



Development and validation of a machine learning–based prognostic model for gastric cancer: a multicenter retrospective study

Xiao Guan¹ · Jinsong Liu² · Lei Xu³ · Chengfeng Wang¹

Received: 18 February 2025 / Accepted: 16 September 2025
© The Author(s) 2025

Abstract

Background Machine learning has emerged as a promising tool for survival prediction in various diseases; however, its application and external validation in real-world gastric cancer populations remain limited.

Methods Clinical data of patients diagnosed with gastric cancer between 2000 and 2018 were obtained from the SEER database, supplemented with data from two Chinese medical centers (2005–2018). Three feature selection methods and four modeling algorithms—including Cox, RSF, CoxBoost, and DeepSurv_Cox—were employed to construct prediction models for overall survival (OS) and cancer-specific survival (CSS). Model performance was evaluated using the concordance index (C-index), integrated Brier score (IBS), and mean area under the curve (AUC). The two best-performing base models were subsequently integrated into a stacked model and compared against the traditional TNM staging system using decision curve analysis (DCA) and time-dependent ROC curves at 3, 5, and 10 years.

Results A total of 21,559 patients from the SEER database and 3,805 patients from two Chinese centers were included. In independent testing, the integrated model achieved a C-index/IBS/mean AUC of 0.693/0.158/0.829 for OS and 0.719/0.171/0.819 for CSS. For 3-, 5-, and 10-year survival prediction, the AUCs were 0.705/0.747/0.851 for OS and 0.734/0.779/0.830 for CSS, outperforming the TNM staging system across all metrics. Superior calibration and clinical utility of the integrated model were further confirmed by calibration curves and DCA.

Conclusion The integrated machine learning model outperformed both traditional TNM staging and deep learning approaches, offering improved predictive accuracy for survival outcomes in patients with gastric cancer.

Keywords Gastric cancer · Machine learning · Deep learning · Survival analysis

Introduction

Gastric cancer is one of the most prevalent malignancies of the digestive system, ranking fifth in global incidence and fourth in cancer-related mortality worldwide [1]. Surgical resection remains the primary and only potentially curative treatment for this disease. However, even after radical surgery, prognosis varies considerably among patients due to multiple clinical and pathological factors, including tumor stage, histological subtype, and treatment modalities [2, 3]. Accurate prediction of postoperative survival is therefore essential for individualized treatment planning and optimized clinical decision-making. The heterogeneous nature and multifactorial pathogenesis of gastric cancer further complicate prognosis estimation and therapeutic strategies [4].

Currently, clinicians typically assess survival rates based on the American Joint Committee on Cancer (AJCC) staging

Xiao Guan, Jinsong Liu and Lei Xu contributed equally to this work.

✉ Chengfeng Wang
drwangcf1962@163.com

¹ State Key Lab of Molecular Oncology and Department of Pancreatic and Gastric Surgery, National Cancer Center/ National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

² Department of VIP Medical, National Cancer Center/ National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

³ Department of General Surgery, Lu'an Hospital of Anhui Medical University, Anhui, China

system combined with their own medical experience, while overlooking the impact of other survival factors. The staging system is widely used and is an effective method for guiding gastric cancer treatment decisions [5]. However, it does not take into account various factors such as gender, age, tumor size, and histological type, all of which can significantly affect survival prognosis. Moreover, traditional survival analysis methods also have limitations, including the assumption of proportional hazards and the linearity of continuous variables. These limitations may restrict their applicability in complex scenarios. However, machine learning-based prognostic models represent a significant advancement as they effectively address these issues. They can handle non-proportional hazards and model non-linear relationships between variables and outcomes, making survival predictions in various clinical settings more flexible and accurate.

Machine learning excels at extracting information from high-dimensional, complex data, automatically learning and making predictions under supervised or unsupervised modes, playing a crucial role in disease prognosis [6]. Compared to staging models, machine learning models may be more suitable for guiding clinical decision-making in clinical settings. To our knowledge, there is currently a lack of effective predictive models to assess the correlation between multiple clinical factors and postoperative prognosis in gastric cancer patients. Deep learning, a specialized machine learning model incorporating multiple neural networks, can process more complex information [7]. Compared to traditional machine learning models—including multitask logistic regression and random forest models—deep learning methods offer several advantages. They can learn intricate patterns and representations from large datasets, automatically extract relevant features from raw data, and thus achieve superior performance over conventional algorithms. Several studies have already employed deep learning models for surgical oncology research and analysis [8]. However, most of these studies focus on diagnostic applications, such as automated quantification of radiological images, digital histopathological image interpretation, or biomarker analysis [9–11]. To the best of our knowledge, published research utilizing machine learning models for prognostic prediction in surgical oncology remains limited, particularly in the field of gastric cancer. Therefore, machine learning based survival analysis provides a valuable reference for predicting postoperative survival in gastric cancer.

In this study, we aimed to develop and validate robust predictive models for postoperative survival in gastric cancer patients by leveraging machine learning and deep learning algorithms. We utilized large-scale population data from the SEER database as well as clinical data from two

independent Chinese medical centers. Our objectives were to compare the prognostic performance of multiple survival models, identify important prognostic factors for overall survival (OS) and cancer-specific survival (CSS), and construct an integrated predictive framework to assist clinicians in tailoring follow-up strategies and treatment decisions for gastric cancer patients.

Methods

Data sources

Data for this study were obtained from patients with gastric cancer recorded in the Surveillance, Epidemiology, and End Results (SEER) database (2000–2018), maintained by the National Cancer Institute (NCI). Patient information was retrieved using SEER*Stat software. Additionally, patients diagnosed with gastric cancer between 2005 and 2018 at the Cancer Hospital, Chinese Academy of Medical Sciences (Approval No. 23/122–3864), and the Second Affiliated Hospital of Nanjing Medical University (Approval No. 2022-KY-113-01) were included, forming a multicenter Chinese dataset. Written informed consent was obtained from all patients or, when applicable, from their legal representatives, and all personally identifiable information was anonymized.

