

Original Article

Development and validation of a prediction model for myelosuppression in lung cancer patients after platinum-based doublet chemotherapy: a multifactorial analysis approach

Xueyan Li¹, Linyu Li¹, Lu Zhang²

¹Department of Respiratory and Critical Care Medicine, The Affiliated Hospital of Qingdao University, Qingdao 266000, Shandong, China; ²Department of Radiation Oncology, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan 250011, Shandong, China

Received January 13, 2024; Accepted February 13, 2025; Epub February 15, 2025; Published February 28, 2025

Abstract: Objective: To develop an individualized prediction model for myelosuppression risk in lung cancer patients undergoing platinum-based doublet chemotherapy and validate its predictive efficacy. Methods: A retrospective analysis was conducted on the clinical data of 584 lung cancer patients who received platinum-based doublet chemotherapy at The Affiliated Hospital of Qingdao University between January 2016 and December 2020. Patients were randomly assigned to a training cohort (n=391) and a validation cohort (n=193). Myelosuppression occurred in 280 (71.6%) patients in the training cohort and 132 (68.4%) in the validation cohort. Univariate analysis and LASSO regression were used to identify independent risk factors for myelosuppression. Prediction models were developed using Support Vector Machine (SVM), Random Forest, Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (Adaboost). Model performance was evaluated using receiver operating characteristic (ROC) curves, calibration curves, and Decision Curve Analysis (DCA). The SHAP algorithm was employed to evaluate feature importance, and a nomogram was developed for individual risk prediction. Results: LASSO regression identified 10 independent risk factors for myelosuppression: age, body mass index (BMI), white blood cell count, neutrophil count, platelet count, total protein, gender, treatment regimen, targeted therapy, and first chemotherapy cycle. In the training cohort, the XGBoost model exhibited the best performance, with an area under the curve (AUC) of 0.855 (95% CI: 0.813-0.897), while the AUC in the validation cohort was 0.793. SHAP analysis identified white blood cell count, platelet count, neutrophil count, BMI, and age as the most influential predictors. The SHAP analysis based on the XGBoost model demonstrated substantial value. Conclusion: This study successfully developed an individualized prediction model for myelosuppression risk in lung cancer patients following platinum-based doublet chemotherapy, with the XGBoost model achieving high predictive accuracy and clinical utility. The model provides a valuable tool for guiding precision medicine.

Keywords: Lung cancer, myelosuppression, platinum-based doublet chemotherapy, machine learning, LASSO regression, random forest, risk prediction, nomogram, SHAP algorithm

Introduction

Lung cancer remains one of the leading causes of cancer-related deaths globally [1]. According to the International Agency for Research on Cancer (IARC), there were 2.3 million new lung cancer cases worldwide in 2022, with approximately 1.9 million deaths, accounting for 18% of all cancer-related fatalities [2]. Despite advancements in early screening, targeted therapy, and immunotherapy that have extend-

ed survival for some patients, the prognosis for advanced lung cancer patients remains poor [3]. Platinum-based doublet chemotherapy remains a cornerstone treatment across various lung cancer subtypes, particularly for advanced-stage patients and those resistant to targeted therapies [4].

Myelosuppression, a common side effect of chemotherapy, results from the cytotoxic impact of chemotherapy drugs on the bone mar-

row hematopoietic system. This significantly reduces white blood cells, red blood cells, and platelets, leading to weakened immunity, increased infection risk, and elevated bleeding tendency [5]. In severe cases, myelosuppression can be life-threatening, requiring treatment interruptions or dosage reductions, thereby compromising the efficacy of chemotherapy [6]. Given its clinical significance, predicting myelosuppression risk before chemotherapy and developing individualized prevention and treatment strategies have become a crucial topic in clinical research.

Numerous studies have investigated risk factors for myelosuppression. For example, elderly patients are more susceptible to myelosuppression due to their decreased bone marrow function, while patients with a low body mass index (BMI) may have reduced chemotherapy tolerance and heightened toxicity sensitivity [7]. Furthermore, comorbidities such as diabetes, hypertension, and coronary heart disease can exacerbate the risk by affecting metabolism and immune function [8]. Laboratory indicators, including baseline white blood cell, neutrophil, and platelet levels, have also been strongly associated with myelosuppression risk [9]. However, most existing studies rely on univariate analysis or traditional regression models, which fail to capture complex interactions among multiple risk factors, particularly in high-dimensional data where traditional methods have limited predictive ability.

In predictive modeling, machine learning algorithms have gained attention due to their ability to handle multidimensional, nonlinear data [10]. Unlike traditional statistical methods, machine learning models can capture complex relationships among variables, offering superior predictive performance [10]. Commonly used algorithms in medical research include Support Vector Machine (SVM), Random Forest, Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (Adaboost) [11]. However, their clinical application remains challenging, primarily due to limited model interpretability. Machine learning models are often perceived as “black box”, making it difficult for clinicians to understand how risk assessments are generated [12]. Recently, the SHAP (SHapley Additive Explanations) algorithm has emerged as a method for improving model interpretability by

quantifying each variable's contribution to predictions [13-15].

This study aims to develop and validate an individualized risk prediction model for myelosuppression in lung cancer patients undergoing platinum-based doublet chemotherapy, using multifactorial analysis and machine learning methods. The objectives of this study are: (1) to identify independent risk factors for myelosuppression using univariate analysis and least absolute shrinkage and selection operator (LASSO) regression; (2) to construct high-performance prediction models using machine learning algorithms and evaluate their calibration, classification performance, and clinical utility; and (3) to enhance model interpretability using the SHAP algorithm and develop nomograms for risk visualization and individualized application. This research aims to provide a reference for precision treatment and individualized management of lung cancer patients.

Methods and materials

Sample size calculation

The required sample size was estimated using the ‘pmsampsize’ package in R software, based on a literature review and a reported 77% incidence rate of myelosuppression in lung cancer patients following treatment [16]. The sample size calculation utilized a logistic regression model with 12 candidate predictors and an AUC (C-statistic) of 0.878. The analysis determined that a minimum of 296 patients was required to develop a robust predictive model, including at least 228 events (cases of myelosuppression) to ensure model stability and minimize overfitting. This calculation corresponds to approximately 18.99 events per predictor parameter (EPP), meeting established criteria for reliable prediction. The actual sample size was determined by the number of patients collected clinically.

Sample source

This retrospective study analyzed the clinical data of 584 lung cancer patients who received platinum-based doublet chemotherapy between January 2016 and December 2020 at The Affiliated Hospital of Qingdao University (**Figure 1**). The study was approved by the

Prediction model for myelosuppression in lung cancer chemotherapy

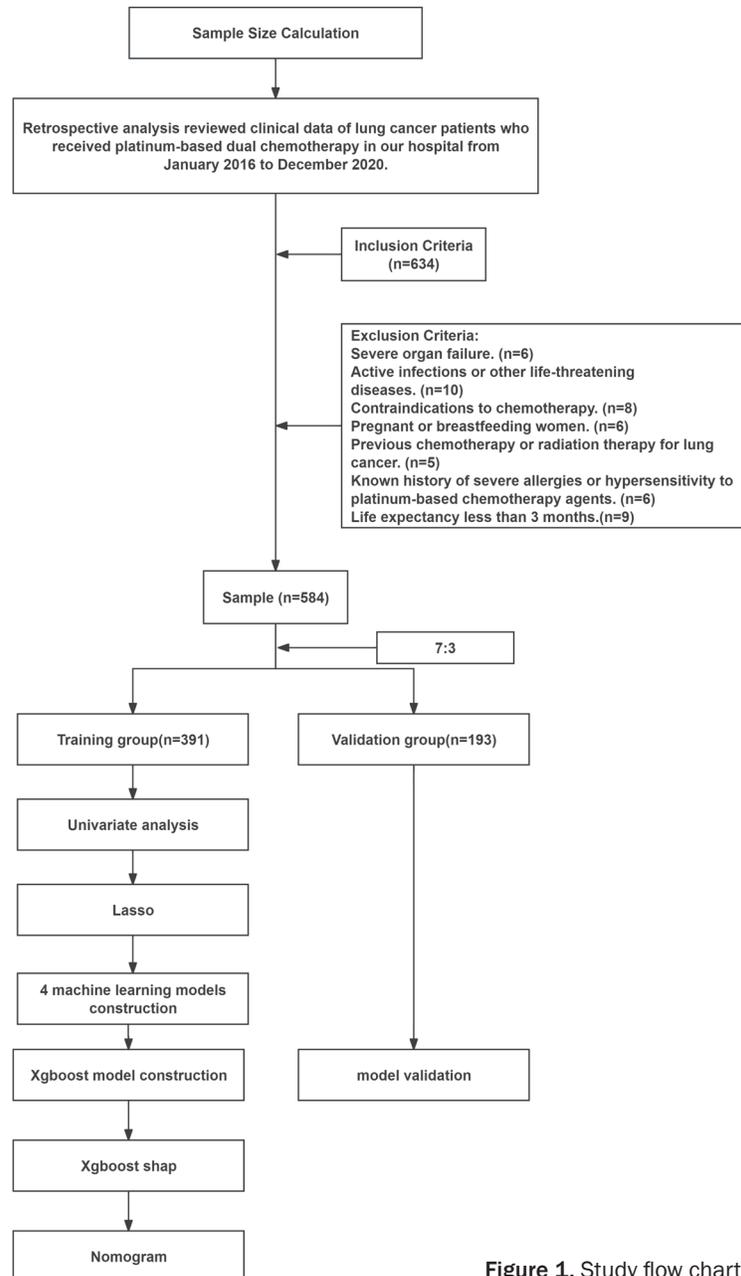


Figure 1. Study flow chart.

