



## OPEN ACCESS

## EDITED BY

Enrico Pozzo,  
Humanitas Research Hospital, Italy

## REVIEWED BY

Paulo Rodrigues-Santos,  
University of Coimbra, Portugal  
Wang Dong,  
Central South University, China

## \*CORRESPONDENCE

Nan Zhou

✉ fcczhoun@zzu.edu.cn

Xinchang Lu

✉ luc999@163.com

<sup>†</sup>These authors have contributed  
equally to this work

RECEIVED 27 October 2025

REVISED 03 January 2026

ACCEPTED 05 January 2026

PUBLISHED 28 January 2026

## CITATION

Li L, Liu J, Pang S, Zhao Y, Wang Y, Wen J,  
Liu Y, Zhang Y, Zhang Y, Li J, Zhou N and  
Lu X (2026) A machine learning-driven  
prognostic model based on peripheral blood  
lymphocyte subsets in osteosarcoma.  
*Front. Immunol.* 17:1733518.  
doi: 10.3389/fimmu.2026.1733518

## COPYRIGHT

© 2026 Li, Liu, Pang, Zhao, Wang, Wen, Liu,  
Zhang, Zhang, Li, Zhou and Lu. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License  
\(CC BY\)](#). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# A machine learning-driven prognostic model based on peripheral blood lymphocyte subsets in osteosarcoma

Longqing Li<sup>1†</sup>, Jinlei Liu<sup>1†</sup>, Songtao Pang<sup>1</sup>, Yuan Zhao<sup>1</sup>,  
Yimeng Wang<sup>2</sup>, Jia Wen<sup>1</sup>, Yongkui Liu<sup>1</sup>, Yi Zhang<sup>1</sup>, Yan Zhang<sup>1</sup>,  
Jiazhen Li<sup>1</sup>, Nan Zhou<sup>1\*</sup> and Xinchang Lu<sup>1\*</sup>

<sup>1</sup>Department of Orthopedics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China,

<sup>2</sup>Department of Obstetrics and Gynecology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

**Background:** The prognosis of osteosarcoma (OS) remains heterogeneous, and the prognostic value of peripheral blood lymphocyte subsets, analyzed through machine learning (ML), is not fully explored. This study aimed to develop an ML-based prognostic model using lymphocyte subset data to improve risk stratification for OS patients.

**Methods:** We retrospectively analyzed data from 65 high-grade OS patients. Peripheral blood lymphocyte subsets were quantified by flow cytometry prior to treatment. Seven algorithms, including stepwise Cox, LASSO, and five ML models (RSF, GBM, XGBoost, SVM, KNN), were employed to construct prognostic models. Model performance was evaluated using the C-index, AUC, and validated via bootstrap and cross-validation.

**Results:** The Gradient Boosting Machine (GBM) algorithm yielded the optimal two-variable model, incorporating CD3<sup>-</sup>CD56<sup>+</sup> NK cells and CD8<sup>+</sup>HLA-DR<sup>+</sup> activated cytotoxic T cells (AUC = 0.959). The resulting gbm\_risk\_score was an independent prognostic factor (HR = 14.516, P = 0.012) and effectively stratified patients into significantly divergent survival groups (P < 0.001). Importantly, the gbm\_risk\_score demonstrated superior predictive performance for 3-year OS compared to traditional inflammatory indices, neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR). A nomogram integrating the GBM risk group and primary metastasis status demonstrated excellent predictive accuracy (C-index: 0.883) and clinical utility, successfully identifying a high-risk subgroup among initially non-metastatic patients.

**Conclusion:** We developed and validated a robust ML-driven prognostic model based on peripheral blood lymphocyte subsets. This model, demonstrating superior prognostic value over conventional inflammatory markers, provides a novel and practical tool for personalized risk assessment in OS, potentially guiding more tailored treatment strategies.

## KEYWORDS

machine learning, osteosarcoma, peripheral blood lymphocyte subsets, prognostic model, risk stratification

## 1 Introduction

Osteosarcoma (OS) is a primary malignant bone tumor characterized by the production of osteoid tissue, most commonly occurring in the metaphysis of long bones in adolescents aged 10–20 years (1). Prior to the 1970s, treatment relied primarily on surgery alone, with poor outcomes; up to 80% of patients ultimately died from pulmonary metastases, and the 5-year overall survival (OS) rate was less than 20%. The landscape changed significantly by the 1980s. A comprehensive treatment protocol incorporating preoperative chemotherapy, surgery, and postoperative chemotherapy was established as the standard of care. This multimodal approach increased the 5-year OS rate to over 60% for patients without metastases at initial diagnosis (2, 3). However, the management of OS continues to pose significant challenges. Approximately 15% of patients present with metastases at diagnosis, conferring a very poor prognosis. Furthermore, about 30% of patients without metastases initially may develop metastases during treatment, leading to a substantial decline in survival (4). While novel therapies, such as targeted treatments and immunotherapy, have been applied for OS in recent years, their efficacy in improving outcomes for patients with metastatic disease remains limited (5, 6). Consequently, the Children's Oncology Group (COG) suggests that the early identification of high-risk patients and the development of individualized treatment plans may represent another crucial pathway for improving prognosis (7).

Lymphocytes, as core components of the immune system, play an increasingly recognized role in tumor initiation, progression, and control. Multiple studies have shown that the presence of tumor-infiltrating lymphocytes (TILs), particularly CD8+ T cells, in tumor tissue is generally associated with a more favorable prognosis (8, 9). Similarly, biomarkers derived from peripheral blood lymphocyte counts, such as the neutrophil-to-lymphocyte ratio (NLR), have been established as independent prognostic factors for various solid tumors, including OS, with a lower pre-treatment NLR often correlating with better survival outcomes (10–12). Advances in lymphocyte research have revealed the complex and often opposing functions of different lymphocyte subsets in tumor progression. For instance, while CD8+ T cells can directly kill tumor cells, regulatory T cells (Tregs) may suppress immune responses and potentially promote tumor growth (13, 14). However, conventional blood tests typically report only the total lymphocyte count and cannot differentiate between subsets such as T cells, B cells, and natural killer (NK) cells. Therefore, peripheral blood lymphocyte subset enumeration holds significant potential. It can provide a more comprehensive reflection of a patient's systemic immune status, enabling a more authentic and objective assessment of overall immune activity. This is crucial for precise prognosis identification and guiding treatment strategies (15).

Although peripheral blood lymphocyte subset analysis can simultaneously measure over ten lymphocyte subsets in a patient's blood, providing rich data for assessing systemic immune status, the processing and interpretation of the resulting multidimensional and complex data remain a significant challenge (16). In this context, machine learning (ML), as a powerful artificial

intelligence (AI) method, has been demonstrated to significantly enhance the analysis and interpretation of complex cancer datasets (17, 18). A key advantage of ML methods is their ability to utilize relatively limited data to perform in-depth analysis of the complex biology of diseases and achieve personalized risk stratification, thereby holding promise for improving patient treatment outcomes (19, 20). Despite this clear potential, the application of ML for the deep analysis of OS-specific lymphocyte subset data has not yet been fully explored. This gap extends to the subsequent development of novel biomarkers. Such an endeavor highlights substantial research potential and value in this field.

Based on the aforementioned background, this study aims to construct a prognostic prediction model by integrating peripheral blood lymphocyte subset data with clinicopathological characteristics of OS patients using an ML algorithm. This research seeks to systematically evaluate the independent and synergistic prognostic value of different lymphocyte subsets and explore their combined utility with established clinical indicators. Ultimately, we intend to develop an effective tool capable of accurately identifying high-risk patients and informing individualized treatment strategies, thereby contributing to the advancement of precision medicine practice.

## 2 Patients and methods

### 2.1 Patients

We retrospectively collected the clinical data of OS patients treated at the Musculoskeletal Tumor Center of The First Affiliated Hospital of Zhengzhou University between May 2020 and May 2023. Patient inclusion and exclusion were based on the following criteria: Inclusion criteria were 1) patients with histopathologically confirmed high-grade OS; 2) patients who underwent lymphocyte subset testing before neoadjuvant chemotherapy (NAC); and 3) patients who completed the standard treatment protocol at our hospital. Exclusion criteria were 1) patients with histopathologically confirmed low-grade OS (intramedullary or surface variants) or parosteal osteosarcoma; 2) patients who did not receive the standard treatment (e.g., misdiagnosed, improperly treated, or failed to complete postoperative chemotherapy); 3) patients with hematological diseases; and 4) patients with other malignancies. Ultimately, 65 patients who met all inclusion criteria and did not meet any exclusion criteria were enrolled in the study. All enrolled patients were regularly followed up until death or September 2025, whichever occurred first.

