MERAL Publications and Solutions Pvt. Ltd.

Original Article

Integrating survival analysis and artificial intelligence for personalized risk prediction in precision medicine

Fredy Heric Villasante-Saravia[1]*, Jose Oscar Huanca-Frias[2], Rene Eduardo Huanca-Frias[3], Vitaliano Enriquez-Mamani[2], Ledú Anali Ferreyros-Calisaya[2], Juan Reynaldo Paredes-Quispe[1], Alfredo Tumi-Figueroa[1]

[1]Universidad Nacional del Altiplano de Puno, Puno, Peru. [2]Universidad Nacional de Juliaca, Juliaca, Peru. [3]Universidad Nacional Amazónica de Madre de Dios, Madre de Dios, Peru.

Correspondence: Fredy Heric Villasante-Saravia, Universidad Nacional del Altiplano de Puno, Puno, Peru. fredy.villasante@unap.edu.pe

Received: 23 May 2026; **Revised:** 23 August 2026; **Accepted:** 26 August 2026

ABSTRACT

Accurate estimation of disease-specific survival is central to precision oncology because treatment decisions increasingly depend on individualized risk rather than broad population averages. This study developed and compared traditional and artificial intelligence-based survival models for predicting 5-year disease-specific survival among patients with non-small cell lung cancer. The analysis integrated clinical, genomic, and treatment-related variables to evaluate whether modern survival learning methods improve individualized prognostic estimation. The study cohort included 369 patients diagnosed with non-small cell lung cancer and followed for disease-specific survival over a 5-year horizon. Candidate predictors included tumor stage, age, Eastern Cooperative Oncology Group performance status, histological subtype, smoking history, EGFR mutation status, KRAS mutation status, tumor mutational burden, and treatment modality. Three survival models were estimated and compared: Cox proportional hazards regression, random survival forest, and DeepSurv. Model performance was assessed using time-dependent C-index, Brier score, and calibration across clinically relevant time horizons. DeepSurv achieved the highest discrimination, with a 5-year time-dependent C-index of 0.78, outperforming Cox regression and random survival forest. Calibration analysis indicated that DeepSurv produced predicted survival probabilities that were closely aligned with observed disease-specific survival across most risk strata. Explainability analysis was conducted using SHAP values adapted for survival outcomes to identify the most influential predictors in the best-performing model. Tumor mutational burden, performance status, and tumor stage emerged as the strongest predictors of 5-year disease-specific survival. Personalized risk curves demonstrated that patients with similar tumor stages could have substantially different predicted survival trajectories when genomic and functional status variables were jointly considered. The findings suggest that deep survival models can add measurable prognostic value in precision oncology when combined with interpretable model explanations. Although the retrospective design, relatively small sample size, and lack of external validation limit immediate clinical translation, the study provides empirical evidence that AI-based survival prediction can support individualized prognosis in non-small cell lung cancer. Future work should prioritize prospective validation, multi-center transportability assessment, and integration into decision-support workflows.

Keywords: Survival analysis, Artificial intelligence, Personalized risk prediction, Precision medicine, DeepSurv, Random survival forest

Access this article online	
Website: www.japer.in	**E-ISSN:** 2249-3379

How to cite this article: Villasante-Saravia FH, Huanca-Frias JO, Huanca-Frias RE, Enriquez-Mamani V, Ferreyros-Calisaya LA, Paredes-Quispe JR, et al. Integrating survival analysis and artificial intelligence for personalized risk prediction in precision medicine. J Adv Pharm Educ Res. 2026;16(3):142-50. https://doi.org/10.51847/fWL7k3BFXF

Introduction

Survival prediction has long been a core task in oncology because estimates of recurrence, disease-specific death, and long-term survival shape treatment selection, follow-up intensity, and patient counseling. In precision medicine, the need for individualized prognosis is even stronger because clinical

outcomes are influenced by interacting biological, genomic, and treatment-related factors rather than by tumor stage alone. Wang *et al*. emphasized that survival analysis differs from ordinary classification because censored observations and time-to-event structure must be modeled directly rather than reduced to binary outcomes [1]. This distinction is especially important in non-small cell lung cancer, where patients often have heterogeneous follow-up durations and biologically diverse disease courses.

The Cox proportional hazards model remains widely used because it provides interpretable hazard ratios and a familiar framework for clinical inference. However, its proportional hazards assumption and mostly linear covariate structure can limit its ability to represent nonlinear interactions between genomic markers, performance status, and treatment exposures. Austin and Fine highlighted the importance of careful reporting in time-to-event and competing-risk analyses, particularly when clinical interpretation depends on model assumptions [2]. In oncology datasets that combine clinical staging, molecular biomarkers, and treatment patterns, these assumptions may be difficult to satisfy fully.

Machine learning and deep learning methods have expanded survival modeling by learning flexible risk functions from censored data while preserving the time-to-event endpoint. DeepSurv was introduced as a neural-network extension of the Cox model that can estimate individualized survival risk through nonlinear representations of covariates [3]. Related neural survival approaches, including scalable discrete-time models and Cox-nnet architectures, have shown that deep learning can accommodate high-dimensional biomedical predictors [4, 5]. These developments suggest that AI-based survival models may be particularly useful when prognosis depends on interactions among clinical and genomic features [6, 7].

