



OPEN Exploring the influence of age on the causes of death in advanced nasopharyngeal carcinoma patients undergoing chemoradiotherapy using machine learning methods

Mengni Zhang, Shipeng Zhang, Xudong Ao, Lisha Liu & Shunlin Peng✉

The present study analyzed the impact of age on the causes of death (CODs) in patients with nasopharyngeal carcinoma (NPC) undergoing chemoradiotherapy (CRT) using machine learning approaches. A total of 2841 patients (1037 classified as older, ≥ 60 years and 1804 as younger, < 60 years) were enrolled. Variations in the CODs between the two age groups were analyzed before and after applying inverse probability of treatment weighting (IPTW). Additionally, seven different machine learning models were employed as predictive tools to identify key variables and assess the therapeutic outcomes in NPC patients receiving CRT. The younger group exhibited a significantly longer overall survival (OS) than the older group, both before the IPTW adjustment (140 vs. 50 months, $P < 0.001$) and after the adjustment (137 vs. 53 months, $P < 0.001$). After IPTW, the older group was associated with worse 5-, 10-, and 15-year cumulative incidences in terms of NPC-related deaths (30, 34, and 38% vs. 21, 27, and 30%; $P < 0.001$), cardiovascular disease (CVD; 4.1, 7.2, and 8.8% vs. 0.5, 1.8, and 3.0%; $P < 0.001$), and other causes (8.3, 17, and 24% vs. 4.1, 8.7, and 12%; $P < 0.001$). However, cumulative incidences of secondary malignant neoplasms were comparable between the two groups ($P = 0.100$). The random forest (RF) model demonstrated the highest concordance index of 0.701 among all models. Time-dependent variable importance plots indicated that age was the most influential factor affecting 3-, 5-, and 10-year survival, followed by metastasis and tumor stage. Younger patients had significantly longer OS than their older counterparts. Older patients had a higher likelihood of dying from non-NPC-related causes, particularly CVDs. The RF model showed the best predictive accuracy, identifying age as the most critical factor influencing OS in NPC patients undergoing CRT.

Keywords Machine learning models, Age, Chemoradiotherapy, Cause of death, Nasopharyngeal carcinoma

Nasopharyngeal carcinoma (NPC), a malignancy arising from the epithelial lining of the nasopharynx, accounts for a significant portion of head and neck cancers, particularly in Southeast Asia and Southern China^{1,2}. Despite its relatively rare occurrence in other parts of the world, NPC is characterized by its distinct epidemiology, aggressive nature, and sensitivity to chemoradiotherapy (CRT)^{3–5}. However, due to obscure anatomical sites and insidious symptoms, over 70% of patients with NPC are diagnosed at an advanced stage, leading to a poor prognosis^{6,7}. Notably, survival rates can vary even among NPC patients with the same TNM stage and receiving similar or identical treatment regimens. Approximately 30–40% of patients eventually develop distant metastasis following curative treatment^{8,9}.

Mounting evidence has shown that age is an important prognostic factor in patients with NPC, influencing both treatment response and survival outcomes^{10–12}. Older patients often face additional challenges, including a higher incidence of comorbidities, reduced immune response, and diminished capacity for recovery, which can affect the overall treatment strategy and its effectiveness¹³. Younger patients typically demonstrate better overall survival (OS) rates than their older counterparts, potentially due to better treatment tolerance and fewer age-related complications¹⁴. Therefore, understanding the relationship between age and the causes of death (CODs) in NPC patients is crucial for optimizing personalized treatment strategies and improving patient outcomes.

Department of Otolaryngology, Hospital of Chengdu University of Traditional Chinese Medicine, No.39, Shierqiao Road, Jinniu District, Chengdu, Sichuan, China. ✉email: shunlinpeng123@163.com

However, large-scale studies comprehensively investigating the impact of age on the prognosis and CODs in NPC patients receiving CRT are scarce.

With the increasing use of machine learning techniques, predictive models such as least absolute shrinkage and selection operator (LASSO), random forest (RF), Cox regression, support vector machine (SVM), eXtreme Gradient Boosting (XGBoost), decision tree (DT), and gradient-boosting machine (GBM) are becoming instrumental in identifying complex patterns compared with conventional statistical methods^{15,16}. These models can integrate a wide range of patient-specific data to better understand the effects of age on mortality in NPC patients, allowing for the development of more personalized treatment approaches and improving survival rates^{17,18}. Consequently, the current study sought to leverage data from the Surveillance, Epidemiology, and End Results (SEER) database to explore the influence of age on CODs in patients with advanced NPC receiving CRT and develop predictive models using machine learning to evaluate patient prognosis.

Materials and methods

Patient selection

Data of 2841 patients with advanced NPC undergoing CRT between 2000 and 2020 were extracted from the SEER-17 database. Patients were classified into two groups based on age: older group (< 60 years, $n = 1037$) and younger group (≥ 60 years, $n = 1804$). Inclusion criteria: (a) patients diagnosed with squamous cell carcinoma; (b) those who received CRT; (c) those with stage III/IV NPC; (d) those with clear CODs. Exclusion criteria: (a) patients with stage I/II or unknown NPC; (b) those aged < 18 years; (c) those with incomplete data. Since SEER data are publicly available for research, approval from the local ethics committee was waived.

Definition of CODs

CODs were grouped into four categories: (1) NPC-related deaths (NRDs), including those resulting from NPC-related conditions; (2) secondary malignant neoplasms (SMNs), covering fatalities caused by secondary cancers; (3) cardiovascular diseases (CVDs), encompassing deaths from heart disease, atherosclerosis, aortic aneurysm and dissection, cerebrovascular conditions, or other arterial issues; and (4) other causes.

