



Full Length Article

Predictive function of tumor burden-incorporated machine-learning algorithms for overall survival and their value in guiding management decisions in patients with locally advanced nasopharyngeal carcinoma

Yang Liu^{1,†}, Shiran Sun^{1,†}, Ye Zhang¹, Xiaodong Huang¹, Kai Wang¹, Yuan Qu¹, Xuesong Chen¹, Runye Wu¹, Jianghu Zhang¹, Jingwei Luo¹, Yexiong Li¹, Jingbo Wang^{1,*}, Junlin Yi^{1,2,*}

¹ Department of Radiation Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

² Department of Radiation Oncology, National Cancer Center/National Clinical Research Center for Cancer/Hebei Cancer Hospital, Chinese Academy of Medical Sciences (CAMS), Langfang 065001, China



ARTICLE INFO

Keywords:

Nasopharyngeal carcinoma
Machine learning
Tumor burden
Prognostic model
Induction chemotherapy

ABSTRACT

Objective: Accurate prognostic predictions and personalized decision-making on induction chemotherapy (IC) for individuals with locally advanced nasopharyngeal carcinoma (LA-NPC) remain challenging. This research examined the predictive function of tumor burden-incorporated machine-learning algorithms for overall survival (OS) and their value in guiding treatment in patients with LA-NPC.

Methods: Individuals with LA-NPC were reviewed retrospectively. Tumor burden signature-based OS prediction models were established using a nomogram and two machine-learning methods, the interpretable eXtreme Gradient Boosting (XGBoost) risk prediction model, and DeepHit time-to-event neural network. The models' prediction performances were compared using the concordance index (C-index) and the area under the curve (AUC). The patients were divided into two cohorts based on the risk predictions of the most successful model. The efficacy of IC combined with concurrent chemoradiotherapy was compared to that of chemoradiotherapy alone.

Results: The 1 221 eligible individuals, assigned to the training ($n = 813$) or validation ($n = 408$) set, showed significant respective differences in the C-indices of the XGBoost, DeepHit, and nomogram models (0.849 and 0.768, 0.811 and 0.767, 0.730 and 0.705). The training and validation sets had larger AUCs in the XGBoost and DeepHit models than the nomogram model in predicting OS (0.881 and 0.760, 0.845 and 0.776, and 0.764 and 0.729, $P < 0.001$). IC presented survival benefits in the XGBoost-derived high-risk but not low-risk group.

Conclusion: This research used machine-learning algorithms to create and verify a comprehensive model integrating tumor burden with clinical variables to predict OS and determine which patients will most likely gain from IC. This model could be valuable for delivering patient counseling and conducting clinical evaluations.

1. Introduction

Nasopharyngeal carcinoma (NPC) is a prevalent endemic disease predominantly found in Southeast Asia.¹ Over 70% of the patients are diagnosed with locally advanced NPC (LA-NPC) on their first visit.² Induction chemotherapy (IC) is critically important to improving tumor control because the leading failure cause in LA-NPC is distant metastasis.^{3, 4} The National Comprehensive Cancer Network (NCCN) guidelines recommended with 2A evidence level that patients diagnosed with LA-NPC undergo IC along with concurrent chemoradiotherapy (IC+CCRT).⁵ However, LA-NPC affects patients with a wide range of T and N cate-

gories, and it is unclear who would benefit from IC. Therefore, it is crucial to have an appropriate method to identify high-risk patients who would gain survival benefits from IC while preventing overtreatment of those at low risk.

Nomograms have been widely studied for including various prognostic determinants and potentially providing improved accuracy in individualized survival estimation and treatment selection.^{6–8} Many nomogram models have been developed based on the traditional Cox regression method that hypothesizes linear relationships between covariates. However, this approach might hinder capturing intricate nonlinear associations between features and outcomes. In contrast,

* Corresponding authors.

E-mail addresses: wangjingbo201001@163.com (J. Wang), yijunlin1969@163.com (J. Yi).

† These authors contributed equally to this work.

machine learning (ML) algorithms have demonstrated remarkable capabilities in analyzing complex and diverse datasets, and their medical application has shown superior performance to conventional risk prediction scoring systems.^{9–11}

Precise tumor mapping and T and N staging can be made using information obtained from multiplanar magnetic resonance imaging (MRI). Previous studies have shown that in addition to the tumor's primary invasion location, size, position, and lymph node (LN) laterality, included in the American Joint Committee on Cancer (AJCC) staging system as tumor load prognostic factors, other variables, such as tumor and nodal volumes, size, and number, were also closely associated with NPC prognosis.^{12–15} It was reported that metastatic LN spread distance (SD) directly reveals the possible microscopic spread pathways to the LNs and accurately quantifies the potential diffused metastatic capacity, both closely associated with NPC prognosis.¹⁶ However, few studies have investigated the role of SD in a comprehensive survival model.

This research aimed to create and compare a nomogram and two ML models (the interpretable eXtreme Gradient Boosting [XGBoost] risk prediction model and DeepHit time-to-event neural network) for their predicted overall survival (OS) in patients with LA-NPC using their MRI-derived tumor load profiles. We used the best model's predictions to divide the patients into low- and high-risk groups and compared their survival outcomes following IC+CCRT or CCRT, aiming to determine which patients were more likely to benefit from IC+CCRT.

2. Materials and methods

2.1. Study design and participants

This study included 1 221 patients consecutively treated in our hospital between January 2010 and December 2017. The inclusion criteria were: (i) histologically confirmed nonmetastatic NPC of WHO types II and III; (ii) stage III-IVA disease; (iii) completed baseline nasopharynx-neck MRI scans. Supplementary Fig. 1 presents a depiction of the patient selection process. The patients were reevaluated and staged following the 8th AJCC staging system. The patients were assigned to the training (those treated during 2010–2015, $n = 813$) or the validation (those treated in 2016–2017, $n = 408$) cohort.

2.2. Imaging protocols

CT (Philips Brilliance) and MRI (GE Discovery) scans were performed during pre-radiation simulation. All of the patients were in supine position with the thermoplastic mask immobilization at the head, neck, and shoulder. The scans were captured in 3-mm slices from the head to 2 cm below the sternoclavicular joint.

2.3. Baseline clinical and tumor burden characteristics

Fifteen sociodemographic and baseline characteristics were evaluated, including age, sex, Karnofsky score (KPS), alcohol consumption, smoking, AJCC 8th T and N stages, clinical stage, plasma Epstein-Barr virus (EBV) DNA copy number, lactate dehydrogenase (LDH), hemoglobin (Hb), albumin (ALB), C-reactive protein (CRP), neutrophilic granulocytes (NC), and lymphocytes (LC). The study also assessed seven tumor burden-related factors, including nasopharynx tumor volume (Vp), regional LN volume (Vln), extranodal extension (ENE), central nodal necrosis (CNN) status, spread distance (SD) from the atlantoaxial spine lateral process, maximum diameter (MD), and positive LN count. The diagnostic criteria for positive LN followed a previously reported consensus.¹⁷ Detailed information on tumor burden variable collection is presented in Supplementary Methods.

2.4. Treatment, follow-up, and study endpoint

Intensity-modulated radiation therapy (IMRT)-based simultaneous integrated boost was administered to all patients. The radiotherapy and

chemotherapy regimens were previously described.¹⁸ The patients received IMRT, CCRT, or CCRT+IC. Concomitant with radiation therapy, the patients underwent 2–3 chemotherapy cycles of 100 mg/m² cisplatin (every three weeks, on D1, 22, 43) and 40 mg/m² cisplatin weekly, with or without 2–3 IC cycles of gemcitabine 1.0 g/m² and D1 and D8 + cisplatin 80 mg/m² on D1.

The regular follow-up visits included at least four at 3-month intervals during the first couple of years, twice a year during the subsequent two years, and once a year after that. Evaluations included complete physical examination, fiberoptic nasopharyngoscopy, MRI of the nasopharynx and neck, chest X-ray/CT, and abdomen sonography/CT with or without bone scans. Other additional investigations (e.g., FDG PET/CT) would also be scheduled if clinically necessary. OS, measured as the time from treatment initiation to all-cause death or last follow-up, was the study's primary endpoint.

