Discover Oncology

Analysis

Construction and interpretation of weight-balanced enhanced machine learning models for predicting liver metastasis risk in colorectal cancer patients

Qunzhe Ding[1] · Chenyang Li[2,3,4] · Chendong Wang[2,3,4] · Qunzhe Ding[1]

Received: 8 November 2024 / Accepted: 3 February 2025
Published online: 12 February 2025

Abstract

Background Colorectal cancer (CRC) is a major contributor to cancer-related mortality, with liver metastases developing in approximately 25% of affected individuals. The presence of liver metastasis significantly deteriorates the prognosis for patients. The objective of this study is to predict liver metastasis in CRC patients by developing machine learning (ML)-based models, thereby aiding clinicians in the decision-making process for appropriate interventions.

Methods Retrospective analysis was performed using the Surveillance, Epidemiology, and End Results (SEER) database, and cases with CRC from 2010 to 2015 were extracted to the downstream analysis. Logistic regression (LR), Random Forest (RF), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), and LightGBM are applied to develop machine learning (ML) models to predict liver metastasis of CRC patient. To optimize the models, an improved weight-balancing algorithm was employed, enhancing the performance of the classifiers. The six models were tenfold cross-validated, and the optimal model was selected based on a combination of performance metrics. Shapley additive explanation (SHAP) was utilized to interpret the best-performing ML models globally, locally, and interactively. To ensure the model's reliability and generalizability, an external validation cohort of CRC cases from 2018 to 2021, obtained from a separate SEER database, was used for external evaluation.

Results In total, 50,062 patients with CRC were included in the analysis, with 5604 patients occurring liver metastasis. Among the six models evaluated, the CatBoost model showed excellent performance with the highest AUC of 0.8844. Moreover, the CatBoost model also outperformed the others in terms of recall (0.8060) and F1-score (0.6736). SHAP-based summary and force plots were used to interpret the CatBoost model. The interpretability analysis revealed that elevated carcinoembryonic antigen (CEA) levels, systemic therapy, N and T stages, and chemotherapy performed were the most significant indicators for predicting liver metastasis according to the optimal model. Furthermore, systemic therapy was suggested to increase liver metastasis risk in N0 stage patients, while it appeared to be beneficial in patients with lymph node metastasis. Preoperative radiation therapy was found to be more effective than postoperative radiation therapy.

Qunzhe Ding, Chenyang Li and Chendong Wang have contributed equally to this work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-01871-2.

✉ Qunzhe Ding, dingqunzhe1@163.com | [1]School of Information Management, Wuhan University, Wuhan, Hubei 430072, People's Republic of China. [2]Hepatic Surgery Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. [3]Clinical Medical Research Center of Hepatic Surgery at Hubei Province, Wuhan, Hubei, China. [4]Hubei Key Laboratory of Hepato-Pancreatic-Biliary Diseases, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China.

Validation using an external cohort of CRC cases from 2018 to 2021 further confirmed the robustness and stability of the CatBoost model, as its overall performance remained consistent with the internal validation results.

Conclusion Elevated levels of carcinoembryonic antigen (CEA) have been identified as a crucial clinical predictor for liver metastasis in CRC patients. Furthermore, the administration of systemic therapy to patients who do not exhibit lymph node involvement has been found to increase the risk of liver metastasis. In terms of radiation therapy, preoperative radiation appears to be more efficacious in controlling the risk of liver metastasis compared to postoperative radiation. This finding underscores the importance of optimizing treatment strategies based on the specific clinical context and patient characteristics.

Keywords Colorectal cancer · Liver metastasis · Machine learning · SHAP · SEER

1 Introduction

Colorectal cancer (CRC) is the third most common cancer globally and the second leading cause of cancer-related deaths, resulting in over 935,000 fatalities annually, according to the Global Cancer Statistics 2020 [1]. Despite significant advances in CRC diagnosis and treatment in recent years, the five-year survival rate remains relatively low. This is largely because most patients are diagnosed at an advanced stage or with distant metastases. Studies indicated that approximately 25% of CRC patients have developed metastases by the time of their initial diagnosis [2]. Over the course of the disease, up to 50% of patients will experience distant metastases, with liver metastasis being the most prevalent [3]. Even after radical resection, there remains a 30% chance of developing liver metastases [4]. For patients with untreated liver metastases, the median survival is about 6 months [5]. In contrast, surgical resection of liver metastases or achieving no evidence of disease (NED) can extend median survival to 39 months and increase the five-year survival rate to 37.1%, according to a recent multicenter study [6].