The following variables were collected from patient records: age, sex, income, city size, number of examined and positive regional lymph nodes, primary tumor site, time from diagnosis to treatment (months), surgery, chemotherapy, surgical origin, radiotherapy, T stage, N stage, M stage, overall stage, CSS, OS, and survival status. OS and CSS were defined as the primary outcomes. Only cases with active follow-up and histologically confirmed gastric cancer were included in the analysis. Patients classified as “not applicable—non-first primary tumor” or as having a “missing/unknown cause of death (COD)” in the SEER cause-specific death classification were excluded. Missing data were addressed using multiple imputation: polynomial regression for continuous variables, multinomial logistic regression for categorical variables, and predictive mean matching for mixed-type features.

Feature selection

The SEER dataset was randomly divided into training and testing sets in a 9:1 ratio. The Chinese multicenter dataset was used as an independent external validation cohort. Feature selection and model development were performed using the training cohort only.

Three feature selection methods were employed: traditional Cox regression, Random Survival Forest (RSF), and CoxBoost. Univariate Cox regression was first applied, followed by multivariate Cox regression to identify variables significantly associated with OS and CSS ($p < 0.05$). The strength of association was quantified using hazard ratios (HRs) and 95% confidence intervals (CIs). For RSF, a variable importance ranking algorithm was implemented to determine the most prognostic features. For CoxBoost, an augmented likelihood-based boosting approach was used to select survival-associated variables, also with a significance threshold of $p < 0.05$.

Model construction

Four predictive models were developed to estimate OS and CSS: three machine learning models (Cox, RSF, and CoxBoost) and one deep learning model (DeepSurv_Cox). All models were developed using five-fold cross-validation on the training cohort.

The semiparametric Cox model estimated the log-risk as a linear combination of covariates. RSF, a non-parametric ensemble method, constructed multiple decision trees to estimate the cumulative hazard function. CoxBoost applied likelihood-based boosting to fit a Cox model, making it suitable for high-dimensional survival data. DeepSurv_Cox was a feed-forward neural network whose output layer performed Cox regression, enabling the modeling of complex non-linear relationships between variables and survival outcomes.

Model evaluation

Model performance was evaluated using the concordance index (C-index), integrated Brier score (IBS), and mean area under the receiver operating characteristic curve (AUC). The C-index measures the concordance between predicted and actual survival outcomes, with higher values indicating better predictive discrimination. IBS reflects the accuracy of survival predictions over time, with lower values indicating better performance. The mean AUC was calculated by integrating the time-dependent AUC curve over the entire follow-up period and dividing by the length of the interval.

The final integrated model was constructed by stacking the best-performing individual models identified during independent testing, using all three metrics (C-index, IBS, and mean AUC) as selection criteria. This ensemble model was compared to the traditional TNM staging system using decision curve analysis (DCA) and time-dependent receiver

operating characteristic (ROC) curves for 3-, 5-, and 10-year survival prediction.

Results

Study population and baseline characteristics

A total of 25,364 patients with gastric cancer were included in the study, comprising 21,559 from the SEER database and 3,805 from two Chinese medical centers. The study flowchart was presented in Fig. 1. Baseline demographic and clinical characteristics differed significantly between cohorts (Table 1; all $p < 0.01$), underscoring the need for model validation across diverse populations.

Feature selection

Three feature selection methods—Cox regression, RSF, and CoxBoost—were applied to identify key prognostic variables. For OS prediction, Cox, RSF, and CoxBoost selected 13, 14, and 12 variables, respectively; for CSS prediction, 11, 13, and 11 variables were selected. The selected variables were subsequently used to train three traditional machine learning algorithms (Cox, RSF, CoxBoost) and one deep learning model (DeepSurv_Cox).

This multi-method feature selection strategy ensured robust identification of prognostic factors consistently relevant across algorithms.

OS prediction

Across all three cohorts, the machine learning models consistently outperformed the deep learning approach in predicting OS. Among the tested models, RSF achieved the highest performance, particularly when combined with CoxBoost or RSF-based feature selection.

Specifically, RSF with CoxBoost-based feature selection achieved C-index values of 0.758 (training), 0.741 (test), and 0.742 (validation); corresponding IBS values were 0.120, 0.132, and 0.136, with AUCs of 0.879, 0.837, and 0.840 (Fig. 2A and I).

RSF+RSF selection produced slightly better results in training cohort (C-index=0.762, AUC=0.874), but similar performance overall (test/validation C-index=0.740/0.744, AUC=0.824/0.831). Importantly, the integrated model, combining the strengths of the best-performing algorithms and feature sets, significantly outperformed the conventional TNM staging system. The integrated model achieved training/test/validation C-indexes of 0.762/0.738/0.743, IBS of