2.5. Machine learning model building workflow

The DeepHit model's advantage is that it combines a recurrent neural network and a multilayer perceptron to learn time-dependent and nonlinear relationships between covariates and events.¹⁹ The recurrent neural network processes the observed sequence of time-dependent covariates, while the multilayer perceptron extracts the static covariate features. The recurrent neural network and multilayer perceptron outputs are linked and fed into a softmax layer to obtain the probabilities of competing events. Multilayer perceptron was developed using pyTorch and pyCox frameworks (Python) to construct the DeepHit neural network. The DeepHit architecture is shown in Supplementary Fig. 2. Supplementary Table 1 shows the hyperparameters of the proposed DeepHit model. Comprehensive details on establishing the DeepHit model are provided in the Supplementary Methods.

XGBoost has demonstrated exceptional performance in numerous predictive works and recently found applications in medical research.^{11, 20, 21} It produces a sequence of tree models constructed iteratively, each built upon the previous one. The desired outcome of the XGBoost survival version includes the OS rate and the survival time. We combined a precise and intricate ML model with interpretable explanations to create a survival model based on the tumor burden and clinical factors. Supplementary Table 2 presents the XGBoost model's hyperparameters after performing 5-fold cross-validation calibration. Detailed information on establishing the XGBoost model is presented in the Supplementary methods. Shapley additive interpretation (SHAP) was used to interpret the prediction results of the ML models and the intricate nonlinear relationship between features and overall mortality. Two examples were used to explain the XGBoost predictions.

Evaluations were made to compare the ML models with the nomogram. We used the Harrell concordance index (C-index) and the area under the receiver operating characteristic curve (AUC) to compare the performance of the three models.

The X-tile method calculated the optimal cutoff values based on the best-performing model's predictions, splitting the study population into low- and high-risk cohorts. The IC therapeutic efficacy was evaluated in both cohorts.

2.6. Statistical analysis

Survival estimates for nomogram model development were calculated using the Kaplan-Meier method and compared using the log-rank test. We identified factors significantly associated with the OS probability and calculated the hazard ratios (HRs) using univariate and multivariable analyses (MVAs) and the Cox proportional hazards regression. A nomogram was created with the rms package in R to calculate individual OS probability based on the MVA results.

We used Python (version 3.4.3) for the ML methods. Other statistical analyses were performed using IBM SPSS Statistics for Macintosh, Version 26.0, and R (version 3.5.2). Two-sided statistical significance was set at $P < 0.05$.

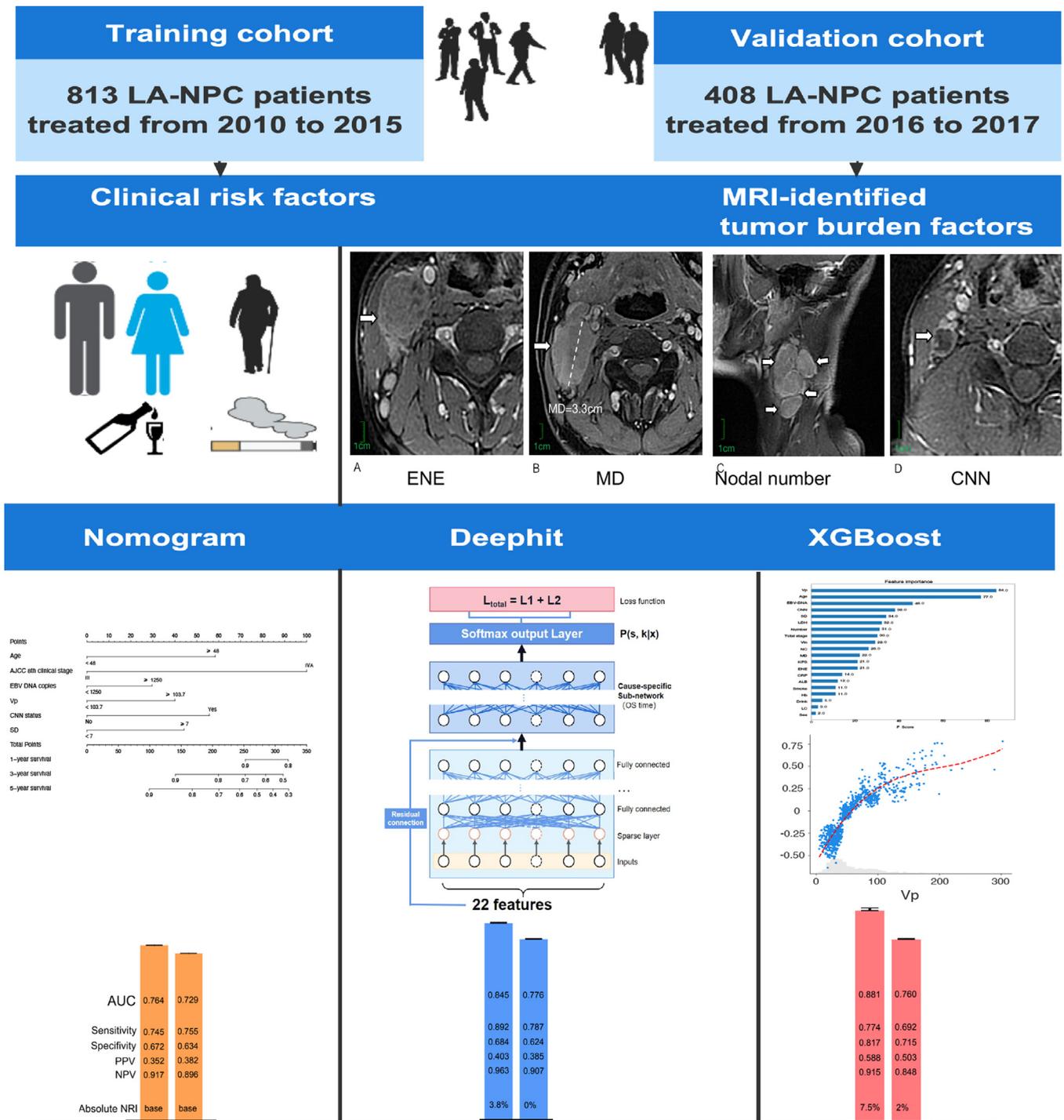


Fig. 1. The study flow diagram. AUC, area under the curve; CNN, central nodal necrosis; ENE, extranodal extension; LA-NPC, locally advanced nasopharyngeal carcinoma; MD, maximum diameter; MRI, magnetic resonance imaging; NPV, negative predictive value; NRI, net reclassification index; OS, overall survival; PPV, positive predictive value; Vp, volume of the primary tumor; XGBoost, eXtreme Gradient Boosting.

3. Results

3.1. General characteristics

The patients in the training and validation sets had similar general characteristics (Supplementary Table 3). Fig. 1 presents the study outline. The participants' median age was 48 (18–86) years, and the male-to-female ratio was 3:1. Of the patients, 679 (55.6%) harbored AJCC III tumors, and 542 (44.4%) had AJCC IVA tumors. Over half of the patients received CCRT (742, 60.8%), 311 patients (25.5%)

received IMRT alone, 150 patients (12.3%) received IC+CCRT, and 28 (2.3%) received IC+IMRT. The 5-year OS was 78.4% in the entire cohort, 79.2% in the training set, and 76.3% in the validation set.