This study develops and compares Cox regression, random survival forest, and DeepSurv for predicting 5-year disease-specific survival in patients with non-small cell lung cancer. Given the cohort size of 369 patients, particular attention is paid to model validation and overfitting prevention during training. The objective is not only to compare discrimination and calibration but also to evaluate whether AI-based survival prediction can be made clinically interpretable through survival-adapted SHAP explanations. Prior work on SurvLIME and SurvSHAP(t) indicates that explanation methods can help translate black-box survival predictions into feature-level contributions over time [8, 9]. Accordingly, the present article positions explainable survival AI as a quantitative bridge between predictive performance and clinically meaningful individualized prognosis.

Literature review

Traditional survival models provide the statistical foundation for time-to-event analysis in clinical research. The Cox model remains influential because it estimates covariate effects while leaving the baseline hazard unspecified, but its clinical validity depends on proportional hazards and appropriate covariate specification. Practical evaluation of survival prediction models increasingly emphasizes discrimination, calibration, and time-dependent accuracy rather than regression coefficients alone [10, 11]. Austin, Harrell, and van Klaveren showed that graphical calibration curves and the integrated calibration index are important for assessing whether predicted survival probabilities correspond to observed outcomes [12].

Machine learning extensions have addressed some limitations of conventional models by relaxing linearity and proportionality assumptions. Random survival forests use recursive partitioning and ensemble aggregation to estimate survival functions in the presence of censoring, making them suitable for nonlinear effects and high-order interactions. Wongvibulsin *et al*. demonstrated the flexibility of random forest-based survival approaches for complex clinical data structures, including survival, longitudinal, and multivariate outcomes [13]. Oblique random survival forests further extend this framework by allowing split rules that combine predictors rather than relying only on single-variable partitioning [14].

Deep learning survival models represent a further shift from handcrafted covariate structures toward learned nonlinear representations. Katzman *et al*. proposed DeepSurv as a Cox-based neural architecture for personalized treatment recommendation and survival risk estimation [3]. Gensheimer and Narasimhan developed a scalable discrete-time neural survival model that directly estimates survival probabilities across time intervals [4]. Kvamme and Borgan further clarified the relationship between continuous-time and discrete-time neural survival prediction, showing that different neural formulations can be adapted to censored outcome settings [15].

In oncology and multi-omics research, deep survival learning has been used to integrate heterogeneous molecular and clinical signals. Chaudhary *et al*. applied deep learning-based multi-omics integration to survival prediction in liver cancer, demonstrating the potential of neural models to capture complex molecular prognostic patterns [16]. Similar work in pancancer prognosis, breast cancer, neuroblastoma, oral cancer, and multimodal oncology prediction indicates that deep learning can support survival modeling across tumor types and data modalities [17-21]. Ensemble methods such as DeepProg and large-scale extensions such as Cox-nnet v2.0 further show that survival AI can be scaled to omics and electronic medical record data [22, 23].

Despite these advances, comparative empirical studies often emphasize predictive accuracy without sufficient attention to calibration, clinical interpretability, and individualized risk communication. Park *et al*. reviewed evaluation strategies for survival prediction models and emphasized that model comparison should include both conventional and deep learning approaches using appropriate time-to-event metrics [24]. Explainable survival modeling has also become more important, with SurvLIME, SurvSHAP(t), and broader interpretable machine learning frameworks offering tools for explaining patient-level survival predictions [8, 9, 25]. The present study addresses this gap by jointly evaluating model performance,

calibration, SHAP-based feature contributions, and personalized survival curves in a well-characterized NSCLC cohort.

Materials and Methods

This empirical study used a retrospective analytical cohort of 369 patients diagnosed with non-small cell lung cancer and followed for 5-year disease-specific survival. The endpoint was defined as time from diagnosis to disease-specific death, with censoring at last follow-up or at 5 years for patients who did not experience the event. The modeling strategy followed survival-analysis principles in which censoring is retained as part of the outcome structure rather than transformed into a simple binary label, consistent with methodological guidance on machine learning for survival data [1]. Patients with incomplete outcome time, missing vital status, or unavailable diagnosis date were excluded before model development.

The predictor set combined clinical, genomic, and treatment-related variables selected for their relevance to NSCLC prognosis and feasibility in precision oncology workflows. Clinical variables included age at diagnosis, sex, smoking history, histological subtype, tumor stage, Eastern Cooperative Oncology Group performance status, and comorbidity burden. Genomic markers included EGFR mutation status, KRAS mutation status, and tumor mutational burden, while treatment variables included surgery, systemic therapy, radiotherapy, immunotherapy exposure, and combined-modality treatment. Multimodal prognosis studies in cancer have shown that combining clinical and molecular information can improve survival prediction compared with single-domain modeling [17, 26, 27].