Machine learning algorithms

In our study, the following 11 variables were included in the LASSO regression: sex, race, marital status, year of diagnosis, household income, grade, histology, T, N, M, and stage. After LASSO regression, several key variables influencing OS were identified and used to develop machine learning models. Patients were divided into a training cohort and a validation cohort in a 7:3 ratio. Six models, including Cox regression, SVM (with Radial Basis Function kernel), GBM, XGBoost, RF, and DT, were developed to predict OS in NPC patients. The predictive performance of these models was compared using the concordance index (C-index), Brier score, and receiver operating characteristic (ROC) curves. Decision curve analysis (DCA) and calibration curves were further applied to identify the optimal model.

SurvSHAP, utilizing a global explanation approach, was applied to calculate the mean SHapley Additive exPlanations (SHAP) value for each feature across the entire dataset. Time-dependent importance plots were generated to highlight the most influential features for predicting 3-, 5-, and 10-year OS. Additionally, partial dependence plots (PDPs) were employed to illustrate the impact of feature value changes on predicted outcomes.

Statistical analysis

The chi-square test was employed to compare categorical variables between the younger and older groups. Inverse probability of treatment weighting (IPTW) was applied to adjust for differences in covariate distributions to balance the two groups. The variables included: sex, race, marital status, year of diagnosis, household income, grade, histology, T, N, M, and stage. Competing risk analysis was performed using Gray's test to accurately calculate the cumulative incidence. Kaplan-Meier curves were plotted to analyze OS outcomes, with comparisons made through the log-rank test. All statistics were conducted using R software version 3.3.2, with a significance threshold set at $P < 0.05$.

Results

Patient characteristics

The younger and older groups showed significant differences in terms of race ($P < 0.001$), marital status ($P = 0.002$), year of diagnosis ($P = 0.001$), tumor grade ($P < 0.001$), histology ($P < 0.001$), T stage ($P = 0.037$), and N stage ($P < 0.001$) before IPTW (Table 1). However, no statistically significant differences in baseline characteristics were found between the two groups after IPTW (Supplementary Table 1, Supplementary Fig. 1).

OS and CODs

The younger group exhibited a significantly longer median OS than the older group before IPTW adjustment (140 vs. 50 months, $P < 0.001$, Fig. 1A) and after IPTW adjustment (137 vs. 53 months, $P < 0.001$, Fig. 1B).

Before IPTW, the older group had higher 5-, 10-, and 15-year cumulative incidences of NRD (31, 35, and 37% vs. 21, 27, and 30%; $P < 0.001$, Fig. 2A), SMNs (11, 14, and 15% vs. 8.4, 10, and 11%; $P = 0.006$, Fig. 2B), CVDs (4.2, 7, and 8.7% vs. 0.5, 1.5, and 2.7%; $P < 0.001$, Fig. 2C), and other causes (8.7, 19, and 25% vs. 3.9, 8.1, and 11%; $P < 0.001$, Fig. 2D) than the younger group.

After IPTW, the older group had worse 5-, 10-, and 15-year cumulative incidences of NRD (30, 34, and 38% vs. 21, 27, and 30%; $P < 0.001$, Fig. 3A), CVDs (4.1, 7.2, and 8.8% vs. 0.5, 1.8, and 3.0%; $P < 0.001$, Fig. 3C), and other causes (8.3, 17, and 24% vs. 4.1, 8.7, and 12%; $P < 0.001$, Fig. 3D) than the younger group. However, cumulative incidences of SMNs were comparable between the two groups ($P = 0.100$, Fig. 3B).

Variable	Age < 60 years	Age ≥ 60 years	P	SMD
	No. (%)	No. (%)		
Patients	1804	1037		
Sex			0.286	0.043
Female	508 (28.2)	272 (26.2)		
Male	1296 (71.8)	765 (73.8)		
Race			< 0.001	0.257
White	789 (43.7)	584 (56.3)		
Black	242 (13.4)	120 (11.6)		
Other/Unknown	773 (42.8)	333 (32.1)		
Marital status			0.002	0.121
Married	791 (43.8)	393 (37.9)		
Unmarried/Unknown	1013 (56.2)	644 (62.1)		
Year of diagnosis			0.001	0.138
2000–2010	625 (34.6)	293 (28.3)		
2011–2020	1179 (65.4)	744 (71.7)		
Median household income			0.434	0.032
< \$75,000	737 (40.9)	440 (42.4)		
≥ \$75,000	1067 (59.1)	597 (57.6)		
Grade			< 0.001	0.185
I/II	158 (8.8)	145 (14.0)		
III/IV	816 (45.2)	403 (38.9)		
Unknown	830 (46.0)	489 (47.2)		
Histology			< 0.001	0.316
KSCC	805 (44.6)	608 (58.6)		
NKC	821 (45.5)	373 (36.0)		
BSCC	25 (1.4)	17 (1.6)		
Other	153 (8.5)	39 (3.8)		
T			0.037	0.114
1	391 (21.7)	184 (17.7)		
2	294 (16.3)	158 (15.2)		
3	537 (29.8)	344 (33.2)		
4	582 (32.3)	351 (33.8)		
N			< 0.001	0.300
0	207 (11.5)	216 (20.8)		
1	355 (19.7)	239 (23.0)		
2	894 (49.6)	435 (41.9)		
3	348 (19.3)	147 (14.2)		
M			0.910	0.007
0	1632 (90.5)	936 (90.3)		
1	172 (9.5)	101 (9.7)		
Stage			0.908	0.006
III	847 (47.0)	490 (47.3)		
IV	957 (53.0)	547 (52.7)		

Table 1. Clinical characteristics of NPC patients before IPTW. NPC, nasopharyngeal carcinoma; IPTW, inverse probability of treatment weighting; KSCC, keratinizing squamous cell carcinoma; NKCC, non-keratinizing carcinoma; BSCC, basaloid squamous cell carcinoma.