Ethics Committee of The Affiliated Hospital of Qingdao University (24-12117KY).

Definition of myelosuppression

Myelosuppression is a clinical condition characterized by suppressed bone marrow hematopoiesis, leading to reduced blood cell production. It manifests as significant decrease in white blood cells (WBC), red blood cells (RBC), and platelets. According to the Common Terminology Criteria for Adverse Events (CTCAE)

[17], myelosuppression includes neutropenia (Absolute neutrophil count (ANC) $<1.5 \times 10^9/L$, severe cases $<0.5 \times 10^9/L$), leukopenia (WBC count $<4.0 \times 10^9/L$), thrombocytopenia (platelet count $<100 \times 10^9/L$), and anemia (hemoglobin $<10 \text{ g/dL}$). These parameters are graded from 1 to 4, with higher grades indicating more severe suppression. Myelosuppression can impair immunity, increase infection risk, and lead to bleeding tendencies, potentially becoming life-threatening in severe cases. Its occurrence is closely related to chemotherapy drug dosage, patient's baseline health status, and individual variability. Monitoring peripheral blood counts is critical for early identification and timely intervention to mitigate complications.

Inclusion and exclusion criteria

Inclusion criteria: (1) Patients aged ≥ 18 years with pathologically confirmed lung cancer (including non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC)). (2) Received at least one cycle of platinum-based doublet chemotherapy (e.g., docetaxel + platinum, gemcitabine + platinum, pemetrexed + platinum). (3) Available complete clinical and laboratory records before chemotherapy. (4) Complete follow-up data for clear assessment of myelosuppression occurrence. (5) No severe organ failure, able to tolerate chemotherapy.

Exclusion criteria: (1) Severe organ failure (e.g., liver, kidney, heart). (2) Active infections or other life-threatening diseases. (3) Contraindications to chemotherapy. (4) Pregnant or breastfeeding women. (5) Previous chemotherapy or radiation therapy for lung cancer. (6) A known history of severe allergies or hypersensitivity to platinum-based chemotherapy

Prediction model for myelosuppression in lung cancer chemotherapy

agents. (7) Life expectancy of less than 3 months.

Clinical data collection

This study retrospectively analyzed patient data from electronic medical records and outpatient follow-up data. Collected clinical information included demographics (gender, BMI, age, smoking history), medical history (hypertension, diabetes, coronary heart disease), tumor pathology (NSCLC, SCLC), treatment regimens (e.g., docetaxel + platinum, gemcitabine + platinum, pemetrexed + platinum), additional therapies (targeted therapy, first-line chemotherapy, and adjuvant chemotherapy data). Laboratory tests conducted prior to chemotherapy included WBC count, neutrophil count, RBC count, hemoglobin, platelet count, total protein, and albumin levels. This comprehensive data collection ensured accurate baseline characterization for subsequent analysis.

Laboratory indicators

Blood routine tests were performed using the Sysmex automated blood analyzer (e.g., XN-1000 model). Blood samples were collected before each chemotherapy cycle to establish baseline values, and 7 to 14 days after chemotherapy, to evaluate the severity of myelosuppression. These time points were critical for assessing hematological toxicity and guiding clinical management.

Follow-up

Patients underwent follow-up 1-2 weeks after chemotherapy for blood tests to assess bone marrow suppression. During follow-up, clinical reviews and complete blood counts (CBC) were performed, mainly monitoring white blood cell count, neutrophil count, and platelet count. Adverse events, including myelosuppression, were recorded and graded according to CTCAE 4.0 criteria. The primary endpoint was the occurrence of myelosuppression, particularly leukopenia, granulocytopenia, thrombocytopenia, and anemia.

Outcome measures

Primary outcome: The predictive efficacy of machine learning models in assessing myelosuppression risk in lung cancer patients undergoing platinum-based doublet chemotherapy was evaluated.

Secondary outcomes: (1) To assess the influence of laboratory indicators and clinical characteristics on the occurrence of myelosuppression, particularly the dynamic changes in key indicators such as white blood cell count, neutrophil count, and platelet count before and after chemotherapy. (2) To evaluate the clinical utility of the prediction model through calibration curves and Decision Curve Analysis (DCA). (3) To assess the interpretability of the prediction model through SHAP analysis. (4) To develop a nomogram for individualized risk prediction, providing precise myelosuppression risk assessment tools tailored for different patient groups.

Statistical analysis

Statistical analysis was performed using SPSS 26.0 and R 4.3.3 software. The Kolmogorov-Smirnov (K-S) test was used for normality testing, with normally distributed data presented as mean \pm standard deviation (Mean \pm SD), and group comparisons conducted using the independent sample t-test. Non-normally distributed data were presented as median (interquartile range) and compared using the rank-sum test. Categorical variables were expressed as frequency and percentage, with group comparisons conducted using chi-square or Fisher's exact test. The significance level was set at two-sided $P < 0.05$. ROC curve analysis was performed using the pROC package to assess the discriminatory ability of continuous variables and calculate the area under the curve (AUC). LASSO regression for variable selection was performed using the glmnet package. Machine learning models, including Support Vector Machine (kernlab package), Random Forest (randomForest package), Extreme Gradient Boosting (xgboost package), and Adaptive Boosting (ada package), were developed and compared based on performance metrics (AUC, sensitivity, specificity, precision, F1-score). Precision-Recall (PR) curves were generated using the pROC package, and nomograms and calibration curves were drawn using the rms package. Decision Curve Analysis (DCA) was used to assess the clinical net benefit of the models, and confusion matrices were used for visualization of model classification performance. Feature interpretability was analyzed using SHAP (shapviz package), and plots were generated using the ggplot2 package.

Prediction model for myelosuppression in lung cancer chemotherapy

Table 1. Comparison of baseline data between the model and validation groups

Variable	Total	Training Group (n=391)	Validation Group (n=193)	Statistic	P-value
Myelosuppression					
Yes	412 (70.55%)	280 (71.61%)	132 (68.39%)	0.644	0.422
No	172 (29.45%)	111 (28.39%)	61 (31.61%)		
Gender					
Male	461 (78.94%)	308 (78.77%)	153 (79.27%)	0.020	0.889
Female	123 (21.06%)	83 (21.23%)	40 (20.73%)		
Treatment Regimen					
Docetaxel + Platinum	173 (29.62%)	119 (30.43%)	54 (27.98%)	1.143	0.767
Gemcitabine + Platinum	114 (19.52%)	75 (19.18%)	39 (20.21%)		
Pemetrexed + Platinum	156 (26.71%)	107 (27.37%)	49 (25.39%)		
Others	141 (24.14%)	90 (23.02%)	51 (26.42%)		
Targeted Therapy					
Yes	69 (11.82%)	48 (12.28%)	21 (10.88%)	0.241	0.623
No	515 (88.18%)	343 (87.72%)	172 (89.12%)		
First Chemotherapy					
Yes	318 (54.45%)	210 (53.71%)	108 (55.96%)	0.264	0.608
No	266 (45.55%)	181 (46.29%)	85 (44.04%)		
Age (years)	62.90±7.77	62.95±7.73	62.80±7.87	0.224	0.823
BMI (kg/m ²)	22.20±3.18	22.24±3.00	22.11±3.53	0.460	0.646
White Blood Cell Count (10 ⁹ /L)	6.45±2.53	6.41±2.49	6.51±2.63	0.451	0.652
Neutrophil Count (10 ⁹ /L)	4.19 [2.76, 5.54]	4.14 [2.88, 5.36]	4.21 [2.61, 5.97]	0.414	0.679
Hemoglobin (g/L)	124.34±17.74	124.09±17.64	124.84±17.96	0.479	0.632
Platelet Count (10 ⁹ /L)	227.61±81.88	228.27±82.23	226.29±81.37	0.274	0.784
Total Protein (g/L)	67.10±6.16	67.10±6.32	67.11±5.84	0.021	0.983

Note: BMI = Body Mass Index; "Others" in Treatment Regimen includes etoposide + platinum and paclitaxel + platinum.

Results

Comparison of baseline characteristics between myelosuppression and non-myelosuppression groups after platinum-based doublet chemotherapy

A comparison of baseline characteristics between patients with myelosuppression and those without myelosuppression after platinum-based doublet chemotherapy revealed statistically significant differences ($P < 0.05$) in the following variables: gender, age, BMI, treatment regimen, targeted therapy, first chemotherapy, WBC count, neutrophil count, platelet count, and total protein levels. Non-significant variables included smoking history, hypertension history, diabetes history, coronary artery disease history, pathology type, postoperative adjuvant chemotherapy, and bone metastasis ($P > 0.05$, [Table S1](#)).

Comparison of baseline data between the model and validation groups

Based on the significant differences identified in the univariate analysis (as shown in [Table S1](#)). Patients were randomly divided in a 0.67:0.33 ratio, with the variables having differences selected as the factors for assigning them to a training cohort ($n=391$) and a validation cohort ($n=193$) for model construction. The baseline characteristics of patients in the model group (training group) and validation group were compared. The results showed no significant differences between the two groups in terms of myelosuppression occurrence, gender, treatment regimen, targeted therapy, first chemotherapy, age, BMI, white blood cell count, neutrophil count, hemoglobin, platelet count, and total protein levels ($P > 0.05$, [Table 1](#)). These findings suggest that the two groups were comparable.