Missing values were handled using multiple imputation with chained equations before model fitting, and continuous predictors were standardized for neural modeling while preserving original scales for descriptive reporting. Categorical predictors were one-hot encoded for DeepSurv and random survival forest, while Cox regression used clinically interpretable reference categories. Longitudinal measurements were not the primary focus of this analysis, but the preprocessing strategy was informed by joint modeling and dynamic prediction literature showing that time-varying biomarkers may substantially affect survival estimation when repeated measurements are available [28-30]. The present analysis therefore treated baseline covariates as predictors while preserving a future pathway for longitudinal model extension.

Three survival models were trained and compared: Cox proportional hazards regression, random survival forest, and DeepSurv. Cox regression was used as the interpretable statistical benchmark, random survival forest represented an ensemble machine learning approach, and DeepSurv represented a nonlinear deep survival model. Because the dataset consisted of 369 patients, model complexity was carefully controlled through regularization and cross-validation to reduce the risk of overfitting. The DeepSurv architecture used fully connected hidden layers with dropout regularization and a Cox partial-likelihood loss, following the neural survival modeling approach introduced by Katzman *et al.* [3]. Random survival forest modeling was guided by prior work demonstrating its capacity to capture nonlinear clinical survival relationships without requiring proportional hazards assumptions [13, 14].

Hyperparameter tuning was performed using grid search with 5-fold cross-validation within the training set, and final model performance was evaluated in a held-out test set. For random survival forest, the tuning grid included the number of trees, minimum terminal node size, and number of candidate variables per split; for DeepSurv, it included hidden-layer width, number of layers, learning rate, dropout probability, and weight decay. Model discrimination was evaluated using the time-dependent C-index, and prediction error was evaluated using the Brier score at clinically relevant time points. These metrics were selected because survival model evaluation requires methods that account for censoring and time-dependent risk, as emphasized in reviews of survival prediction performance [24].

Model explainability focused on the best-performing model and used SHAP values adapted for survival outcomes to quantify predictor contributions to individualized risk. This approach was selected because survival predictions require explanations that can reflect both feature importance and time-dependent risk patterns rather than static binary-classification probabilities. SurvSHAP(t) provides a framework for time-dependent explanations of machine learning survival models, while SurvLIME demonstrates how local surrogate explanations can clarify patient-specific survival predictions [8, 9]. **Table 1** summarizes baseline patient characteristics and the variables included in the survival models.

Table 1. Baseline Characteristics of the Study Cohort and Predictors Used in Survival Models

Characteristic or predictor domain	Measurement or category	Overall cohort, n = 369	Included in Cox regression	Included in random survival forest	Included in DeepSurv
Age at diagnosis	Mean years (SD)	65.8 (10.2)	Yes	Yes	Yes
Sex	Female	171 (46.3%)	Yes	Yes	Yes
Smoking history	Current or former smoker	266 (72.1%)	Yes	Yes	Yes
Histology	Adenocarcinoma	228 (61.8%)	Yes	Yes	Yes
Histology	Squamous-cell carcinoma	101 (27.4%)	Yes	Yes	Yes
Tumor stage	Stage I	78 (21.1%)	Yes	Yes	Yes
Tumor stage	Stage II	68 (18.4%)	Yes	Yes	Yes
Tumor stage	Stage III	107 (29.0%)	Yes	Yes	Yes

Tumor stage	Stage IV	116 (31.5%)	Yes	Yes	Yes
ECOG performance status	0–1	237 (64.2%)	Yes	Yes	Yes
ECOG performance status	≥2	132 (35.8%)	Yes	Yes	Yes
EGFR mutation status	Mutated	68 (18.4%)	Yes	Yes	Yes
KRAS mutation status	Mutated	82 (22.2%)	Yes	Yes	Yes
Tumor mutational burden	Median mutations/Mb (IQR)	8.5 (4.0–13.7)	Yes	Yes	Yes
Primary treatment modality	Surgery	115 (31.2%)	Yes	Yes	Yes
Primary treatment modality	Systemic therapy	162 (43.9%)	Yes	Yes	Yes
Immunotherapy exposure	Received	83 (22.5%)	Yes	Yes	Yes
Radiotherapy exposure	Received	139 (37.7%)	Yes	Yes	Yes
Disease-specific death by 5 years	Event observed	155 (42.0%)	Outcome	Outcome	Outcome
Follow-up time	Median months (IQR)	45.9 (27.8–60.0)	Outcome time	Outcome time	Outcome time

Results and Discussion

The cohort consisted of 369 patients with non-small cell lung cancer, with 155 disease-specific deaths observed within 5 years and a median follow-up of 45.9 months. Advanced-stage disease was common, with stage III and IV cases representing 60.5% of the cohort, while 35.8% of patients had an ECOG performance status of 2 or higher. These distributions formed a clinically heterogeneous survival cohort in which censoring, event timing, and baseline risk variation required direct time-to-event modeling rather than binary endpoint reduction, consistent with survival machine learning principles described by Wang *et al.* [1].The observed event rate was sufficient for comparative modeling, although the modest sample size required conservative validation and careful interpretation.