3.2. Construction and validation of the nomogram

Using the OS of the training set as the endpoint, the X-tile identified the following optimal cutoffs in the training cohort: 1 250 for EBV-DNA copies, 103.7 mL for Vp, 13.7 mL for Vln, 7 for nodal number, 7.0 cm

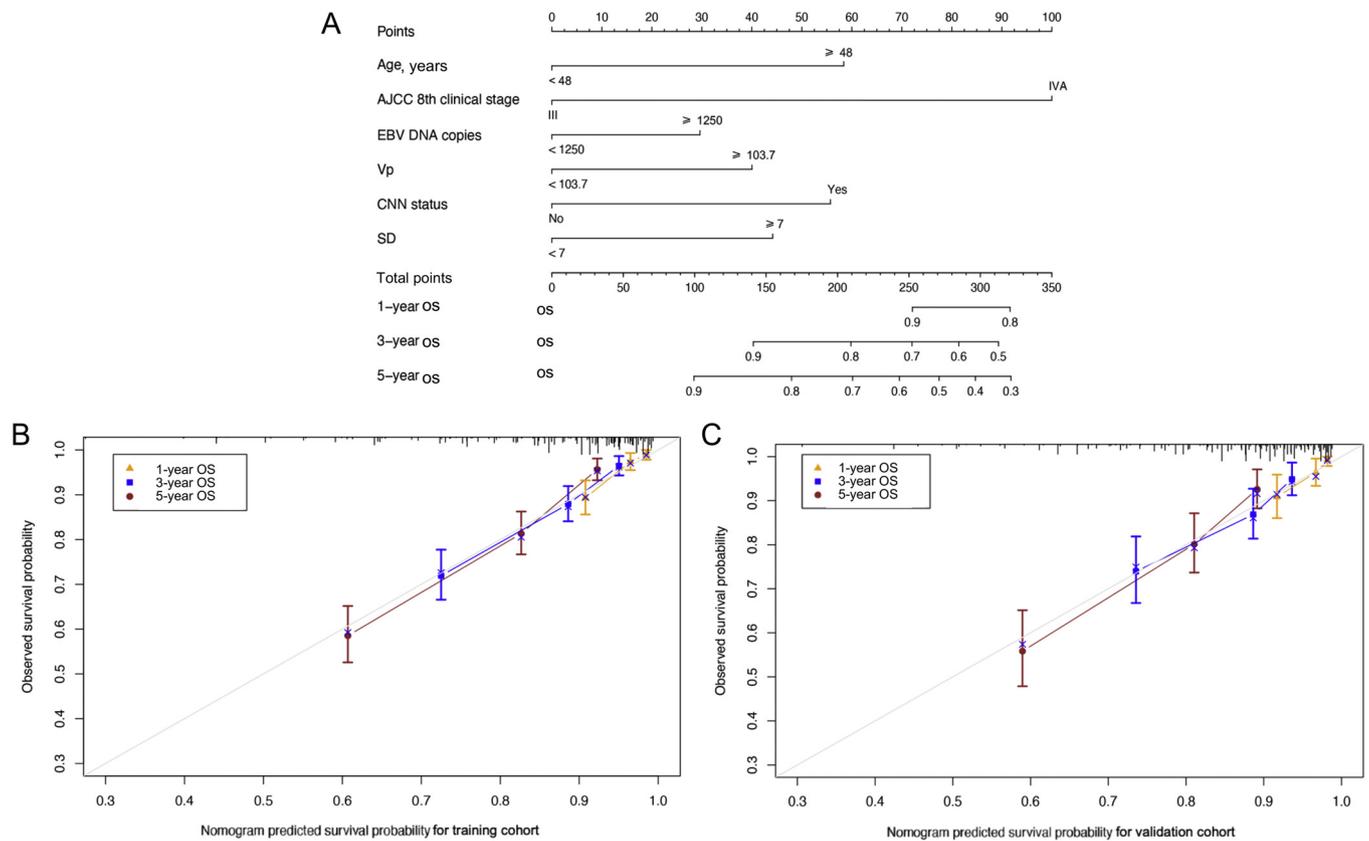


Fig. 2. A nomogram based on the training set. (A) and calibration plots based on the training (B) and validation (C) sets. AJCC, American Joint Committee on Cancer; CNN, central nodal necrosis; EBV, Epstein-Barr virus; OS, overall survival; SD, spread distance; Vp, primary tumor volume.

for SD, 3.0 cm for MD, 158 U/L for LDH, 47.1 g/L for ALB, 162 g/L for Hb, 0.01 mg/dL for CRP, 1.5×10^9 /L for LC, and 3.9×10^9 /L for NC.

Univariable analysis indicated that age, KPS, alcohol drinking, smoking, AJCC 8th T and N stages, clinical stage, EBV-DNA copies, Vp, nodal number, ENE, CNN status, SD, LDH, ALB, Hb, LC, and NC exhibited significant associations with OS. MVA demonstrated that age, clinical stage, EBV-DNA copy number, primary tumor volume, CNN, and SD were independent factors associated with OS (Table 1). A nomogram was developed using all the independent prognostic factors for OS identified by MVA based on the training set. The 1-, 3-, and 5-year OS rates predicted by the nomogram are presented in Fig. 2A. The OS nomogram’s C-index was 0.730 (95% CI, 0.713–0.747) for the training set and 0.705 (95% CI, 0.683–0.727) for the validation set. The calibration curves for the nomogram exhibited good agreement between the observed and predicted probabilities for 1-, 3-, and 5-year OS in the training and validation sets, closely following the 45° diagonal line (Figs. 2B and C).

3.3. Construction and validation of the DeepHit model

The C-index of the DeepHit-based survival models was 0.811 (95% CI, 0.808–0.814) for the training set and 0.767 (95% CI, 0.763–0.771) for the validation set, respectively, resulting in respective 5-year AUCs of 0.845 (95% CI, 0.842–0.848) and 0.776 (95% CI, 0.772–0.780), respectively. The AUCs for the two cohorts at the three time intervals are presented in Supplementary Table 4.

3.4. Construction and validation of an interpretable XGBoost model

The XGBoost survival model’s C-index was 0.849 (95% CI, 0.843–0.855) for the training set and 0.768 (95% CI, 0.761–0.775) for the validation set. The model performance resulted in respective 5-year AUCs of 0.881 (95% CI, 0.877–0.885) and 0.760 (95% CI, 0.749–0.771). The

AUCs for the two cohorts at the three time intervals are presented in Supplementary Table 5. The 20 most influential features, those showing the highest “gain” and used in the XGBoost prediction model, are depicted in Fig. 3A and included Vp, CNN status, and SD (tumor burden-related) among the top five parameters. Age and EBV-DNA copy number also played important predictive roles.

Two explicit examples are shown in Fig. 3B and C to further illustrate this model. The figures explain how the patients were assigned a predicted 5-year overall mortality risk given the status of the 22 variables. For example, the XGBoost survival model predicted a relative 5-year mortality risk of 0.266 (base value: -0.711) for the high-risk patient in Fig. 3C, while it predicted a mortality risk of -1.752 for the low-risk patient (Fig. 3B). The main factors responsible for the high risk in the first patient were Vp of 168.6 mL, age of 27, 6887 EBV-DNA copies, SD of 11.0 cm, LDH of 204.6 U/L, stage of IVA, and regional LN volume of 29.8 mL. The predicted overall mortality risk in the other patient was low because the patient had N1 stage, no LNs with CNN, and the LDH level was 168.10 U/L.

The SHAP method provided an intuitive explanation for the nonlinear impact of features on OS, as shown in Fig. 4A–F. The overall mortality risk increased with increasing Vp (Fig. 4A). The curve showed that the mortality risk increased with the increased number of EBV-DNA copies (Fig. 4C). An SD below 7 cm was relatively safe for patients with NPC. The risk started to grow rapidly above that distance (Fig. 4D). The risk increased linearly with the nodal number until it reached 30 (Fig. 4E). Patients with higher LDH concentrations had a significantly higher risk of mortality (Fig. 4F).