### 2.2 Flow cytometric analysis of peripheral blood lymphocyte subsets

The procedure for peripheral blood lymphocyte subset analysis in this study was conducted as follows. First, prior to the initiation of treatment, blood samples were collected via peripheral venipuncture from patients with pathologically confirmed high-grade OS. Subsequently, the whole blood was

treated with a red blood cell (RBC) lysis buffer to remove erythrocytes, resulting in a leukocyte suspension. This leukocyte suspension was then stained strictly following the manufacturer's protocol using a panel of fluorescently labeled monoclonal antibodies (mAbs) to specifically target distinct lymphocyte surface antigens. The stained samples were analyzed using a BD FACS Canto II flow cytometer for detection and data acquisition. Finally, specialized software was employed to analyze the data, obtaining and recording the absolute counts and/or relative percentages of the following key lymphocyte subsets: CD45+ leukocytes, CD3+ T cells, CD3+CD4+ helper T cells (CD4+ T cells), CD3+CD8+ cytotoxic T cells (CD8+ T cells), CD3-CD56+ NK cells, as well as subsets indicative of T-cell activation and functional status, including CD4+CD28+, CD8+CD28+, CD4+CD38+, CD8+CD38+, CD4+HLA-DR+, and CD8+HLA-DR+ cells.

## 2.3 Development, comparison, and selection of prognostic models

This study integrated traditional statistical methods with various ML algorithms to develop a prognostic prediction model. During the model construction phase, we employed not only stepwise Cox regression models and least absolute shrinkage and selection operator (LASSO) Cox regression as baseline methods but also introduced five ML algorithms for comparative analysis: Random Survival Forest (RSF), Gradient Boosting Machine (GBM), XGBoost, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). The predictive accuracy of these seven algorithms was systematically evaluated using the concordance index (C-index) and area under the curve (AUC) metrics. Particular emphasis was placed on the clinical utility of the models—prioritizing the model with the fewest variables to balance complexity and practicality, provided predictive performance was maintained. To ensure model reliability, a multiple validation strategy was adopted: Bootstrap validation with 100 resampling iterations was used to assess model stability, 5-fold cross-validation (CV) was employed to evaluate generalizability, and time-dependent receiver operating characteristic (ROC) curve analysis was applied to examine predictive performance at different time points. Ultimately, based on the three dimensions of predictive accuracy, clinical utility, and model reliability, the optimal algorithm for analyzing and interpreting peripheral blood lymphocyte subset data in OS patients was selected. Following model development, the continuous predictive scores were dichotomized into 'low-risk' and 'high-risk' groups using the median value as the cutoff to facilitate clinical interpretation and application. Partial R code is provided in the [Supplementary Materials](#).

## 2.4 Development and validation of the predictive nomogram

A nomogram was constructed based on the independent risk factors identified. First, its discriminatory ability was quantified using the concordance index (C-index). Subsequently, calibration curves were plotted to assess the agreement between predicted

probabilities and observed outcomes. Finally, decision curve analysis (DCA) was applied to estimate the clinical net benefit and, consequently, the potential clinical utility of the nomogram.

## 2.5 Comparison of the model's predictive performance with NLR and PLR

To evaluate and compare prognostic indicators, we calculated NLR and PLR for all patients. Kaplan-Meier survival analysis was used to compare overall survival between high- and low-risk groups (stratified by the median value) for each inflammatory index. This was followed by univariate Cox regression analysis to assess the association of NLR and PLR (as continuous variables and as risk groups) with overall survival in osteosarcoma. Finally, ROC analysis was performed to compare the predictive capabilities of these traditional markers with our model-derived score.

## 2.6 Statistical analysis

All statistical analyses and visualizations in this study were performed using R software (version 4.4.2). The normality of continuous variables was assessed using the Shapiro-Wilk test. Data conforming to a normal distribution are presented as mean  $\pm$  standard deviation (mean  $\pm$  SD), while non-normally distributed (skewed) data are expressed as median (range). For group comparisons, appropriate statistical tests were selected based on the distribution characteristics of the data: the independent samples t-test was used for normally distributed data, and the Mann-Whitney U test was employed for non-normally distributed data. Categorical data are presented as actual counts (frequencies), and comparisons between groups were performed using the Chi-square test or Fisher's exact test, as appropriate. Overall survival (OS) curves were generated using the Kaplan-Meier (K-M) method and compared using the log-rank test. A two-tailed test was adopted for all analyses, and a P-value  $< 0.05$  was considered statistically significant.

# 3 Results

## 3.1 Patient characteristics

A total of 65 patients with OS were included in this study, comprising 30 males and 35 females, with a median age of 16 years (range: 7–65 years). Regarding tumor characteristics, the primary tumor was located in the lower limbs in 59 patients and in the upper limbs in 6 patients. At initial diagnosis, 43 patients had a tumor size  $< 8$  cm, while 22 patients had a tumor size  $\geq 8$  cm. Metastasis was present at diagnosis in 13 patients, whereas 52 patients showed no evidence of metastasis. Pathological fracture was identified at initial presentation in 6 patients. Based on body mass index (BMI) assessment, 37 patients were classified as having a normal BMI, and 28 were classified as abnormal. By the last follow-up, 17 patient deaths had been recorded ([Table 1](#)).

## 3.2 Characterization of peripheral blood lymphocyte subsets

Through flow cytometric analysis, this study classified lymphocytes into 11 subsets based on their surface marker expression profiles, comprising the total lymphocyte percentage, T-cell percentage, NK-cell percentage, and eight functional T-cell subset percentages. **Figure 1A** illustrates the correlations among the 11 lymphocyte subsets. Comparative analysis revealed that the proportions of CD4<sup>+</sup>CD38<sup>+</sup> and CD8<sup>+</sup>CD38<sup>+</sup> subsets were higher in the deceased group than in the survival group, whereas the majority of the remaining lymphocyte subsets were lower in the deceased group (**Figures 1B–D**). It is noteworthy that, among the subsets showing numerical differences between the two groups, the differences in five subsets—CD45<sup>+</sup>, CD3<sup>+</sup>, CD3<sup>+</sup>CD4<sup>+</sup>, CD4<sup>+</sup>CD38<sup>+</sup>, and CD8<sup>+</sup>CD38<sup>+</sup>—did not reach statistical significance. Patients were stratified into high-expression and low-expression groups based on the median value of each lymphocyte subset, and K-M survival curves were plotted for comparison. The analysis revealed that patients with higher proportions of CD3<sup>+</sup>CD56<sup>+</sup> NK cells, CD4<sup>+</sup>HLA-DR<sup>+</sup> activated helper T cells, CD8<sup>+</sup>CD28<sup>+</sup> proliferative cytotoxic T cells, and CD8<sup>+</sup>HLA-DR<sup>+</sup> activated cytotoxic T cells in peripheral blood exhibited significantly prolonged overall survival (**Figures 1B–D**,  $P < 0.05$ ).

## 3.3 Feature selection and performance of the GBM prognostic model

This study initially employed stepwise regression and LASSO regression as baseline models, which were compared against five

**TABLE 1** Demographic and clinical characteristics of osteosarcoma patients (n=65).

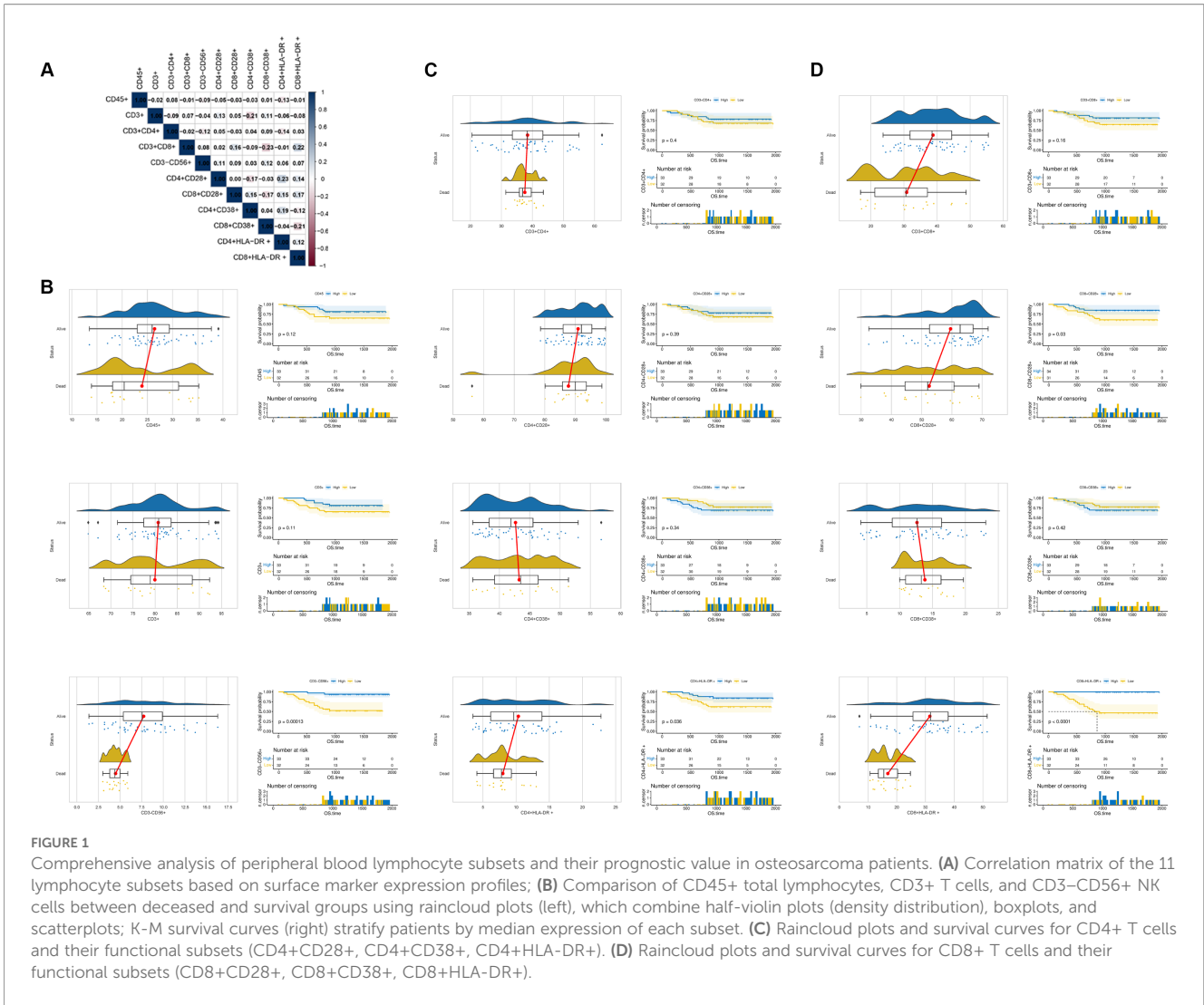
Characteristic	Category	Patients (n, %)
Gender	Male	30 (46.2%)
	Female	35 (53.8%)
Median Age (Range)		16 years (7–65)
Location	Lower Limbs	59 (90.8%)
	Upper Limbs	6 (9.2%)
Tumor size	< 8 cm	43 (66.2%)
	≥ 8 cm	22 (33.8%)
Primary metastases	Yes	13 (20.0%)
	No	52 (80.0%)
Pathological fracture	Yes	6 (9.2%)
	No	59 (90.8%)
BMI	Normal	37 (56.9%)
	Abnormal	28 (43.1%)