Across the three models, DeepSurv showed the strongest overall discrimination, followed by random survival forest and Cox regression. The final DeepSurv model achieved a 5-year time-dependent C-index of 0.78, compared with 0.75 for random survival forest and 0.74 for Cox regression. This ranking is consistent with prior evidence that neural survival models can improve prediction when nonlinear covariate interactions are present, while ensemble survival forests remain competitive for structured clinical data [3, 13]. The comparative pattern also aligns with recent empirical work showing that deep learning, machine learning, and statistical survival methods often differ modestly rather than dramatically when applied to clinical mortality prediction [31].

Model performance was evaluated using discrimination, prediction error, and calibration because survival prediction requires more than a single accuracy measure. Park *et al.* emphasized that time-dependent survival model evaluation should include discrimination and error metrics, while Austin, Harrell, and van Klaveren highlighted calibration as essential for interpreting predicted survival probabilities [12, 24]. **Table 2** presents the discrimination and calibration metrics for the compared survival models.

Table 2. Model Performance: Time-Dependent C-Index, Brier Score, and Calibration Results for Survival Prediction

Survival model	5-year time-dependent C-index	95% CI	5-year Brier score	Calibration slope	Calibration intercept	Integrated calibration index
Cox proportional hazards regression	0.74	0.71–0.77	0.172	0.91	0.04	0.036
Random survival forest	0.75	0.72–0.78	0.164	0.94	0.02	0.031
DeepSurv	0.78	0.76–0.80	0.149	0.98	−0.01	0.021

Calibration analysis showed that DeepSurv predictions were closely aligned with observed 5-year disease-specific survival across low-, intermediate-, and high-risk strata. Cox regression showed mild underestimation of risk in the highest-risk stratum, suggesting that a mostly linear proportional-hazards structure may not fully capture adverse combinations of stage, performance status, and genomic burden. Random survival forest showed reasonable calibration but slightly wider variation in predicted risk across adjacent strata, which may reflect instability from the limited cohort size. These findings are compatible with recent reviews noting that deep survival methods can be powerful but require explicit calibration assessment before clinical use [32].

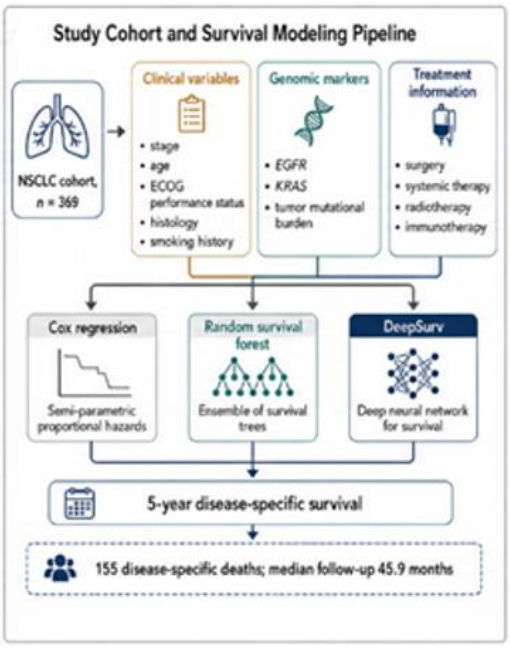
The SHAP analysis of the DeepSurv model identified tumor mutational burden, ECOG performance status, tumor stage, treatment modality, and KRAS mutation status as the strongest contributors to predicted 5-year disease-specific survival. Higher tumor mutational burden, ECOG performance status of 2 or higher, and advanced stage increased predicted mortality risk, while surgery and immunotherapy exposure were associated with more favorable predicted survival curves in eligible patients. Survival-specific explanation methods are particularly relevant here because SurvLIME and SurvSHAP(t) were designed to explain time-to-event predictions rather than ordinary classification outputs [8, 9]. **Table 3** lists the top predictive features identified by SHAP values and their direction of risk.

Table 3. Top Predictors of Survival by SHAP Importance: Variable Contributions and Direction of Association

