doi.org/10.36721/PJPS.2026.39.12.338.1

Machine learning prognostic model and drug survival analysis for lung adenocarcinoma in the context of radiotherapy

Zaishuang Ju[1], Yutong Wu[2], Lili Tian[3], Dachuan Shen[2] and Ruoyu Wang[1]*

[1]Dalian Medical University, No. 9, West Section of Lushun South Road, Dalian 116044, Liaoning Province, China.
[2]Department of Oncology, Affiliated Zhongshan Hospital of Dalian University, No.6 Jie Fang Street, Zhongshan District, Dalian 116001, Liaoning Province, China.
[3]Department of Oncology, First Affiliated Hospital of Dalian Medical University, No.222 Zhongshan Street, Xigang District, Dalian 116021, Liaoning, China.

Abstract: **Background**: Patients with lung adenocarcinoma (LUAD) receiving radiotherapy represent an important but underexplored clinical subgroup. These patients often undergo concomitant pharmacologic treatments, yet the prognostic impact and underlying determinants of such combined regimens remain poorly understood. **Objective**: This retrospective observational study aimed to develop and validate a radiotherapy-specific machine learning prognostic model for LUAD and to compare survival across concomitant pharmacologic regimens. **Methods**: In this retrospective observational study, using genomic and clinical data from TCGA, a radiotherapy-specific prognostic model for LUAD was developed and validated through ten machine learning algorithms. Survival analyses were conducted across distinct concomitant pharmacologic strategies, followed by functional enrichment to elucidate molecular mechanisms underlying differential outcomes. **Results**: Demonstrating robust prognostic abilities, the model efficiently sorted patients into high- and low-risk categories. Both treatment type and risk score independently predicted overall survival, with significant interaction effects. Low-risk patients receiving targeted or combination therapy—mainly erlotinib, gefitinib, or bevacizumab—exhibited substantially improved survival compared with those receiving conventional chemotherapy. Enrichment of "Exogenous peptide presentation," "MHC class II assembly," "Peptide–MHC II assembly," and "Symbiotic interaction" pathways indicated immune modulation and host–tumor crosstalk as key mediators of treatment efficacy. **Conclusion**: This study establishes a radiotherapy-specific prognostic model for lung adenocarcinoma, demonstrating distinct molecular and therapeutic heterogeneity and highlighting the superior survival benefit of targeted combination therapy in low-risk patients.

Keywords: LUAD; Machine learning; Prognostic model; Radiotherapy; Targeted therapy

Submitted on 11-11-2025 – Revised on 12-12-2025 – Accepted on 24-12-2025

INTRODUCTION

Lung adenocarcinoma (LUAD) represents the predominant histological form of non–small cell lung cancer (NSCLC) and accounts for over 40% of lung cancer cases globally(Xu *et al*., 2024; Zhao *et al*., 2023). Advances in diagnosis and treatment have improved survival outcomes for LUAD; however, its incidence and mortality rates continue to rise, making it the leading cause of lung cancer globally (Ma *et al*., 2023; Zhang *et al*., 2023). Radiotherapy plays a crucial role in the management of locally advanced or unresectable LUAD, offering effective local tumor control and prolonged survival (Horinouch, 2018; van Diessen, 2019). Nevertheless, even under standardized conditions of staging, radiation dose and technique, there remains substantial heterogeneity in long-term survival outcomes (Li *et al*., 2021; Liu *et al*., 2017; Yue *et al*., 2024). Such variability suggests that traditional clinicopathological factors are insufficient to explain differences in outcomes among patients receiving radiotherapy, implying an underlying biological diversity within this subgroup. This heterogeneity underscores the importance of conducting in-depth molecular investigations to refine risk stratification and optimize treatment strategies.

**Corresponding author:* e-mail: juzaishuang@sina.com

With the growing understanding of tumor biology and advances in data mining techniques in bioinformatics, molecular biomarkers have emerged as increasingly important tools for prognostic evaluation of lung cancer. Accumulating evidence suggests that differentially expressed messenger RNAs (mRNAs) are not merely byproducts of tumor heterogeneity but active participants in cancer progression, therapeutic resistance and immune regulation. Specific mRNA signatures have been shown to reflect tumor aggressiveness, radiosensitivity and the immune microenvironment, offering potential guidance for individualized therapy (Han *et al*., 2020). However, most previously established prognostic models have been derived from analyses of the general LUAD population without focusing specifically on patients treated with radiotherapy. Therefore, these models often perform suboptimally when applied to radiotherapy subgroups (Lender *et al*., 2025; Liu *et al*., 2024; Sun *et al*., 2025). Moreover, LUAD patients selected for radiotherapy often represent a biologically distinct subset—characterized by different tumor burden, microenvironmental context and prior treatment exposure—so the risk patterns captured in general LUAD cohorts may not accurately reflect prognosis in this setting (Wang *et al*., 2025). Consequently, the predictive power and clinical applicability of these models in the radiotherapy setting remain limited and

precise tools tailored to radiotherapy-associated molecular characteristics are still lacking.

In clinical practice, LUAD patients undergoing radiotherapy frequently receive concurrent systemic therapies, including cytotoxic chemotherapy, targeted therapy (such as EGFR or ALK inhibitors), or combinations thereof (Wrona *et al*., 2021; Xu *et al*., 2021; Ying *et al*., 2020; Zia *et al*., 2019). Evidence suggests that molecularly targeted agents may modulate radiosensitivity and interact with DNA damage response pathways. Despite the existing mechanistic evidence, systematic studies comparing the survival benefits of different pharmacological combinations in the context of radiotherapy remain limited. This gap constrains the development of optimal combined radiotherapy–systemic therapy strategies in clinical practice. Therefore, constructing a prognostic model based on radiotherapy-associated mRNA profiles and evaluating survival differences across groups stratified by actual therapeutic regimens holds significant clinical translational value.

In summary, the present study aims to develop a robust survival prediction model centered on radiotherapy-associated mRNA expression profiles, integrating both molecular and clinical treatment data. By incorporating real-world information on systemic therapy regimens, the model seeks to quantify survival benefits across treatment subgroups. This study aims to establish a theoretical framework for precise stratification and personalized decision-making in LUAD patients receiving combined radiotherapy and pharmacotherapy. Such an approach is expected to advance the paradigm of precision oncology, improve patient prognosis and optimize individualized treatment strategies for LUAD.