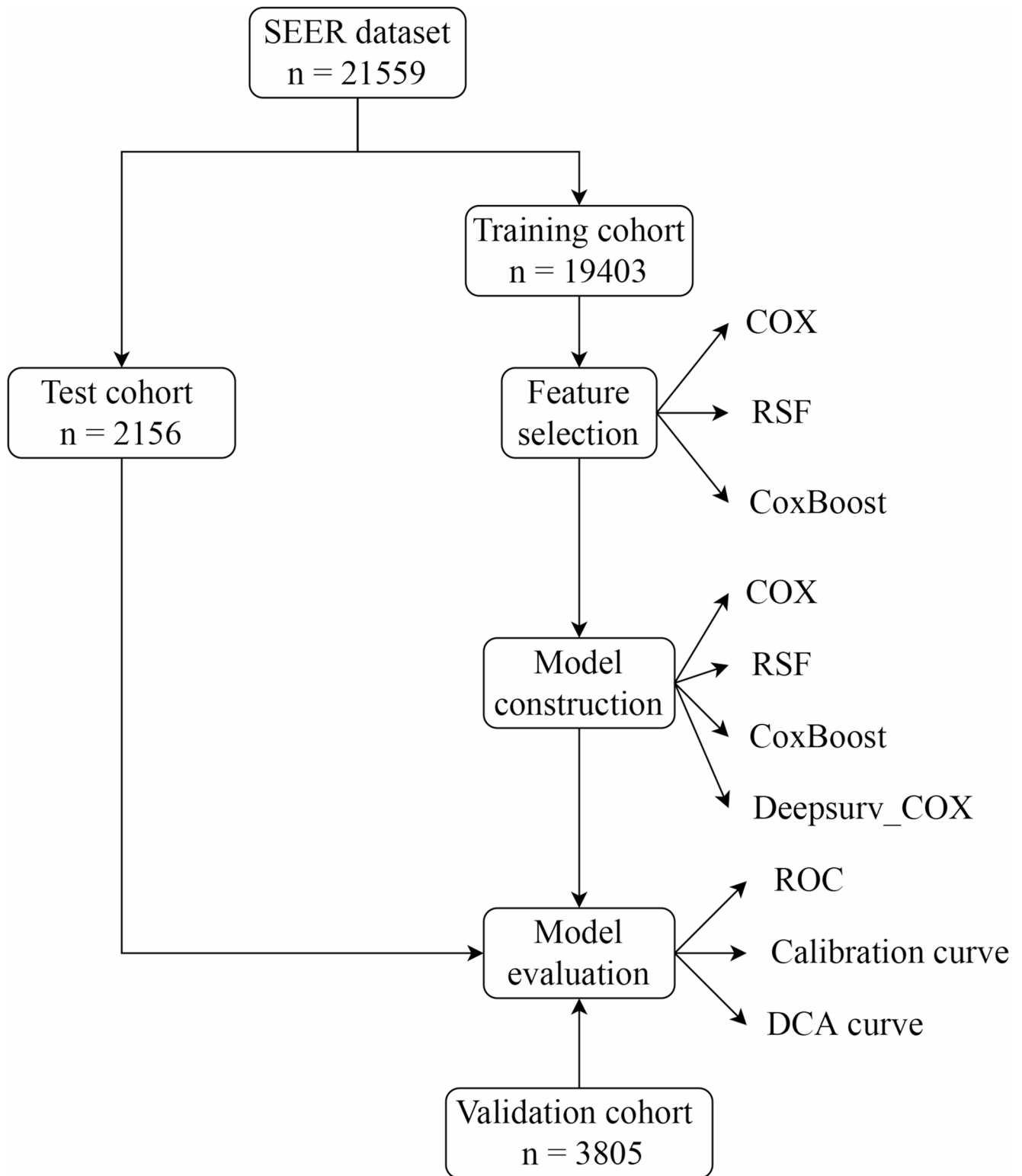


Fig. 1 Flowchart of this study

Table 1 Patient characteristics

Characteristics	SEER datasets	Chinese datasets	<i>p</i>
Age	65.93±11.886	72.44±7.958	<0.01
Gender			<0.01
Male	14,758 (68.5)	2823 (74.2)	
Female	6801 (31.5)	982 (25.8)	
Income			<0.01
<45,000	4849 (22.5)	1846 (48.5)	
45,000–60,000	8359 (38.8)	1626 (42.7)	
>60,000	8351 (38.7)	333 (8.8)	
Rural			<0.01
>1 million	2266 (10.5)	74 (2)	
0.25–1.25 million	1515 (7)	583 (15.3)	
<0.25 million	4442 (20.6)	609 (16)	
Nonmetropolitan counties	13,336 (61.9)	2539 (66.7)	
Site	3 (0, 5)	2 (0, 4)	<0.01
Examined	9 (0, 19)	12 (1, 23)	<0.01
Positive	5 (0, 98)	10 (2, 98)	<0.01
Totreatment	1 (0, 2)	1 (1, 2)	<0.01
Surgery			<0.01
Yes	15,825 (26.6)	2276 (59.8)	
No	5734 (73.4)	1529 (40.2)	
Chemotherapy			<0.01
Yes	12,567 (41.7)	2105 (55.3)	
No	8992 (58.3)	1700 (44.7)	
Radiotherapy			<0.01
Yes	5163 (23.9)	630 (16.6)	
No	16,369 (76.1)	3175 (83.4)	
OS			<0.01
Alive	5279 (24.5)	824 (21.7)	
Dead	16,280 (75.5)	2981 (78.3)	
CSS			<0.01
Alive	8961 (41.6)	1355 (35.6)	
Dead	12,598 (58.4)	2450 (64.4)	
Months	23 (8, 77)	17 (7, 58)	<0.01
SurgSite	32 (0, 50)	30 (0, 50)	<0.01
T			<0.01
T1	6369 (29.5)	749 (19.6)	
T2	8956 (41.6)	1926 (50.6)	
T3	3809 (17.7)	693 (18.2)	
T4	2425 (11.2)	441 (11.6)	
N			<0.01
N0	9635 (44.7)	1344 (35.3)	
N1	8573 (39.8)	1474 (38.7)	
N2	2477 (11.4)	703 (18.5)	
N3	874 (4.1)	284 (7.5)	
M			<0.01
M0	16,862 (78.2)	2870 (75.4)	
M1	4697 (21.8)	935 (24.6)	
Stage			<0.01
I	7847 (36.4)	1018 (26.8)	
II	4162 (19.3)	1094 (28.8)	
III	3464 (16.1)	581 (15.3)	
IV	6086 (28.2)	1112 (29.2)	