ML algorithms: RSF, XGBoost, GBM, SVM, and KNN. By systematically applying five machine learning algorithms to evaluate the variable importance of 11 peripheral blood lymphocyte subsets, this study elucidated the prognostic value of key subsets. Notably, CD8<sup>+</sup>HLA-DR<sup>+</sup> activated cytotoxic T cells consistently ranked first in variable importance across all five algorithms, underscoring their robustness as a core prognostic biomarker. The CD3<sup>+</sup>CD56<sup>+</sup> NK cell subset ranked second in four algorithms (GBM, KNN, SVM, and RSF) but third in the XGBoost algorithm. Furthermore, the CD45<sup>+</sup> total leukocyte subset ranked second in the XGBoost algorithm, third in the GBM, KNN, and SVM algorithms, and dropped to fourth in the RSF algorithm, reflecting fluctuations in its importance across different models (**Figures 2A–F**). These findings collectively highlight the critical role of multi-algorithm consensus in screening robust feature subsets to enhance the reliability of prognostic models.

Through systematic evaluation of predictive performance from single-variable to five-variable combination models, it was found that the two-variable models constructed by LASSO, SVM, XGBoost, GBM, and KNN all achieved AUC values above 0.90. Among these, the GBM two-variable model demonstrated the best performance (AUC = 0.9591), whereas the two-variable models for RSF and stepwise regression yielded AUC values of 0.837 and 0.7897, respectively (**Figure 3A**). Further analysis indicated that, except for Akaike information criterion (AIC) and RSF, increasing the number of variables provided limited improvement in AUC for the other algorithms. Consequently, balancing predictive power and model simplicity, the GBM two-variable model was ultimately selected for analyzing OS lymphocyte subset data. The features incorporated into the model were the CD3<sup>+</sup>CD56<sup>+</sup> cell subset and the CD8<sup>+</sup>HLA-DR<sup>+</sup> cell subset, with the model parameters set as  $n.trees = 50$ ,  $interaction.depth = 3$ , and  $shrinkage = 0.1$ . Both Bootstrap validation and 5-fold CV confirmed the model's good stability (**Figures 3B, C**). Furthermore, the *gbm\_risk\_score* derived from the model showed a significant negative correlation relationship with overall patient survival time (**Figure 3D**,  $R = -0.68$ ,  $P < 0.001$ ). To evaluate the clinical relevance of the *gbm\_risk\_score*, this study further investigated its relationship with key clinical variables. The analysis revealed that patients with metastases at initial diagnosis had a significantly higher *gbm\_risk\_score* compared to those without metastases ( $P = 0.001$ ). However, no statistically significant differences in the score were observed across subgroups defined by other clinical variables, such as age, gender, or tumor location (**Figures 3E, F**).

## 3.4 The GBM risk score as a powerful and independent prognostic indicator

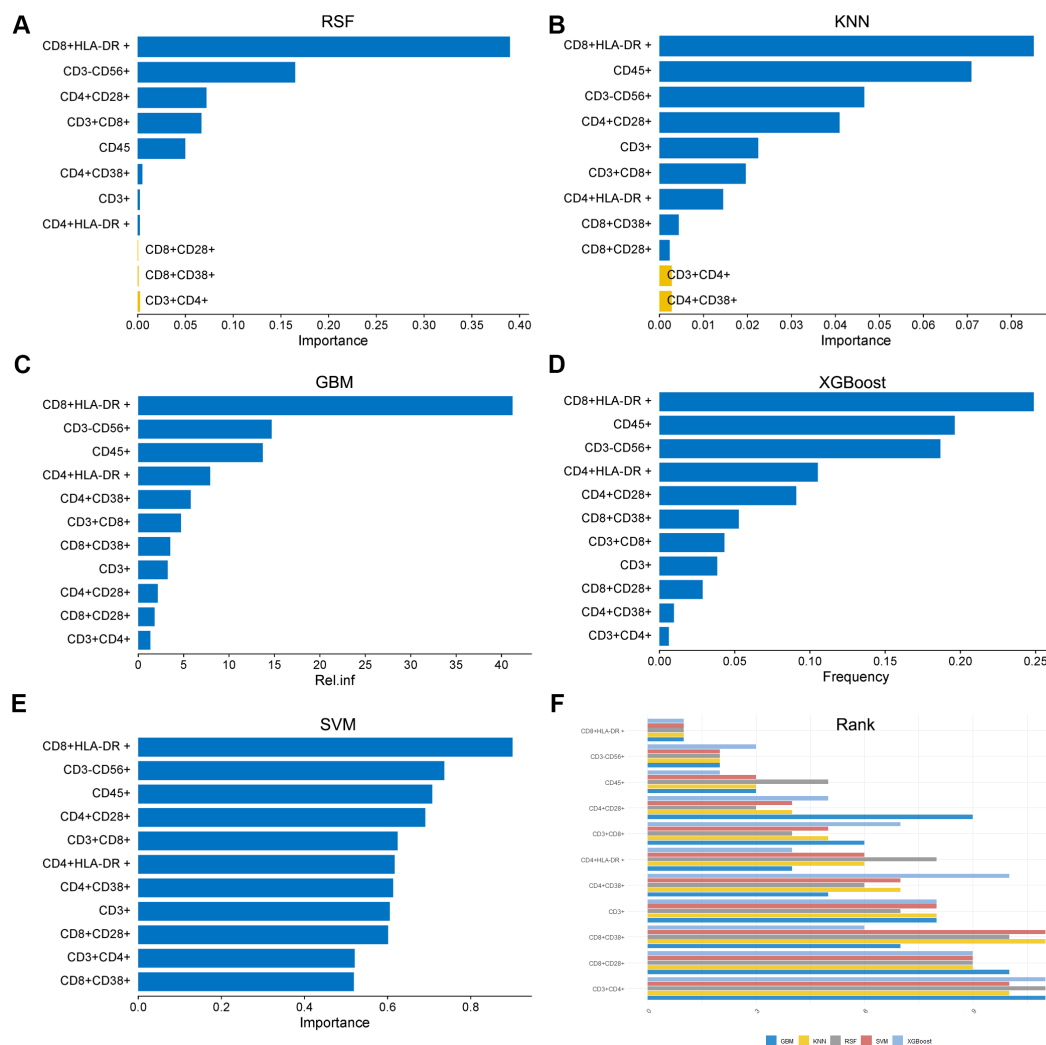
Analysis using restricted cubic spline interpolation demonstrated a linear relationship between the *gbm\_risk\_score* and the risk of overall survival in osteosarcoma patients ( $P = 0.133$ ). Subsequently, patients were stratified into low-risk and high-risk groups based on the median value of this score. K-M



survival analysis results clearly indicated that the overall survival time was significantly shorter in the high-risk group compared to the low-risk group ( $P < 0.001$ ), validating the definite discriminatory ability of the *gbm\_risk\_score* in clinical prognostic stratification (Figures 4A, B). To explore the potential clinical application value of the GBM risk score, this study conducted a comprehensive analysis integrating it with various clinical characteristics. Univariate Cox regression analysis revealed that the GBM risk grouping (hazard ratio [HR] = 21.437,  $P = 0.003$ ), metastasis status at initial diagnosis (HR = 7.934,  $P < 0.001$ ), and pathological fracture (HR = 3.257,  $P = 0.039$ ) were significantly associated with the OS of OS patients. Subsequently, multivariate Cox regression analysis further confirmed that the GBM risk grouping (HR = 14.516,  $P = 0.012$ ) and metastasis status at initial diagnosis (HR = 7.831,  $P = 0.002$ ) were independent risk factors affecting patient prognosis (Figures 4C, D). These findings indicate that the GBM risk score, based on lymphocyte subsets, holds prognostic predictive value independent of traditional clinical indicators.