MATERIALS AND METHODS

Data processing and differential expression analysis

The TCGA-LUAD cohort was obtained from the public TCGA repository in this retrospective observational study, and both the mRNA expression matrix and the associated clinical annotations were extracted for subsequent analyses. Patients who were treated with radiotherapy and possessed full clinical information, including overall survival (OS), OS time, survival status and major clinicopathological characteristics, with available mRNA expression profiles were included in this study. To reduce potential selection and information bias in survival analyses, only patients with complete key clinical covariates (age, sex, pathological stage, AJCC TNM classification and treatment information) were included in prognostic modeling and multivariable analyses. Eligible patients were randomly divided into a training cohort and a testing cohort at a 2:1 ratio.

mRNA expression profiles were log-transformed and normalized using R software (v4.5.1) and Bioconductor packages (Huber *et al*., 2015). Genes with different expression levels between tumor and adjacent normal tissues were identified through the DESeq2 package. The selection thresholds were set at |log2 fold change| > 1 and Padj < 0.01, applying the Benjamini–Hochberg correction for multiple testing (Love *et al*., 2014). A univariate Cox regression analysis was employed to determine the prognostic relevance of the identified DEGs with respect to OS in LUAD cases from TCGA (Prentice and Zhao, 2021).

Prognostic model construction and feature selection

To reduce the dimensionality of the identified DEGs, genes with expression levels > 0 in more than 10% of samples were retained, while genes exhibiting identical expression across all samples were removed. Subsequently, the median absolute deviation (MAD) algorithm was employed to adaptively identify genes with high expression variability, while retaining up to 400 genes. Gene-gene correlations were then calculated and highly correlated genes were removed to mitigate multicollinearity. Finally, the gene set was further truncated to 200 genes and then standardized. A model for prognosis was built by integrating ten different algorithms, including random survival support vector machine (survivalSVM), Elastic Net (Enet), StepCox, Ridge, plsRcox, Lasso, SuperPC and Survival Forest (RSF) (Chen, 2016; Fouodo *et al*., 2018; Taylor, 2011; Zou and Hastie, 2005). These algorithms were combined to form various distinct algorithm combinations for model building. Model development was performed exclusively in the training cohort and the held-out testing cohort was used for independent validation to assess generalizability and monitor potential overfitting. To further limit overfitting, the number of input features was constrained by the above filtering steps and highly correlated genes were removed prior to model fitting. The concordance index (C-index) and other performance metrics were compared across the training and testing cohorts to assess the stability of each algorithm and each combination. Computational efficiency during this process was enhanced through parallel processing using GPU-accelerated XGBoost. The overall workflow of prognostic model construction is illustrated in fig. 1.

Model evaluation and validation

For model evaluation and validation, a prognostic score was computed for each patient using the formula:

Risk score = $\Sigma(Exp_i \times \beta_i)$

$$\text{Risk score} = \sum_{i=1}^{n} Exp_i \times \beta_i$$

After Z-score normalization, patients were categorized into high- and low-risk groups. The optimal cutoff for risk stratification was selected by maximizing the log-rank statistic for survival separation. Internal validation was conducted in the training cohort, and model reproducibility was further assessed in the independent testing cohort.

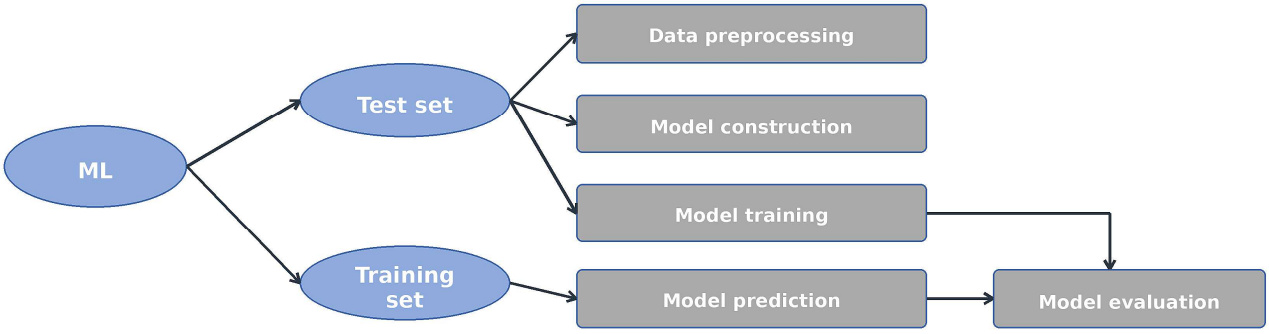

Fig. 1: Workflow of prognostic model construction, including TCGA-LUAD data selection, preprocessing, feature screening, model training, validation and evaluation.