Prediction model for myelosuppression in lung cancer chemotherapy

Table 2. Univariate analysis of baseline data related to myelosuppression

Variable	Total	Myelosuppression Group (n=391)	Non-Myelosuppression Group (n=193)	Statistic	P-value
Gender					
Male	308 (78.77%)	229 (81.79%)	79 (71.17%)	5.356	0.021
Female	83 (21.23%)	51 (18.21%)	32 (28.83%)		
Treatment Regimen					
Docetaxel + Platinum	119 (30.43%)	79 (28.21%)	40 (36.04%)	14.310	0.003
Gemcitabine + Platinum	75 (19.18%)	66 (23.57%)	9 (8.11%)		
Pemetrexed + Platinum	107 (27.37%)	69 (24.64%)	38 (34.23%)		
Others	90 (23.02%)	66 (23.57%)	24 (21.62%)	14.310	0.003
Targeted Therapy					
Yes	48 (12.28%)	42 (15.00%)	6 (5.41%)	6.795	0.009
No	343 (87.72%)	238 (85.00%)	105 (94.59%)		
First Chemotherapy					
Yes	210 (53.71%)	162 (57.86%)	48 (43.24%)	6.828	0.009
No	181 (46.29%)	118 (42.14%)	63 (56.76%)		
Age (years)	62.95±7.73	63.35±7.30	61.95±8.69	1.622	0.106
BMI (kg/m ²)	22.24±3.00	21.84±2.99	23.23±2.79	4.216	<0.001
White Blood Cell Count (10 ⁹ /L)	6.41±2.49	6.00±2.28	7.46±2.69	5.406	<0.001
Neutrophil Count (10 ⁹ /L)	4.19±1.93	4.00±1.76	4.67±2.25	3.123	0.002
Hemoglobin (g/L)	124.09±17.64	125.29±17.43	121.08±17.87	2.135	0.033
Platelet Count (10 ⁹ /L)	228.27±82.23	220.26±80.94	248.46±82.38	3.091	0.002
Total Protein (g/L)	67.10±6.32	66.78±6.29	67.91±6.36	1.604	0.109

Note: BMI = Body Mass Index; "Others" in Treatment Regimen includes etoposide + platinum and paclitaxel + platinum.

Univariate analysis of myelosuppression-related factors in the training group

Univariate analysis identified several variables that were significantly associated with the occurrence of myelosuppression after platinum-based doublet chemotherapy ($P < 0.05$), including gender, treatment regimen, targeted therapy, first chemotherapy, BMI, white blood cell count, neutrophil count, platelet count, and hemoglobin levels. Specifically, male gender, receipt of targeted therapy or first chemotherapy, lower BMI, lower WBC count, lower neutrophil count, lower platelet count, and lower hemoglobin levels were associated with a higher risk of myelosuppression. Other variables, such as smoking history, coronary artery disease history, age, and total protein levels, showed no significant association with the occurrence of myelosuppression ($P > 0.05$). Details are shown in **Table 2**.

Categorization of continuous variables with significant differences

LASSO regression (family = "binomial") was applied to convert continuous variables into

categorical variables for analysis. ROC curves analysis was plotted for continuous variables with significant differences to assess their discriminatory power. The results showed that BMI, WBC count, neutrophil count, and platelet count exhibited strong discriminatory power, with AUCs of 0.851, 0.656, 0.958, and 0.606, respectively. The optimal cutoff values for these variables were 23.765, 7.705, 5.340, and 180.500. In contrast, age and total protein demonstrated lower discriminatory ability, with AUCs of 0.557 and 0.570, respectively. In conclusion, low BMI, low white blood cell count, low neutrophil count, and low platelet count were significantly associated with an increased risk of myelosuppression (**Figure 2**).

LASSO regression for myelosuppression factors after platinum-based doublet chemotherapy

All significant variables identified in univariate analysis were assigned values, as shown in **Table 3**. In LASSO regression analysis, when $\lambda_{\min} = 0.0039$, 10 characteristic vari-

Prediction model for myelosuppression in lung cancer chemotherapy

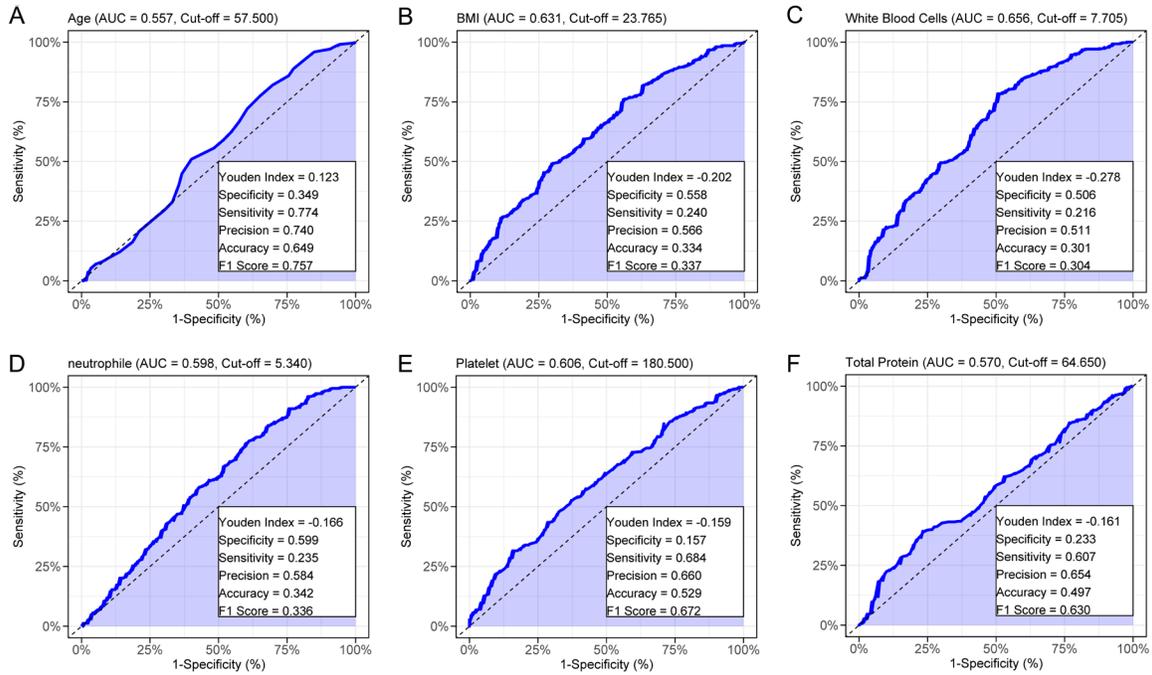


Figure 2. ROC curve analysis results for continuous variables. A: ROC curve for age in predicting myelosuppression. B: ROC curves of BMI in predicting myelosuppression. C: ROC curve of white blood cell count in predicting myelosuppression. D: ROC curves of neutrophil count in predicting myelosuppression. E: ROC curve of platelet count in predicting myelosuppression. F: ROC curve for total protein in predicting myelosuppression. Note: ROC, Receiver Operating Characteristic; AUC, Area Under Curve; BMI, Body Mass Index.

Table 3. Assignment table

Variable	Type	Content
Age (years)	(X)	<57.5=0, ≥57.5=1
BMI (kg/m ²)	(X)	<23.765=0, ≥23.765=1
White Blood Cell Count (10 ⁹ /L)	(X)	<7.705=0, ≥7.705=1
Neutrophil Count (10 ⁹ /L)	(X)	<5.340=0, ≥5.340=1
Platelet Count (10 ⁹ /L)	(X)	<180.5=0, ≥180.5=1
Total Protein (g/L)	(X)	<64.65=0, ≥64.65=1
Sex	(X)	Male =1, and female =0
Treatment Options	(X)	Docetaxel + platinum =1, gemcitabine + platinum =0, pemetrexed + platinum =3, and other =4
Targeted therapy	(X)	Yes =1, No =0
First chemotherapy	(X)	Yes =1, No =0
Myelosuppression	(Y)	Yes =1, No =0

Note: BMI, Body Mass Index.

ables associated with myelosuppression after platinum-based doublet chemotherapy were selected: age, BMI, WBC count, neutrophil count, platelet count, total protein, gender, treatment regimen, targeted therapy, and first chemotherapy (**Figure 3**). These variables were selected through cross-validation at the optimal penalty parameter and are closely related to the occurrence of myelosuppression, representing important characteristics.

Machine learning model construction based on LASSO-selected features

Based on the features selected by LASSO regression, four machine learning models - SVM, Random Forest, XGBoost, and Adaboost - were evaluated in both the training and validation groups. In the training group, XGBoost had the highest AUC of 0.855 (95% CI: 0.813-0.897), followed by Adaboost (0.853), SVM