Patients based on age groups and spline curve analysis

All patients were categorized into seven age groups: 18–30, 30–40, 40–50, 50–60, 60–70, 70–80, and 80+ years old (Supplementary Table 2). Detailed comparisons of OS among these groups showed significant differences, with the 18–30 age group demonstrating the longest OS and the 80+ age group exhibiting the worst OS (Fig. 4). Additionally, we performed spline regression analysis for the entire cohort and for each COD category, including NRD, SMN, CVD, and other causes. Our results confirm that as age increases, the risk of mortality due to all causes (Supplementary Fig. 2), NRDs (Supplementary Fig. 3A), SMNs (Supplementary Fig. 3B), and other causes increases (Supplementary Fig. 3D) in a continuous fashion. However, for CVD-related mortality

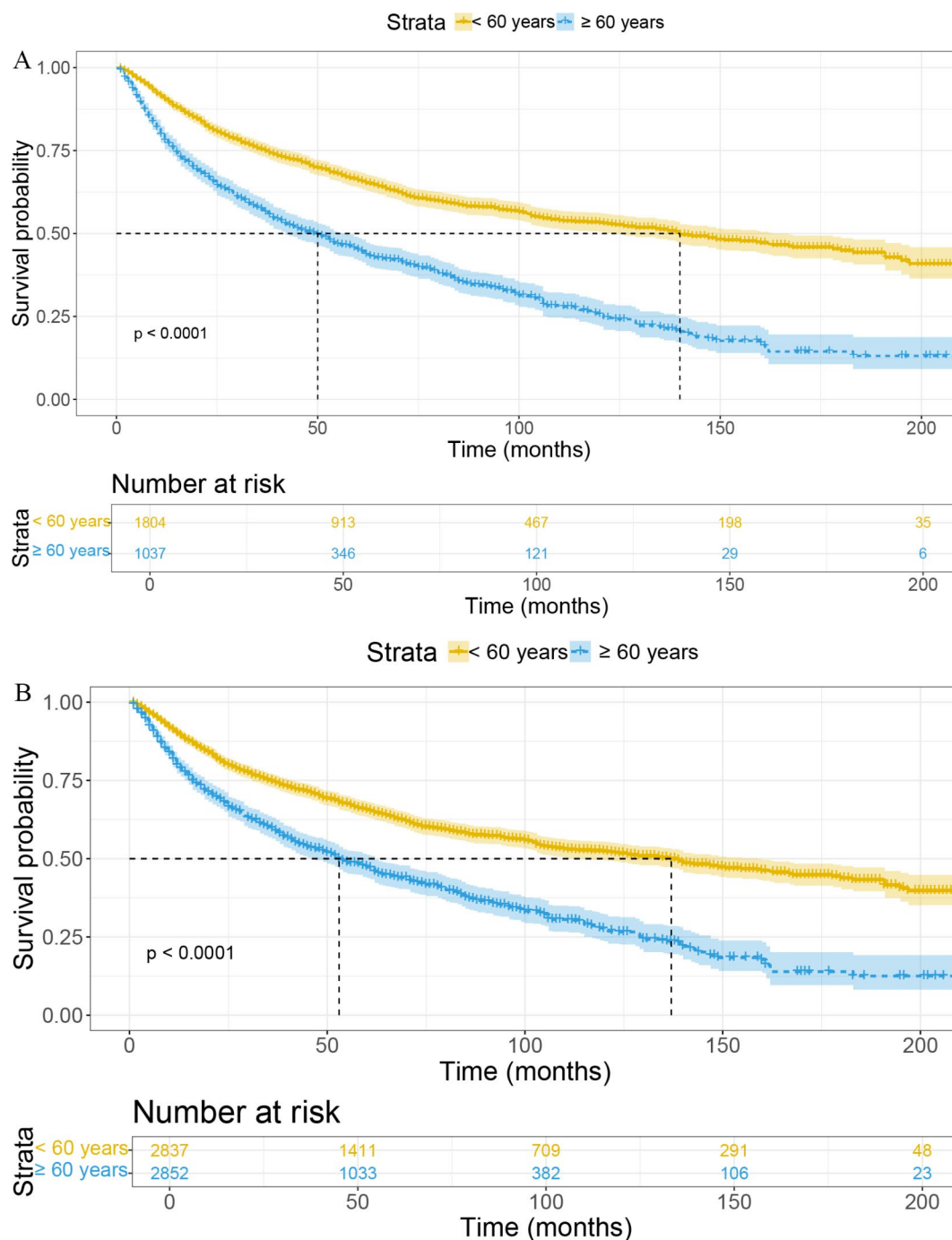


Fig. 1. Overall survival (OS) based on age groups before and after the inverse probability of treatment weighting.

(Supplementary Fig. 3C), we observed a distinct pattern: the risk remained similar up to the age of 55, after which the risk increased significantly.

Supplementary Table 3 presents the distribution of SMNs in the cohort, categorized by age groups. Multivariable Cox regression analysis shows that sex, age, marital status, year of diagnosis, household income, grade, histology, and M stage are associated with SMNs (Supplementary Table 4).

Evaluation of the machine learning models

Age, metastasis, stage, marital status, histology, year of diagnosis, and household income were identified as factors affecting OS through LASSO analysis (Supplementary Fig. 4). The RF model demonstrated the highest C-index among all models, reaching 0.701 (Fig. 5). The six models were evaluated in the validation set by comparing their Brier scores and areas under the ROC curves (AUCs, Table 2; Fig. 6). Furthermore, the DCA (Supplementary