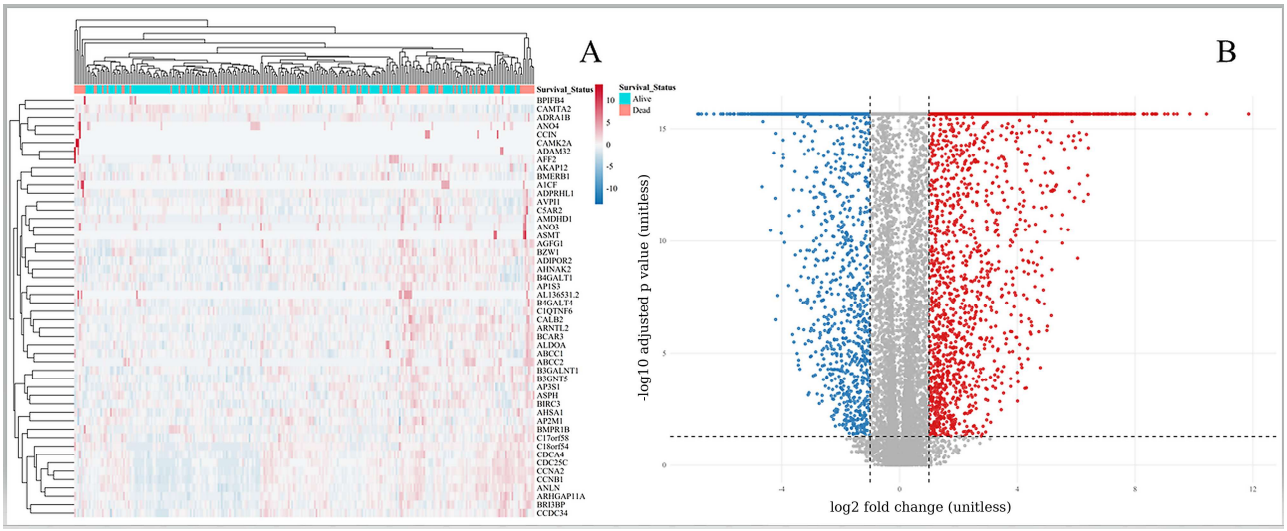

Fig. 2: TCGA-LUAD differential expression analysis. (A) Heatmap of the top 50 DEGs showing LUAD/normal sample clustering. (B) Volcano plot of differentially expressed genes.

Kaplan-Meier (KM) survival analysis was performed, with log-rank tests used to identify statistical differences among the groups. Time-dependent receiver operating characteristic (ROC) curves were used to analyze the temporal prediction accuracy and discriminatory potential of the prognostic model in both the training and testing cohorts, providing an additional assessment of potential overfitting (Heagerty *et al*., 2000). No formal sensitivity analyses were prespecified; model stability was assessed through internal training performance and independent testing-cohort validation.

Drug stratification and prognostic analysis

Patients with LUAD receiving radiotherapy were divided into two groups based on concomitant medication: a chemo group (conventional chemotherapy) and a combo group (targeted or combination therapy). Multivariate Cox regression was conducted using the model-derived risk score, age, sex, treatment, pathological stage and AJCC TNM classification (T, N, M), including an interaction between treatment and risk score. These variables were included because they are clinically relevant covariates and potential confounders of survival. No imputation was performed; complete-case analysis was used. These multivariable models were used to adjust for major clinical confounders and to evaluate whether the prognostic effect of the risk score was independent of these factors. Validation was performed in the independent testing cohort and 95% confidence intervals (CIs), hazard ratios (HRs) and p-values were determined. The optimal risk score cutoff was used to classify patients into low- and high-risk subgroups, namely low-chemo, low-combo, high-chemo and high-combo. KM analysis and log-rank tests, with a significance level of $p<0.05$, were utilized to evaluate prognostic performance and survival differences.

Additionally, functional enrichment analysis of the genes identified by the prognostic model was conducted using clusterProfiler and ranked according to their enrichment scores (Yu *et al*., 2012).

RESULTS

Identification of DEGs in LUAD

Initially, raw count data for TCGA-LUAD patients encompassed 541 LUAD tumor samples and 59 adjacent

normal control samples. Following the application of inclusion criteria, a total of 438 LUAD samples with radiotherapy records and complete clinical and transcriptomic data were included for analysis.

Differential expression analysis identified 2544 DEGs (absolute $\log_2$ fold change > 1, adjusted $p < 0.01$) between LUAD tumor tissues and adjacent normal tissues. The differential expression patterns between LUAD and normal samples were visualized using a heatmap (Fig. 2A) and a volcano plot (Fig. 2B).

Machine learning-based prognostic model development

After initial gene filtering and dimensionality reduction, the 2544 identified DEGs were analyzed. Ten machine learning algorithms were employed in the TCGA-LUAD training set to construct predictive models. This process resulted in the identification of 78 non-zero coefficient feature genes and the creation of 101 machine learning algorithm combinations. Following the exclusion of null values, the optimization results for these machine learning algorithm combinations are shown in fig. 3. table 1 presents the top 10 best-performing machine learning models and their corresponding C-indices. StepCox [forward] was removed even though it had the highest C-index of 0.97 in the training set, as its C-index was under 0.5 in the validation set. Ultimately, the ENet model with an alpha of 0.1, which had the highest average C-index, was picked as the final prognostic model (Table 1).

Evaluating and validating the performance of the prognostic model

The model's discriminative ability was well-demonstrated in the training cohort through time-dependent ROC analysis, with AUC values of 0.80, 0.77 and 0.73 for 1, 3 and 5-year survival, respectively (Fig. 4A). These results indicated robust performance of the model in predicting both short-term and long-term outcomes. Consistent performance was also observed in the independent validation cohort, where the AUCs reached 0.61, 0.65 and 0.68 at 1, 3 and 5 years (Fig. 4B). Patients were separated into high- and low-risk groups according to the optimal risk score threshold. KM analysis showed that individuals in the high-risk group exhibited markedly worse overall survival than those in the low-risk group in the training set (log-rank $p < 0.001$), indicating the model's effective risk-stratification ability (Fig. 4C). In the validation cohort, although the separation between survival curves was slightly reduced, the high-risk group still had significantly poorer survival (log-rank $p = 0.002$), further confirming the model's stability and reproducibility (Fig. 4D).

Survival effects of drug treatment regimens

Patients with LUAD who received radiotherapy were stratified into two groups based on concomitant medication types: chemo group and combo group. The combo group predominantly received erlotinib, gefitinib and bevacizumab, with these patients primarily concentrated in T1 and T2 stages. In contrast, the chemo group mainly received cisplatin, carboplatin, pemetrexed and paclitaxel and these agents were used across T1-T4 stages without a particular predominance in any specific stage. The prognostic model was utilized to evaluate the treatment effects in the two stratified groups. As depicted in Fig. 5A, both treatment and risk score emerged as strong independent predictors (HR=13.43, $p < 0.05$; HR = 2.11, $p < 0.001$ for risk score), further supporting the prognostic value of the selected ENet model. A significant interaction effect (Trt×T2: HR = 0.02, $p < 0.05$) indicated that the treatment efficacy varied depending on disease stage, suggesting therapeutic heterogeneity (Table 2). Kaplan-Meier survival analysis performed across low-risk and high-risk subgroups for different concomitant medication regimens revealed significant differences in prognostic performance and survival ($p<0.05$). Patients in the combo group exhibited significantly prolonged survival compared to the Chemo group, maintaining a 1.0 survival rate for an extended period and demonstrating significantly higher ultimate survival rates than the chemo group. Further analysis showed that the low-combo subgroup consistently presented superior prognostic performance compared to the high-combo subgroup, suggesting a significant impact of risk stratification on patient prognosis. This result further confirmed the effectiveness and clinical applicability of the constructed prognostic model (Fig. 5B).