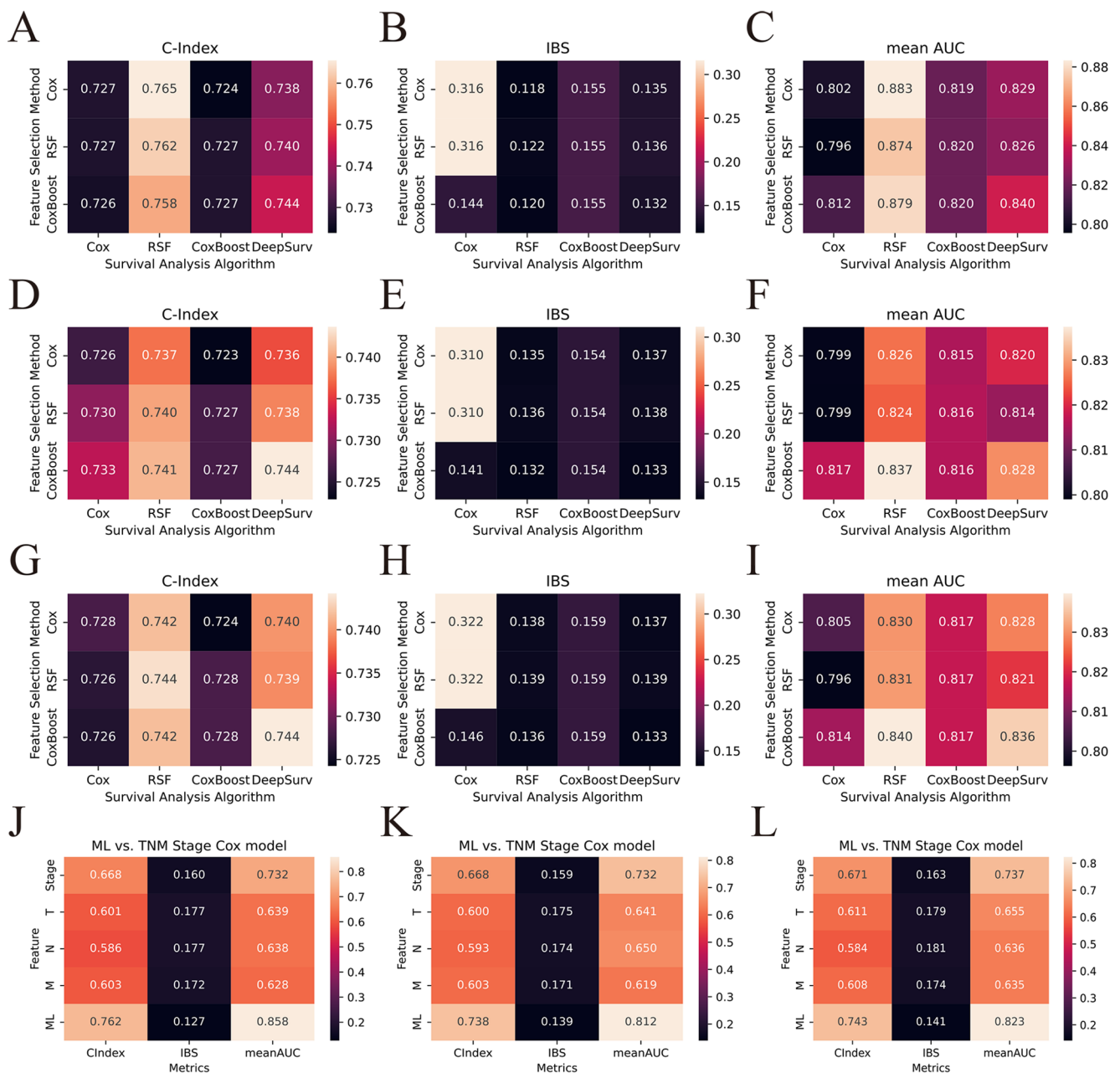


Fig. 2 Model evaluation for OS. (A-I) Evaluation of different prediction models. C-index, IBS and mean AUC on the training (A-C), test (D-F), and validation (G-I) cohorts. (J-L) The integrated model versus

traditional TNM staging model comparison. C-index, IBS, and mean AUC on the training (J), test (K), and validation (L) cohorts

0.127/0.139/0.141, and mean AUCs of 0.858/0.812/0.823 (Fig. 2J and L).

It also accurately predicted 3-, 5-, and 10-year OS across all cohorts, with AUCs consistently exceeding 0.80 (Fig. 3A and I). Calibration and DCA demonstrated good calibration and clinical applicability of the model (Figures S1 and S2).

The results showed that the RSF-based integrated model significantly improves OS prediction over traditional TNM staging, with robust generalization across populations.

CSS prediction

The prediction of CSS yielded similar trends. Machine learning models, particularly RSF, demonstrated superior performance compared to deep learning. Again, the best results were obtained with RSF models combined with either CoxBoost or RSF-based feature selection.

RSF+CoxBoost achieved C-indexes of 0.781 (training), 0.774 (test), and 0.767 (validation); IBS values of 0.130,