Prediction model for myelosuppression in lung cancer chemotherapy

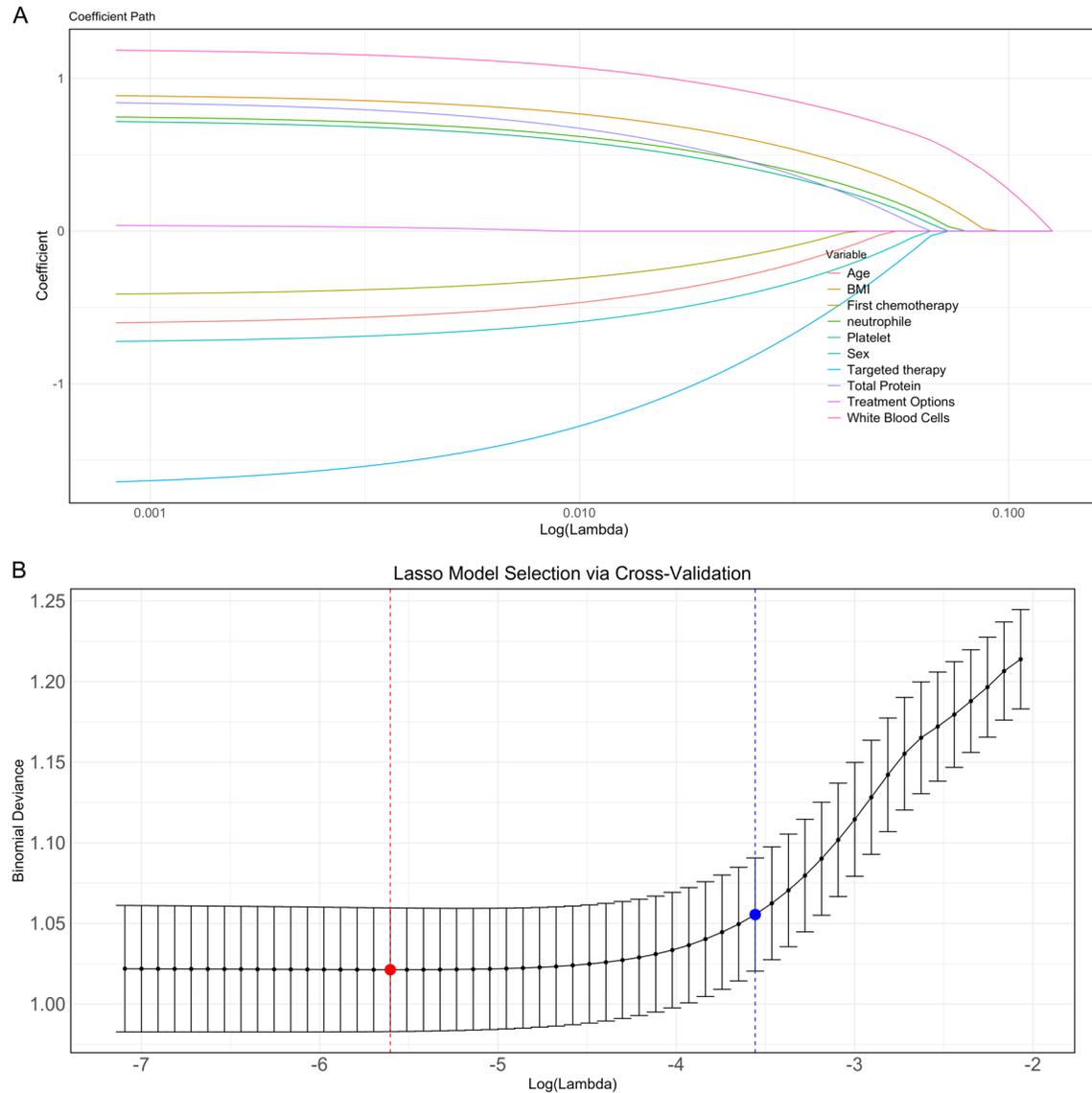


Figure 3. Selection of key predictors using LASSO regression analysis. A: Coefficient path diagram of LASSO regression. B: Selection of optimal Lambda values through cross-validation. Note: LASSO, Least Absolute Shrinkage and Selection Operator; BMI, Body Mass Index.

(0.815), and Random Forest (0.801) (**Figure 4A**). XGBoost achieved the highest sensitivity (87.77%) and precision (87.77%), while Random Forest showed the highest specificity (73.45%) and sensitivity (74.10%). SVM and Adaboost had lower sensitivity (72.66% and 82.73%) but balanced precision (72.66% and 82.73%). AUC significance comparisons (**Figure 4C**) indicated that Random Forest had a significantly higher AUC than the other models. In the validation group, AUCs decreased in all models, with Adaboost and XGBoost showing the highest AUCs at 0.794 and 0.793, respectively. SVM

and Random Forest had AUCs of 0.763 and 0.736, respectively (**Figure 4B**). The AUC differences between models in the validation group were not significant ($P=0.113$) (**Figure 4D**). A comparison of performance metrics in the training group (**Figure 4E**) showed Random Forest had the highest sensitivity (92.31%) and specificity (82.20%). In the validation group (**Figure 4F**), performance differences were smaller, with Adaboost and SVM showing more balanced performance, having higher specificity (75.93% and 90.74%) and moderate sensitivity (61.15% and 47.48%).

Prediction model for myelosuppression in lung cancer chemotherapy

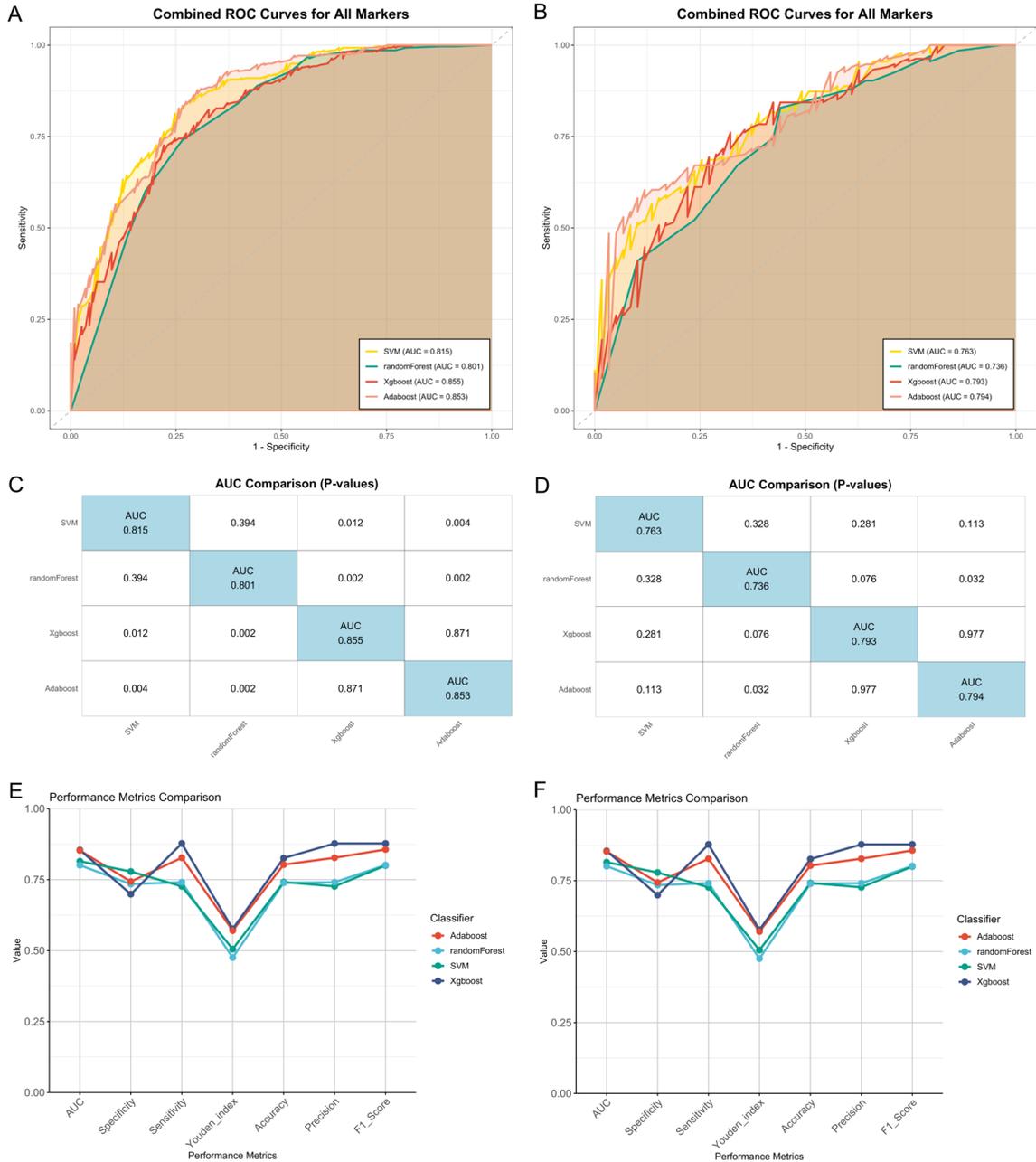


Figure 4. Performance of machine learning models based on LASSO filtered features. A: ROC curves of different models in the training group. B: ROC curves of different models in the validation group. C: Comparison of AUCs between different models in the training group. D: Comparison of AUCs between different models in the validation group. E: Comparison of performance indicators of different models in the training group. F: Comparison of performance indicators of different models in the validation group. Note: Receiver Operating Characteristic (ROC), Area Under Curve (AUC), Sensitivity, Specificity, Accuracy, Precision, F1-score, Youden Index, Support Vector Machine (SVM) Support Vector Machine (SVM), Random Forest, Extreme Gradient Boosting (XGBoost), Adaptive Boosting (Adaboost).