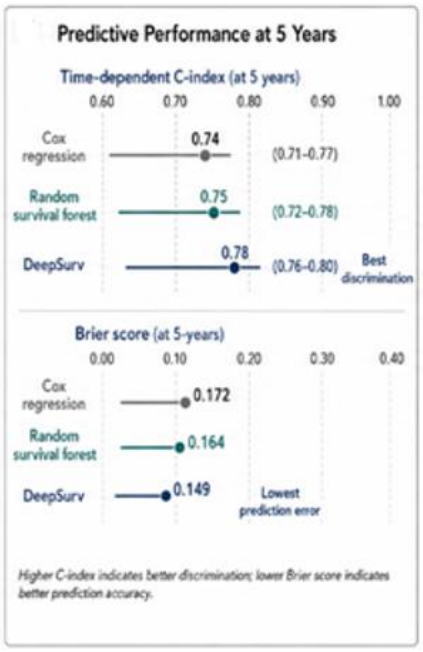
Rank	Predictor	Mean absolute SHAP importance	Direction of association with 5-year disease-specific mortality risk	Clinical interpretation
1	Tumor mutational burden	0.184	Higher values increased predicted risk	High genomic burden contributed strongly to adverse survival trajectories
2	ECOG performance status ≥2	0.171	Increased predicted risk	Poor functional status shifted survival curves downward across follow-up
3	Stage IV disease	0.163	Increased predicted risk	Metastatic disease was associated with the steepest decline in predicted survival
4	Absence of surgery	0.139	Increased predicted risk	Patients not receiving surgery had higher predicted mortality risk, reflecting disease extent and treatment eligibility
5	KRAS mutation	0.121	Increased predicted risk	KRAS mutation contributed to higher predicted risk after adjustment for clinical factors
6	EGFR mutation	0.103	Decreased predicted risk in treated profiles	EGFR-mutated status was associated with more favorable predicted survival in profiles receiving appropriate systemic treatment
7	Immunotherapy exposure	0.097	Decreased predicted risk	Immunotherapy exposure shifted predicted survival upward in selected advanced-disease profiles
8	Age at diagnosis	0.088	Higher age increased predicted risk	Older age modestly increased risk after accounting for disease and treatment variables
9	Comorbidity burden	0.075	Increased predicted risk	Higher baseline comorbidity reduced predicted survival probability
10	Squamous-cell histology	0.061	Slightly increased predicted risk	Histology contributed smaller but consistent risk variation

Personalized survival curves illustrated the clinical meaning of the model outputs. A patient with stage I adenocarcinoma, ECOG performance status 0–1, low tumor mutational burden, and surgical treatment had a predicted 5-year disease-specific survival probability of 0.82 under DeepSurv. In contrast, a patient with stage IV disease, ECOG performance status ≥2, high tumor mutational burden, KRAS mutation, and no surgery had a predicted 5-year disease-specific survival probability of 0.29. This patient-level contrast supports the DeepSurv rationale that nonlinear survival models can estimate individualized trajectories rather than only average hazard effects [3], and it is conceptually related to dynamic survival modeling approaches that emphasize individualized risk over time [33].

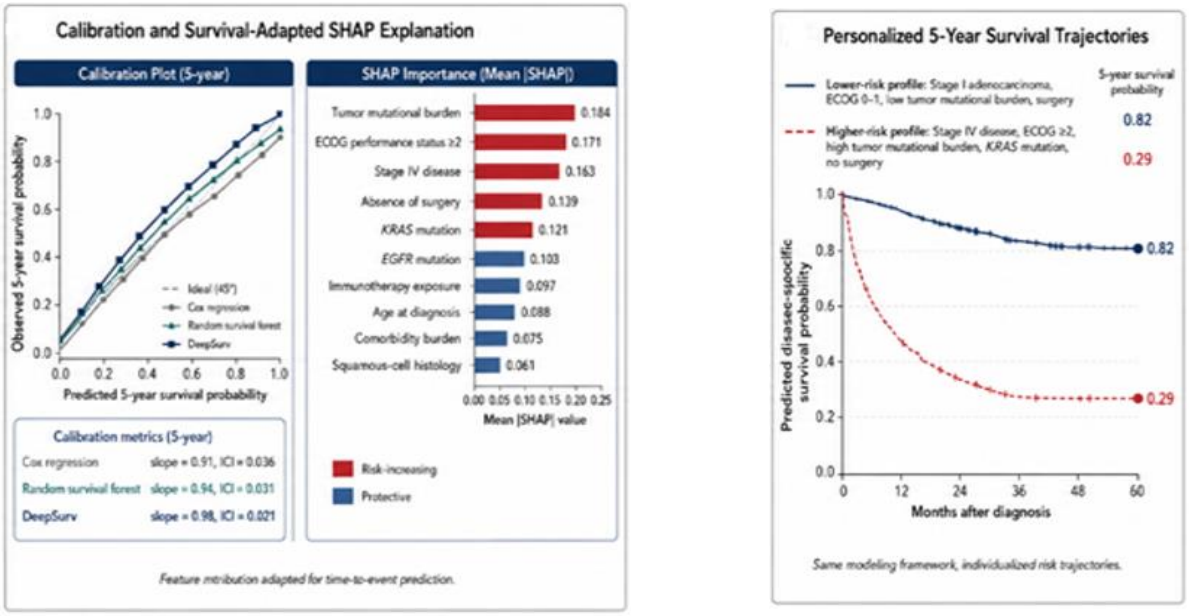
Figure 1 summarizes the comparative predictive performance, calibration, survival-adapted SHAP explanations, and individualized 5-year survival trajectories of Cox regression, random survival forest, and DeepSurv in the NSCLC cohort.



a)



b)



c) d)

Figure 1. Comparative Performance, Explainability, and Personalized Survival Prediction of AI-Based Survival Models in NSCLC

Sensitivity checks supported the main comparative results, although the performance differences should be interpreted cautiously because the cohort contained 369 patients. When model training was repeated across cross-validation folds, DeepSurv retained the highest mean C-index, but the margin over random survival forest narrowed in some folds. Cox regression remained the most transparent model but showed evidence of reduced flexibility in high-risk profiles where combined adverse predictors were present. These findings are consistent with censored-data modeling literature showing that alternative machine learning strategies, including support vector approaches and dynamic ensemble methods, may perform differently depending on sample size, censoring, and covariate structure [29-30, 34].