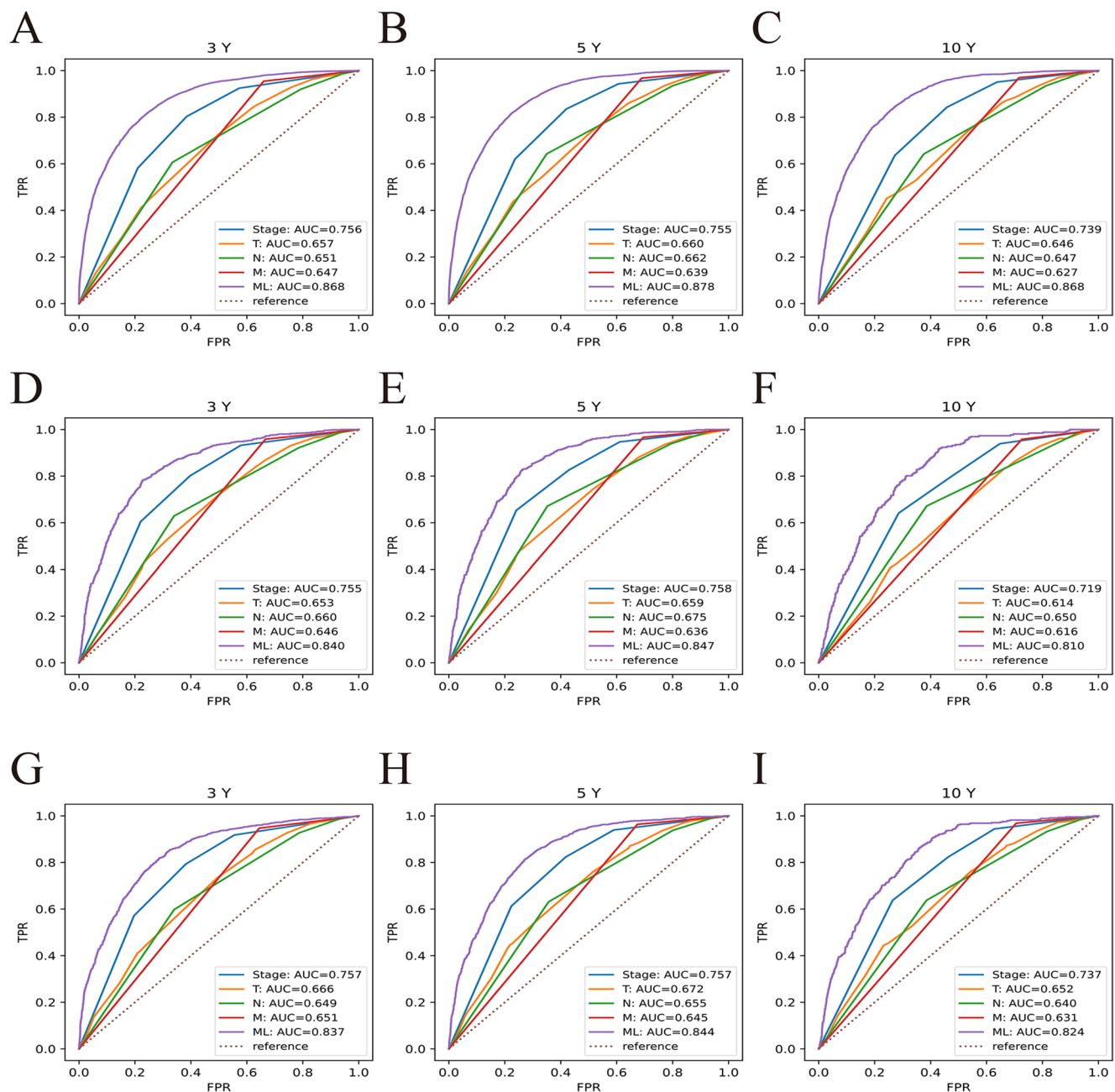


Fig. 3 Integrate model evaluation for OS. AUC of the 3-, 5-, and 10-year ROC curves of the integrated model compared with the traditional TNM staging model on the training (A-C), test (D-F), and validation (G-I) cohorts

0.157, and 0.144; and AUCs of 0.885, 0.841, and 0.856 (Fig. 4A and I).

RSF+RSF selection yielded even better C-indices of 0.789, 0.777, and 0.770, with corresponding AUCs of 0.896, 0.845, and 0.861.

The final integrated model for CSS again outperformed the TNM system, with C-index/IBS/AUC of 0.790/0.132/0.888 (training), 0.772/0.156/0.839 (test), and 0.765/0.148/0.854

(validation) (Fig. 4J and L). Its predictive ability for 3-, 5-, and 10-year CSS remained stable across cohorts, with AUCs consistently > 0.85 (Fig. 5A and I). Calibration and DCA curves confirmed its accuracy and decision-making benefit (Figure S3, S4).

The results showed that the integrated model provides highly accurate, generalizable CSS predictions, substantially outperforming the TNM staging system.

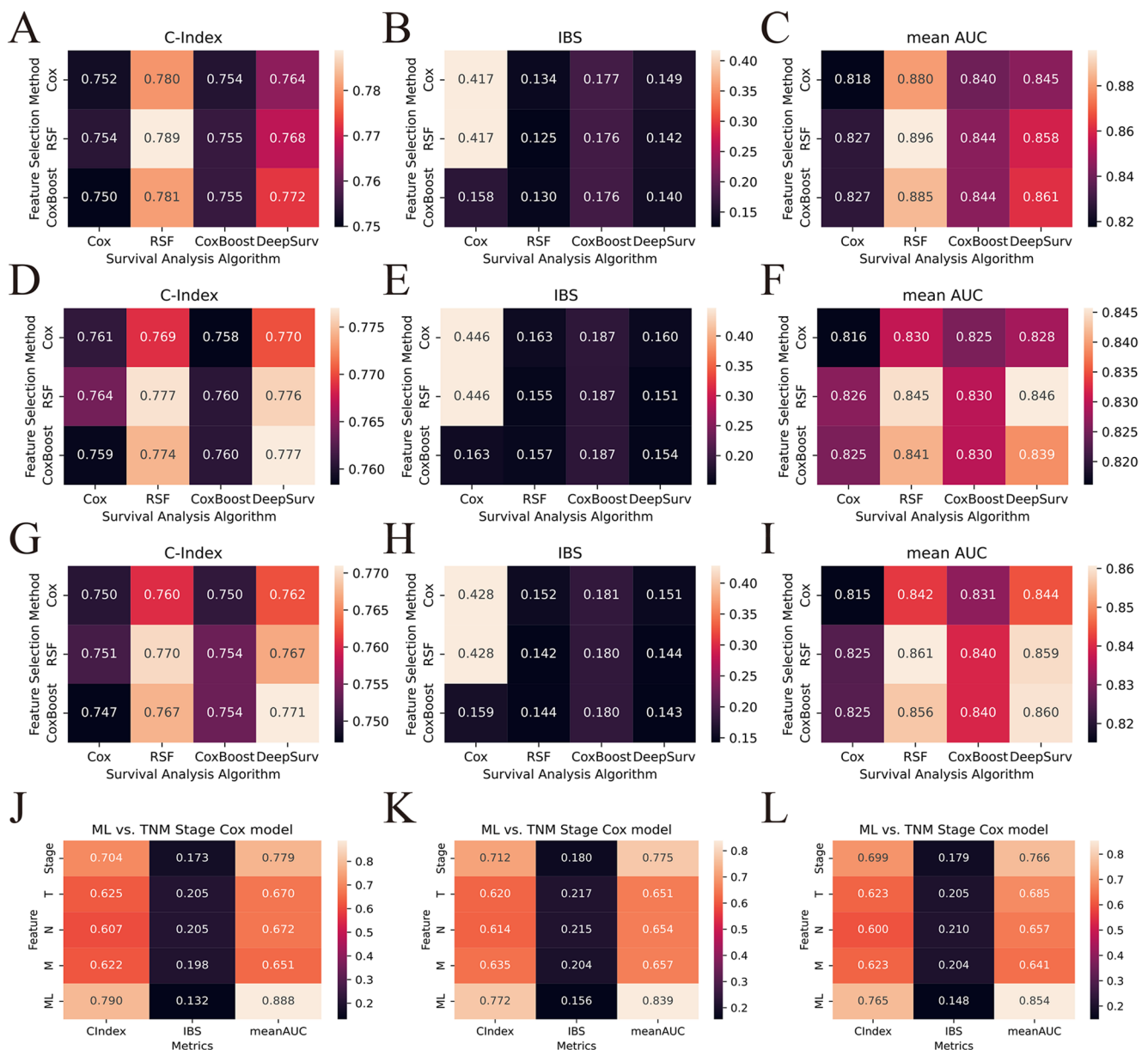


Fig. 4 Model evaluation for CSS. C-index, IBS and mean AUC on the training (A-C), test (D-F), and validation (G-I) cohorts. (J-L) The integrated model versus traditional TNM staging model comparison. C-index, IBS, and mean AUC on the training (J), test (K), and validation (L) cohorts