### 3.5 Development and clinical validation of a prognostic nomogram

Subsequently, a prognostic nomogram prediction model was developed based on the GBM risk grouping and primary metastasis status (Figure 5A). This model demonstrated excellent predictive performance, achieving a C-index of 0.883, indicating high discriminative accuracy (Figure 5B). Furthermore, the calibration curve confirmed that the nomogram’s predicted 3-year OS probabilities showed the best agreement with the actual observed outcomes (Figure 5C). DCA revealed that the incorporation of the GBM risk grouping provided significant clinical net benefit across a wide range of threshold probabilities (Figures 5D, E). Additionally, by combining the GBM risk grouping with primary metastasis status in K-M survival curves, a patient subgroup with a higher risk of mortality was successfully identified among those without metastases at initial diagnosis, thereby highlighting the model’s refined risk-stratification capability (Figure 5F).



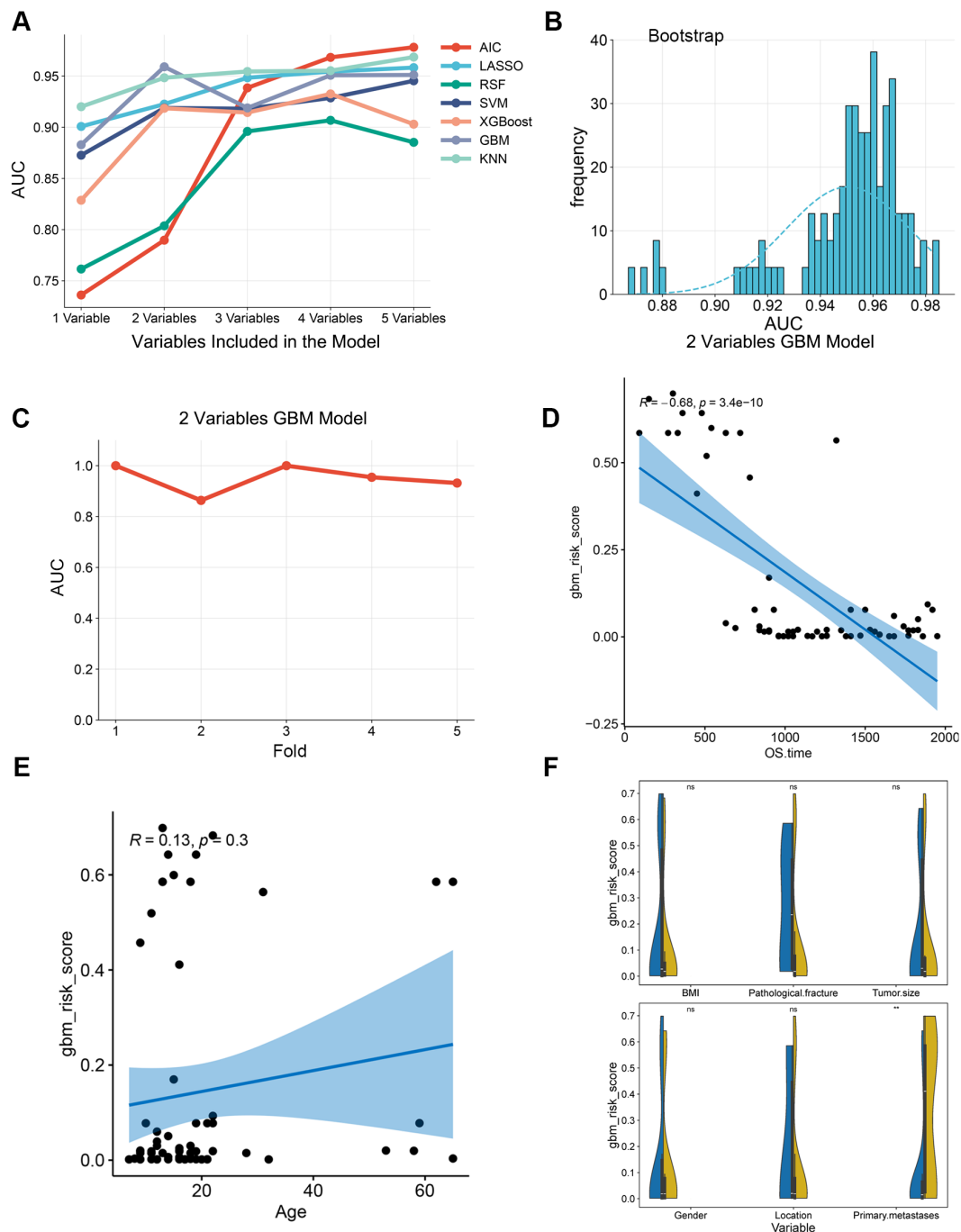
**FIGURE 2** Variable importance of peripheral blood lymphocyte subsets evaluated by five ML algorithms for prognostic prediction in osteosarcoma. (A) RSF; (B) KNN; (C) GBM; metric: relative influence [Rel.inf]; (D) XGBoost (metric: Frequency); (E) SVM. In bar plots (A–E), blue bars indicate positive importance values, while yellow bars (where applicable) represent negative contributions. All subsets in panels (C) showed exclusively positive values. (F) Consolidated ranking of the 11 lymphocyte subsets across all five algorithms.

### 3.6 The GBM risk score demonstrates superior predictive performance to NLR and PLR

As shown in Figures 6A, B, Kaplan-Meier analysis confirmed that patients in the high-risk groups for both NLR and PLR exhibited shorter OS than those in the corresponding low-risk groups (NLR:  $P = 0.012$ ; PLR:  $P = 0.018$ ). Univariate Cox analysis indicated that NLR ( $HR = 1.37$ ,  $P = 0.002$ ), PLR ( $HR = 1.01$ ,  $P = 0.023$ ), and their dichotomized risk categories were all significantly associated with OS in patients with osteosarcoma (Figure 6C). ROC analysis revealed that both the continuous *gbm\_risk\_score* and its dichotomized risk classification had significantly better predictive performance for OS than either the NLR or the PLR in predicting 3-year OS (Figure 6D). These results suggest that the lymphocyte subset-based GBM risk score may offer superior prognostic value compared to traditional peripheral blood inflammatory indices.

## 4 Discussion

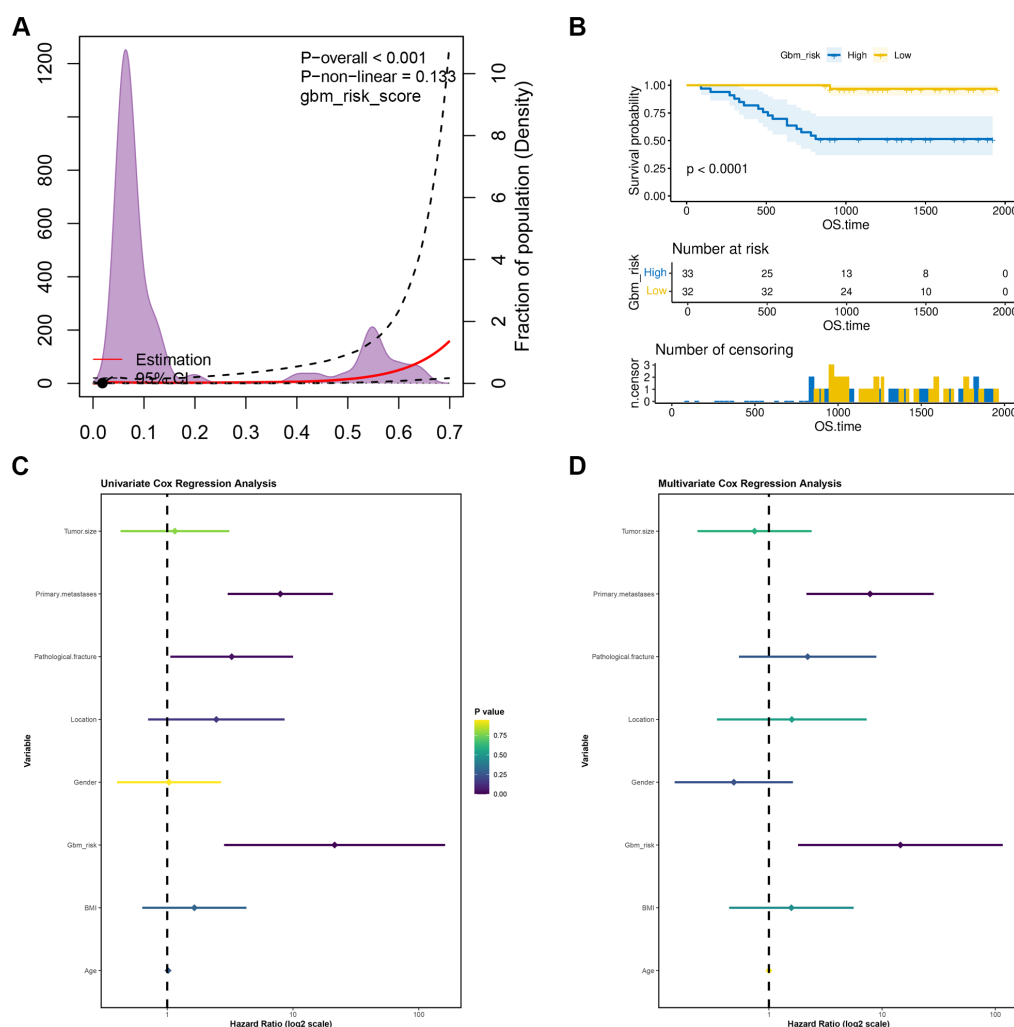
Through a systematic analysis of peripheral blood lymphocyte subsets in 65 OS patients, this study revealed significant associations between immune characteristics of lymphocytes and patient prognosis. It was found that the distribution of lymphocyte subsets differed significantly between deceased and surviving patient groups. Specifically, the proportions of six subsets— $CD3^+CD8^+$ ,  $CD3^-CD56^+$ ,  $CD4^+CD28^+$ ,  $CD8^+CD28^+$ ,  $CD4^+HLA-DR^+$ , and  $CD8^+HLA-DR^+$ —were significantly lower in the deceased group compared to the survival group. Notably, although an increasing trend was observed in the proportions of the  $CD4^+CD38^+$  and  $CD8^+CD38^+$  subsets within the deceased group, this difference did not reach statistical significance. By comparing various ML algorithms, a prognostic prediction model based on the GBM was established. This model, utilizing only two variables— $CD3^-CD56^+$  NK cells and  $CD8^+HLA-DR^+$  activated cytotoxic T