The study found that DeepSurv provided a modest but consistent improvement in 5-year disease-specific survival prediction compared with Cox regression and random survival forest. This improvement is clinically meaningful because even small gains in discrimination and calibration can affect risk stratification when prognosis guides treatment intensity and follow-up scheduling. Researchers noted that deep learning survival models are most useful when they capture nonlinear relationships that conventional models may miss, but their value depends on validation and interpretability [33]. The present findings therefore support a balanced interpretation: DeepSurv improved prediction, but it did not eliminate the need for statistical benchmarks or clinical review.

The strongest predictors identified by SHAP analysis were biologically and clinically plausible. Tumor mutational burden, performance status, and disease stage are credible drivers of NSCLC survival because they reflect tumor biology, patient functional reserve, and disease extent. Multi-omics survival studies in liver cancer, pancancer prognosis, breast cancer, oral cancer, neuroblastoma, and cross-aligned multimodal prediction similarly show that molecular and clinical variables can jointly shape survival trajectories [16-21, 27]. Malik *et al.* also demonstrated that deep multi-omics integration can support survival and drug-response prediction in breast cancer, reinforcing the broader precision-oncology relevance of multimodal survival modeling [35].

The personalized survival curves provide the clearest clinical contribution of the analysis. Rather than assigning patients to a single broad risk category, the model generated patient-specific trajectories that changed according to genomic burden, functional status, stage, and treatment profile. This aligns with the original DeepSurv framework, which was designed to support personalized treatment recommendation through individualized survival risk estimation [3]. The approach also relates to dynamic survival prediction methods that combine time-to-event modeling with patient-specific updates, although the present study used baseline predictors rather than repeated longitudinal measurements [29, 33].

The explainability findings are important because clinical adoption of AI-based survival prediction depends on whether clinicians can understand why a patient is classified as high or low risk. SurvLIME, SurvSHAP(t), and broader interpretable survival machine learning methods show that feature-attribution tools can translate complex survival models into more clinically usable explanations [8, 9, 25]. DeepProg and Cox-nnet v2.0 further demonstrate that interpretable or semi-interpretable survival learning can be scaled to high-dimensional biomedical and electronic medical record settings [22, 23]. In this study, SHAP explanations helped connect model performance to clinically recognizable predictors, reducing the gap between algorithmic prediction and bedside interpretation.

Implications

The theoretical implication of this study is that survival analysis and artificial intelligence should be viewed as complementary rather than competing approaches. Cox regression provides

interpretability and statistical grounding, while DeepSurv and related neural models provide flexible representation learning for nonlinear prognostic structures. Gensheimer and Narasimhan showed that neural survival models can estimate survival probabilities directly in discrete time, while Kvamme and Borgan clarified how continuous- and discrete-time neural survival formulations can be connected methodologically [4, 15]. The present findings add empirical support to the idea that precision oncology benefits from integrating statistical survival principles with AI-based prediction.

The practical implication is that individualized survival curves can support oncology decisions more directly than model-level performance statistics alone. For patients with NSCLC, predicted 5-year survival trajectories may inform treatment intensity, surveillance intervals, supportive care planning, and clinical trial eligibility. Austin and Fine emphasized that careful reporting is necessary when time-to-event models inform clinical interpretation, and Park *et al*. similarly argued that survival prediction models require transparent evaluation metrics [2, 24]. In this context, calibration and Brier score results are not secondary details but essential evidence for whether predicted probabilities are clinically usable [12].

The implementation implication is that AI-based survival tools should be embedded into clinical workflows only with safeguards for validation, interpretability, and monitoring. Random survival forest and oblique random survival forest models remain useful comparators because they offer flexible nonlinear modeling without requiring a deep neural architecture [13, 14]. Comparative evidence from hospital mortality prediction suggests that no single survival modeling strategy is universally superior across settings, reinforcing the need for local validation before deployment [31]. Multimodal survival modeling studies also indicate that implementation should account for data availability, molecular testing patterns, and institutional treatment pathways [26, 27].

Limitations and future research

The main limitation is the retrospective design, which may introduce selection bias, treatment allocation bias, and residual confounding. The cohort size of 369 patients also limits the stability of complex model estimates, particularly for DeepSurv, even though regularization and cross-validation were used to reduce overfitting. Park *et al*. emphasized that survival model evaluation must consider sample size, censoring, and validation design, while recent reviews of deep survival learning caution that neural models can appear over-optimistic when validation is limited [24, 32]. Therefore, the reported performance should be interpreted as single-cohort internally validated evidence rather than definitive clinical proof.