3.5. Prediction performance comparison between the models

As depicted in Table 2, the XGBoost, DeepHit, and nomogram models differed significantly in their training and validation set C-indices (0.849 and 0.768, 0.811 and 0.767, and 0.730 and 0.705, respectively);

Table 1
Univariable and multivariable analysis on OS in Training Cohort (n = 813)

Characteristics	Univariable			Multivariable		
	HR	95% CI	P value	HR	95% CI	P value
Age, years						
< 48	Ref			Ref		
≥ 48	1.57	1.26–1.96	<0.001	1.89	1.42–2.52	<0.001
Sex						
Male	Ref					
Female	0.83	0.64–1.08	0.159			
KPS						
90	Ref			Ref		
80	0.66	0.33–1.29	0.224	0.82	0.41–1.68	0.594
70	0.49	0.25–0.98	0.044	0.78	0.37–1.64	0.519
Smoke						
No	Ref			Ref		
Yes	0.79	0.63–0.98	0.032	1.00	0.72–1.37	0.985
Drink						
No	Ref			Ref		
Yes	1.75	1.32–2.32	<0.001	1.12	0.75–1.68	0.589
AJCC 8 th T stage	1.88	1.60–2.21	<0.001			
T1	Ref					
T2	1.19	0.51–2.74	0.691	0.80	0.34–1.88	0.605
T3	1.06	0.53–2.13	0.868	1.03	0.51–2.10	0.931
T4	1.49	1.49–5.79	0.002	0.94	0.32–2.78	0.905
AJCC 8 th N stage						
N0	Ref					
N1	0.95	0.58–1.56	0.949	0.90	0.54–1.50	0.677
N2	1.14	0.72–1.79	0.684	1.02	0.61–1.69	0.950
N3	1.73	1.07–1.39	0.043	0.66	0.32–1.34	0.246
AJCC 8 th clinical stage						
III	Ref			Ref		
IV	1.82	1.62–2.05	<0.001	2.98	2.16–4.11	<0.001
EBV-DNA copies						
< 1 250	Ref			Ref		
≥ 1 250	1.52	1.21–1.91	<0.001	1.41	1.05–1.88	0.021
Vp, mL						
< 103.7	Ref			Ref		
≥ 103.7	2.75	2.14–3.52	<0.001	1.52	1.08–2.14	0.016
Vln, mL				/	/	/
< 13.7	Ref					
≥ 13.7	1.25	1.00–1.56	0.051			
Nodal number						
< 7	Ref			Ref		
≥ 7	1.24	1.08–1.42	0.002	0.86	0.61–1.22	0.402
ENE status						
No	Ref			Ref		
Yes	1.37	1.10–1.71	0.006	1.02	0.68–1.52	0.924
CNN status						
No	Ref			Ref		
Yes	1.68	1.35–2.10	<0.001	1.79	1.35–2.37	<0.001
SD (cm)						
< 7.0	Ref			Ref		
≥ 7.0	1.54	1.10–2.16	0.011	1.55	1.00–2.39	0.048
MD, cm						
< 3.0	Ref			Ref		
≥ 3.0	1.26	1.01–1.57	0.042	1.06	0.75–1.50	0.747
LDH, U/L						
< 158	Ref			Ref		
≥ 158	1.63	1.25–2.11	<0.001	1.36	0.98–1.89	0.065
ALB, g/L						
< 47.1	Ref			Ref		
≥ 47.1	0.64	0.44–0.93	0.018	0.93	0.57–1.52	0.766
Hb, g/L						
< 162	Ref			Ref		
≥ 162	0.62	0.40–0.98	0.040	0.79	0.45–1.37	0.401
CRP, mg/dl						
< 0.01	Ref			Ref		
≥ 0.01	0.85	0.67–1.08	0.183	0.80	0.58–1.09	0.158
LC, × 10 ⁹ /L						
< 1.5	Ref			Ref		
≥ 1.5	1.41	1.11–1.79	0.005	1.22	0.85–1.75	0.292
NC, × 10 ⁹ /L						
< 3.9	Ref			Ref		
≥ 3.9	1.45	1.1–1.91	0.008	1.14	0.78–1.68	0.498

Abbreviations: AJCC, American Joint Committee on Cancer; ALB, albumin; CRP, C-reactive protein; CNN, central nodal necrosis; ENE, extranodal extension; Hb, hemoglobin; HR, hazard ratio; KPS, Karnofsky score; LC, lymphocyte; LDH, lactate dehydrogenase; MD, maximum diameter; NC, neutrophilic granulocytes; OS, overall survival; Ref, reference; SD, spread distances; Vp, volume of the primary tumor; Vln, volume of the regional lymph nodes.

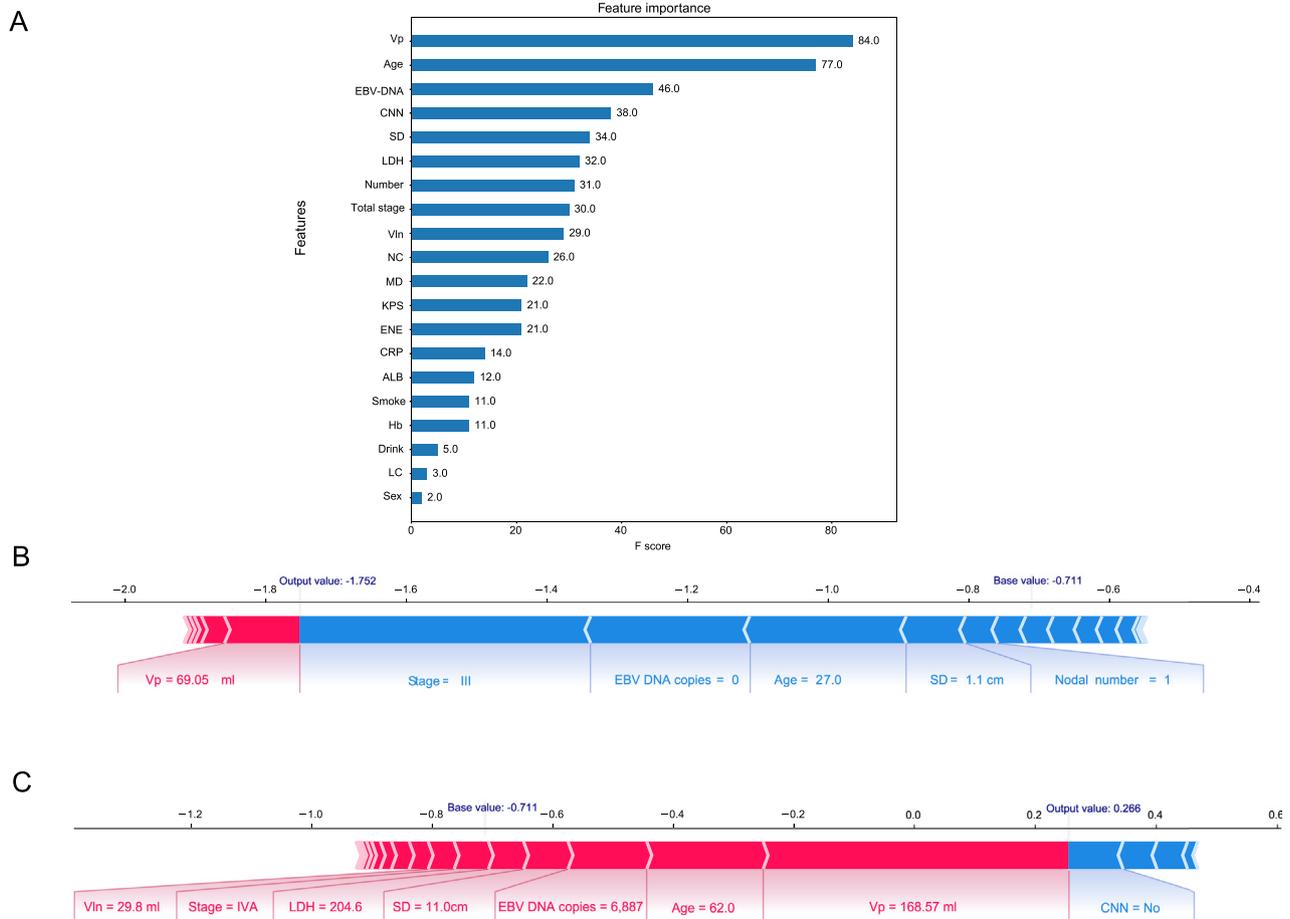


Fig. 3. Interpretation of XGBoost model. (A) The 20 features with the highest “gain” used for prediction by the XGBoost model. (B, C) Explained risk for two explicit examples. Blue bars: low-risk variables; red bars: high-risk variables. The contributing variables are arranged along the horizontal line and sorted by the absolute value of their impact. The output value is the predicted relative risk of overall mortality. The base value is the expected value in the XGBoost model based on the training dataset. ALB, albumin; CNN, central nodal necrosis; CRP, C-reactive protein; EBV, Epstein-Barr virus; ENE, extranodal extension; Hb, hemoglobin; KPS, Karnofsky score; LC, lymphocyte; LDH, lactate dehydrogenase; MD, maximum diameter; NC, neutrophilic granulocytes; SD, spread distance; Vln, regional lymph nodes volume; Vp, primary tumor volume; XGBoost, eXtreme Gradient Boosting.