**FIGURE 3** Model selection process and comprehensive evaluation of the final GBM prognostic model for osteosarcoma. (A) Line graph comparing the AUC values of models constructed by LASSO, SVM, XGBoost, GBM, and KNN algorithms across variable combinations (1 to 5 features). (B) Histogram with fitted curve showing the frequency distribution of AUC values from 100 bootstrap validation iterations (C) AUC value derived from 5-fold cross-validation. (D) Scatter plot demonstrating a significant negative correlation between the gbm\_risk\_score and overall patient survival time. (E) Scatter plot showing no significant correlation between gbm\_risk\_score and patient age. (F) Violin plots comparing gbm\_risk\_score across subgroups defined by key clinical variables (e.g., age, gender, tumor location).

cells—achieved excellent predictive performance, with an AUC of 0.9591. Furthermore, the gbm\_risk\_score, constructed based on these two lymphocyte subsets, showed a significant negative correlation with overall patient survival and was confirmed as an independent prognostic factor in multivariate analysis, distinct

from traditional clinical factors such as metastasis status at initial diagnosis. The subsequently developed nomogram model demonstrated good accuracy and clinical utility in predicting the 3-year OS rate, successfully identifying high-risk subgroups among patients without initial metastasis. These findings provide new



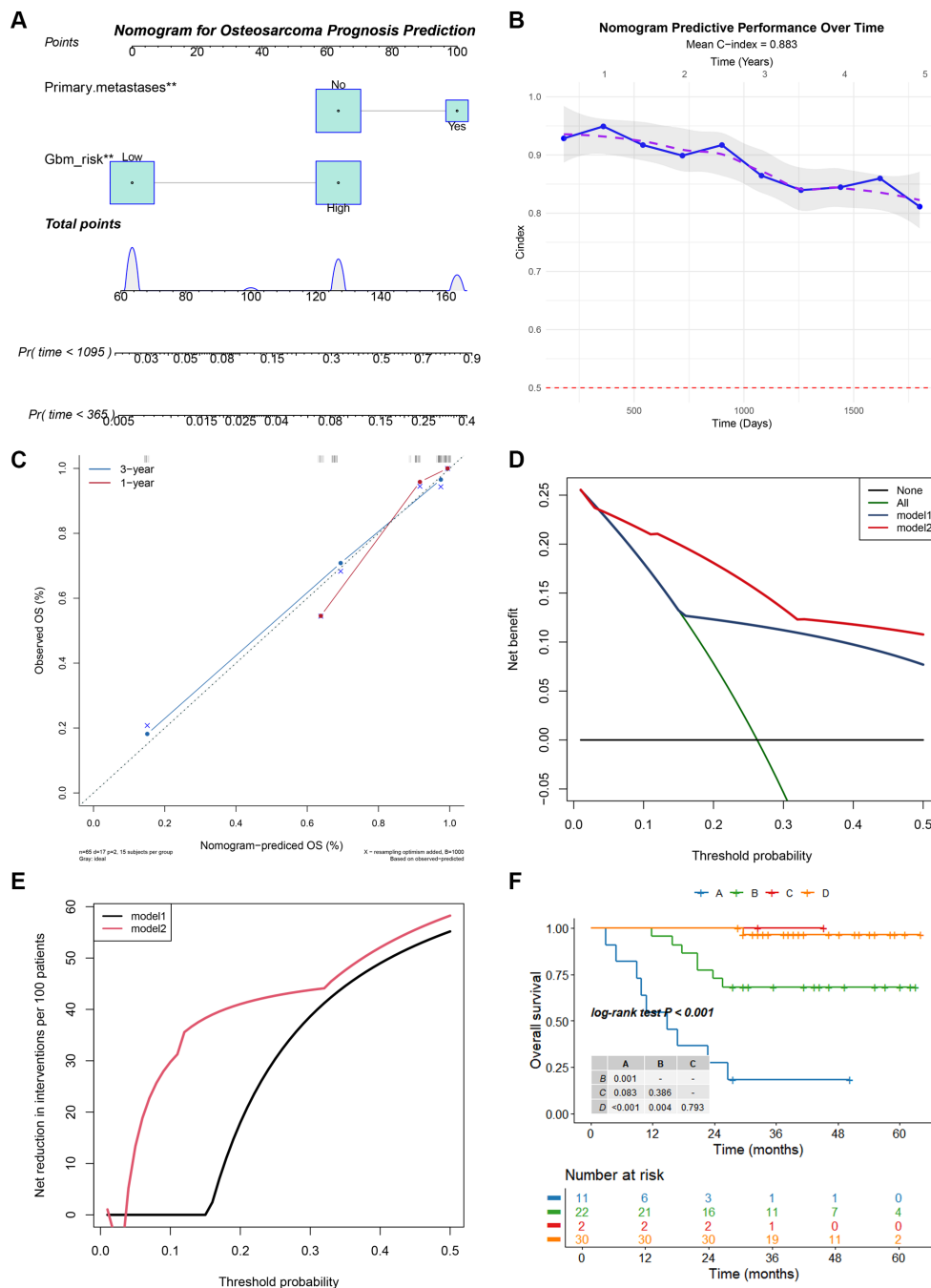
**FIGURE 4** Prognostic value of the GBM risk group in predicting overall survival of osteosarcoma patients. **(A)** Restricted cubic spline analysis of the gbm\_risk\_score for overall survival. **(B)** K-M curves of low- and high-risk groups stratified by the median gbm\_risk\_score. **(C)** Forest plot of univariate Cox analysis for osteosarcoma overall survival. **(D)** Forest plot of multivariate Cox analysis for osteosarcoma overall survival.

perspectives and methodologies for the immune prognostic assessment of OS.

Previous studies have extensively explored the role of total peripheral blood lymphocyte count and its derived biomarkers, such as the NLR, in prognostic assessment across various cancers (21–23). For instance, in extensive-stage small cell lung cancer (ES-SCLC), research has demonstrated that patients with a low pretreatment total lymphocyte count or a high NLR exhibit shorter OS (24). Similarly, a meta-analysis focusing on gastric cancer (GC) patients revealed that elevated NLR and platelet-to-lymphocyte ratio (PLR) were associated with poorer OS and progression-free survival (PFS) following treatment with immune checkpoint inhibitors (ICIs), whereas a high lymphocyte-to-monocyte ratio (LMR) was correlated with improved survival outcomes (25). Furthermore, in OS research, NLR has been confirmed as an independent risk factor for predicting patient survival, and dynamic monitoring of NLR changes can further enhance the accuracy of prognostic evaluation (10). Collectively,

these findings underscore the significant value of total peripheral blood lymphocyte count and its derived ratios in pan-cancer prognostic prediction and even in the evaluation of treatment response.

