clinical outcomes and healthcare costs after liver transplant. *Clin Nutr ESPEN*. 2022;48:210-219. doi:10.1016/j.clnesp.2022.02.013.
16. Pravisani R, Mocchegiani F, Isola M, et al. Postoperative trends and prognostic values of inflammatory and nutritional biomarkers after liver transplantation for hepatocellular carcinoma. *Cancers (Basel)*. 2021;13(3):513. doi:10.3390/cancers13030513.
17. Min JY, et al. Predictive impact of modified-prognostic nutritional index for acute kidney injury within 1 week after living donor liver transplantation. *Int J Med Sci*. 2020;17(1):82-88. doi:10.7150/ijms.39014.
18. Kornberg A, Kaschny L, Kornberg J, Friess H. Preoperative prognostic nutritional index may be a strong predictor of hepatocellular carcinoma recurrence following liver transplantation. *J Hepatocell Carcinoma*. 2022;9:649-660. doi:10.2147/JHC.S366107.
19. Olthoff KM, Kulik L, Samstein B, et al. Validation of a current definition of early allograft dysfunction in liver transplant recipients and analysis of risk factors. *Liver Transpl*. 2010;16(8):943-949. doi:10.1002/lt.22091.

20. Liu J, Pang L, Ng KTP, et al. Mechanistic insight and clinical implications of ischemia/reperfusion injury post liver transplantation. *Cell Mol Gastroenterol Hepatol*. 2023;15(6):1463-1474. doi:10.1016/j.jcmgh.2023.03.003.
21. Zant R, Melter M, Knoppke B, Ameres M, Kunkel J. Kinetics of interleukin-6, procalcitonin, and C-reactive protein after pediatric liver transplantation. *Transplant Proc*. 2014;46(10):3507-3510. doi:10.1016/j.transproceed.2014.08.048.
22. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378.
23. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*. 2025;388:e082505. doi:10.1136/bmj-2024-082505.
24. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441. doi:10.1136/bmj.m441.
25. Riley RD, Snell KIE, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: part II - binary and time-to-event outcomes. *Stat Med*. 2019;38(7):1276-1296. doi:10.1002/sim.7992.
26. 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. doi:10.1016/j.jclinepi.2019.02.004.
27. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2.
28. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17:230. doi:10.1186/s12916-019-1466-7.
29. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J*. 2021;21(10):1643-1648. doi:10.1016/j.spinee.2021.02.024.
30. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol*. 1996;58(1):267-288.
31. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32. doi:10.1023/A:1010933404324.
32. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20:273-297. doi:10.1007/BF00994018.
33. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016:785-794. doi:10.1145/2939672.2939785.
34. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. 2017;30.
35. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56-67. doi:10.1038/s42256-019-0138-9.
36. Nitta H, Mitsuura C, Shiraishi Y, et al. Predictive model for postoperative pleural effusion after hepatectomy. *Ann Gastroenterol Surg*. 2021;5(3):373-380. doi:10.1002/ags3.12417.
37. Zhao JY, Lin HY, Gong CF, et al. Establishment and validation of a predictive nomogram for severe pleural effusion in liver cancer patients after hepatectomy. *Medicine (Baltimore)*. 2024;103(10):e36556. doi:10.1097/MD.00000000000036556.

38. Fan ZQ, Diao YK, Li C, et al. Severe pleural effusion after hepatectomy for hepatocellular carcinoma: multicentre retrospective study. *BJS Open*. 2023;7(6):zrad118. doi:10.1093/bjsopen/zrad118.
39. Chu KJ, Yao XP, Fu XH. Factors related to pleural effusion following hepatectomy for primary liver cancer. *Hepatobiliary Pancreat Dis Int*. 2007;6(1):58-62.
40. Xu C, Zhang W, Wang X, et al. Correlation analysis of pleural effusion and lung infection after liver transplantation. *Clin Lab*. 2021;67(9). doi:10.7754/Clin.Lab.2020.201037.