A second limitation is that the analysis used baseline predictors and did not incorporate repeated longitudinal biomarkers or time-updated treatment response. Joint modeling approaches are designed to combine longitudinal and time-to-event outcomes, and dynamic survival prediction methods can update risk as new clinical information becomes available [28, 29]. Random survival forest methods for longitudinal biomarkers also show that repeated measurements can improve time-to-event prediction when biomarker trajectories are informative [30]. Future NSCLC survival models should therefore incorporate serial laboratory values, imaging response, toxicity profiles, and treatment modifications over time.

Future research should prioritize multi-center prospective validation, treatment-effect heterogeneity, and broader genomic integration. Dynamic-DeepHit and other deep competing-risk frameworks may be useful when disease-specific death, non-cancer death, recurrence, and treatment discontinuation must be modeled jointly [33]. Multi-omics survival studies such as DeepProg and cross-aligned multimodal representation learning suggest that future models could combine genomics, transcriptomics, pathology images, and electronic medical record variables in a unified prognostic architecture [16, 17, 22, 27]. Additional research should also evaluate whether survival AI can support treatment selection, as suggested by deep learning-assisted multi-omics survival and drug-response prediction frameworks [35-37].

Conclusion

This empirical study developed and compared Cox regression, random survival forest, and DeepSurv for individualized 5-year disease-specific survival prediction in patients with non-small cell lung cancer. DeepSurv achieved the strongest discrimination, lowest Brier score, and best calibration among the evaluated models. The findings show that AI-based survival modeling can add prognostic value when clinical, genomic, and treatment-related variables are jointly analyzed.

The study also demonstrates that predictive performance alone is insufficient for clinical translation. SHAP-based explanations identified tumor mutational burden, performance status, stage, treatment modality, and mutation profile as major contributors to predicted survival risk. These explanations helped connect patient-level risk curves to clinically meaningful prognostic factors.

Overall, the analysis supports the integration of survival analysis and artificial intelligence for personalized prognosis in precision oncology. External validation, prospective testing, and workflow-based implementation studies are needed before such models can guide routine clinical decision-making. Future research should focus on robust validation, longitudinal updating, and real-world deployment in multidisciplinary oncology care.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Wang P, Li Y, Reddy CK. Machine learning for survival analysis: A survey. ACM Comput Surv. 2019;51(6):110. doi:10.1145/3214306
2. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Stat Med. 2017;36(27):4391-400. doi:10.1002/sim.7501
3. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol. 2018;18(1):24. doi:10.1186/s12874-018-0482-1
4. Gensheimer MF, Narasimhan B. A scalable discrete-time survival model for neural networks. PeerJ. 2019;7:e6257. doi:10.7717/peerj.6257
5. Ching T, Zhu X, Garmire LX. Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. PLoS Comput Biol. 2018;14(4):e1006076. doi:10.1371/journal.pcbi.1006076
6. Kruger EA, Schulz M, Vogel P. Key lessons gained on the journey to becoming a clinical professor. Ann Pharm Educ Saf Public Health Advocacy. 2024;4:181-4. doi:10.51847/RJU3tSFfse
7. Chhim S, Chea V. Co-creating CURA: A clinical ethics support tool for palliative care. Asian J Ethics Health Med. 2025;5:175-86. doi:10.51847/59jc1Bjb91
8. Kovalev MS, Utkin LV, Kasimov EM. SurvLIME: A method for explaining machine learning survival models. Knowl Based Syst. 2020;203:106164. doi:10.1016/j.knosys.2020.106164
9. Krzyziński M, Spytek M, Baniecki H, Biecek P. SurvSHAP(t): Time-dependent explanations of machine learning survival models. Knowl Based Syst. 2023;262:110234. doi:10.1016/j.knosys.2022.110234
10. Karthikeyan V, Muthupriya P, Gopikrishna M, Sivakumar K. Effects of electromagnetic radiation and radio frequency on freshwater calanoid and cyclopoid copepods. World J Environ Biosci. 2024;13(2):1-5. doi:10.51847/YYlqFBgHxk
11. Gogoi P, Kamle M, Kumar P. Endophytic bacteria associated with rice: Role in biotic and abiotic stress protection and plant growth promotions. World J Environ Biosci. 2023;12(1):1-9. doi:10.51847/ELxLUdbokK
12. Austin PC, Harrell FE Jr, van Klaveren D. Graphical calibration curves and the integrated calibration index (ICI) for survival models. Stat Med. 2020;39(21):2714-42. doi:10.1002/sim.8570
13. Wongvibulsin S, Wu KC, Zeger SL. Clinical risk prediction with random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis. BMC Med Res Methodol. 2019;20(1):1. doi:10.1186/s12874-019-0863-0
14. Jaeger BC, Long DL, Long DM, Sims M, Szychowski JM, Min YI, et al. Oblique random survival forests. Ann Appl Stat. 2019;13(3):1847-83. doi:10.1214/19-AOAS1261
15. Kvamme H, Borgan Ø. Continuous and discrete-time survival prediction with neural networks. Lifetime Data Anal. 2021;27(4):710-36. doi:10.1007/s10985-021-09532-6
16. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. Clin Cancer Res. 2018;24(6):1248-59. doi:10.1158/1078-0432.CCR-17-0853
17. Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. Bioinformatics. 2019;35(14):i128-i136. doi:10.1093/bioinformatics/btz342
18. Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. Deep learning-based survival prediction of oral cancer patients. Sci Rep. 2019;9(1):6994. doi:10.1038/s41598-019-43372-7
19. Kim S, Kim K, Choe J, Lee I, Kang J. Improved survival analysis by learning shared genomic information from pan-cancer data. Bioinformatics. 2020;36(Suppl 1):i389-i397. doi:10.1093/bioinformatics/btaa462
20. Huang Z, Zhan X, Xiang S, Johnson TS, Helm B, Yu CY, et al. SALMON: Survival analysis learning with multi-omics neural networks on breast cancer. Front Genet. 2019;10:166. doi:10.3389/fgene.2019.00166
21. Zhang L, Lv C, Jin Y, Cheng G, Fu Y, Yuan D, et al. Deep learning-based multi-omics data integration reveals two prognostic subtypes in high-risk neuroblastoma. Front Genet. 2018;9:477. doi:10.3389/fgene.2018.00477
22. Poirion OB, Jing Z, Chaudhary K, Huang S, Garmire LX. DeepProg: An ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. Genome Med. 2021;13(1):112. doi:10.1186/s13073-021-00930-x
23. Wang D, Jing Z, He K, Garmire LX. Cox-nnet v2.0: Improved neural-network-based survival prediction extended to large-scale EMR data. Bioinformatics. 2021;37(17):2772-4. doi:10.1093/bioinformatics/btab046
24. Park SY, Park JE, Kim H, Park SH. Review of statistical methods for evaluating the performance of survival or other time-to-event prediction models (from conventional to deep learning approaches). Korean J Radiol. 2021;22(10):1697-707. doi:10.3348/kjr.2021.0223
25. Langbein SH, Krzyziński M, Spytek M, Baniecki H, Biecek P, Wright MN. Interpretable machine learning for survival analysis. Biom J. 2025;67(6). doi:10.1002/bimj.70089
26. Vale-Silva LA, Rohr K. Long-term cancer survival prediction using multimodal deep learning. Sci Rep. 2021;11(1):13505. doi:10.1038/s41598-021-92799-4
27. Wu X, Shi Y, Wang M, Li A. CAMR: Cross-aligned multimodal representation learning for cancer survival