Drug treatment related functional pathway analysis

In the context of radiotherapy combined with drug treatment for LUAD, Gene Ontology (GO) functional enrichment of the 78 DEGs was conducted. In total, 43 GO pathways showed significant enrichment. Fig. 6 illustrates the top 12 GO enrichment pathways, among which "Exogenous peptide presentation," "MHC class II assembly," "Peptide–MHC II assembly," and "Symbiotic interaction" pathways were highly significantly enriched. These pathways primarily involved exogenous antigen processing and presentation, as well as immune responses related to host-symbiont interaction. Additionally, other pathways were mainly associated with cell apoptosis, membrane structure and migration and protein folding.

DISCUSSION

This study systematically evaluated the prognostic characteristics of LUAD patients who received radiotherapy, a clinically relevant yet insufficiently characterized subgroup. Using large-scale genomic and clinical data from TCGA, a robust machine learning–based prognostic model was developed and validated to specifically address irradiated LUAD. Based on this model, the survival effects of different post-radiotherapy pharmacologic strategies and their potential molecular mechanisms were further investigated.

Fig. 3: C-index heatmap of machine-learning model combinations in training and testing cohorts.

Table 1: Top ten models and c-index values.

Model name	Training dataset (C-index)	Validation dataset (C-index)
StepCox[forward]	0.9733	0.4744
StepCox[both]+Enet[alpha=0.1]	0.7703	0.5765
StepCox[backward]+Enet[alpha=0.1]	0.7634	0.5835
StepCox[both]+Enet[alpha=0.2]	0.773	0.5724
Enet[alpha=0.1]	0.7675	0.5817
StepCox[backward]+Enet[alpha=0.2]	0.7532	0.5853
plsRcox	0.7513	0.5711
CoxBoost+Enet[alpha=0.9]	0.7379	0.4507
CoxBoost+StepCox[forward]	0.7388	0.5465
CoxBoost+Enet[alpha=0.2]	0.7369	0.5732

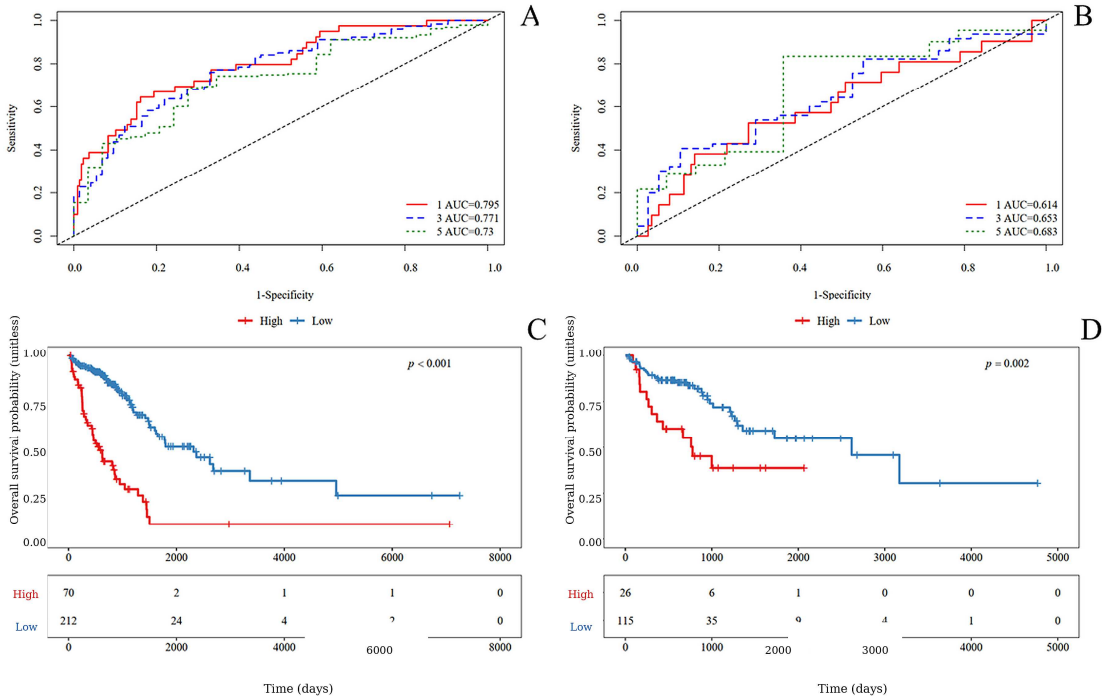

Fig. 4: Prognostic-model performance. (A) Training-set ROC at 1, 3 and 5 years (x-axis: 1-specificity, unitless; y-axis: sensitivity, unitless); (B) Validation-set ROC with the same axes; (C) Training-set KM curves (x-axis: time, days; y-axis: OS probability, unitless); (D) Validation-set KM curves with the same axes.