With the deepening of research on the tumor immune microenvironment (TIME), the functional diversity of lymphocyte subsets and their complex roles in tumorigenesis and development have become increasingly clear (26). Among them, CD4+ T cells are crucial for initiating and regulating adaptive immunity; they effectively promote anti-tumor immune responses by supporting the clonal expansion, differentiation, and memory formation of cytotoxic T lymphocytes (CTLs) (27, 28). CD8+ T cells, as direct effector cells, kill tumor cells through the release of cytotoxic substances such as perforin and granzymes (29). In contrast, Tregs play an immunosuppressive role in the TIME by inhibiting the function of effector T cells, thereby promoting tumor immune escape and progression (30); studies have shown that the proportion of Tregs in the peripheral blood of some cancer patients



**FIGURE 5** Integration of the GBM risk group with clinical variables for prognostic prediction in osteosarcoma. (A) Prognostic nomogram developed by combining the GBM risk grouping and primary metastasis status. (B) C-index of the nomogram. (C) Calibration curves for 1- and 3-year OS. (D, E) DCA showing the clinical net benefit and net reduction (F) K-M survival curves of patient subgroups stratified by the GBM risk group and primary metastases.

is significantly higher than that in healthy individuals (31). Therefore, compared to measuring the total lymphocyte count alone, precise analysis of these functionally diverse and often antagonistic lymphocyte subsets in peripheral blood provides a more comprehensive reflection of the overall immune status of cancer patients and demonstrates greater potential for clinical application in prognostic prediction.

CD56+ NK cells, one of the cornerstones of the gbm\_risk\_score, constitute a central component of innate anti-tumor immunity, which can directly lyse tumor cells without the need for prior activation, notably those with low or absent MHC-I expression (32). Accumulating evidence indicates that the presence of NK cells is generally associated with a favorable prognosis. For instance, in patients with advanced hepatocellular carcinoma treated with ICIs,

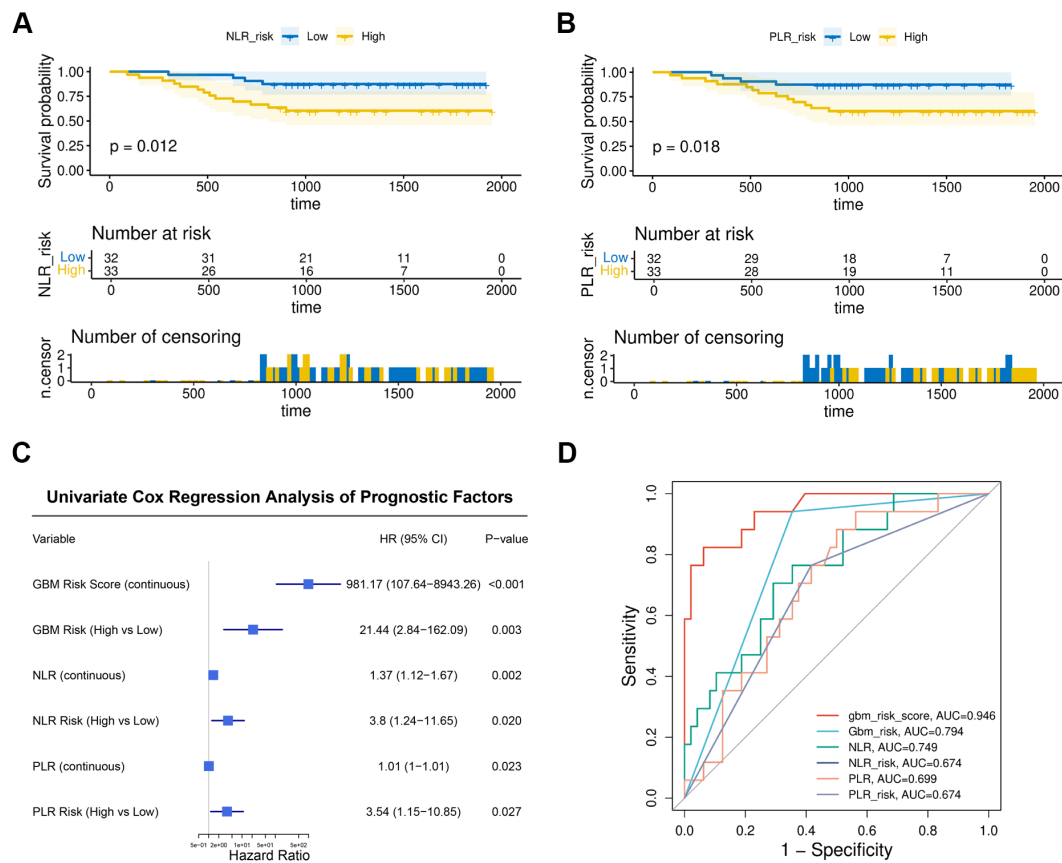


FIGURE 6

The GBM prognostic model exhibited superior predictive capability compared to NLR and PLR. (A) K-M curves of low- and high-risk groups stratified by the median NLR. (B) K-M curves of low- and high-risk groups stratified by the median PLR. (C) Forest plot of univariate Cox analysis for osteosarcoma overall survival. (D) The ROC curves illustrate the performance of GBM, NLR, and PLR in predicting the 3-year overall survival of patients.

an increased proportion of peripheral blood NK cells at week 3 post-treatment is an independent predictor of objective response and long-term survival (PFS and OS) (33). Similarly, in advanced gastric cancer patients treated with apatinib, a peripheral blood NK cell proportion below 17% is associated with poorer PFS and OS (34). In osteosarcoma, a dual-center retrospective study of 106 patients demonstrated that a high proportion of NK cells is correlated with improved survival (35). Furthermore, preclinical strategies aimed at enhancing NK cell migration and infiltration into osteosarcoma tumors have yielded promising preliminary results (36). These findings align with our own data, thereby highlighting the critical role of NK cells in tumor immune surveillance via direct cytotoxicity.

CD8<sup>+</sup>HLA-DR<sup>+</sup> T cells represent a subset of CTLs commonly regarded as markers of activated, effector memory T cells (37). HLA-DR is a class II MHC molecule whose expression on T cells typically denotes an activated state associated with antigen presentation or effector functions. Existing research suggests that CD8<sup>+</sup>HLA-DR<sup>+</sup> T cells can have dual prognostic significance in cancer patients. For example, studies in breast cancer show that high levels of HLA-DR<sup>+</sup> CTLs are associated with a favorable response to neoadjuvant chemotherapy and improved PFS, indicating a positive prognostic role (38, 39). However, other

research indicates that in primary central nervous system lymphoma, elevated levels of CD8<sup>+</sup>HLA-DR<sup>+</sup> T cells has been linked to T cell exhaustion and an immunosuppressive microenvironment, suggesting a poor prognosis (37). In contrast, the prognostic value of CD8<sup>+</sup>HLA-DR<sup>+</sup> T cells in osteosarcoma has not been definitively established. Only one study has proposed that increased HLA-DR expression in osteosarcoma likely reflects an ongoing anti-tumor immune response (40). In our study, a higher proportion of peripheral blood CD8<sup>+</sup>HLA-DR<sup>+</sup> T cells was associated with prolonged OS. This implies that within the context of osteosarcoma, this cell population may represent an effective anti-tumor immune response rather than an exhausted phenotype. Further studies are warranted to validate this finding.

While numerical differences were observed for several lymphocyte subsets (including CD45<sup>+</sup>, CD3<sup>+</sup>, CD3<sup>+</sup>CD4<sup>+</sup>, CD4<sup>+</sup>CD38<sup>+</sup>, and CD8<sup>+</sup>CD38<sup>+</sup>) between deceased and surviving patients, these differences did not attain statistical significance in our cohort. Kaplan-Meier analysis also revealed no significant difference in OS between the defined risk groups. Several factors may account for these non-significant findings. First, the relatively limited sample size may have reduced the statistical power to detect associations with modest clinical effect sizes. Furthermore, the use of the median value as a uniform cutoff for all biomarkers—a

conservative strategy adopted to mitigate overfitting in a small cohort—may have obscured the predictive utility of some markers. Consequently, validation of these results will require future multi-center studies with larger sample sizes or well-designed prospective trials.

In contrast to the non-significant markers, our analysis identified several lymphocyte subsets with clear prognostic value. Patients with higher proportions of CD8<sup>+</sup>CD28<sup>+</sup> T cells had significantly longer OS, a finding consistent with reports by Liu et al. in non-small cell lung cancer (NSCLC) that affirm the central role of these effector T cells in anti-tumor immunity (41, 42). Furthermore, activated CD4<sup>+</sup>HLA-DR<sup>+</sup> T cells were significantly enriched in the survival group and independently predicted a better prognosis. The importance of such highly activated T cells has also been validated in studies on NSCLC and in melanoma cohorts receiving anti-PD-1 therapy, further emphasizing the key value of activated T cells in anti-tumor immunity (43, 44). Collectively, these results underscore the prognostic significance of specific activated T-cell populations.

The biological interpretation of complex omics data has long been a primary challenge in clinical translational research (16). The emergence of ML algorithms provides an effective tool for discovering potential biomarkers from these high-dimensional datasets and constructing predictive models (45–47). For instance, a 10-metabolite GC diagnostic model developed by Chen et al. using ML analysis demonstrated performance significantly superior to models based on traditional clinical parameters (16). In OS research, Zhao et al. successfully developed a diagnostic model for identifying high-risk subtypes by integrating transcriptomic and methylation data with ML algorithms, which showed good diagnostic performance (48). To optimize the model and avoid overfitting, this study initially employed seven different algorithms to construct models containing combinations of 1 to 5 variables. The results revealed that algorithms such as GBM and XGBoost achieved AUC values exceeding 0.90 with only two variables, substantially outperforming the traditional AIC algorithm. This highlights the advantage of ML in data dimensionality reduction and building streamlined yet robust prognostic models. The stability of the model was further confirmed through internal Bootstrap validation and 5-fold CV. Ultimately, the risk grouping defined by integrating lymphocyte subsets via ML was confirmed as an independent prognostic factor for OS patients. Its significant clinical value lies in its ability to identify a subgroup of patients without detectable metastasis at initial diagnosis who nonetheless harbor a high risk of mortality.