Evaluation of model performance

The performance of the XGBoost model was evaluated in both the training and validation

sets. In the training set, the PR curve (Figure 5A) showed the relationship between Precision and Recall, indicating good classification performance. The calibration curve (Figure 5B)

Prediction model for myelosuppression in lung cancer chemotherapy

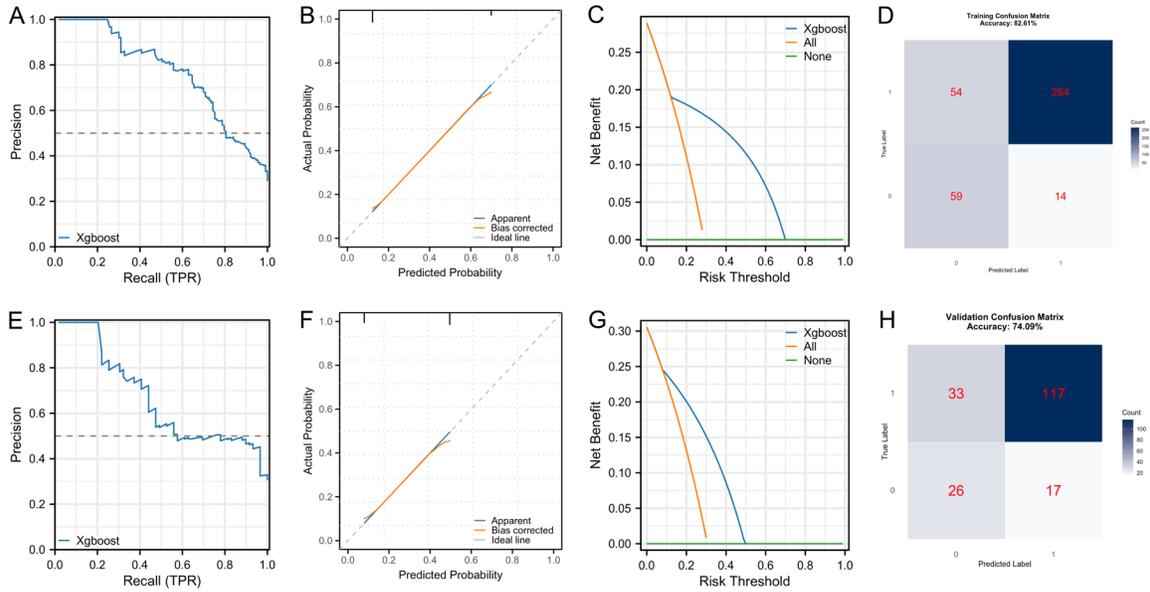


Figure 5. Performance evaluation of random forest model based on training and validation sets. A: PR curve of the random forest model in the training set. B: Calibration curve of the random forest model in the training set. C: Decision curve analysis (DCA) of the random forest model in the training set. D: Confusion matrix of the training set. E: PR curve of the random forest model in the validation set. F: Calibration curve of the random forest model in the validation set. G: Decision curve analysis (DCA) of the random forest model in the validation set. H: Confusion matrix of the validation set. Note: Precision-Recall (PR) Curve, Calibration Curve, Decision Curve Analysis (DCA), Area Under Curve (AUC), Confidence Interval (CI), Akaike Information Criterion (AIC), C-index.

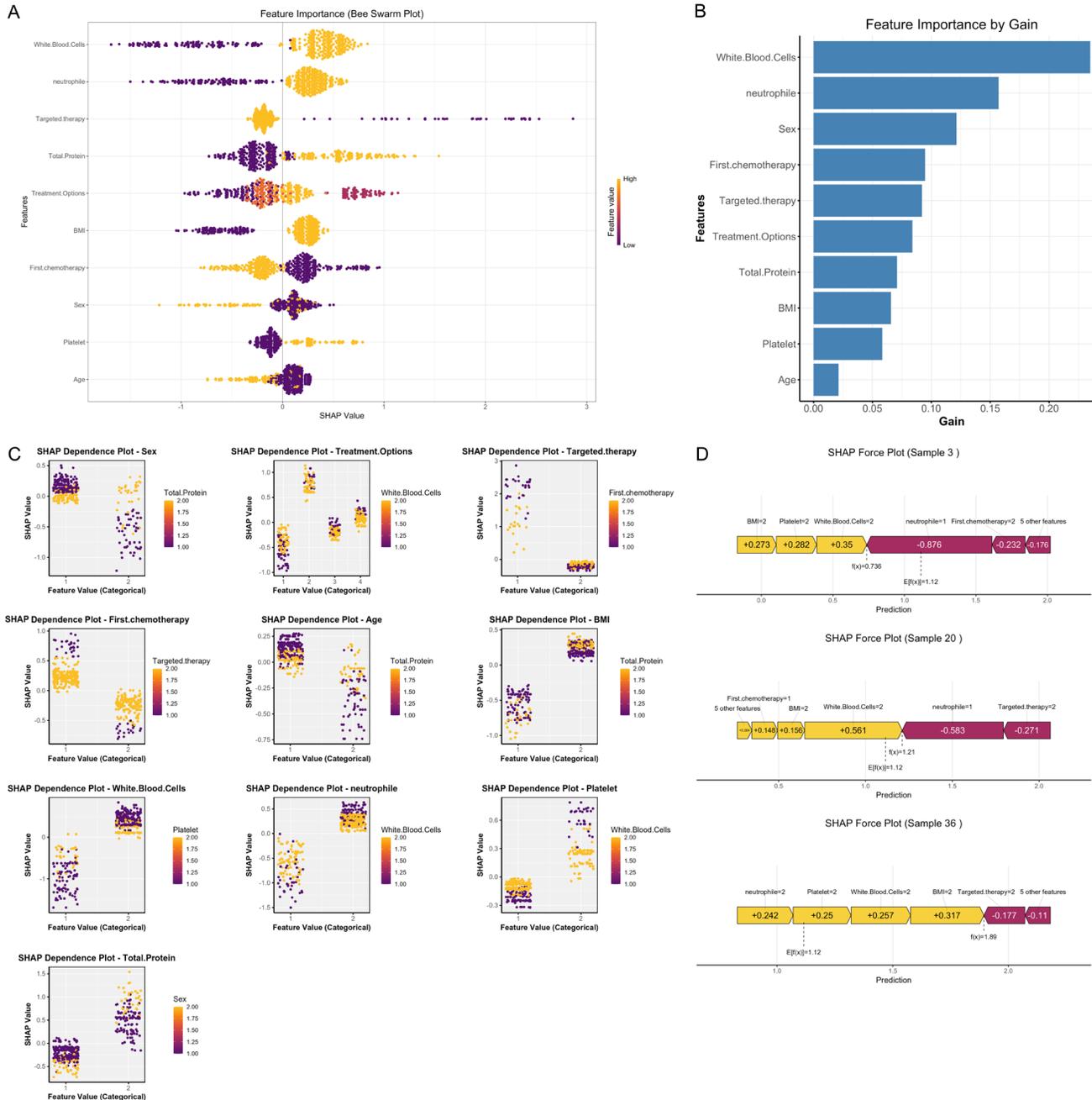
showed the C-index for the XGBoost model to be 0.788 (95% CI: 0.741-0.836, $P < 0.0001$), with a chi-square goodness-of-fit test value of $8.4055e-20$ ($P=1$), demonstrating excellent calibration ability. The AIC value of 262.587 also indicated a good fit of the model. The decision curve analysis (Figure 5C) showed a net benefit across a risk threshold range of 0%-69%, with the maximum net benefit reaching 28.90%. The confusion matrix (Figure 5D) showed a prediction accuracy of 94.3% in the training set. In the validation set, the model's performance declined slightly (Figure 5E-H). The PR curve (Figure 5E) showed a relationship between Precision and Recall, with slightly lower classification performance in the test set than in the training set. The calibration curve (Figure 5F) showed the C-index to be 0.743 (95% CI: 0.683-0.803, $P < 0.0001$), with a chi-square goodness-of-fit test value of $7.9134e-26$ ($P=1$), indicating that the model's calibration ability in the validation set remained useful. The AIC value was 212.627. Decision curve analysis (Figure 5G) showed that the model was beneficial across a risk threshold range of 0%-49%, with the maximum net benefit reaching 30.56%. The confusion matrix (Figure 5H)

showed a prediction accuracy of 66.4% in the validation set.

SHAP analysis of random forest model contributions and key variable interpretation

The Xgboost model was subjected to variable contribution analysis and feature importance evaluation using the SHAP algorithm. The SHAP values displayed the positive and negative influences of each variable on the model's prediction results, as well as the ranking of variable importance at both the individual sample and overall data levels (Figure 6A). In the overall feature importance analysis (Figure 6B), white blood cell count, neutrophil count, platelet count, BMI, and age were found to be the most influential variables on the prediction results. The SHAP dependence plot (Figure 6C) further analyzed the specific effect patterns of these variables. For example, lower WBC counts contributed to increased risk of myelosuppression. The SHAP force plot (Figure 6D) displayed the direction and magnitude of each variable's effect in a randomly selected individual patient, providing a clear explanation of the model's prediction mechanism at the individual level.

Prediction model for myelosuppression in lung cancer chemotherapy



Prediction model for myelosuppression in lung cancer chemotherapy

Figure 6. Contribution analysis and interpretation of the Random Forest Model. A: SHAP Bee Swarm Plot. B: Random Forest Model Feature Importance Ranking. C: SHAP Dependence Plot. D: SHAP Force Plot. Note: SHAP, Shapley Additive Explanations; BMI, Body Mass Index.