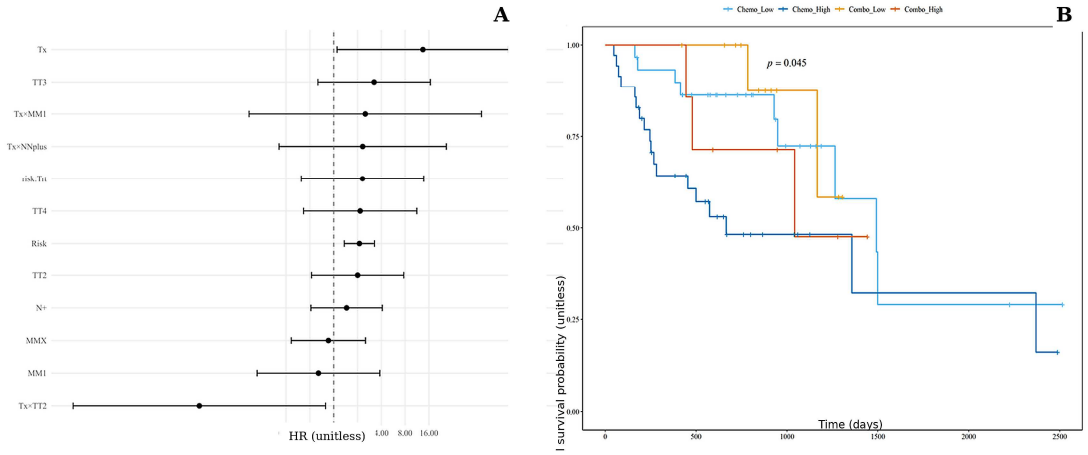

Fig. 5: Treatment-effect analyses. (A) Cox forest plot (x-axis: HR, unitless; y-axis: covariates). (B) Risk-stratified KM curves by medication group (x-axis: time, days; y-axis: OS probability, unitless).

Table 2: Significant differences in multivariate Cox regression results.

Variable	HR(95% CI)	p
Treatment	13.43(1.10-163.72)	0.04
Risk	2.11 (1.36-3.28)	<0.001
Trt×T2	0.02 (0.0005-0.789)	0.04

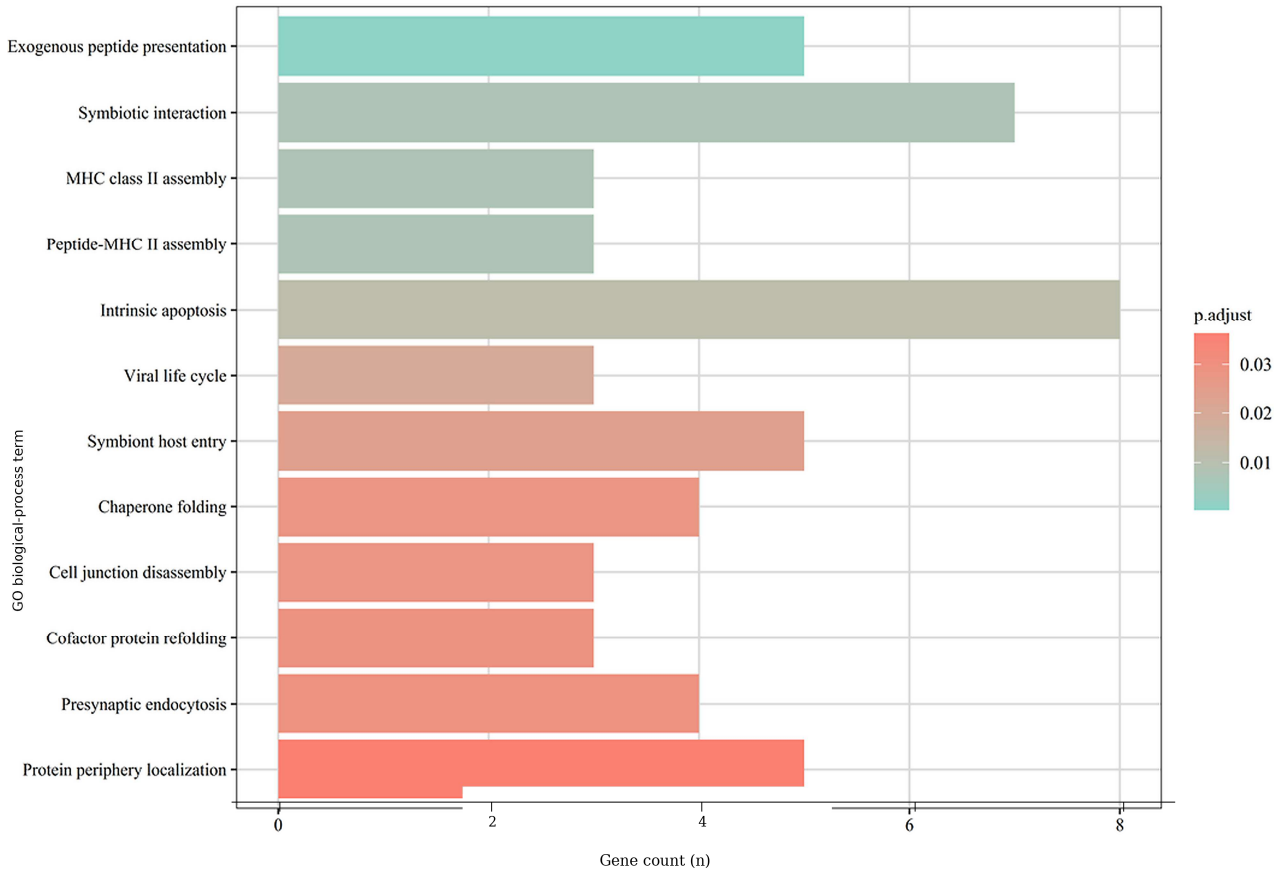

Fig. 6: GO enrichment analysis (x-axis: gene count, n; y-axis: GO terms; color: adjusted p value, unitless).

The findings showed a strong link between the risk stratification defined by the model and patient outcomes, with various therapeutic regimens showing notably different survival benefits. Among these, patients in the low-risk group receiving combined targeted therapy achieved the most favorable prognosis. These observations highlight pronounced heterogeneity in treatment response to radiotherapy and provide valuable clues regarding the molecular processes influencing LUAD prognosis, underscoring the potential clinical relevance of the model for guiding individualized therapy.

The prognostic model effectively distinguished patients into separate low-risk and high-risk groups. Compared with models designed for unselected LUAD populations (Li *et al*., 2019; Xie *et al*., 2022; Zhang *et al*., 2022). This model was specifically optimized for those who underwent radiotherapy, a subgroup facing unique biological challenges and treatment responses. Such specialization enhances its clinical applicability by providing more precise risk assessments for this specific therapeutic context. Subgroup analysis further indicated that low-risk patients receiving targeted combination therapy consistently exhibited better survival outcomes than high-risk patients, highlighting the critical role of accurate risk stratification in optimizing treatment efficacy.