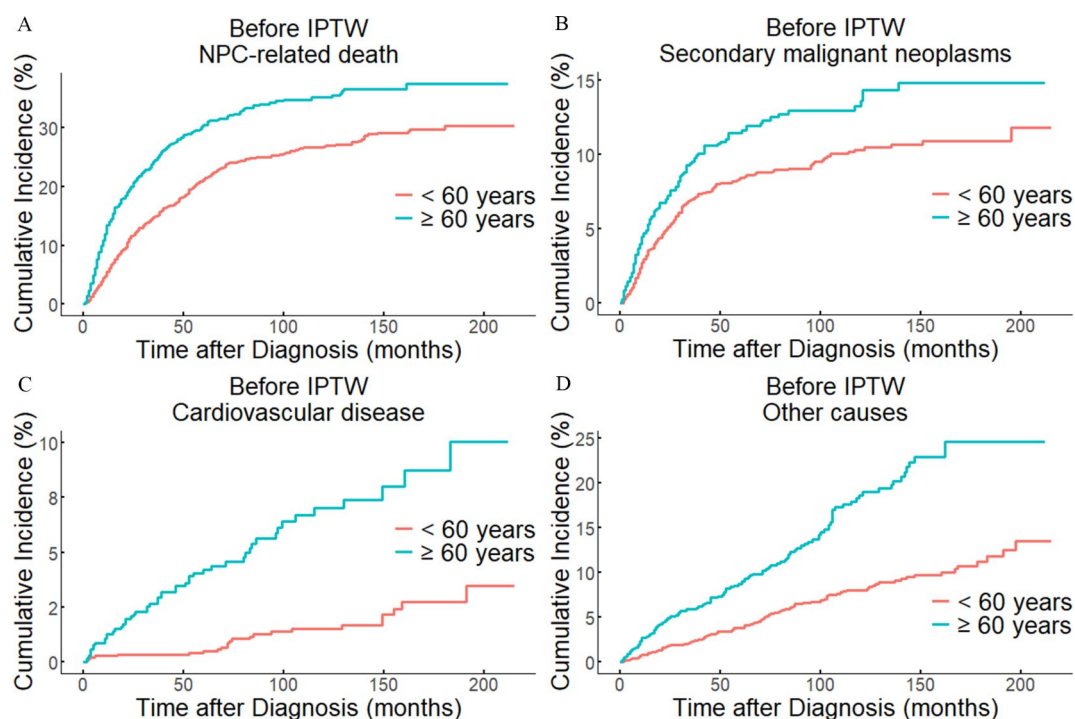


Fig. 2. Comparison of cumulative incidences of NRDs (A), SMNs (B), CVDs (C), and other causes (D) between the older and younger groups at 5, 10, and 15 years before IPTW. IPTW, inverse probability of treatment weighting; NPC, nasopharyngeal carcinoma; NRD, NPC-related deaths; SMN, secondary malignant neoplasm; CVD, cardiovascular disease.

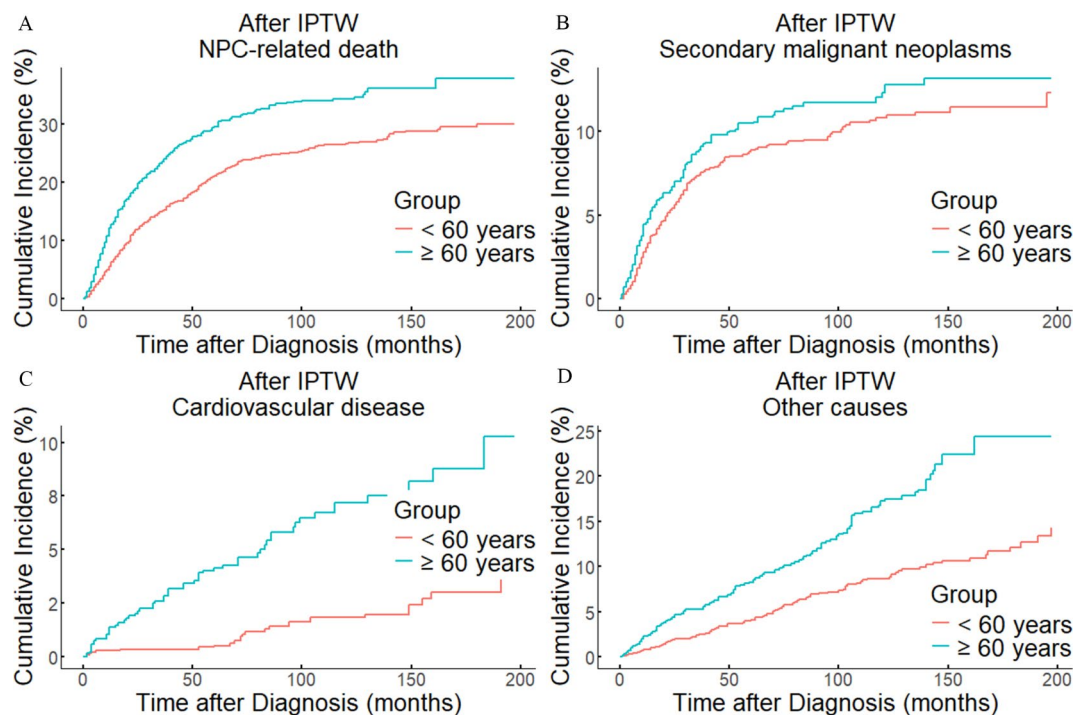


Fig. 3. Comparison of cumulative incidences of NRDs (A), SMNs (B), CVDs (C), and other causes (D) between the older and younger groups at 5, 10, and 15 years after IPTW. IPTW, inverse probability of treatment weighting; NPC, nasopharyngeal carcinoma; NRD, NPC-related deaths; SMN, secondary malignant neoplasm; CVD, cardiovascular disease.