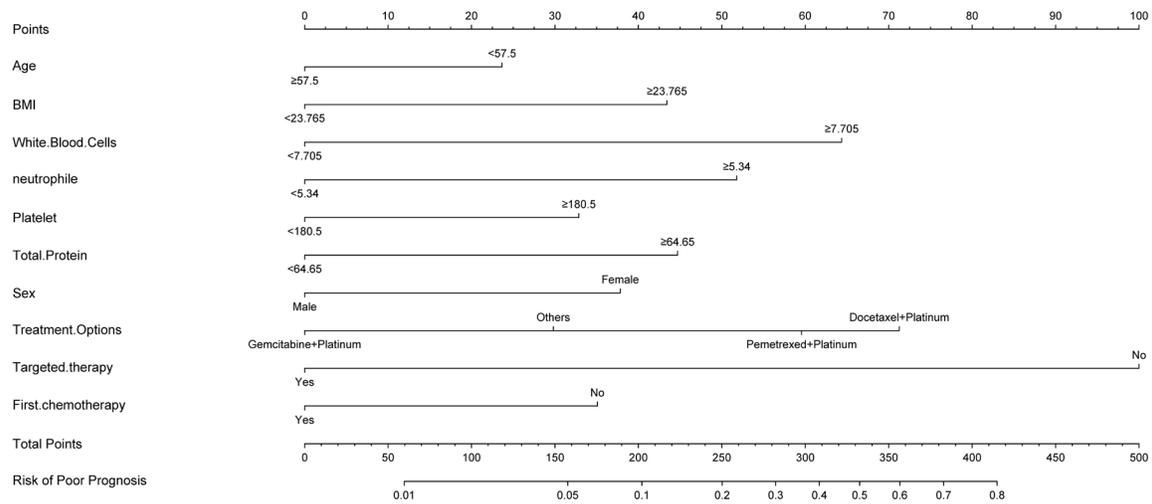


Figure 7. Nomogram based on Random Forest model in predicting myelosuppression risk. Note: BMI, Body Mass Index.

We randomly selected samples 3, 20, and 36 for display, where sample 3 and sample 36 were myelosuppression patients, and sample 20 was a non-myelosuppression patient. The SHAP force plot illustrated each variable's specific contribution and direction, offering an intuitive interpretation for clinical predictions.

Nomogram construction and clinical usability of the model

A nomogram was developed based on the Random Forest model to provide an individualized predictive tool for assessing the risk of myelosuppression after platinum-based doublet chemotherapy. The nomogram included the following variables: age, BMI, WBC count, neutrophil count, platelet count, total protein, gender, treatment regimen (gemcitabine + platinum, other chemotherapy regimens, docetaxel + platinum, pemetrexed + platinum), targeted therapy, and first chemotherapy. Based on the variable contribution in the model, WBC count, platelet count, and BMI were strongly correlated with the risk of myelosuppression, while gender showed weaker correlations (**Figure 7**). By calculating the total score for each patient based on their variable scores, the risk of myelosuppression can be quantified, providing valuable reference for clinical decision-making.

Discussion

Myelosuppression is a prevalent and severe adverse reaction during chemotherapy in cancer patients, increasing the risk of complications like infection and bleeding and potentially leading to chemotherapy interruptions that adversely affect treatment outcomes and patient survival rates [18]. The mechanisms underlying myelosuppression are multifaceted, closely associated with patient's overall condition, chemotherapy regimen, and various individualized factors [18]. Consequently, accurately predicting the risk and implementing timely interventions is crucial. Literature indicates that the occurrence of febrile neutropenia (FN) in breast cancer patients is closely associated with the number of underlying risk factors [19]. Building on this, our study employed LASSO regression and identified 12 variables significantly linked to myelosuppression, encompassing demographic characteristics, hematological indicators, nutritional and metabolic status, medical history, and chemotherapy-related factors, thereby laying the groundwork for exploring the potential mechanisms of myelosuppression.

Demographic risk factors: age and gender

Age and gender are significant demographic risk factors for myelosuppression. Additionally,

Prediction model for myelosuppression in lung cancer chemotherapy

diminished liver and kidney functions in older individuals can impair the metabolism and excretion of chemotherapy drugs, exacerbating their toxic effects. Studies have shown that high-dose chemotherapy or radioactive iodine treatment may induce mild myelosuppression in the elderly, typically manageable at appropriate doses [20]. Chronic inflammatory conditions, such as “inflammaging”, further suppress bone marrow hematopoietic function in older patients.

Regarding gender, males are at a higher risk of myelosuppression compared to females. This disparity may stem from differences in hormone levels and drug metabolism. Male hormones might negatively affect the repair capacity of the hematopoietic microenvironment. Notably, certain drugs like Linezolid can induce myelosuppression in elderly patients, along with rare adverse reactions like hypoglycemia and hyponatremia, necessitating careful monitoring during treatment [21]. Comprehensive assessments for elderly patients, such as evaluating the suitability of HCT-ASCT regimens, can aid in determining the appropriateness of more intensive treatments, thereby optimizing outcomes and mitigating myelosuppression risk [22].

Nutritional and metabolic status: BMI and total protein

BMI and total protein levels are indicators of a patient’s nutritional status and are closely linked to myelosuppression risk. Patients with low BMI are more likely to experience malnutrition, which compromises immune function and bone marrow reserve capacity. Insufficient body weight can decrease the number of hematopoietic stem cells in the bone marrow, while inadequate protein and energy intake may impede bone marrow repair. Although chemotherapy dosages are typically weight-based, low-BMI patients may still experience excessive drug exposure, exacerbating myelosuppression despite dose reductions [23]. Oral nutritional supplements (ONS) have been shown to significantly reduce myelosuppression incidence in esophageal cancer patients with $BMI \leq 18.5$ kg/m² [24], highlighting the importance of optimizing nutritional support to enhance patient tolerance and reduce bone marrow toxicity.

Low total protein levels reflect poor nutritional status, impacting fluid balance, drug transport,

immune regulation, and tissue repair. Reduced total protein can increase the free fraction of chemotherapy drugs, heightening their toxic effects. Additionally, protein deficiency may impair the body’s ability to repair bone marrow damage, elevating myelosuppression risk [25]. Enteral nutrition interventions have demonstrated benefits in chemotherapy and radiotherapy patients by lowering myelosuppression rates and improving treatment completion rates and overall survival [26]. Furthermore, high BMI has been associated with reduced clonogenic potential of bone marrow-derived mesenchymal stem cells (BMSC) and altered osteogenic potential, affecting bone marrow function [27]. In medulloblastoma patients undergoing whole-brain and whole-spinal cord radiotherapy, low BMI was identified as a significant risk factor for myelosuppression, closely linked to treatment interruptions and severe bone marrow toxicity [28].

Hematological indicators: white blood cell, neutrophil, and platelet counts

WBC count, neutrophil count, and platelet count are critical hematological indicators for assessing bone marrow reserve function. Patients with lower baseline levels of these cells generally have poorer bone marrow reserves and are more susceptible to chemotherapy-induced suppression. In breast cancer studies, severe myelosuppression has been associated with better disease-free survival [29]. Chemotherapy agents primarily target high-proliferation bone marrow cells, and patients with already low baseline blood cell levels may experience profound reductions. Low WBC and neutrophil counts increase infection risks, while low platelet counts heighten bleeding risks, collectively reducing chemotherapy tolerance [29]. Therefore, dynamic monitoring of these indicators is essential for the timely identification and intervention of high-risk patients.

Baseline blood cell assessments before chemotherapy, coupled with preventive measures such as hematopoietic growth factor administration, can effectively mitigate chemotherapy-related myelosuppression risks. Pharmacokinetic prediction tools have been developed to assess neutropenia risk post-chemotherapy in NSCLC patients, guiding clinical interventions [30]. Studies have also linked the frequency of severe neutropenia (Grade 4) with adverse clin-

Prediction model for myelosuppression in lung cancer chemotherapy

ical outcomes like FN and infections, emphasizing the need for appropriate threshold settings to identify high-risk individuals [31]. In B-cell acute lymphoblastic leukemia, specific antibody therapies like blinatumomab, although generally not causing significant myelosuppression, have been associated with severe neutropenia, particularly in the second treatment cycle [32]. Additionally, radiotherapy combined with hormone therapy have shown hematological toxicity in prostate cancer patients, including neutropenia and thrombocytopenia, which increases the risk of myelosuppression when multiple treatment modalities are used simultaneously [33].

Chemotherapy regimens and their mechanisms on myelosuppression

Different chemotherapy regimens exert distinct mechanisms of action on myelosuppression. The docetaxel-platinum regimen, for instance, can significantly suppress hematopoietic stem cells by disrupting microtubule function and DNA synthesis. Platinum-based drugs, known for inducing DNA damage and oxidative stress, intensify bone marrow toxicity. Therefore, selecting chemotherapy regimens requires a careful balance between efficacy and toxicity, especially for patients with compromised baseline health. Zhang et al. [34] demonstrated variability in myelosuppression severity among esophageal cancer patients treated with different chemotherapy regimens. Specifically, combinations with oxaliplatin significantly reduced myelosuppression incidence, whereas combinations with nedaplatin and paclitaxel increased the risk. Adjusting platinum-based chemotherapy dosages and selecting appropriate combination drugs are effective strategies to mitigate myelosuppression.

Targeted therapies combined with chemotherapy can enhance tumor treatment outcomes but may also elevate myelosuppression risk. Some targeted agents inhibit angiogenesis or directly affect the bone marrow microenvironment, indirectly impairing hematopoietic function. For example, Lu et al. [35] found that the bevacizumab-like monoclonal antibody IBI305, when combined with immunotherapy and chemotherapy, significantly prolonged progression-free survival (PFS) in NSCLC patients but also increased neutropenia incidence. Similarly, Fan

et al. [36] reported that NSCLC patients unresponsive to EGFR-TKI therapy primarily experienced neutropenia when treated with pemetrexed-based chemotherapy. These findings underscore the necessity of optimizing treatment regimens based on patients' baseline conditions and adjusting dosages accordingly. When combining targeted therapy with chemotherapy, careful monitoring of bone marrow function and timely dosage adjustments, or the use of protective agents, are imperative. Additionally, Zhang et al. [37] observed that the Endostar-chemotherapy regimen could induce myelosuppression, with varying toxicity impacts depending on administration methods.