Discussion

Accurate prediction of OS and CSS in patients with gastric cancer is essential for guiding treatment strategies, planning follow-up, and facilitating informed patient counseling. Previous studies have identified a range of prognostic factors—including age, histological subtype, tumor grade and size, and the presence of metastases—as being significantly associated with survival outcomes in gastric cancer. However, traditional Cox proportional hazards models are inherently limited by their assumption of linear relationships, which may not hold in high-dimensional clinical

datasets. In this context, machine learning and deep learning approaches have emerged as powerful tools for survival analysis due to their ability to capture complex and non-linear relationships within the data [12]. These techniques have shown promising results in analyzing clinical, imaging, and genomic data across various cancer types. Despite their potential, the application of machine learning and deep learning to survival prediction in gastric cancer remains limited. To our knowledge, no comprehensive studies have directly compared the performance of machine learning and deep learning algorithms in predicting OS and CSS in this patient population. Moreover, determining the most suitable

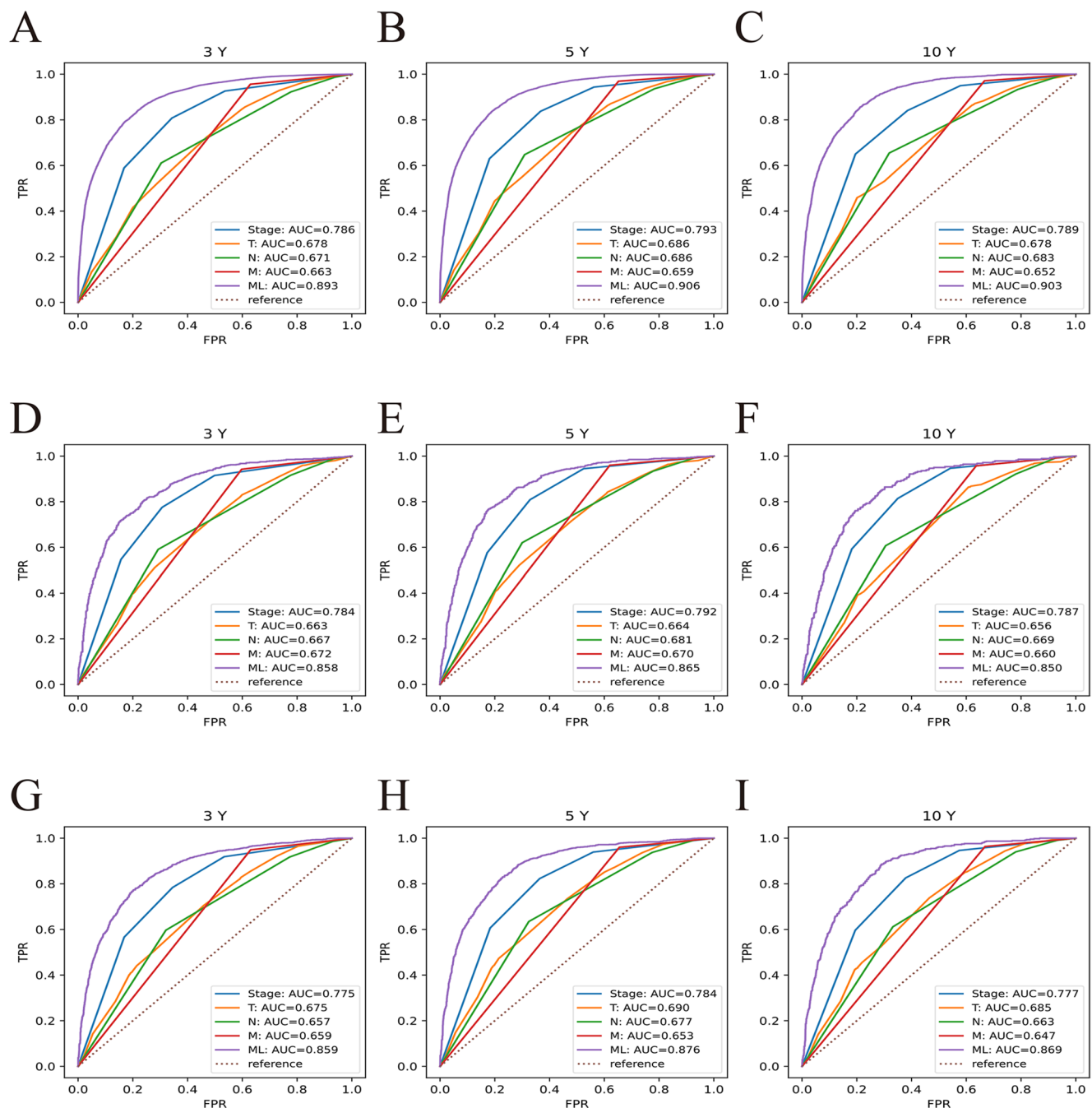


Fig. 5 Model evaluation for CSS. (A-C) AUC of the 3-, 5-, and 10-year ROC curves of the integrated model compared with the traditional TNM staging model on the training (A-C), test (D-F), and validation (G-I) cohorts

approach for survival estimation continues to be a key challenge in clinical oncology. In light of this, we developed and validated multiple survival prediction models using both machine learning and deep learning algorithms, based on large-scale population data from the SEER database and real-world clinical data from two Chinese medical centers. This study aimed to identify the most accurate modeling approach for prognostic assessment and to provide

clinicians with a reliable tool for the personalized management of gastric cancer.

In this study, we developed and validated a machine learning-based survival prediction framework that significantly outperforms traditional prognostic models for gastric cancer. By analyzing large, multi-institutional cohorts from both Western (SEER) and Eastern (Chinese centers) populations, we ensured the generalizability and robustness

of our findings across diverse clinical settings. Our results consistently showed that ensemble-based machine learning models, particularly the RSF algorithm combined with CoxBoost or RSF-based feature selection, achieved superior performance in predicting both OS and CSS compared to conventional TNM staging and deep learning models such as DeepSurv_Cox. This advantage was evident in higher C-index values, better calibration, improved AUCs for 3-, 5-, and 10-year survival, and greater clinical utility, as demonstrated by DCA. Notably, deep learning models did not demonstrate a consistent performance advantage, suggesting that in structured clinical datasets with a moderate number of variables, traditional machine learning models may be more effective and interpretable. We acknowledge that our study did not include direct head-to-head comparisons with previously published machine learning algorithms for survival prediction in gastric cancer. Most existing models have been developed on single-center or region-specific datasets, often with different inclusion criteria, predictor sets, and outcome definitions, which makes exact reproduction and comparison challenging. Our primary objective was to perform a systematic comparison of multiple machine learning and deep learning algorithms within a unified multicenter framework, using harmonized variables and standardized evaluation metrics across training, internal testing, and external validation cohorts. While this design ensured methodological rigor and generalizability, it limited our ability to directly benchmark against existing published algorithms. In future work, we plan to reproduce selected published models using our multicenter dataset to enable a more comprehensive and fair performance comparison in a uniform clinical setting. Our feature selection strategy, combining three distinct methods, not only reinforced the stability and relevance of the prognostic factors identified but also contributed to reducing model overfitting and enhancing interpretability.