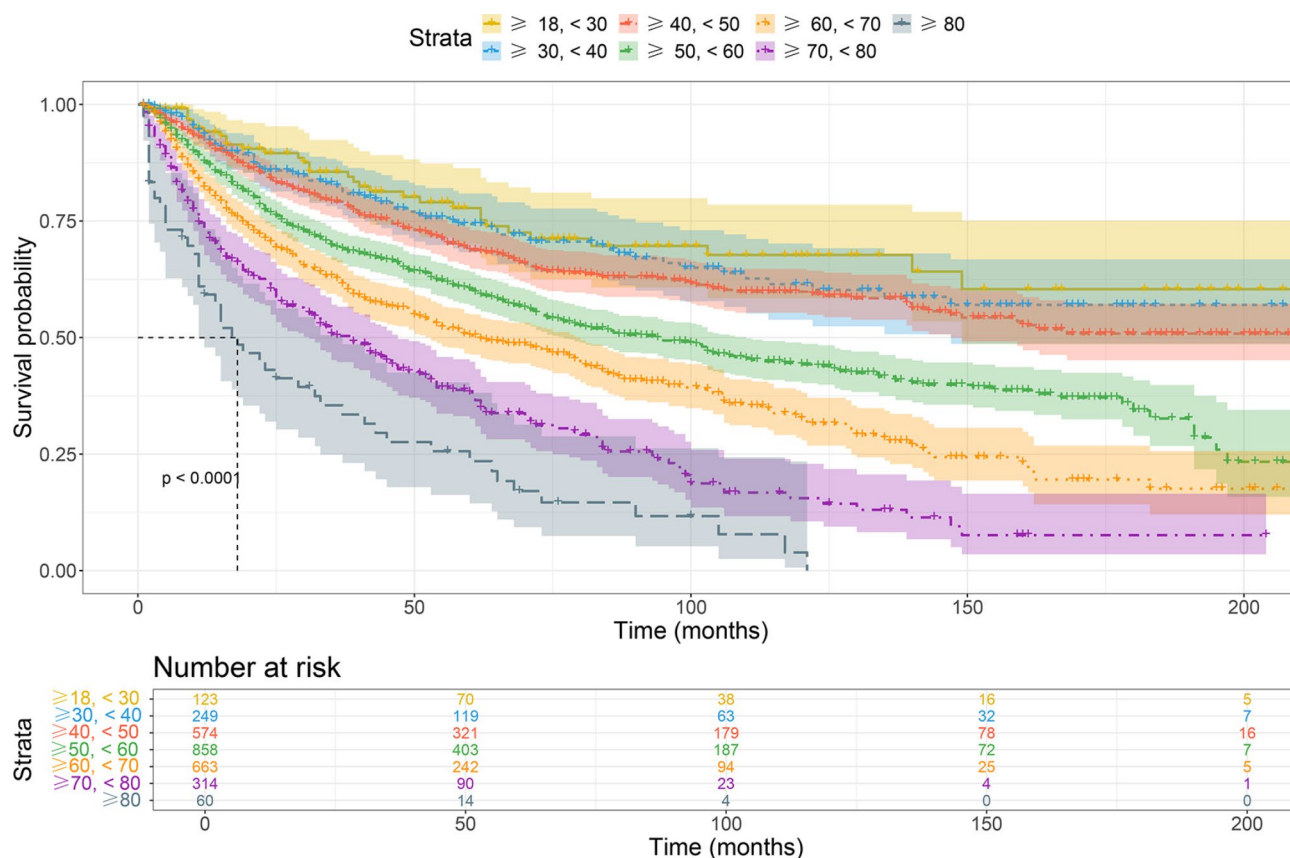


Fig. 4. Survival analysis based on different age subgroups.

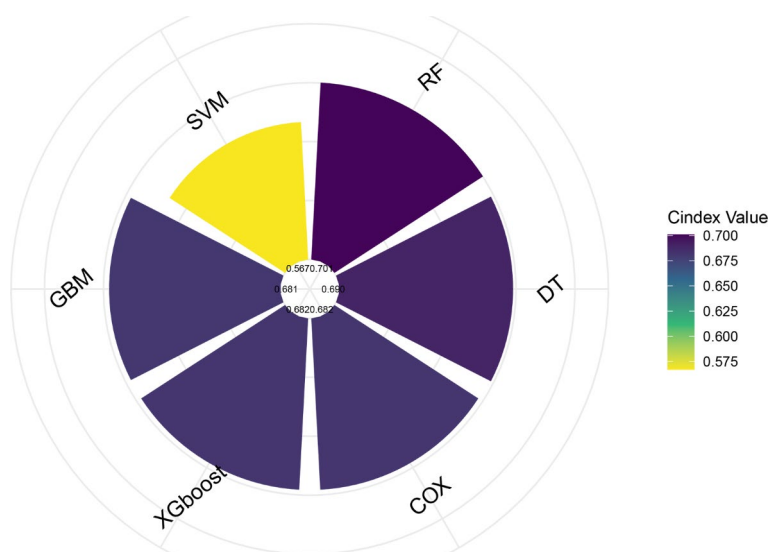


Fig. 5. Concordance index ranking graph.

Fig. 5A-C) and calibration curves (Supplementary Fig. 5D) for the RF model showed strong predictive accuracy in estimating 3-, 5-, and 10-year survival rates.

Model interpretation

The time-dependent variable importance bar plots revealed that age had the greatest influence on 3-, 5-, and 10-year survival, followed by metastasis and tumor stage (Fig. 7). PDPs further indicated that advanced stages, increased age, and the presence of M were associated with worse survival (Supplementary Fig. 6). Likewise,

Model	C	Brier						AUC		
		3-year	P	5-year	P-value	10-year	P-value	3-year	5-year	10-year
COX	0.682	0.201	0.006	0.222	0.004	0.217	0.002	0.671	0.680	0.705
DT	0.69	0.318	<0.001	0.349	<0.001	0.303	<0.001	0.659	0.659	0.700
RF	0.701	0.201	0.007	0.224	0.008	0.215	<0.001	0.732	0.751	0.748
GBM	0.681	0.208	0.100	0.230	0.059	0.239	0.607	0.667	0.680	0.711
SVM	0.567	0.222	0.063	0.251	0.223	0.262	0.002	0.544	0.527	0.527
XGboost	0.682	0.210	0.374	0.233	0.325	0.220	0.015	0.671	0.680	0.706

Table 2. Performance comparison of machine learning models. AUC, area under the ROC curve.