It is important to acknowledge the limitations of this study. First, the retrospective and single-center nature of the cohort (n=65 high-grade osteosarcoma patients) introduces inherent risks of selection bias and limits the external generalizability of the findings. While osteosarcoma is a rare disease, making large cohorts challenging to assemble, and the institution's peripheral blood lymphocyte subset analysis was only recently implemented, these factors underscore the need for cautious interpretation.

Furthermore, despite employing machine learning algorithms (GBM) and rigorous internal validation strategies (bootstrap and 5-fold cross-validation) to enhance model robustness and mitigate overfitting risks in this modest dataset, the absence of an independent external validation cohort remains a critical constraint. Future validation through larger, multicenter prospective studies is indispensable to confirm the generalizability and clinical utility of the prognostic model. Second, regarding the lymphocyte subset analysis, the panel used in our institution's flow cytometry did not include Tregs, a crucial subset, preventing the assessment of their role in OS prognosis. Furthermore, with the growing understanding of the TIME, characterizing T-cell function using only dual markers such as CD4+CD38+ or CD8+CD38+ may be insufficient. For instance, studies indicate that CD8+CD38+ programmed cell death protein 1-positive (PD-1+) and PD-1–subpopulations may possess distinct functional properties (49).

Nevertheless, by integrating lymphocyte subset data with ML algorithms, this study effectively revealed the potential value of peripheral blood lymphocyte subsets in prognostic assessment for OS. The findings suggest that functional lymphocyte subsets (e.g., activated or exhausted phenotypes) might reflect a patient's immune status and disease progression risk more accurately than the simple CD4/CD8 ratio alone. Future research should integrate a broader panel of markers (e.g., PD-1, cytotoxic T-lymphocyte-associated protein 4 [CTLA-4]) to deeply characterize lymphocyte functional subsets and clarify their specific roles in personalized prognosis prediction and treatment strategies for OS.

## 5 Conclusion

In conclusion, this research introduces a pioneering framework for OS prognosis by leveraging ML to decode peripheral blood lymphocyte subset signatures. This model addresses the shortcomings of conventional prognostic markers and provides a robust theoretical and practical basis for future investigations into the evolving TIME, ultimately aiming to improve clinical decision-making and patient survival.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by The First Affiliated Hospital of Zhengzhou University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

## Author contributions

LL: Conceptualization, Data curation, Writing – review & editing, Writing – original draft, Methodology, Investigation, Visualization, Formal Analysis, Software. JLL: Conceptualization, Writing – original draft, Formal Analysis, Data curation. SP: Writing – original draft, Formal Analysis, Data curation. YuZ: Writing – original draft, Formal Analysis, Data curation. YW: Software, Writing – original draft. JW: Software, Methodology, Writing – original draft. YL: Writing – original draft, Project administration, Investigation. YiZ: Funding acquisition, Project administration, Writing – original draft, Investigation. YaZ: Writing – original draft, Supervision, Validation. JZL: Validation, Supervision, Writing – original draft. NZ: Writing – review & editing, Resources, Supervision, Writing – original draft, Project administration, Validation. XL: Writing – review & editing, Funding acquisition, Writing – original draft, Supervision, Resources, Project administration, Validation.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the Key Research Program for University of Henan Province (Grant No. 23A320047) awarded to Yi Zhang and the Key Projects of Medical Science and Technology Research in Henan Province (Grant No. SBGJ202402042) awarded to Xinchang Lu.

## Acknowledgments

Thanks to the support of The First Affiliated Hospital of Zhengzhou University for the research. The authors acknowledge the InsightPaper LLM system for its contribution to the language refinement of this manuscript.

## References

1. Beird HC, Bielack SS, Flanagan AM, Gill J, Heymann D, Janeway KA, et al. Osteosarcoma. *Nat Rev Dis Primers*. (2022) 8:77. doi: 10.1038/s41572-022-00409-y
2. Whelan JS, Davis LE. Osteosarcoma, chondrosarcoma, and chordoma. *J Clin Oncol*. (2017) 36:188–93. doi: 10.1200/JCO.2017.75.1743
3. Isakoff MS, Bielack SS, Meltzer P, Gorlick R. Osteosarcoma: current treatment and a collaborative pathway to success. *J Clin Oncol*. (2015) 33:3029–35. doi: 10.1200/jco.2014.59.4895
4. Aljubran AH, Griffin A, Pintilie M, Blackstein M. Osteosarcoma in adolescents and adults: survival analysis with and without lung metastases. *Ann Oncol*. (2009) 20:1136–41. doi: 10.1093/annonc/mdn731
5. Italiano A, Mir O, Mathoulin-Pelissier S, Penel N, Piperno-Neumann S, Bompas E, et al. Cabozantinib in patients with advanced Ewing sarcoma or osteosarcoma (CABONE): a multicentre, single-arm, phase 2 trial. *Lancet Oncol*. (2020) 21:446–55. doi: 10.1016/s1470-2045(19)30825-3
6. Tawbi HA, Burgess M, Bolejack V, Van Tine BA, Schuetz SM, Hu J, et al. Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *Lancet Oncol*. (2017) 18:1493–501. doi: 10.1016/s1470-2045(17)30624-1
7. Roberts RD, Lizardo MM, Reed DR, Hingorani P, Glover J, Allen-Rhoades W, et al. Provocative questions in osteosarcoma basic and translational biology: A report from the Children's Oncology Group. *Cancer*. (2019) 125:3514–25. doi: 10.1002/cncr.32351
8. Mazzaschi G, Madeddu D, Falco A, Bocchialini G, Goldoni M, Sogni F, et al. Low PD-1 expression in cytotoxic CD8(+) tumor-infiltrating lymphocytes confers an immune-privileged tissue microenvironment in NSCLC with a prognostic and predictive value. *Clin Cancer Res*. (2018) 24:407–19. doi: 10.1158/1078-0432.Ccr-17-2156
9. Kervarrec T, Gaboriaud P, Berthon P, Zaragoza J, Schrama D, Houben R, et al. Merkel cell carcinomas infiltrated with CD33+ myeloid cells and CD8+ T cells are associated with improved outcome. *J Am Acad Dermatol*. (2018) 78:973–982.e8. doi: 10.1016/j.jaad.2017.12.029
10. Li L, Li Y, Lu M, Wang Y, Li Z, Hu X, et al. The combination of baseline neutrophil to lymphocyte ratio and dynamic changes during treatment can better predict the survival of osteosarcoma patients. *Front Oncol*. (2023) 13:1235158. doi: 10.3389/fonc.2023.1235158
11. Li L, Wang Y, He X, Li Z, Lu M, Gong T, et al. Hematological prognostic scoring system can predict overall survival and can indicate response to immunotherapy in patients with osteosarcoma. *Front Immunol*. (2022) 13:879560. doi: 10.3389/fimmu.2022.879560

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. During the revision of the manuscript, the language-related sections were refined using LLM systems in accordance with the reviewers' suggestions.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2026.1733518/full#supplementary-material>