Unlike TNM staging, which is limited to anatomical criteria, our model incorporates a broader range of patient-specific factors, enabling personalized risk stratification. This allows patients to be classified more accurately into low- or high-risk groups, enabling oncologists to tailor treatment intensity and follow-up schedules accordingly. For instance, patients predicted to have poor long-term survival may benefit from more aggressive adjuvant therapy or closer surveillance, whereas those with favorable predictions may avoid unnecessary treatment-related toxicity. Furthermore, the model's ability to predict 3-, 5-, and 10-year survival offers valuable information during patient counseling and shared decision-making, helping clinicians to set realistic expectations and guide individualized care plans. Importantly, the model maintained high performance across both Western and Eastern populations, supporting its potential use in diverse healthcare settings. Beyond its

statistical performance, the model provides practical value by enhancing clinical decision-making, improving resource allocation, and advancing precision oncology in gastric cancer management. While gastric cancer demonstrated well documented ethnic disparities, including higher prevalence of genomically stable subtypes in Asian populations and divergent treatment paradigms, our model validation strategy intentionally leveraged these differences to stress test model generalizability [13, 14]. The maintained AUC (0.861 in internal validation) suggested successful capture of universal prognostic factors despite population variability, while avoiding overfitting risks inherent in single dataset studies. This approach found precedent in the MAGIC trial's subgroup analyses, where neoadjuvant chemotherapy benefited persisted across ethnic groups despite differential baseline risks, and aligned with current NCCN guidelines that increasingly incorporate Asian trial data alongside Western studies [15, 16]. Despite these promising results, several limitations should be acknowledged. These include the retrospective nature of the study and the lack of treatment-specific variables and molecular data, which may further improve prognostic precision in future studies. Nevertheless, our findings highlight the transformative potential of machine learning approaches in prognostic modeling for gastric cancer and provide a solid foundation for future prospective validation and clinical implementation.

Compared with previous studies on survival prediction in gastric cancer patients, the present study demonstrates notable advantages in terms of flexibility and generalizability. Prior research has suggested that increased lymph node dissection and radiotherapy may contribute to improved survival in gastric cancer patients [17]. Additionally, numerous prognostic factors—such as gender, age, smoking and alcohol consumption, histologic subtype, and TNM stage—have been associated with survival outcomes, supporting the clinical feasibility and necessity of building prognostic models based on multifactorial clinical data [18, 19]. For instance, Wang et al. constructed a nomogram to predict OS using data from SEER patients diagnosed between 2014 and 2015, yielding a C-index of 0.707 in the test cohort [20]. Similarly, Chen et al. proposed an online prognostic interface based on a nomogram with a validation cohort C-index of only 0.569 [21]. Although these models achieved acceptable performance, they share a key limitation: they rely on traditional Cox regression, which assumes that the risk of death is a linear function of covariates—an assumption that often does not hold in real-world clinical settings.

This limitation was also evident in our study, where both the RSF and DeepSurv_COX models outperformed the conventional Cox model, underscoring the superior ability of machine learning and deep learning algorithms to capture complex, nonlinear interactions among variables.

Moreover, although Zeng *et al.* applied a deep learning approach to predict survival in gastric cancer patients [22], and Liu *et al.* constructed a nomogram with a calibration index of 0.726 [23], these studies were restricted to single-center data and lacked external validation. In contrast, our study incorporated external validation from two independent clinical centers, enhancing the robustness and generalizability of the proposed models. When compared with traditional Cox regression and TNM staging systems, our machine learning-based models demonstrated significantly improved predictive performance. These findings suggest that individualized survival prediction for gastric cancer patients can be substantially enhanced through the application of advanced machine learning methodologies.

Despite these promising results, several limitations should be acknowledged. First, the number of patient characteristics available in the SEER database was limited, particularly for predicting overall survival. Although the Chinese hospital cohorts contained more detailed clinical information, many of these additional variables—such as laboratory data, performance status, comorbidities, and molecular biomarkers—were not available in the SEER dataset. For consistency across training and validation cohorts, we therefore restricted model inputs to variables that were recorded in both datasets. This strategy was adopted to ensure that model development and external validation were performed on the same set of predictors, thereby avoiding bias due to heterogeneous feature availability. We recognize that incorporating more comprehensive prognostic variables could further improve model performance, and in future studies, we plan to develop and validate models using extended datasets from Chinese centers, incorporating additional clinical, treatment-related, and molecular factors.

Conclusion

In summary, this study systematically evaluated and compared the performance of three machine learning algorithms and one deep learning algorithm in predicting OS and CSS in patients with gastric cancer. Machine learning models, particularly RSF and CoxBoost, demonstrated superior discrimination, calibration, and robustness across training, testing, and external validation cohorts. These findings suggest that machine learning-based models have significant potential as reliable and practical prognostic tools, enabling clinicians to make more informed and individualized treatment decisions for patients with gastric cancer.

Acknowledgements We were very grateful for the data provided by the SEER databases.

Author contributions Xiao Guan and Jinsong Liu designed the study and was involved in database search and statistical analyses. Lei Xu was involved in the collection and organization of clinical data. Xiao Guan wrote the manuscript. Chengfeng Wang was responsible for the submission of the final version of the paper. All authors approved the final version. All authors agreed to be accountable for all aspects of the work.