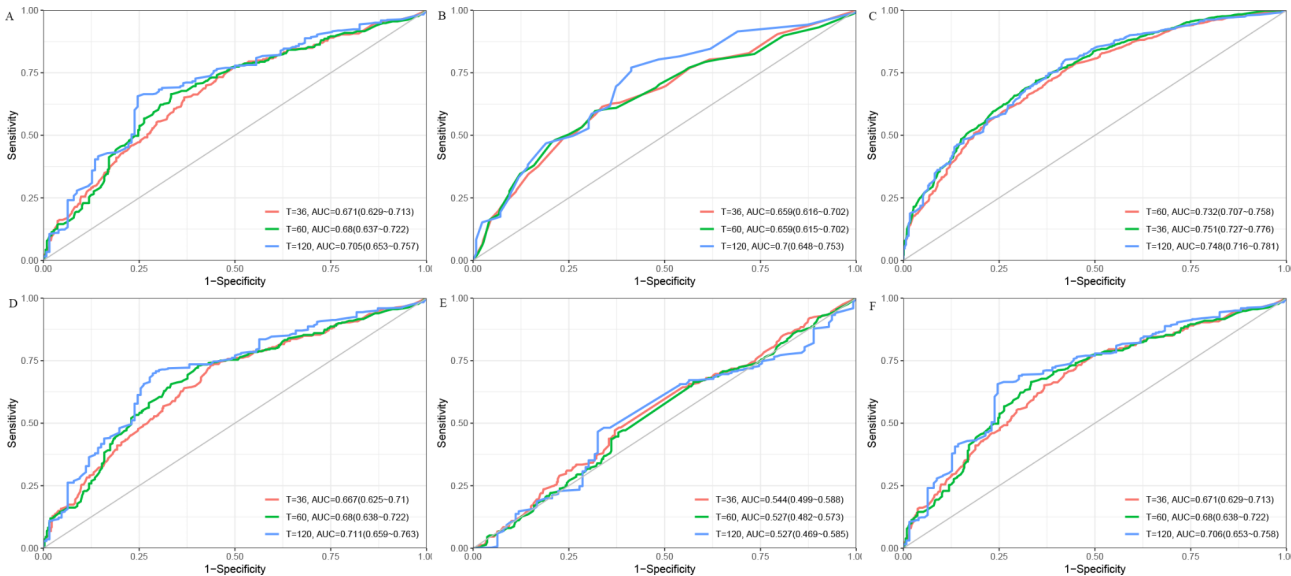


Fig. 6. Receiver operating characteristic curves for predicting 3-, 5, and 10-year survival based on seven different models: Cox regression (A), DT (B), DT (C), GBM (D), SVM (E), and XGBoost (F). ROC, receiver operating characteristic; DT, decision tree; GBM, gradient-boosting machine; SVM, support vector machine; XGBoost, eXtreme Gradient Boosting.

SHAP value-based box plots revealed a strong association between older age and reduced survival rates (Supplementary Fig. 7).

Discussion

The present study comprehensively explored the influence of age on the prognosis and CODs in NPC patients undergoing CRT. Although previous studies have highlighted age as a critical factor in NPC prognosis, only a handful of reports have examined its impact using detailed and large-scale analysis^{19,20}. By leveraging data from the SEER database and applying advanced machine learning models, the current study provides new insights into the role of age in determining survival outcomes and CODs in NPC patients²¹. This study significantly adds to the existing knowledge by offering a nuanced understanding of age-related prognostic factors and how they can be integrated into more personalized treatment approaches for NPC patients.

Our findings showed that younger patients had longer OS than their older counterparts both before and after IPTW adjustment. This difference in survival may be attributed to several age-related factors, such as a reduced capacity to tolerate aggressive treatment, the presence of multiple comorbidities, and a diminished immune response in older patients. The impact of age was further validated by time-dependent variable importance plots, which identified age as the most significant predictor of survival, followed by M and tumor stage. This underscores the critical need to consider age as a key factor when evaluating the prognosis and management strategies for patients with NPC^{22,23}.

Moreover, older patients exhibited higher cumulative incidences of NRDs and deaths from other causes, highlighting the complexity of managing NPC in an aging population. Further analysis of CODs revealed that older NPC patients had a higher incidence of CVDs and SMNs as a cause of mortality than their younger counterparts. This finding is consistent with the notion that older patients are more vulnerable to non-cancer-related causes of death, which may be exacerbated by the strain of cancer treatments. These findings underscore the need for a multidisciplinary approach that addresses not only the cancer itself but also the overall health and comorbid conditions of older patients.