A major insight from this study is the marked heterogeneity in therapeutic outcomes across different concomitant drug categories in radiotherapy-treated LUAD. Both treatment and model-derived risk score emerged as strong independent predictors of survival, with significant interaction between them. The superior outcomes observed in the low-risk combination subgroup underscore the importance of integrating molecular risk stratification with treatment selection to identify patients most likely to benefit from intensified or personalized regimens. This study revealed that patients receiving targeted or combination therapy—mainly erlotinib, gefitinib and bevacizumab—exhibited significantly prolonged survival

compared with those treated with conventional chemotherapy, consistent with previous evidence in LUAD (Deng *et al*., 2021; Greenhalgh *et al*., 2021; Lee *et al*., 2017; Nan *et al*., 2017). These targeted agents primarily act against *EGFR* mutations, which occur in approximately 30–50% of LUAD patients (Chen *et al*., 2023; Metro and Crinò, 2012). Earlier studies demonstrated that erlotinib and gefitinib sensitize LUAD cells to radiation (Liu *et al*., 2020; Nymoen *et al*., 2024; Xu *et al*., 2016), whereas bevacizumab can normalize the tumor vasculature to improve oxygenation, thereby enhancing radiosensitivity (Guo *et al*., 2024; Zhou *et al*., 2024). Synchronous or sequential administration of these agents may further potentiate the radiation effect and suppress post-radiotherapy proliferative recovery, leading to significant prolongation of overall survival (Motta-Guerrero *et al*., 2024). Although several studies have established targeted therapy as a standard approach for advanced LUAD (Cheng *et al*., 2021; Ohe *et al*., 2019), its use in the setting of radiotherapy remained limited in historical TCGA datasets, largely reflecting earlier therapeutic practices. In contrast, conventional chemotherapy was prescribed more frequently, reflecting conservative decision-making patterns in clinical management.

Functional enrichment analysis provided preliminary molecular context for the observed treatment effects. Significant enrichment of pathways related to "Exogenous peptide presentation," "MHC class II assembly," and "Peptide–MHC II assembly," is consistent with the hypothesis that immune modulation may contribute to post-radiotherapy outcomes, particularly through antigen processing and presentation pathways. Multiple studies have identified enhanced expression of these pathways in irradiated tumor cells, suggesting that radiotherapy not only exerts direct cytotoxicity through DNA damage but also induces *MHC II*-mediated exogenous peptide presentation, leading to immunogenic cell death. In this context, the addition of targeted therapy might further influence antigen exposure and presentation, although the present results do not directly demonstrate such effects and should be regarded as hypothesis-generating (Sun *et al*., 2023; Yang *et al*., 2023). Furthermore, enrichment of immune- and virus-like response pathways—such as "Symbiotic interaction" and " Viral life cycle "—may reflect activation of antiviral or virus-mimicking transcriptional programs in some patients. Recent studies have shown that such “viral mimicry” responses, driven by endogenous retroelements and interferon-related signaling, can enhance antitumor immunity in experimental models (Chiappinelli *et al*., 2016; Ishak and De Carvalho, 2020; Jansz and Faulkner, 2021; Vitiello *et al*., 2021). These findings are consistent with prior observations but remain associative and the mechanisms linking these pathways to clinical outcomes will require further experimental and clinical validation.

Several limitations should be acknowledged. The relatively limited use of targeted therapy within the historical TCGA cohort led to a smaller number of patients in the combination therapy group and lack of representation of newer targeted agents. In addition, the radiotherapy annotation in TCGA is incomplete and inconsistently documented, which may affect treatment-group classification and limit the reproducibility of radiotherapy-specific analyses. Potential confounding effects in enrichment analyses may also affect mechanistic interpretation. Although the present study demonstrates that combination therapy provides greater benefit in low-risk patients, outcomes in high-risk patients have not been directly evaluated. The genomic and clinical data available in TCGA could support such analyses; however, the limited sample size and incomplete treatment annotation constrain robust assessment in this subgroup. Future investigations based on larger, prospectively collected datasets, including contemporary targeted and immunotherapeutic regimens, are warranted to confirm and extend these observations.

In summary, this study formulated and authenticated a gene-based prognostic model for LUAD patients undergoing radiotherapy, achieving effective risk stratification and revealing marked heterogeneity in treatment responses. The findings demonstrate that targeted combination therapy provides superior survival benefit, particularly for low-risk individuals and that immune-related pathways play a pivotal role in determining prognosis. These findings enhance the understanding of the molecular basis underlying radiotherapy responses in LUAD and offer support for developing risk-tailored and personalized therapeutic approaches. Prospective validation and mechanistic exploration are required to translate these insights into clinical practice.

CONCLUSION

This study established and validated a machine learning–based prognostic model tailored for LUAD patients undergoing radiotherapy. The model demonstrated robust performance in stratifying patients by survival risk and revealed substantial heterogeneity in therapeutic responses across concomitant pharmacologic regimens. Notably, low-risk individuals achieved the most significant survival benefits when treated with combined targeted therapy. Functional enrichment analyses indicated that immune-related pathways—particularly those involving antigen presentation and viral mimicry signaling—play central roles in mediating post-radiotherapy outcomes. Collectively, these findings emphasize the clinical value of integrating molecular risk stratification with targeted combination strategies to optimize radiotherapeutic outcomes and guide precision management in LUAD.

Acknowledgments
The authors thank the Cancer Genome Atlas (TCGA) Research Network and all investigators who made the LUAD genomic and clinical datasets publicly available.

Authors' contribution
Zaishuang Ju: Conceived and supervised the study; Zaishuang Ju and Yutong Wu: Collected and preprocessed the data; Lili Tian and Dachuan Shen: Developed and optimized the machine learning models; Zaishuang Ju and Ruoyu Wang: performed model validation and visualization; Zaishuang Ju and Yutong Wu: Drafted the manuscript. All authors contributed to the review and approval of the final version.

Funding
There was no funding.

Data availability statement
The datasets analyzed during the current study are publicly available from the Cancer Genome Atlas (TCGA) data portal. The processed data generated during the current study are available from the corresponding author on reasonable request.