Table 2

Overall survival prediction performance in the training and validation cohort of traditional nomogram versus machine learning models.

	Nomogram		DeepHit		XGBoost	
	Training (n = 813)	Validation (n = 408)	Training (n = 813)	Validation (n = 408)	Training (n = 813)	Validation (n = 408)
C-index (95% CI)	0.730 (0.713–0.747)	0.705 (0.683–0.727)	0.811 (0.808–0.814)	0.767 (0.763–0.771)	0.849 (0.843–0.855)	0.768 (0.761–0.775)
AUC at 5y (95% CI)	0.764 (0.760–0.768)	0.729 (0.725–0.733)	0.845 (0.842–0.848)	0.776 (0.772–0.780)	0.881 (0.877–0.885)	0.760 (0.749–0.771)
Sensitivity at 5y	0.745	0.755	0.892	0.787	0.774	0.692
Specificity at 5y	0.672	0.634	0.684	0.624	0.817	0.715
PPV at 5y	0.352	0.382	0.403	0.385	0.588	0.503
NPV at 5y	0.917	0.896	0.964	0.907	0.915	0.848
Absolute net reclassification index at 5y	base	base	0.038	0.000	0.075	0.020

Abbreviations: AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value; XGBoost, Xtreme Gradient Boosting; y, years.

nomogram vs. XGBoost or DeepHit, both $P < 0.001$). Similarly, the AUCs of the XGBoost and DeepHit models for predicting the 5-year OS in both cohorts were significantly larger than the respective AUCs of the nomogram model (0.881 and 0.760, 0.845 and 0.776, and 0.764 and 0.729; Table 2). Furthermore, when using the nomogram model as the baseline, both ML models showed positive net reclassification indices, indicating that they correctly classified more patients to the corresponding risk category than the nomogram.

We divided the training set’s patients into low- and high-risk groups based on the XGBoost model-generated predictions, using a prediction cutoff value of 0.39 identified by X-tile. Kaplan-Meier analysis revealed that the respective 5-year OS of the two subgroups in the training (86.8% and 54.2%, HR= 3.874 [95% CI, 2.940-5.103]), $P < 0.001$; Fig. 5A) and validation (83.8% and 53.5%, HR=3.730 [95% CI, 2.624-5.300]), $P < 0.001$; Fig. 5B) sets differed significantly.

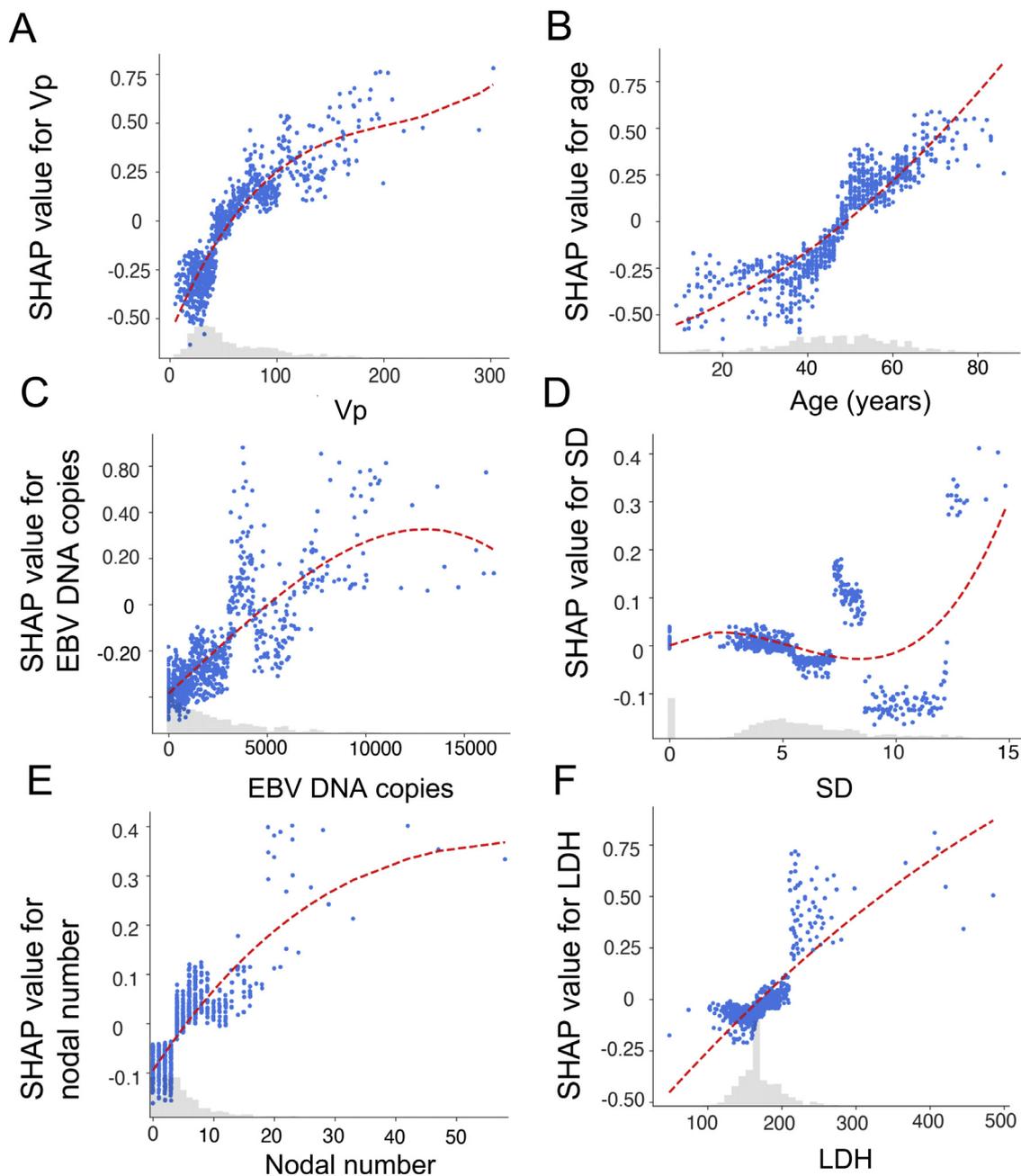


Fig. 4. (A–F) Effect of various variables on the predicted overall survival in the eXtreme Gradient Boosting model. Each point represents a SHAP value of an individual patient (Y-axis) in association with the variables on the X-axis. SHAP values above zero indicate a positive association with all-cause mortality, while values below zero indicate a negative association (i.e., the higher the SHAP value, the higher the risk of mortality). The gray histograms in each plot show the distribution of values for the variables in the training set. EBV, Epstein-Barr virus; LDH, lactate dehydrogenase; SD, spread distances; SHAP, Shapley additive interpretation; Vp, primary tumor volume.

3.6. XGBoost model-based risk stratification and adaptive usage of induction chemotherapy

We analyzed the IC clinical efficacy in the low- and high-risk groups derived from the XGBoost model. Compared with CCRT, IC+CCRT presented no OS benefit to the low-risk group (Fig. 6A,B) but was beneficial to the high-risk group (training set: HR = 0.526 [95% CI, 0.324–0.855], $P = 0.008$; validation set: HR = 0.526 [95% CI, 0.292–0.949], $P = 0.029$; Figs. 6C,D).

4. Discussion

It is crucial to accurately predict the prognosis of patients with NPC as it can help make treatment decisions. This large population-based study used multidimensional tumor burden feature quantifications to demonstrate remarkably accurate predictive ML algorithms, greatly outperforming the Cox regression-based nomogram model. These results paved the way for personal risk prediction using multidimensional data. Of greater significance, the ML-based prognostic stratification helped identify patients for whom IC was most likely unnecessary. Management of these individuals can be appropriately tailored considering the treatment-related risks and benefits.