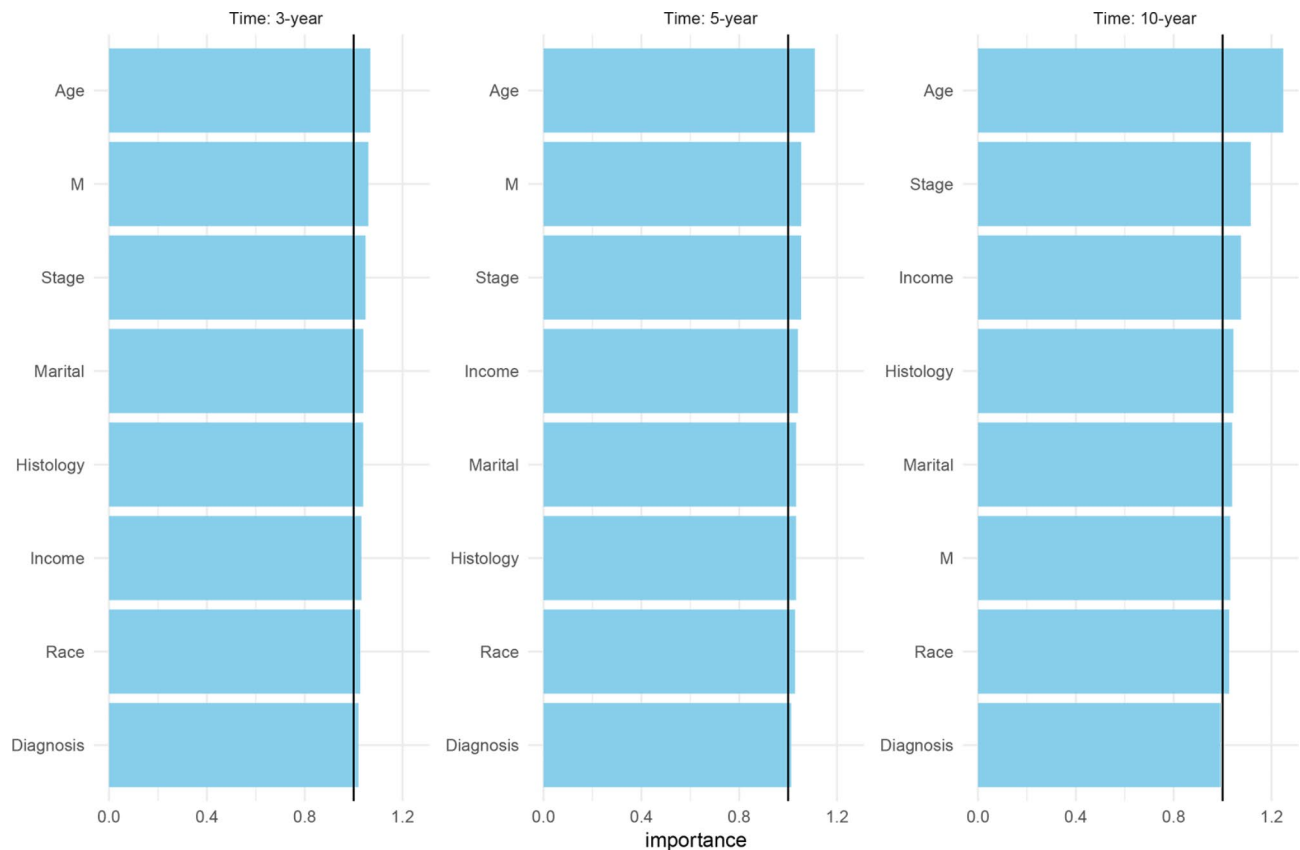


Fig. 7. Time-dependent variable importance bar plots display the ranking of significant features affecting the 3-, 5-, and 10-year survival.

The use of machine learning models in this study allowed for a more precise identification of factors influencing survival outcomes in NPC patients²⁴. LASSO regression identified age, metastasis, stage, marital status, histology, year of diagnosis, and household income as significant predictors of OS. The RF model showed the highest predictive accuracy among all models, with a C-index of 0.701. The superior performance of the RF model is likely due to its ability to handle large datasets with complex, nonlinear relationships between variables, making it an ideal tool for predicting survival outcomes in NPC patients^{25–28}. DCA and calibration curves further confirmed the strong predictive performance of the RF model, particularly in estimating 3-, 5-, and 10-year survival.

PDPs and SHAP value-based analyses provided additional insights into the impact of age and other factors on survival. Advanced tumor stages, older age, and the presence of metastasis were all associated with poorer survival outcomes. These findings highlight the importance of integrating machine learning models into clinical decision-making to improve risk stratification and tailor treatment strategies for NPC patients^{29,30}.

Nonetheless, this study has several limitations. First, the retrospective nature of the study and the use of the SEER database may introduce potential biases related to data accuracy and completeness. Second, while machine learning models provide valuable insights, they require further validation in broader and independent populations to ensure their generalizability. The lack of an external validation cohort in this study limits the robustness of the predictive models and their applicability to diverse clinical settings. Finally, other factors not included in this analysis, such as treatment model, lifestyle, and environmental influences, may also influence NPC prognosis and thus warrant further investigation.

Conclusions

In summary, this study has demonstrated that age plays a critical role in the prognosis of patients with NPC undergoing CRT, with younger patients demonstrating significantly better survival outcomes than older patients. Machine learning models, particularly the RF model, proved to be powerful tools for predicting survival and identifying key prognostic factors. These findings emphasize the importance of age-specific management strategies and personalized treatment approaches in NPC patients, especially for older individuals who may benefit from enhanced cardiovascular monitoring and closer follow-up care. Future research should further validate these models and explore additional factors that may influence the survival of patients with NPC.

Data availability

All data generated or analyzed during this study are included in this article. Further enquiries can be directed to the corresponding author (shunlinpeng123@163.com).