Ethical approval
This study used publicly available, de-identified TCGA-LUAD data and did not involve direct contact with human participants or collection of new human specimens. Therefore, institutional ethical approval was not required and no institutional ethical approval number was assigned for this secondary analysis. The original TCGA studies were conducted with appropriate institutional review board approval and informed consent. This study was performed in adherence with the STROBE guidelines. See supplementary file for the STROBE checklist.

Conflict of interest
The authors declare no conflict of interest.

Supplementary data
https://pjps.pk/uploads/2026/08/SUP1787723491.pdf

REFERENCES

Chen CL, Wang ST, Liao WC, Chen CH, Tu CY, Hsia TC, Cheng WC and Chen HJ (2023). A real-world study comparing perioperative chemotherapy and EGFR-tyrosine kinase inhibitors for treatment of resected stage III EGFR-mutant adenocarcinoma. *BMC Cancer*, **23**: 847.

Chen T and Guestrin C (2016). XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, pp. 785-794.

Cheng Y, He Y, Li W, Zhang H, Zhou Q, Wang B, Liu C, Walding A, Saggese M and Huang X (2021). Osimertinib versus comparator EGFR TKI as first-line treatment for EGFR-mutated advanced NSCLC: FLAURA China, a randomized study. *Target. Oncol*., **16**(2): 165-176.

Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, Hein A, Rote NS, Cope LM and Snyder A (2016). Abstract B32: Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cancer Res*., **76**(14 Suppl): B32.

Deng Z, Qin Y, Liu Y, Zhang Y and Lu Y (2021). Role of antiangiogenic agents combined with EGFR tyrosine kinase inhibitors in treatment-naive lung cancer: a meta-analysis. *Clin. Lung Cancer*, **22**(1): e70-e83.

Fouodo CJ, Konig IR, Weihs C, Ziegler A and Wright MN (2018). Support vector machines for survival analysis with R. *R J*., **10**(1): 412-423.

Greenhalgh J, Boland A, Bates V, Vecchio F, Dundar Y, Chaplin M and Green JA (2021). First-line treatment of advanced epidermal growth factor receptor mutation-positive non-squamous non-small cell lung cancer. *Cochrane Database Syst. Rev*., **3**: CD010383.

Guo Z, Lei L, Zhang Z, Du M and Chen Z (2024). The potential of vascular normalization for sensitization to radiotherapy. *Heliyon*, **10**(4): e26279.

Han L, Shi H, Luo Y, Sun W, Li S, Zhang N, Jiang X, Gong Y and Xie C (2020). Gene signature based on B cell predicts clinical outcome of radiotherapy and immunotherapy for patients with lung adenocarcinoma. *Cancer Med*., **9**(24): 9581-9594.

Heagerty PJ, Lumley T and Pepe MS (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, **56**(2): 337-344.

Horinouch H (2018). Precision radiotherapy for patients with locally advanced non-small cell lung cancer in the era of immunotherapy and precision medicine. *Transl. Lung Cancer Res*., **7**(Suppl 3): S146-S148.

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L and Girke T (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods*, **12**(2): 115-121.

Ishak CA and De Carvalho DD (2020). Reactivation of endogenous retroelements in cancer development and therapy. *Annu. Rev. Cancer Biol*., **4**: 159-176.

Jansz N and Faulkner GJ (2021). Endogenous retroviruses in the origins and treatment of cancer. *Genome Biol*., **22**: 147.

Lee CK, Davies L, Wu YL, Mitsudomi T, Inoue A, Rosell R, Zhou C, Nakagawa K, Thongprasert S and Fukuoka M (2017). Gefitinib or erlotinib vs chemotherapy for EGFR mutation-positive lung cancer: Individual patient data meta-analysis of overall survival. *J. Natl. Cancer Inst*., **109**(6): djw279.

Lender Y, Givton O, Bornshten R, Azar M, Moscona R, Yarden Y and Rubin E (2025). Immune clustering

reveals molecularly distinct subtypes of lung adenocarcinoma. *Biomedicines*, **13**(4): 849.

Li W, Liu S, Su S, Chen Y and Sun G (2021). Construction and validation of a novel prognostic signature of microRNAs in lung adenocarcinoma. *Peer J.*, **9**: e10470.

Li Y, Ge D, Gu J, Xu F, Zhu Q and Lu C (2019). A large cohort study identifying a novel prognosis prediction model for lung adenocarcinoma through machine learning strategies. *BMC Cancer*, **19**: 886.

Liu J, Cheng H, Zhang C, Dai W, Ge X, Xu L, Zhang Z and Sun X (2017). Meta-analysis of hypofractionated radiotherapy versus conventional radiotherapy in locally advanced non-small-cell lung cancer. *Int. J. Clin. Exp. Pathol*., **10**(6): 6179-6188.

Liu Y, Li Z, Meng Q, Ning A, Zhou S, Li S, Tao X, Wu Y, Chen Q and Tian T (2024). Identification of the consistently differential expressed hub mRNAs and proteins in lung adenocarcinoma and construction of the prognostic signature: A multidimensional analysis. *Int. J. Surg*., **110**(2): 1052-1067.

Liu Z, Han C and Fu YX (2020). Targeting innate sensing in the tumor microenvironment to improve immunotherapy. *Cell. Mol. Immunol*., **17**(1): 13-26.

Love MI, Huber W and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*., **15**: 550.

Ma J, Yang X, Zhang J, Antonoff MB, Wu Q and Ji H (2023). APOBEC mutational signature predicts prognosis and immunotherapy response in nonsmoking patients with lung adenocarcinoma. *Transl. Lung Cancer Res*., **12**(4): 580-593.

Metro G and Crino L (2012). Advances on EGFR mutation for lung cancer. *Transl. Lung Cancer Res*., **1**(1): 5-13.

Motta-Guerrero R, Garrido-Lecca AL, Failoc-Rojas VE, Calle-Villavicencio A, Villacorta-Carranza R, Huerta-Collado Y, Torres-Mera A, Valladares-Garrido MJ, Rivera-Francia V and Carracedo C (2024). Effectiveness and safety of the bevacizumab and erlotinib combination versus erlotinib alone in EGFR mutant metastatic non-small-cell lung cancer: Systematic review and meta-analysis. *Front. Oncol*., **13**: 1335373.