Funding This work was supported by the National Natural Science Foundation of China (81972314 and 81802463), CAMS Innovation Fund for Medical Sciences (CIFMS), 2022-I2M-1-010.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

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/>.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 71(3):209–249
2. Kim YJ, Chung WC, Cho IH, Kim J, Kim S (2019) Prognostic effect of different etiologies in patients with gastric cardia cancer. *Medicine (Baltimore)* 98(50):e18397
3. Yamada S, Hatta W, Shimosegawa T, Takizawa K, Oyama T, Kawata N, Takahashi A, Oka S, Hoteya S, Nakagawa M, Hirano M, Esaki M, Matsuda M, Nakaya N, Gotoda T (2019) Different risk factors between early and late cancer recurrences in patients without additional surgery after noncurative endoscopic submucosal dissection for early gastric cancer. *Gastrointest Endosc* 89(5):950–960
4. Eusebi LH, Telese A, Marasco G, Bazzoli F, Zagari RM (2020) Gastric cancer prevention strategies: a global perspective. *J Gastroenterol Hepatol* 35(9):1495–1502
5. Jiang Y, Jin C, Yu H, Wu J, Chen C, Yuan Q, Huang W, Hu Y, Xu Y, Zhou Z, Fisher GA, Li J, Li G (2021) Development and validation of a deep learning CT signature to predict survival and chemotherapy benefit in gastric cancer: A Multicenter, retrospective study. *Ann Surg* 274(6):e1153–e1161
6. Tu JB, Liao WJ, Long SP, Li MP, Gao XH (2024) Construction and validation of a machine learning model for the diagnosis of

- juvenile idiopathic arthritis based on fecal microbiota. *Front Cell Infect Microbiol* 14:1371371
7. Hashimoto DA, Rosman G, Rus D, Meireles OR (2018) Artificial intelligence in surgery: promises and perils. *Ann Surg* 268(1):70–76
 8. Lee C, Light A, Alaa A, Thurtle D, van der Schaar M, Gnanaprasam VJ (2021) Application of a novel machine learning framework for predicting non-metastatic prostate cancer-specific mortality in men using the Surveillance, Epidemiology, and end results (SEER) database. *Lancet Digit Health* 3(3):e158–e165
 9. Kuntz S, Kriehoff-Henning E, Kather JN, Jutzi T, Höhn J, Kiehl L, Hekler A, Alwers E, von Kalle C, Fröhling S, Utikal JS, Brenner H, Hoffmeister M, Brinker TJ (2021) Gastrointestinal cancer classification and prognostication from histology using deep learning: systematic review. *Eur J Cancer* 155:200–215
 10. Daneshjou R, He B, Ouyang D, Zou JY (2021) How to evaluate deep learning for cancer diagnostics - factors and recommendations. *Biochimica et Biophysica Acta (BBA)* 1875(2):188515
 11. Issa NT, Stathias V, Schürer S, Dakshanamurthy S (2021) Machine and deep learning approaches for cancer drug repurposing. *Semin Cancer Biol* 68:132–142
 12. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI (2015) Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 13:8–17
 13. Comprehensive molecular characterization of gastric adenocarcinoma (2014) *Nature* 513(7517):202–209
 14. Li P, Huang CM, Zheng CH, Russo A, Kasbekar P, Brennan MF, Coit DG, Strong VE (2018) Comparison of gastric cancer survival after R0 resection in the US and China. *J Surg Oncol* 118(6):975–982
 15. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, Scarffe JH, Lofts FJ, Falk SJ, Iveson TJ, Smith DB, Langley RE, Verma M, Weeden S, Chua YJ, Participants MT (2006) Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med* 355(1):11–20
 16. Hodan R, Gupta S, Weiss JM, Axell L, Burke CA, Chen LM, Chung DC, Clayback KM, Felder S, Foda Z, Giardiello FM, Grady W, Gustafson S, Hagemann A, Hall MJ, Hampel H, Idos G, Joseph N, Kassem N, Katona B, Kelly K, Kieber-Emmons A, Kupfer S, Lang K, Llor X, Markowitz AJ, Prats MM, Niell-Swiler M, Outlaw D, Pirzadeh-Miller S, Samadder NJ, Shibata D, Stanich PP, Swanson BJ, Szymaniak BM, Welborn J, Wiesner GL, Yurgelun MB, Dwyer M, Darlow S, Diwan Z (2024) Genetic/Familial High-Risk assessment: Colorectal, Endometrial, and Gastric, version 3.2024, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 22(10):695–711
 17. Ze-Long Y, Guo-Hui M, Lin Z, Wei-Hong Y, Ke-Cheng Z, Yan-Wen J (2019) Survival trends of patients with surgically resected gastric cardia cancer from 1988 to 2015: a population-based study in the United States. *Am J Clin Oncol* 42(7):581–587
 18. Gu J, Xie S, Wang S, Xue L, Zhou J, Li M, Fan Z, Chen R, Middleton DRS, Hao C, Wang J, Li B, Li X, Wei W (2021) Surveillance of premalignant gastric cardia lesions: a population-based prospective cohort study in China. *Int J Cancer* 149(9):1639–1648
 19. Zhai Z, Zhu ZY, Cong XL, Han BL, Gao JL, Yin X, Zhang Y, Lou SH, Fang TY, Wang YM, Li CF, Yu XF, Ma Y, Xue YW (2020) Changing trends of clinicopathologic features and survival duration after surgery for gastric cancer in Northeast China. *World J Gastrointest Oncol* 12(10):1119–1132
 20. Wang CY, Yang J, Zi H, Zheng ZL, Li BH, Wang Y, Ge Z, Jian GX, Lyu J, Li XD, Ren XQ (2020) Nomogram for predicting the survival of gastric adenocarcinoma patients who receive surgery and chemotherapy. *BMC Cancer* 20(1):10
 21. Chen K, Deng X, Yang Z, Yu D, Zhang X, Zhang J, Xie D, He Z, Cheng D (2020) Survival nomogram for patients with metastatic Siewert type II adenocarcinoma of the esophagogastric junction: a population-based study. *Expert Rev Gastroenterol Hepatol* 14(8):757–764
 22. Zeng J, Li K, Cao F, Zheng Y (2023) Development and validation of survival prediction model for gastric adenocarcinoma patients using deep learning: a SEER-based study. *Front Oncol* 13:1131859
 23. Liu X, Jiang Q, Yue C, Wang Q (2021) Clinicopathological characteristics and survival predictions for adenocarcinoma of the esophagogastric junction: a SEER population-based retrospective study. *Int J Gen Med* 14:10303–10314

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