Received: 23 October 2024; Accepted: 8 January 2025

Published online: 13 January 2025

References

- Chen, Y. P. et al. Nasopharyngeal carcinoma. *Lancet* **394**(10192), 64–80 (2019).
- Lee, A. W. M. et al. Management of locally recurrent nasopharyngeal carcinoma. *Cancer Treat. Rev.* **79**, 101890 (2019).
- Tang, L. L. et al. Effect of radiotherapy alone vs radiotherapy with concurrent chemoradiotherapy on survival without disease relapse in patients with low-risk nasopharyngeal carcinoma: A randomized clinical trial. *JAMA* **328**(8), 728–736 (2022).
- Liu, L. T. et al. Concurrent chemoradiotherapy followed by adjuvant cisplatin-gemcitabine versus cisplatin-fluorouracil chemotherapy for N2-3 nasopharyngeal carcinoma: A multicentre, open-label, randomised, controlled, phase 3 trial. *Lancet Oncol.* **24**(7), 798–810 (2023).
- Hennessy, M. A. & Morris, P. G. Induction treatment prior to chemoradiotherapy in nasopharyngeal carcinoma: Triplet or doublet chemotherapy? *Anti-cancer Drugs*. **31**(2), 97–100 (2020).
- Pan, J. J. et al. Prognostic nomogram for refining the prognostication of the proposed 8th edition of the AJCC/UICC staging system for nasopharyngeal cancer in the era of intensity-modulated radiotherapy. *Cancer* **122**(21), 3307–3315 (2016).
- Pan, J. J. et al. Proposal for the 8th edition of the AJCC/UICC staging system for nasopharyngeal cancer in the era of intensity-modulated radiotherapy. *Cancer* **122**(4), 546–558 (2016).
- Hui, E. P. et al. Lung metastasis alone in nasopharyngeal carcinoma: A relatively favorable prognostic group. A study by the Hong Kong Nasopharyngeal Carcinoma study group. *Cancer* **101**(2), 300–306 (2004).
- Mao, Y. P. et al. Re-evaluation of 6th edition of AJCC staging system for nasopharyngeal carcinoma and proposed improvement based on magnetic resonance imaging. *Int. J. Radiat. Oncol. Biol. Phys.* **73**(5), 1326–1334 (2009).
- Guo, X. et al. Long-term trends of nasopharyngeal carcinoma mortality in China from 2006 to 2020 by region and sex: An age-period-cohort analysis. *BMC Public Health* **23**(1), 2057 (2023).
- Li, W. Z. et al. Age-dependent changes of gender disparities in nasopharyngeal carcinoma survival. *Biol. Sex. Differ.* **12**(1), 18 (2021).
- Xu, C., Chen, Y. P. & Ma, J. Clinical trials in nasopharyngeal carcinoma—past, present and future. *Chin. Clin. Oncol.* **5**(2), 20 (2016).
- Zhang, W. M., Mo, Q. Y. & Zhu, X. D. Contribution of age at diagnosis to cancer-specific survival of nasopharyngeal carcinoma patients receiving radiotherapy. *Medicine* **102**(33), e34816 (2023).
- Tsai, A. et al. Using CT texture analysis to differentiate between nasopharyngeal carcinoma and age-matched adenoid controls. *Eur. J. Radiol.* **108**, 208–214 (2018).
- Hosseini, M. P., Hosseini, A. & Ahi, K. A review on machine learning for EEG signal processing in bioengineering. *IEEE Rev. Biomed. Eng.* **14**, 204–218 (2021).
- Deo, R. C. Machine learning in medicine. *Circulation* **132**(20), 1920–1930 (2015).
- Lee, Y. W., Choi, J. W. & Shin, E. H. Machine learning model for predicting malaria using clinical information. *Comput. Biol. Med.* **129**, 104151 (2021).
- Zhang, B. et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett.* **403**, 21–27 (2017).
- Lee, B. C. et al. Pregnancy incidence in female nasopharyngeal carcinoma survivors of reproductive age: A population-based study. *Medicine* **95**(20), e3729 (2016).
- Huang, S. J. et al. Impact of age on survival of locoregional nasopharyngeal carcinoma: An analysis of the Surveillance, Epidemiology, and end results program database, 2004–2013. *Clin. Otolaryngol. Off. J. ENT-UK Off. J. Neth. Soc. Oto-Rhino-Laryngol. Cervico-Facial Surg.* **43**(5), 1209–1218 (2018).
- Sunthakar, S. D. et al. Machine learning to predict interstage mortality following single ventricle palliation: A NPC-QIC database analysis. *Pediatr. Cardiol.* **44**(6), 1242–1250 (2023).
- Xiao, G., Cao, Y., Qiu, X., Wang, W. & Wang, Y. Influence of gender and age on the survival of patients with nasopharyngeal carcinoma. *BMC Cancer* **13**, 226 (2013).
- Zhang, X. et al. Identification of markers for predicting prognosis and endocrine metabolism in nasopharyngeal carcinoma by miRNA-mRNA network mining and machine learning. *Front. Endocrinol.* **14**, 1174911 (2023).
- Zhong, L. et al. A deep learning-based radiomic nomogram for prognosis and treatment decision in advanced nasopharyngeal carcinoma: A multicentre study. *EBioMedicine* **70**, 103522 (2021).
- Hu, J. & Szymczak, S. A review on longitudinal data analysis with random forest. *Brief. Bioinform.* **24**(2), 1253711 (2023).
- Ganaie, M. A., Tanveer, M., Suganthan, P. N. & Snasel, V. Oblique and rotation double random forest. *Neural Netw. Off. J. Int. Neural Netw. Soc.* **153**, 496–517 (2022).
- Song, S. et al. Using machine learning methods to investigate the impact of age on the causes of death in patients with early intrahepatic cholangiocarcinoma who underwent surgery. Clinical & translational oncology: Official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico (2024).
- Su, K. et al. Effect of external beam radiation therapy versus transcatheter arterial chemoembolization for non-diffuse hepatocellular carcinoma (≥ 5 cm): A multicenter experience over a ten-year period. *Front. Immunol.* **14**, 1265959 (2023).
- Gu, B. et al. Multi-task deep learning-based radiomic nomogram for prognostic prediction in locoregionally advanced nasopharyngeal carcinoma. *Eur. J. Nucl. Med. Mol. Imaging* **50**(13), 3996–4009 (2023).
- Chen, J. W. et al. Early detection of nasopharyngeal carcinoma through machine-learning-driven prediction model in a population-based healthcare record database. *Cancer Med.* **13**(7), e7144 (2024).

Acknowledgements

The authors would like to thank all the reviewers who participated in the review and MJEditor (www.mjeditor.com) for its linguistic assistance during the preparation of this manuscript.

Author contributions

Mengni Zhang and Shipeng Zhang participated in the writing, data curation, supervision. Xudong Ao and Lisha Liu participated in the validation, and formal analysis. Xudong Ao participated in the methodology and software. Shunlin Peng designed the this study. All authors approved the final version of the manuscript.

Funding

None.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

Approval from the local research ethics committee was deemed unnecessary for this study because the SEER data is publicly accessible for research purposes.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-86178-6>.

Correspondence and requests for materials should be addressed to S.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025