Model performance

In this study, the Random Forest model exhibited excellent predictive performance, achieving an AUC of 0.924 in the training group and 0.710 in the validation group, surpassing other models. The high AUC in the training group reflects the model's robust ability to differentiate between high-risk and low-risk patients. Although the AUC in validation group was lower, it remained within an acceptable range, indicating the model's generalizability. Dong et al. [38] similarly reported Random Forest models for predicting chemotherapy-induced myelosuppression with AUCs of 0.878 and 0.885 in training and validation groups, respectively, reinforcing Random Forest's efficacy in this domain. Zheng et al. [39] developed a Random Forest-based risk prediction model for chemotherapy-induced myelosuppression in esophageal cancer patients, achieving an AUC of 0.883 in the training group and demonstrating high sensitivity and specificity in the validation group. However, the lower AUC in the validation group suggests that further optimization is necessary to improve performance with new datasets.

Moreover, our model demonstrated a sensitivity of 92.31% and specificity of 82.20% in the training group, indicating high accuracy. In the validation group, AdaBoost and SVM models showed higher specificity but relatively lower sensitivity, highlighting performance differences across datasets. This suggests that in practical applications, balancing different models' strengths and weaknesses is essential. Li et al. [40] developed an XGBoost model that also

Prediction model for myelosuppression in lung cancer chemotherapy

showed outstanding predictive performance in both training and validation groups, with AUCs of 0.981 and 0.896, respectively, supporting the potential of machine learning models in predicting myelosuppression.

Clinical implications

This study utilized advanced machine learning algorithms, including Random Forest and XGBoost, to construct a predictive model for myelosuppression risk. This approach overcomes the limitations of traditional statistical methods by effectively managing complex, multidimensional data, thereby significantly enhancing predictive accuracy and stability. The incorporation of SHAP algorithms further improved model interpretability, enabling clinicians to understand each variable's contribution to the predictions, thus increasing the model's clinical applicability.

The comprehensive inclusion of demographic characteristics, nutritional and metabolic indicators, hematological indices, and chemotherapy-related factors allowed for a systematic evaluation of multiple risk factors. However, the study's retrospective, single-center design may introduce patient selection bias and incomplete data, limiting external validation and generalizability. Additionally, the exclusion of dynamic blood cell count changes and drug dosage adjustments may reduce the model's comprehensiveness and accuracy.

Future research should focus on expanding the sample size, conducting multi-center studies, and incorporating more dynamic variables to enhance the model's generalizability and predictive performance. The predictive model developed herein holds potential as a valuable tool for clinical decision-making, facilitating precise patient risk assessments, optimizing chemotherapy regimens, improving treatment outcomes, and advancing personalized and precision medicine.

Conclusion

This study developed an individualized risk prediction model for myelosuppression in lung cancer patients undergoing platinum-based doublet chemotherapy using machine learning techniques such as XGBoost. The model demonstrated robust predictive performance in

both training and validation groups, underscoring its potential utility in clinical settings for enhancing patient care and treatment efficacy.

Disclosure of conflict of interest

None.

Address correspondence to: Lu Zhang, Department of Radiation Oncology, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, No. 42 Wenhua West Road, Jinan 250011, Shandong, China. E-mail: zhanglu9874@outlook.com

References

- [1] Batra U, Prabhash K, Noronha V, Deshpande R, Khurana S, Bhat GM, Mistry R, Agarwala V, Rajpurohit S, Poladia B, Sharma M and Banday SZ. Prevalence of EGFR mutations in patients with resected stage I to III nonsquamous non-small cell lung cancer: results of India cohort. *JCO Glob Oncol* 2025; 11: e2400353.
- [2] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024; 74: 229-263.
- [3] Chen L, Zhu XZ, Zhao SJ and Yang QW. PD-L1 as a predictive factor for non-small-cell lung cancer prognosis. *Lancet Oncol* 2024; 25: e233.
- [4] Akinboro O, Drezner N, Amatya A, Runyan J, Fourie-Zirkelbach J, Zhao M, Bi Y, Korsah K, Mixer B, Tang S, Larkins E, Pazdur R, Beaver JA and Singh H. US food and drug administration approval summary: nivolumab plus platinum-doublet chemotherapy for the neoadjuvant treatment of patients with resectable non-small-cell lung cancer. *J Clin Oncol* 2023; 41: 3249-3259.
- [5] Nian Q, Liu R and Zeng J. Unraveling the pathogenesis of myelosuppression and therapeutic potential of natural products. *Phytomedicine* 2024; 132: 155810.
- [6] Zhigulev A, Norberg Z, Cordier J, Spalinskas R, Bassereh H, Björn N, Pradhananga S, Gréen H and Sahlén P. Enhancer mutations modulate the severity of chemotherapy-induced myelosuppression. *Life Sci Alliance* 2024; 7: e202302244.
- [7] Broekman MMTJ, Coenen MJH, Wanten GJ, van Marrewijk CJ, Klungel OH, Verbeek ALM, Hooymans PM, Guchelaar HJ, Scheffer H, Derijks LJJ, Wong DR and de Jong DJ. Risk factors for thiopurine-induced myelosuppression and

Prediction model for myelosuppression in lung cancer chemotherapy

- infections in inflammatory bowel disease patients with a normal TPMT genotype. *Aliment Pharmacol Ther* 2017; 46: 953-963.
- [8] Jiang N, Chen XC and Zhao Y. Analysis of the risk factors for myelosuppression after concurrent chemoradiotherapy for patients with advanced non-small cell lung cancer. *Support Care Cancer* 2013; 21: 785-791.
- [9] Tu X, Chen R, Huang G, Lu N, Chen Q, Bai X and Li B. Factors predicting severe myelosuppression and its influence on fertility in patients with low-risk gestational trophoblastic neoplasia receiving single-agent methotrexate chemotherapy. *Cancer Manag Res* 2020; 12: 4107-4116.
- [10] Machine-learning model makes predictions about network biology. *Nature* 2023; [Epub ahead of print].
- [11] Maharjan R, Kim KH, Lee K, Han HK and Jeong SH. Machine learning-driven optimization of mRNA-lipid nanoparticle vaccine quality with XGBoost/Bayesian method and ensemble model approaches. *J Pharm Anal* 2024; 14: 100996.
- [12] Ozkan J. Thinking outside the black box: CardioPulse takes a look at some of the issues raised by machine learning and artificial intelligence. *Eur Heart J* 2023; 44: 1007-1009.
- [13] Azodi CB, Tang J and Shiu SH. Opening the black box: interpretable machine learning for geneticists. *Trends Genet* 2020; 36: 442-455.
- [14] Park B, Kim CH, Jun JK, Suh M, Choi KS, Choi IJ and Oh HJ. A machine learning risk prediction model for gastric cancer with SHapley Additive exPlanations. *Cancer Res Treat* 2024; [Epub ahead of print].
- [15] Song Y, Zhang D, Wang Q, Liu Y, Chen K, Sun J, Shi L, Li B, Yang X, Mi W and Cao J. Prediction models for postoperative delirium in elderly patients with machine-learning algorithms and SHapley Additive exPlanations. *Transl Psychiatry* 2024; 14: 57.
- [16] Cao S, Wang S, Ma H, Tang S, Sun C, Dai J, Wang C, Shu Y, Xu L, Yin R, Song X, Chen H, Han B, Li Q, Wu J, Bai C, Chen J, Jin G, Hu Z, Lu D and Shen H. Genome-wide association study of myelosuppression in non-small-cell lung cancer patients with platinum-based chemotherapy. *Pharmacogenomics J* 2016; 16: 41-46.
- [17] Kkf C, Mitchell SA, Chan N, Ang E, Tam W and Kanesvaran R. Linguistic validation of the simplified Chinese version of the US National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE™). *BMC Cancer* 2020; 20: 1153.
- [18] Li Y, Bao Y, Zheng H, Qin Y and Hua B. A nomogram for predicting severe myelosuppression in small cell lung cancer patients following the first-line chemotherapy. *Sci Rep* 2023; 13: 17464.
- [19] Li E, Schroeder BK, Campbell D, Campbell K and Wang W. The impact of baseline risk factors on the incidence of febrile neutropenia in breast cancer patients receiving chemotherapy with pegfilgrastim prophylaxis: a real-world data analysis. *J Health Econ Outcomes Res* 2021; 8: 106-115.
- [20] Duskin-Bitan H, Leibner A, Amitai O, Diker-Cohen T, Hirsch D, Benbassat C, Shimon I and Robenshtok E. Bone-marrow suppression in elderly patients following empiric radioiodine therapy: real-life data. *Thyroid* 2019; 29: 683-691.
- [21] Singhania SVK, Shenoy S and Kapoor D. Linezolid-induced rare triad of hypoglycaemia, bone marrow suppression and hyponatraemia in elderly. *J Clin Pharm Ther* 2020; 45: 376-378.
- [22] Isbell LK, Uibelesen R, Friedl A, Burger E, Dopatka T, Scherer F, Orban A, Lauer E, Malenica N, Semenova I, Vreden A, Valk E, Wendler J, Neumaier S, Fricker H, El Rabih AAH, Gloggengießer C, Hilbig D, Bleul S, Weis J, Gmehlin D, Backenstrass M, Wirtz S, Ihorst G, Finke J, Illerhaus G and Schorb E. Age-adjusted high-dose chemotherapy followed by autologous stem cell transplantation or conventional chemotherapy with R-MP as first-line treatment in elderly primary CNS lymphoma patients - the randomized phase III PRIMA-CNS trial. *BMC Cancer* 2023; 23: 767.
- [23] Wei J, Xiang W, Wei H, Hu X, Lu Y and Dong X. Impact of nutrition risk index, prognostic nutritional index and skeletal muscle index on early myelosuppression of first-line chemotherapy in stage IV gastric cancer patients. *BMC Gastroenterol* 2024; 24: 452.
- [24] Li X, Dai T, Rao Z and Hu W. Impact of oral nutrition supplementation on outcomes of esophageal cancer patients treated with chemotherapy: a retrospective cohort study with propensity score matching. *Front Nutr* 2022; 9: 1004372.
- [25] Bian Y, Xie F, Han J and Ding Y. Nutritional evaluation study based on NRS 2002, OPNI, and their combined use in patients with adverse drug reactions after chemotherapy: a cross-sectional study. *Ann Transl Med* 2022; 10: 180.
- [26] Lyu J, Shi A, Li T, Li J, Zhao R, Zhu S, Wang J, Xing L, Yang D, Xie C, Shen L, Zhang H, Zhu G, Wang J, Pan W, Li F, Lang J and Shi H. Effects of enteral nutrition on patients with oesophageal carcinoma treated with concurrent chemoradiotherapy: a prospective, multicentre,