Nan X, Xie C, Yu X and Liu J (2017). EGFR TKI as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer. *Oncotarget*, **8**(43): 75712-75726.

Nymoen HM, Alver TN, Horndalsveen H, Eide HA, Bjanaes MM, Brustugun OT, Gronberg BH, Haakensen VD and Helland A (2024). Thoracic radiation in combination with erlotinib-results from a phase 2 randomized trial. *Front. Oncol*., **14**: 1412716.

Ohe Y, Imamura F, Nogami N, Okamoto I, Kurata T, Kato T, Sugawara S, Ramalingam SS, Uchida H and Hodge R (2019). Osimertinib versus standard-of-care EGFR-TKI as first-line treatment for EGFRm advanced NSCLC: FLAURA Japanese subset. *Jpn. J. Clin. Oncol*., **49**(1): 29-36.

Prentice RL and Zhao S (2021). Regression models and multivariate life tables. *J. Am. Stat. Assoc*., **116**(535): 1330-1345.

Sun D, Duan X, Li N, Qiao O, Hou Y, Ma Z, Liu S, Gong Y and Liu Z (2025). Construction of ubiquitination-related risk model for predicting prognosis in lung adenocarcinoma. *Sci. Rep*., **15**: 11787.

Sun L, Zhou H, Wu C and Peng Y (2023). Molecular markers that predict response to combined radiotherapy and immunotherapy in patients with lung adenocarcinoma: A bioinformatics analysis. *Transl. Cancer Res*., **12**(10): 2646-2661.

Taylor JMG (2011). Random survival forests. *J. Thorac. Oncol*., **6**(12): 1974-1975.

van Diessen JNA (2019). Locally advanced lung cancer: improved patient selection and treatment. PhD thesis, Vrije Universiteit Amsterdam, Amsterdam.

Vitiello GAF, Ferreira WAS, Cordeiro de Lima VC and Medina TS (2021). Antiviral responses in cancer: boosting antitumor immunity through activation of interferon pathway in the tumor microenvironment. *Front. Immunol*., **12**: 782852.

Wang Y, Shang Y, Hua M, Wang Y, Hui B, Hu W, Zhu M, Zhang X and Li J (2025). Identification and exploration of key genes associated with radioresistance in lung adenocarcinoma. *Cancer Cell Int*., **25**: 155.

Wrona A, Dziadziuszko R and Jassem J (2021). Combining radiotherapy with targeted therapies in non-small cell lung cancer: Focus on anti-EGFR, anti-ALK and anti-angiogenic agents. *Transl. Lung Cancer Res*., **10**(4): 2032-2050.

Xie B, Chen X, Deng Q, Shi K, Xiao J, Zou Y, Yang B, Guan A, Yang S and Dai Z (2022). Development and validation of a prognostic nomogram for lung adenocarcinoma: A population-based study. *J. Healthc. Eng*., 2022: 5698582.

Xu R, Han F, Zhao Y, Liu A, An N, Wang B, Zardo P, Sanz-Santos J, Franssen AJPM, de Loos ER and Zhao M (2024). Role of CENPL, DARS2 and PAICS in determining the prognosis of patients with lung adenocarcinoma. *Transl. Lung Cancer Res*., **13**(12): 2729-2745.

Xu W, Yang Z and Lu N (2016). Molecular targeted therapy for the treatment of gastric cancer. *J. Exp. Clin. Cancer Res*., **35**: 1.

Xu Z, Hao X, Lin L, Li J and Xing P (2021). Concurrent chemotherapy and first-generation epidermal growth factor receptor (EGFR)-tyrosine kinase inhibitors (TKIs) with or without an antiangiogenic agent as first-line treatment in advanced lung adenocarcinoma harboring an EGFR mutation. *Thorac. Cancer*, **12**(16): 2233-2240.

Yang C, Liang Y, Liu N and Sun M (2023). Role of the cGAS-STING pathway in radiotherapy for non-small cell lung cancer. *Radiat. Oncol*., **18**: 145.

Ying X, Liu H, Wang M, Peng M, Ruan P, Verma V and Han G (2020). Clinical response to apatinib combined

with brain radiotherapy in EGFR wild-type and ALK-negative lung adenocarcinoma with multiple brain metastases. *Front. Oncol*., **10**: 517.

Yu G, Wang LG, Han Y and He QY (2012). Cluster Profiler: An R package for comparing biological themes among gene clusters. *OMICS*, **16**(5): 284-287.

Yue J, Guo H, Ma J, Shi W and Wu Y (2024). Novel prognostic signature for lung adenocarcinoma based on immune-related mRNA pairs. *Heliyon*, **10**(3): e25172.

Zhang B, Wang Y, Zhou X, Zhang Z, Ju H, Diao X, Wu J and Zhang J (2022). Construction of a prognostic and early diagnosis model for LUAD based on necroptosis gene signature and exploration of immunotherapy potential. *Cancers*, **14**(20): 5153.

Zhang L, Liu Y, Zhuang JG, Guo J, Li YT, Dong Y and Song G (2023). Identification of key genes and biological pathways in lung adenocarcinoma by integrated bioinformatics analysis. *World J. Clin. Cases*, **11**(23): 5504-5518.

Zhao Y, Shi W and Tang Q (2023). An eleven-gene risk model associated with lymph node metastasis predicts overall survival in lung adenocarcinoma. *Sci. Rep*., **13**: 6852.

Zhou Y, Li J, Li Y, Deng G, Wang Q, Qin H, Li J and Li Z (2024). Bevacizumab improved prognosis for advanced EGFR-mutant lung adenocarcinoma with brain metastasis receiving cerebral radiotherapy. *Clin. Transl. Oncol*., **26**(8): 1968-1975.

Zia NN, Khokhar MA, Qamar S, Hanif A, Goraya AW and Awan NU (2019). Concurrent radiotherapy and chemotherapy with erlotinib followed by maintenance erlotinib in patients with Epidermal Growth Factor mutation-positive adenocarcinoma lung. *J. Pak. Med. Assoc*., **69**(11): 1605-1609.

Zou H and Hastie T (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol*., **67**(2): 301-320.