prediction. Bioinformatics. 2023;39(1):btad025. doi:10.1093/bioinformatics/btad025

28. Papageorgiou G, Mauff K, Tomer A, Rizopoulos D. An overview of joint modeling of time-to-event and longitudinal outcomes. Annu Rev Stat Appl. 2019;6:223-40. doi:10.1146/annurev-statistics-030718-105048
29. Tanner KT, Sharples LD, Daniel RM, Keogh RH. Dynamic survival prediction combining landmarking with a machine learning ensemble: Methodology and empirical comparison. J R Stat Soc Ser A Stat Soc. 2021;184(1):3-30. doi:10.1111/rssa.12611
30. Pickett KL, Suresh K, Campbell KR, Davis S, Juarez-Colunga E. Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker. BMC Med Res Methodol. 2021;21(1):216. doi:10.1186/s12874-021-01375-x
31. Wiegrebe S, Kopper P, Sonabend R, Bischl B, Bender A. Deep learning for survival analysis: A review. Artif Intell Rev. 2024;57(3):65. doi:10.1007/s10462-023-10681-3
32. Goldberg Y, Kosorok MR. Support vector regression for right censored data. Electron J Stat. 2017;11(1):532-69. doi:10.1214/17-EJS1231
33. Lee C, Yoon J, Van Der Schaar M. Dynamic-DeepHit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. IEEE Trans Biomed Eng. 2020;67(1):122-33. doi:10.1109/TBME.2019.2909027
34. Wang Z, Lee JW, Chakraborty T, Ning Y, Liu M, Xie F, et al. Survival modeling using deep learning, machine learning, and statistical methods: A comparative analysis for predicting mortality after hospital admission. Health Data Sci. 2026;6:0449. doi:10.34133/hds.0449
35. Malik V, Kalakoti Y, Sundar D. Deep learning assisted multi-omics integration for survival and drug-response prediction in breast cancer. BMC Genomics. 2021;22(1):214. doi:10.1186/s12864-021-07524-2
36. Mendoza F, Ruiz S, Chávez P. Evolution of pharmaceutical expenditure and cost avoidance in cancer clinical trials: A 10-year retrospective study of adult solid tumor patients (highlights longitudinal "evolution" in drug patterns alongside DCA). Asian J Curr Res Clin Cancer. 2024;4(2):83-96. doi:10.51847/KyDkw3tYd3
37. Morgan PL, Hughes AT, Nolan RJ. Socioeconomic inequalities in substance use: Findings from a household cross-sectional survey in Murang'a County, Kenya. Int J Soc Psychol Asp Healthc. 2024;4:77-86. doi:10.51847/p7WMaooqJU