Prediction model for myelosuppression in lung cancer chemotherapy

- randomised, controlled study. *Front Oncol* 2022; 12: 839516.
- [27] Zong Q, Bundkirchen K, Neunaber C and Nock S. Effect of high BMI on human bone marrow-derived mesenchymal stromal cells. *Cell Transplant* 2024; 33: 9636897241226546.
- [28] Li Z, Lin Z, Liu H, Xiao R, Lin C, Zhu W, Luo J, Xu S, Chi F and He H. Treatment continuity and bone marrow suppression in whole-brain and whole-spinal cord radiotherapy for medulloblastoma patients. *Clin Med Insights Oncol* 2024; 18: 11795549241286431.
- [29] Niu YD, Zhang YW, Zhu RJ, Chu T, Wang L, Wang S, Li YY and Dong Y. The influence of various myelosuppression degrees during neoadjuvant chemotherapy on the curative effect and prognosis of triple-negative breast cancer. *Zhonghua Yi Xue Za Zhi* 2022; 102: 2290-2294.
- [30] Park K, Kim Y, Son M, Chae D and Park K. A Pharmacometric model to predict chemotherapy-induced myelosuppression and associated risk factors in non-small cell lung cancer. *Pharmaceutics* 2022; 14: 914.
- [31] Mohanlal R, Ogenstad S, Lyman GH, Huang L and Blayne DW. Grade 4 neutropenia frequency as a binary risk predictor for adverse clinical consequences of chemotherapy-induced neutropenia: a meta-analysis. *Cancer Invest* 2023; 41: 369-378.
- [32] Chen EC, Flamand Y, Tiao E, DeAngelo DJ and Luskin MR. Incidence, duration, and severity of neutropenia in adults with B-cell acute lymphoblastic leukemia receiving blinatumomab consolidation. *Leuk Lymphoma* 2025; 66: 64-71.
- [33] Katib Y, Tisseverasinghe S, Gerard IJ, Royal-Preyra B, Chaddad A, Sasson T, Bahoric B, Roncarolo F and Niazi T. Evaluating the effects of prostate radiotherapy intensified with pelvic nodal radiotherapy and androgen deprivation therapy on myelosuppression: single-institution experience. *Curr Oncol* 2024; 31: 5439-5451.
- [34] Zhang QL, Wu TT, Han Y, Zheng ZM and Zhang Y. Chemotherapy-induced myelosuppression in esophageal cancer patients: risks and suggestions for its management. *Curr Med Sci* 2022; 42: 530-537.
- [35] Lu S, Wu L, Jian H, Chen Y, Wang Q, Fang J, Wang Z, Hu Y, Sun M, Han L, Miao L, Ding C, Cui J, Li B, Pan Y, Li X, Ye F, Liu A, Wang K, Cang S, Zhou H, Sun X, Ferry D, Lin Y, Wang S, Zhang W and Zhang C. Sintilimab plus bevacizumab biosimilar IBI305 and chemotherapy for patients with EGFR-mutated non-squamous non-small-cell lung cancer who progressed on EGFR tyrosine-kinase inhibitor therapy (ORIENT-31): first interim results from a randomised, double-blind, multicentre, phase 3 trial. *Lancet Oncol* 2022; 23: 1167-1179.
- [36] Fan C, Jiang Z, Teng C, Song X, Li L, Shen W, Jiang Q, Huang D, Lv Y, Du L, Wang G, Hu Y, Man S, Zhang Z, Gao N, Wang F, Shi T and Xin T. Efficacy and safety of intrathecal pemetrexed for TKI-failed leptomeningeal metastases from EGFR+ NSCLC: an expanded, single-arm, phase II clinical trial. *ESMO Open* 2024; 9: 102384.
- [37] Han B, Kang Y, Wang H, Wang J, Shen R, Liu S, Lu L, Sun Z and Zhang N. A retrospective study on the efficacy and safety of endostar with chemotherapy in EGFR-TKI-resistant NSCLC. *BMC Pulm Med* 2023; 23: 437.
- [38] Dong Y, Hu C, Liu J and Lv H. Construction of an auxiliary scoring model for myelosuppression in patients with lung cancer chemotherapy based on random forest algorithm. *Am J Transl Res* 2023; 15: 4155-4163.
- [39] Zheng Z, Zhang Q, Han Y, Wu T and Zhang Y. Predictive model of chemotherapy-induced myelosuppression for patients with esophageal cancer. *Cancer Control* 2022; 29: 10732748221126929.
- [40] Li M, Wang Q, Lu P, Zhang D, Hua Y, Liu F, Liu X, Lin T, Wei G and He D. Development of a machine learning-based prediction model for chemotherapy-induced myelosuppression in children with Wilms' tumor. *Cancers (Basel)* 2023; 15: 1078.

Prediction model for myelosuppression in lung cancer chemotherapy

Table S1. Comparison of baseline characteristics between myelosuppression and non-myelosuppression patients after platinum-based doublet chemotherapy

Variable	Total (N=584)	Myelosuppression (n=412)	Non-Myelosuppression (n=172)	Statistic Value	P Value
Gender					
Male	461 (78.94%)	342 (83.01%)	119 (69.19%)	13.947	<0.001
Female	123 (21.06%)	70 (16.99%)	53 (30.81%)		
Smoking History					
Yes	355 (60.79%)	260 (63.11%)	95 (55.23%)	3.156	0.076
No	229 (39.21%)	152 (36.89%)	77 (44.77%)		
Hypertension History					
Yes	165 (28.25%)	115 (27.91%)	50 (29.07%)	0.080	0.777
No	419 (71.75%)	297 (72.09%)	122 (70.93%)		
Diabetes History					
Yes	60 (10.27%)	45 (10.92%)	15 (8.72%)	0.638	0.424
No	524 (89.73%)	367 (89.08%)	157 (91.28%)		
Coronary Artery Disease History					
Yes	21 (3.60%)	12 (2.91%)	9 (5.23%)	1.884	0.170
No	563 (96.40%)	400 (97.09%)	163 (94.77%)		
Pathology Type					
Non-small cell lung cancer	447 (76.54%)	309 (75.00%)	138 (80.23%)	1.850	0.174
Small cell lung cancer	137 (23.46%)	103 (25.00%)	34 (19.77%)		
Treatment Regimen					
Docetaxel + Platinum	173 (29.62%)	111 (26.94%)	62 (36.05%)	18.069	<0.001
Gemcitabine + Platinum	114 (19.52%)	95 (23.06%)	19 (11.05%)		
Pemetrexed + Platinum	156 (26.71%)	99 (24.03%)	57 (33.14%)		
Other	141 (24.14%)	107 (25.97%)	34 (19.77%)		
Targeted Therapy					
Yes	69 (11.82%)	62 (15.05%)	7 (4.07%)	14.038	<0.001
No	515 (88.18%)	350 (84.95%)	165 (95.93%)		
First Chemotherapy					
Yes	318 (54.45%)	239 (58.01%)	79 (45.93%)	7.139	0.008
No	266 (45.55%)	173 (41.99%)	93 (54.07%)		
Postoperative Adjuvant Chemotherapy					
Yes	250 (42.81%)	173 (41.99%)	77 (44.77%)	0.382	0.536
No	334 (57.19%)	239 (58.01%)	95 (55.23%)		
Bone Metastasis					
Yes	68 (11.64%)	49 (11.89%)	19 (11.05%)	0.085	0.771
No	516 (88.36%)	363 (88.11%)	153 (88.95%)		
Age (years)	62.90±7.77	63.41±7.25	61.69±8.80	-2.450	0.015
BMI (kg/m ²)	22.20±3.18	21.76±3.06	23.25±3.23	5.281	<0.001
White Blood Cells (10 ⁹ /L)	6.45±2.53	6.01±2.30	7.49±2.77	6.647	<0.001
Neutrophils (10 ⁹ /L)	4.19 [2.76, 5.54]	4.01 [2.63, 5.26]	4.64 [3.20, 6.17]	3.740	<0.001
Red Blood Cells (10 ¹² /L)	4.13±0.54	4.10±0.56	4.19±0.48	1.742	0.082
Hemoglobin (g/L)	124.34±17.74	125.10±17.58	122.52±18.02	-1.601	0.110
Platelets (10 ⁹ /L)	227.61±81.88	218.01±80.30	250.62±81.27	4.457	<0.001
Total Protein (g/L)	67.10±6.16	66.66±6.16	68.18±6.06	2.735	0.006
Albumin (g/L)	38.15±4.50	38.09±4.39	38.28±4.75	0.471	0.638

Note: Body Mass Index (BMI); "Other" in treatment regimen includes etoposide + platinum and paclitaxel + platinum.