12. Kazandjian D, Gong Y, Keegan P, Pazdur R, Blumenthal GM. Prognostic value of the lung immune prognostic index for patients treated for metastatic non-small cell lung cancer. *JAMA Oncol.* (2019) 5:1481–5. doi: 10.1001/jamaoncol.2019.1747
13. Yi G, Guo S, Liu W, Wang H, Liu R, Tsun A, et al. Identification and functional analysis of heterogeneous FOXP3(+) Treg cell subpopulations in human pancreatic ductal adenocarcinoma. *Sci bulletin.* (2018) 63:972–81. doi: 10.1016/j.scib.2018.05.028
14. Núñez NG, Tosello Boari J, Ramos RN, Richer W, Cagnard N, Anderfuhren CD, et al. Tumor invasion in draining lymph nodes is associated with Treg accumulation in breast cancer patients. *Nat Commun.* (2020) 11:3272. doi: 10.1038/s41467-020-17046-2
15. Mao F, Yang C, Luo W, Wang Y, Xie J, Wang H. Peripheral blood lymphocyte subsets are associated with the clinical outcomes of prostate cancer patients. *Int Immunopharmacol.* (2022) 113:109287. doi: 10.1016/j.intimp.2022.109287
16. Chen Y, Wang B, Zhao Y, Shao X, Wang M, Ma F, et al. Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. *Nat Commun.* (2024) 15:1657. doi: 10.1038/s41467-024-46043-y
17. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug discovery.* (2019) 18:463–77. doi: 10.1038/s41573-019-0024-5
18. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol.* (2022) 23:40–55. doi: 10.1038/s41580-021-00407-0
19. Andrew TW, Alrawi M, Plummer R, Reynolds N, Sondak V, Brownell I, et al. A hybrid machine learning approach for the personalized prognostication of aggressive skin cancers. *NPJ Digital Med.* (2025) 8:15. doi: 10.1038/s41746-024-01329-9
20. Placido D, Yuan B, Hjaltekin JX, Zheng C, Haue AD, Chmura PJ, et al. A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. *Nat Med.* (2023) 29:1113–22. doi: 10.1038/s41591-023-02332-5
21. He X, Lu M, Hu X, Li L, Zou C, Luo Y, et al. Osteosarcoma immune prognostic index can indicate the nature of indeterminate pulmonary nodules and predict the metachronous metastasis in osteosarcoma patients. *Front Oncol.* (2022) 12:952228. doi: 10.3389/fonc.2022.952228
22. Zhao C, Li LQ, Yang FD, Wei RL, Wang MK, Song DX, et al. A hematological-related prognostic scoring system for patients with newly diagnosed glioblastoma. *Front Oncol.* (2020) 10:591352. doi: 10.3389/fonc.2020.591352
23. Graziano V, Grassadonia A, Iezzi L, Vici P, Pizzuti L, Barba M, et al. Combination of peripheral neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio is predictive of pathological complete response after neoadjuvant chemotherapy in breast cancer patients. *Breast (Edinburgh Scotland).* (2019) 44:33–8. doi: 10.1016/j.breast.2018.12.014
24. Suzuki R, Wei X, Allen PK, Cox JD, Komaki R, Lin SH. Prognostic significance of total lymphocyte count, neutrophil-to-lymphocyte ratio, and platelet-to-lymphocyte ratio in limited-stage small-cell lung cancer. *Clin Lung Cancer.* (2019) 20:117–23. doi: 10.1016/j.clcc.2018.11.013
25. Tan S, Zheng Q, Zhang W, Zhou M, Xia C, Feng W. Prognostic value of inflammatory markers NLR, PLR, and LMR in gastric cancer patients treated with immune checkpoint inhibitors: a meta-analysis and systematic review. (2024) 15. doi: 10.3389/fimmu.2024.1408700
26. Khosravi G-R, Mostafavi S, Bastan S, Ebrahimi N, Gharibvand RS, Eskandari N. Immunologic tumor microenvironment modulators for turning cold tumors hot. *Cancer Communications* (2024) 44:521–53. doi: 10.1002/cac2.12539
27. Laidlaw BJ, Craft JE, Kaech SM. The multifaceted role of CD4(+) T cells in CD8(+) T cell memory. *Nat Rev Immunol.* (2016) 16:102–11. doi: 10.1038/nri.2015.10
28. Ridge JP, Di Rosa F, Matzinger P. A conditioned dendritic cell can be a temporal bridge between a CD4+ T-helper and a T-killer cell. *Nature.* (1998) 393:474–8. doi: 10.1038/30989
29. Reina-Campos M, Scharping NE, Goldrath AW. CD8+ T cell metabolism in infection and cancer. *Nat Rev Immunol.* (2021) 21:718–38. doi: 10.1038/s41577-021-00537-8
30. Imianowski CJ, Chen Q, Workman CJ, Vignali DAA. Regulatory T cells in the tumour microenvironment. *Nat Rev Cancer.* (2025) 25:703–22. doi: 10.1038/s41568-025-00832-9
31. Saleh R, Elkord E. FoxP3(+) T regulatory cells in cancer: Prognostic biomarkers and therapeutic targets. *Cancer letters.* (2020) 490:174–85. doi: 10.1016/j.canlet.2020.07.022
32. Chu J, Gao F, Yan M, Zhao S, Yan Z, Shi B, et al. Natural killer cells: a promising immunotherapy for cancer. *J Trans Med.* (2022) 20:240. doi: 10.1186/s12967-022-03437-0
33. Pan Z, Song S, An Y, Guan L, Liu H, Li W. Predictive value of peripheral blood CD4+ T and NK cells on efficacy and long-term survival in advanced HCC patients receiving immunotherapy. (2025) 16. doi: 10.3389/fimmu.2025.1683328
34. Zhang N, Hu SY, Wang YT, Shan H, Tian GY, Wang Y, et al. Reduced peripheral blood natural killer cell proportion predicts poor overall survival in advanced gastric cancer patients treated with apatinib. *World J Oncol.* (2025) 16:546–54. doi: 10.14740/wjon2655
35. Luo K, Tang H, Yan W, Li S, Luo X, Yang M, et al. Prognostic significance of T cells and NK cells in osteosarcoma: a dual-center retrospective study. *World J Surg Oncol.* (2025) 23:130. doi: 10.1186/s12957-025-03784-4
36. Eguchi S, Luo W, Zhu H, Hoang HM, Xu C, Behbehani GK, et al. CXCL10-induced chemotaxis of ex vivo-expanded natural killer cells combined with NKTR-255 enhances anti-tumor efficacy in osteosarcoma. *Mol Ther Oncol.* (2025) 33:201051. doi: 10.1016/j.omton.2025.201051
37. Wu Y, Lv L, Liu J, Sun X, Gao C, Sun S, et al. Mass cytometric analysis of circulating immune landscape in primary central nervous system lymphoma. (2025) 16. doi: 10.3389/fimmu.2025.1658015
38. Saraiva DP, Azeredo-Lopes S, Antunes A, Salvador R, Borralho P, Assis B, et al. Expression of HLA-DR in cytotoxic T lymphocytes: A validated predictive biomarker and a potential therapeutic strategy in breast cancer. *Cancers* (2021) 13:3841. doi: 10.3390/cancers13153841
39. Osuna-Gómez R, Arqueros C, Galano C, Mulet M, Zamora C, Barnadas A, et al. Effector mechanisms of CD8+ HLA-DR+ T cells in breast cancer patients who respond to neoadjuvant chemotherapy. *Cancers* (2021) 13:6167. doi: 10.3390/cancers13246167
40. Trieb K, Lechleitner T, Lang S, Windhager R, Kotz R, Dirnhöfer S. Evaluation of HLA-DR expression and T-lymphocyte infiltration in osteosarcoma. *Pathology Res practice.* (1998) 194:679–84. doi: 10.1016/s0344-0338(98)80126-x
41. Liu C, Jing W, An N, Li A, Yan W, Zhu H, et al. Prognostic significance of peripheral CD8+CD28+ and CD8+CD28- T cells in advanced non-small cell lung cancer patients treated with chemo(radio)therapy. *J Trans Med.* (2019) 17:344. doi: 10.1186/s12967-019-2097-7
42. Liu C, Hu Q, Hu K, Su H, Shi F, Kong L, et al. Increased CD8+CD28+ T cells independently predict better early response to stereotactic ablative radiotherapy in patients with lung metastases from non-small cell lung cancer. *J Trans Med.* (2019) 17:120. doi: 10.1186/s12967-019-1872-9
43. Wei Z, Zhang W, Gao F, Wu Y, Zhang G, Liu Z, et al. Impact of lymphocyte subsets on chemotherapy efficacy and long-term survival of patients with advanced non-small-cell lung cancer. *Zhongguo yi xue ke xue yuan xue bao Acta Academiae Medicinae Sinicae.* (2017) 39:371–6. doi: 10.3881/j.issn.1000-503X.2017.03.012
44. Chi P, Jiang H, Li D, Li J, Wen X, Ding Q, et al. An immune risk score predicts progression-free survival of melanoma patients in South China receiving anti-PD-1 inhibitor therapy—a retrospective cohort study examining 66 circulating immune cell subsets. (2022) 13. doi: 10.3389/fimmu.2022.1012673
45. Yokoyama S, Hamada T, Higashi M, Matsuo K, Maemura K, Kurahara H, et al. Predicted prognosis of patients with pancreatic cancer by machine learning. *Clin Cancer Res.* (2020) 26:2411–21. doi: 10.1158/1078-0432.Ccr-19-1247
46. Eckardt JN, Hahn W, Ries RE, Chrost SD, Winter S, Stasik S, et al. Age-stratified machine learning identifies divergent prognostic significance of molecular alterations in AML. *HemaSphere.* (2025) 9:e70132. doi: 10.1002/hem3.70132
47. Smith LA, Oakden-Rayner L, Bird A, Zeng M, To MS, Mukherjee S, et al. Machine learning and deep learning predictive models for long-term prognosis in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Lancet Digital Health.* (2023) 5:e872–81. doi: 10.1016/s2589-7500(23)00177-2
48. Zhao W, Meng H, Dai Z, Zhang L, Cheng Z, Song Y, et al. Prediction of patients with high-risk osteosarcoma on the basis of XGBoost algorithm using transcriptome and methylation data from SGH-OS cohort. *JCO Precis Oncol.* (2025) 9:e2400732. doi: 10.1200/po-24-00732
49. Verma V, Shrimali RK, Ahmad S, Dai W, Wang H, Lu S, et al. PD-1 blockade in subprimed CD8 cells induces dysfunctional PD-1(+)CD38(hi) cells and anti-PD-1 resistance. *Nat Immunol.* (2019) 20:1231–43. doi: 10.1038/s41590-019-0441-y