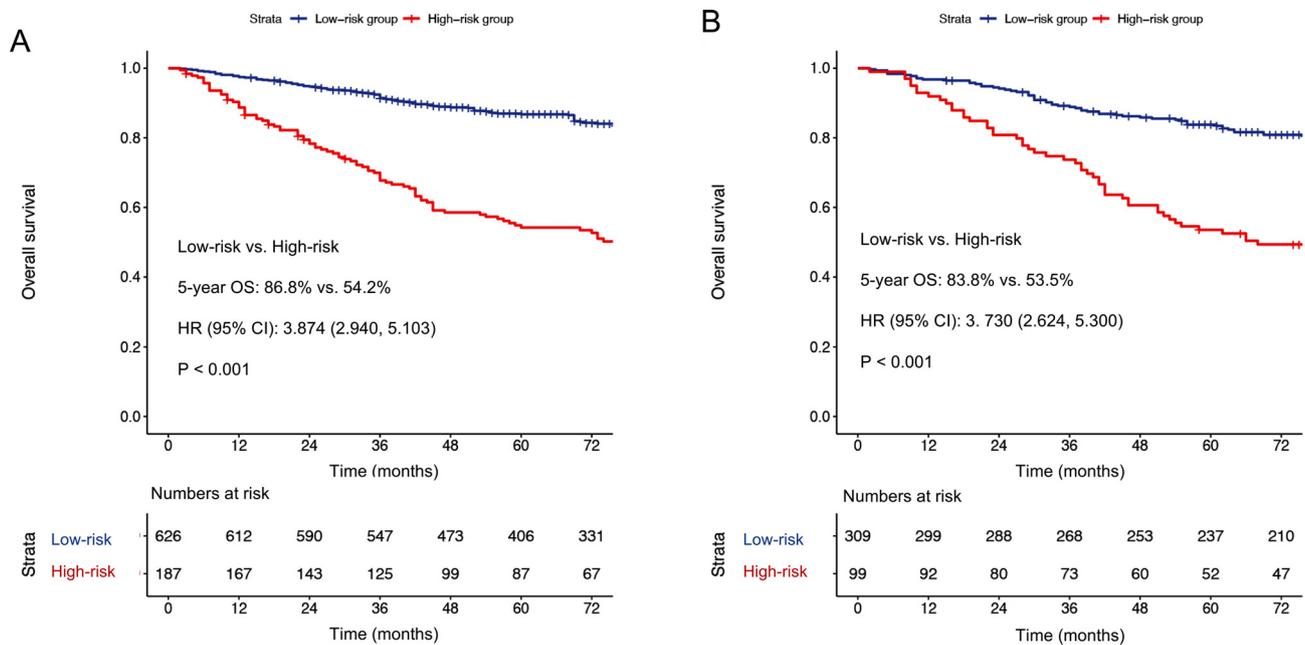


Fig. 5. Survival curves of the low- and high-risk subgroups defined based on the eXtreme Gradient Boosting prediction results in the training (A) and validation (B) cohorts. CI, confidence interval; HR, hazard ratio; OS, overall survival.

MRI-based tumor burden is a crucial determinant influencing prognosis and treatment decisions. The AJCC N classification primarily relies on the location, size, and cervical LN laterality. However, a growing recent trend advocates reporting additional variables, including the positive LN count, ENE, and CNN status, to provide a more comprehensive assessment of the nodal burden profile.^{22–26} Recent studies have shown that the metastatic LN quantity was a significant independent survival predictor in patients with NPC. This parameter is assumed to better represent the aggregate effect of the involved LNs than N staging following the AJCC system.¹⁵ Furthermore, quantitative assessment of metastatic LN regions in NPC was shown to facilitate a prognostic assessment superior to that based on the N classification.²⁷ Moreover, high-grade ENE in patients with NPC has been identified as a valuable predictor of those likely to develop distant metastasis.^{23,28,29} Another critical LN feature in patients with NPC was CNN, which proved to be an independent poor prognostic factor.^{22,26} The present research focused on general nodal features while incorporating a novel and encouraging predictor, the SD.¹⁶ We used MVA to show that the independent variables Vp, CNN status, and SD outperformed the conventional AJCC N staging. Furthermore, three of the top five features identified in the XGBoost method were tumor burden-related features. This finding suggested that tumor burden parameters could offer valuable and detailed information for outcome prediction. Our results demonstrated that selected quantitative primary tumor and regional LNs features more accurately reflect the disease burden and led to improved risk stratification than the traditional approach.

Using the MRI-detected tumor burden in traditional risk prediction models, with its multidimensional and complex interactions, is limited as it assumes either linearity or proportionality. Although traditional survival models such as proportional Cox regression are often employed,^{30–34} these oversimplified models might lead to insufficient prediction accuracy and mislead subsequent decisions because of their heavy reliance on the assumption of linear forms. In contrast, the ML approach allows a flexible mathematical fitting of the data to accurately describe the biological reality, leading to more accurate outcome predictions. Furthermore, ML algorithms can handle many continuous or categorical variables without the need for scaling or modifications.^{35,36} Therefore, we employed the advanced DeepHit and XGBoost techniques to construct predicting models.

The net reclassification index is a statistical indicator that calculates the difference in the proportion of correctly classified as high- or low-risk individuals by a new model relative to an old model.^{37,38} Therefore, it is typically used to assess the degree of improvement in the classification accuracy of a new predictive model relative to an old one. The positive net reclassification indices of the two ML models in this study indicated that they correctly classified more individuals into their corresponding risk categories than the nomogram model did. Moreover, the two ML models achieved better C-indices and AUCs than the conventional nomogram method.

The increasing feasibility of multidimensional quantification of prognostic markers at the individual level highlights the potential clinical utility of ML techniques. Nevertheless, these techniques are frequently criticized for being "black boxes" that produce opaque and uninterpretable predictions, limiting their widespread use. Therefore, we used the SHAP method in this study to interpret the prediction results. Considering the inherent limitations in interpretability associated with ML models, we demonstrated how the XGBoost prediction improved the understandability of the intricate nonlinear relations between features and overall mortality. Our tumor burden-related results align with previous research, demonstrating that larger primary tumors were associated with relapse and metastasis. This association could be attributed to factors such as poor tumor blood and oxygen supply that arise from the aggressive nature of cancer and could lead to radioresistance.³⁹ Furthermore, we assessed the correlation between the numbers of EBV-DNA copies and LNs, among others, and overall mortality. The overall mortality risk was linearly correlated to tumor burden factors but remained constant once a certain value was reached. Visualization of the ML system's prediction showed that the association between the factors and outcomes was nonlinear. The ML models outperformed the Cox method because they can capture such non-linear relationships.

The ML-driven risk classification for overall mortality could additionally function as a resource for making treatment decisions. The significance of chemotherapy in improving disease control was underscored by the failure pattern observed in patients with NPC.^{3,4} However, it is still unclear if all patients should receive IC. Several studies examined the significance of nodal variables in determining the potential advantages of IC.^{26,40} The nomogram Chen et al. reported incorporated MRI-derived tumor burden characteristics and CNN, demonstrating a

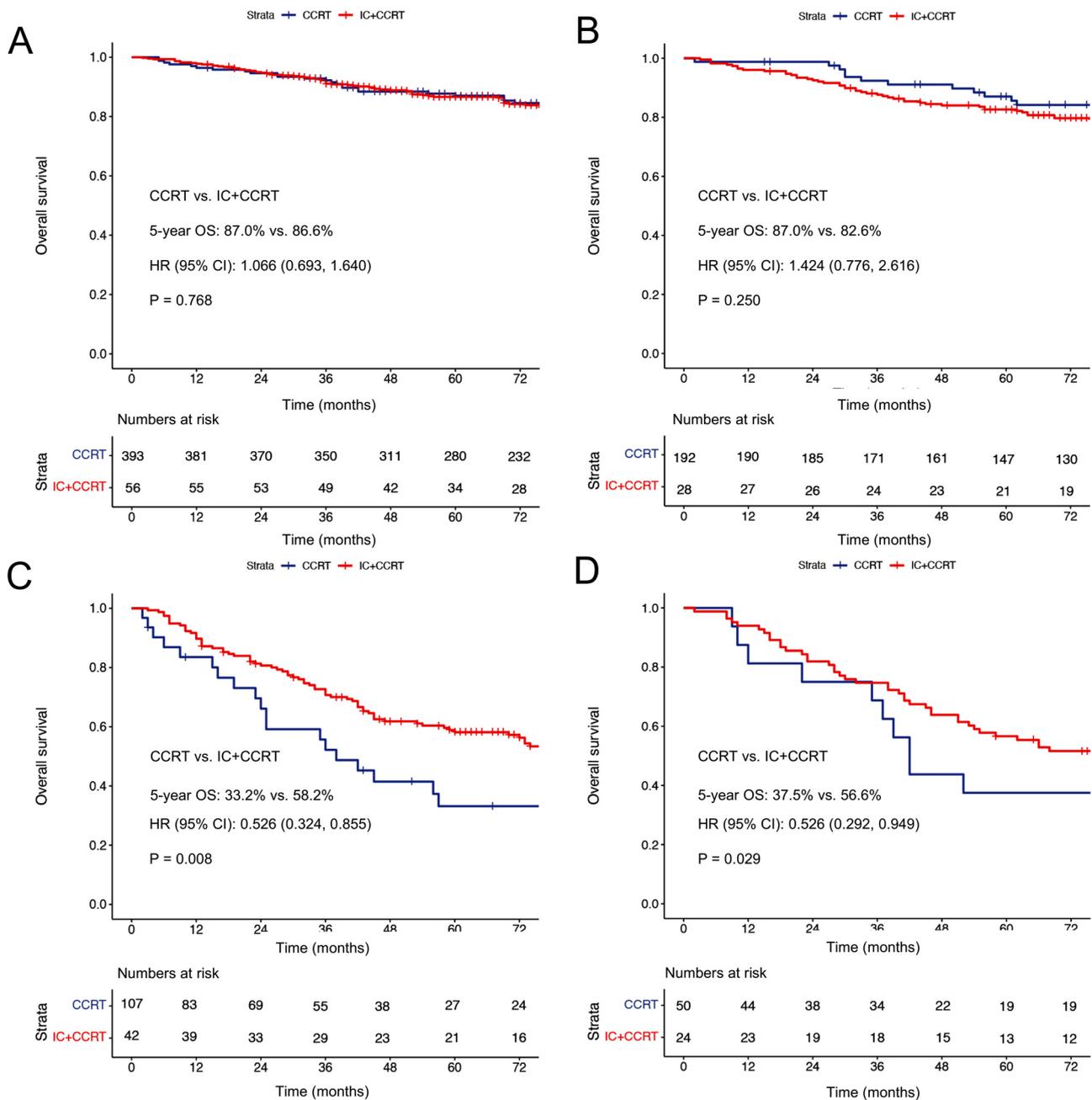


Fig. 6. Impact of IC+CCRT in the Xtreme Gradient Boosting model-based low-risk (A, B) and high-risk subgroups (C, D) in the training (A, C) and validation (B, D) cohorts. CCRT, concurrent chemoradiotherapy; CI, confidence interval; HR, hazard ratio; IC, induction chemotherapy; OS, overall survival.

consistent ability to identify individuals likely to benefit from IC.²⁶ A recent report suggested that patients at AJCC stage II-IV with multiple CNNs could benefit from IC.⁴⁰

We used deep ML algorithms to establish ML-based risk stratification for OS and used a cutoff value based on its predictions to divide the research participants into two risk groups. This risk classification successfully demonstrated the distinct clinical effectiveness of IC+CCRT in these risk groups. It was revealed that IC positively impacted participants in the training and validation sets in the high-risk category, reducing the overall mortality risk by around 60%. Conversely, we observed no beneficial effect in the low-risk group. The findings indicated that ML-based overall mortality risk stratification was valid and could be used by clinicians to decide if to use IC+CCRT or CCRT in patients with NPC.

Certain limitations of this research should be discussed. As a retrospective study, possible selection biases may have influenced its results. Therefore, a multicenter prospective study should be performed to validate our findings. Furthermore, we used single-center data, suggesting that our ML model-derived prognostic and predictive results should be validated independently using a large external cohort.

5. Conclusions

This study used a large population to create and validate ML-based models for risk stratification. These models exhibited excellent performances and yielded significantly better OS predictions than a nomogram in patients with LA-NPC. Moreover, the ML model-derived stratification successfully identified patients likely to gain survival benefits from

CCRT+IC. Hence, the ML-based risk stratification models could benefit patients' consultations and clinical evaluations. Nevertheless, our findings should be externally validated to expand their clinical usefulness.

Declaration of competing interest

The authors declare that they have no conflict of interests.

Ethics statement

The research followed the principles of Good Clinical Practice as specified in the Declaration of Helsinki. Because of the retrospective nature of this study, our institutional review board approved the study and waived the need to obtain the patients' informed consent (approval number: 23/353-4095).

Consent for publication

The patients whose medical images are presented in this manuscript provided consent for publication of their medical images for the manuscript. Only de-identified data and images are used in this article.

Author contributions

Y.L., S.S., J.Y., J.W. provided conception and design of the research; J.Y. and J.W. provided the financial support; J.Z., Y.L., J.Y. and J.W. provided administrative support; Y.L., S.S., X.H., K.W., Y.Q., X.C., R.W., Y.Z., J.L., J.Y. and J.W. provided the study materials or patients; Y.L., S.S., R.W., Y.L., J.Y. and J.W. collected and assembled the data; Y.L., S.S., X.H., K.W., Y.Q., X.C., Y.Z., J.Z., J.L., J.Y. and J.W. conducted data analysis and interpretation. All authors drafted the manuscript, approved the final manuscript, and are accountable for all aspects of the work.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jncc.2023.10.002.

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424. doi:10.3322/caac.21492.
- Tang LL, Chen WQ, Xue WQ, et al. Global trends in incidence and mortality of nasopharyngeal carcinoma. *Cancer Lett*. 2016;374(1):22–30. doi:10.1016/j.canlet.2016.01.040.
- Sun X, Su S, Chen C, et al. Long-term outcomes of intensity-modulated radiotherapy for 868 patients with nasopharyngeal carcinoma: an analysis of survival and treatment toxicities. *Radiother Oncol*. 2014;110(3):398–403. doi:10.1016/j.radonc.2013.10.020.
- Tian YM, Liu MZ, Zeng L, et al. Long-term outcome and pattern of failure for patients with nasopharyngeal carcinoma treated with intensity-modulated radiotherapy. *Head Neck*. 2019;41(5):1246–1252. doi:10.1002/hed.25545.
- National Comprehensive Cancer Network Inc. *NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) Head and Neck Cancers (Version 2.2023, May 15, 2023)* <https://www.NCCN.org>.
- Pan JJ, Ng WT, Zong JF, 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*. 2016;122(21):3307–3315. doi:10.1002/cncr.30198.
- Liang W, Shen G, Zhang Y, et al. Development and validation of a nomogram for predicting the survival of patients with non-metastatic nasopharyngeal carcinoma after curative treatment. *Clin J Cancer*. 2016;35(1):98. doi:10.1186/s40880-016-0160-9.
- Zhong LZ, Fang XL, Dong D, et al. A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0. *Radiother Oncol*. 2020;1:1–9. doi:10.1016/j.radonc.2020.06.050.
- Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319(13):1317–1318. doi:10.1001/jama.2017.18391.
- Chen X, Li Y, Li X, et al. An interpretable machine learning prognostic system for locoregionally advanced nasopharyngeal carcinoma based on tumor burden features. *Oral Oncol*. 2021;118:105335. doi:10.1016/j.oraloncology.2021.105335.
- Unterhuber M, Kresoja KP, Rommel KP, et al. Proteomics-enabled deep learning machine algorithms can enhance prediction of mortality. *J Am Coll Cardiol*. 2021;78(16):1621–1631. doi:10.1016/j.jacc.2021.08.018.
- Chen FP, Wen DW, Li F, et al. The role of post-neoadjuvant chemotherapy tumor volume for prognostication and treatment guidance in loco-regionally advanced nasopharyngeal carcinoma. *Cancers (Basel)*. 2019;11(11):1632. doi:10.3390/cancers11111632.
- Li JY, Huang CL, Luo WJ, et al. An integrated model of the gross tumor volume of cervical lymph nodes and pretreatment plasma Epstein-Barr virus DNA predicts survival of nasopharyngeal carcinoma in the intensity-modulated radiotherapy era: a big-data intelligence platform-based analysis. *Ther Adv Med Oncol*. 2019;11:1758835919877729. doi:10.1177/1758835919877729.
- Huang CL, Chen Y, Guo R, et al. Prognostic value of MRI-determined cervical lymph node size in nasopharyngeal carcinoma. *Cancer Med*. 2020;9(19):7100–7106. doi:10.1002/cam4.3392.
- Ma H, Liang S, Cui C, et al. Prognostic significance of quantitative metastatic lymph node burden on magnetic resonance imaging in nasopharyngeal carcinoma: a retrospective study of 1224 patients from two centers. *Radiother Oncol*. 2020;151:40–46. doi:10.1016/j.radonc.2020.07.023.
- Liu Y, Zhang Y, Wang J, et al. Caudal distribution pattern of metastatic neck lymph nodes in nasopharyngeal carcinoma and prognostic significance of nodal spread distances. *Radiother Oncol*. 2023;179:109443. doi:10.1016/j.radonc.2022.109443.
- van den Brekel MW, Stel HV, Castelijns JA, et al. Cervical lymph node metastasis: assessment of radiologic criteria. *Radiology*. 1990;177(2):379–384. doi:10.1148/radiology.177.2.2217772.
- Huang X, Chen X, Zhao C, et al. Adding concurrent chemotherapy to intensity-modulated radiotherapy does not improve treatment outcomes for stage II nasopharyngeal carcinoma: a phase 2 multicenter clinical trial. *Front Oncol*. 2020;10:1314. doi:10.3389/fonc.2020.01314.
- Lee C, Yoon J, Schaar MV. Dynamic-DeepHit: a deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Trans Biomed Eng*. 2020;67(1):122–133. doi:10.1109/tbme.2019.2909027.
- Raparelli V, Romiti GF, Di Teodoro G, et al. A machine-learning based biopsychosocial model for the prediction of non-obstructive and obstructive coronary artery disease. *Clin Res Cardiol*. 2023;112(9):1263–1277. doi:10.1007/s00392-023-02193-5.
- Rahimi S, Chu C, Grad R, et al. Explainable machine learning model to predict COVID-19 severity among older adults in the province of Quebec. *Ann Fam Med*. 2023;21(21):3619 Suppl 1. doi:10.1370/afm.21.s1.3619.
- Du YY, Luo DH, Sun XS, et al. Combining pretreatment plasma Epstein-Barr virus DNA level and cervical node necrosis improves prognostic stratification in patients with nasopharyngeal carcinoma: a cohort study. *Cancer Med*. 2019;8(16):6841–6852. doi:10.1002/cam4.2481.
- Hu Y, Lu T, Huang SH, et al. High-grade radiologic extra-nodal extension predicts distant metastasis in stage II nasopharyngeal carcinoma. *Head Neck*. 2019;41(9):3317–3327. doi:10.1002/hed.25842.
- Liu Y, Chen S, Dong A, et al. Nodal grouping in nasopharyngeal carcinoma: prognostic significance, N classification, and a marker for the identification of candidates for induction chemotherapy. *Eur Radiol*. 2020;30(4):2115–2124. doi:10.1007/s00330-019-06537-6.
- Lin TY, Lan MY, Tsou HH, et al. Survival impacts of different nodal characteristics and T-classification in N3 nasopharyngeal carcinoma patients. *Oral Oncol*. 2020;108:104820. doi:10.1016/j.oraloncology.2020.104820.
- Chen X, Cao X, Jing B, et al. Prognostic and treatment guiding significance of MRI-based tumor burden features and nodal necrosis in nasopharyngeal carcinoma. *Front Oncol*. 2020;10:537318. doi:10.3389/fonc.2020.537318.
- Zhou X, Ou X, Yang Y, et al. Quantitative metastatic lymph node regions on magnetic resonance imaging are superior to AJCC N classification for the prognosis of nasopharyngeal carcinoma. *J Oncol*. 2018;9172585. doi:10.1155/2018/9172585.
- Ai QY, King AD, Poon DMC, et al. Extranodal extension is a criterion for poor outcome in patients with metastatic nodes from cancer of the nasopharynx. *Oral Oncol*. 2019;88:124–130. doi:10.1016/j.oraloncology.2018.11.007.
- Lu T, Hu Y, Xiao Y, et al. Prognostic value of radiologic extranodal extension and its potential role in future N classification for nasopharyngeal carcinoma. *Oral Oncol*. 2019;99:104438. doi:10.1016/j.oraloncology.2019.09.030.
- Li WZ, Liang H, Lv X, et al. 1095P-Prognostic value of MRI-derived residual retropharyngeal lymph node after intensity-modulated radiotherapy in nasopharyngeal carcinoma and a nomogram for the prediction of it. *Ann Oncol*. 2018;29(8):viii390 suppl. doi:10.1093/annonc/mdy287.051.
- Liang W, Shen G, Wu X, Zhang Y, Zhao C, Zhang L. 1022P-Development and validation of a nomogram for predicting overall survival of patients with non-metastatic nasopharyngeal carcinoma after curative therapy. *Ann Oncol*. 2014;25(4):iv351 suppl. doi:10.1093/annonc/mdu340.37.
- Sun JD, Chen CZ, Chen JZ, et al. Long term outcomes and prognostic factors of n0 stage nasopharyngeal carcinoma: a single institutional experience with 610 patients. *Asian Pac J Cancer Prev*. 2012;13(5):2101–2107. doi:10.7314/apjcp.2012.13.5.2101.
- Cui C, Wang S, Zhou J, et al. Machine learning analysis of image data based on detailed MR image reports for nasopharyngeal carcinoma prognosis. *Biomed Res Int*. 2020;2020:8068913. doi:10.1155/2020/8068913.
- Liang Y, Chen KH, Yang J, et al. Acute toxicities and prognosis of elderly patients with nasopharyngeal carcinoma after intensity-modulated radiotherapy: prediction with nomogram. *Cancer Manag Res*. 2020;12:8821–8832. doi:10.2147/CMAR.S261717.
- Li S, Wan X, Deng YQ, et al. Predicting prognosis of nasopharyngeal carcinoma based on deep learning: peritumoral region should be valued. *Cancer Imaging*. 2023;23:14. doi:10.1186/s40644-023-00530-5.
- Luo X, Liao W, He Y, et al. Deep learning-based accurate delineation of primary gross tumor volume of nasopharyngeal carcinoma on heterogeneous magnetic resonance imaging: a large-scale and multi-center study. *Radiother Oncol*. 2023;180:109480. doi:10.1016/j.radonc.2023.109480.

37. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. *Stat Med.* 2011;30(1):22–38. doi:10.1002/sim.4026.
38. Alba AC, Agoritsas T, Walsh M, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *JAMA.* 2017;318(14):1377–1384. doi:10.1001/jama.2017.12126.
39. Becker A, Hänsgen G, Bloching M, Weigel C, Lautenschläger C, Dunst J. Oxygenation of squamous cell carcinoma of the head and neck: comparison of primary tumors, neck node metastases, and normal tissue. *Int J Radiat Oncol Biol Phys.* 1998;42(1):35–41. doi:10.1016/s0360-3016(98)00182-5.
40. Li J, Zhao Q, Zhang Y, et al. Prognostic value of quantitative cervical nodal necrosis burden on MRI in nasopharyngeal carcinoma and its role as a stratification marker for induction chemotherapy. *Eur Radiol.* 2022;32:7710–7721. doi:10.1007/s00330-022-08785-5.