Therefore, identifying individuals at higher risk for liver metastasis is essential for timely intervention and improved overall prognosis accurately. Contrast-enhanced MRI and contrast-enhanced ultrasound are standard techniques for screening CRC patients for distant metastases [7]. However, due to the high expense and invasiveness associated with enhanced imaging, it is not advisable to use these techniques for colorectal liver metastasis screening in all CRC patients.

Compared to traditional methods, Machine Learning (ML) has emerged as a formidable predictive tool, largely attributed to its proficiency in managing nonlinear data types. Nevertheless, the limited interpretability of ML models has been a significant barrier to their widespread adoption in clinical practice, where transparency and accountability are of paramount importance. Recent strides in interpretability techniques, particularly Shapley Additive Explanations (SHAP), have substantially enhanced the transparency of ML models. This advancement has paved the way for their more extensive application in clinical contexts. SHAP and similar techniques do not merely augment the trustworthiness of ML models; they also empower clinicians and researchers with a deeper comprehension of the underlying mechanisms that shape predictions.

By elucidating the contributions of individual features and their interactions, these interpretability methods facilitate a more nuanced understanding of the model's decision-making process. This, in turn, enables healthcare professionals to make more informed, evidence-based decisions, thereby bridging the gap between the complexity of ML algorithms and the practical needs of clinical decision-making. As a result, the integration of interpretable ML models into clinical workflows becomes more feasible, promising to improve patient care through data-driven insights.

2 Methods

2.1 Data extraction/study population and preprocessing

The dataset employed in this study was derived from the "SEER Research Plus Data, 17 Registries, Nov 2021 Sub (2000–2019)," which was generated utilizing the SEER*Stat software (version 8.4.0.1; National Cancer Institute, USA). The criteria for patient inclusion in the study were stringent and specifically defined as follows: 1. Patients were diagnosed with colorectal cancer, a diagnosis that was confirmed through histopathological examination, in strict adherence to the guidelines set forth by the International Classification of Diseases for Oncology, Third Edition (ICDO-3) and the World Health Organization (WHO) 2008 standards. 2. For each patient included in the study, colorectal cancer was the

sole primary malignancy diagnosed. This criterion ensured that the study cohort was homogenous with respect to the primary cancer type, thereby minimizing potential confounding factors related to multiple primary cancers.

Two types of characteristics were collected from the SEER database: demographic and pathological. The demographic features include age at diagnosis, sex, race, rural–urban continuum, and marital status. The pathological features encompass liver, bone, and brain metastasis, T and N stage, grade, perineural invasion recode, tumor size, carcinoembryonic antigen (CEA) pretreatment interpretation recode, tumor deposits recode, chemotherapy recode, surgery type, surgery and radiation sequence, surgery and systemic therapy sequence, as well as the number and ratio of positive regional nodes. These characteristics were analyzed to understand their impact on the outcomes of colorectal cancer liver metastasis. By examining both patient demographics and disease-specific factors, the analysis provides valuable insights into the interplay between these elements in determining patient prognosis.

The exclusion criteria used for this analysis are illustrated in Fig. 1. As part of preparing the dataset, we also implemented several consolidation strategies to streamline certain categories and address missing values. Given that missing data accounted for only a small proportion of the dataset, imputation techniques were intentionally omitted to preserve data integrity and minimize the risk of introducing biases. Instead, missing values, along with some uncertain categories, were consolidated into broader, predefined categories to ensure consistency and facilitate interpretation. This approach not only reduced sparsity but also maintained the robustness and reliability of subsequent analyses.

For instance, in the "Race" category, "American Indian/Alaska Native" and "Unknown" were merged into "Others" to reduce sparsity. Similarly, for the "Rural–Urban Continuum," records marked as "Unknown," "Missing," and "No Match" were grouped into "Others" to ensure consistency. Additionally, for surgery type, "RX Summ–Surg Prim Site (1998 +)" was utilized

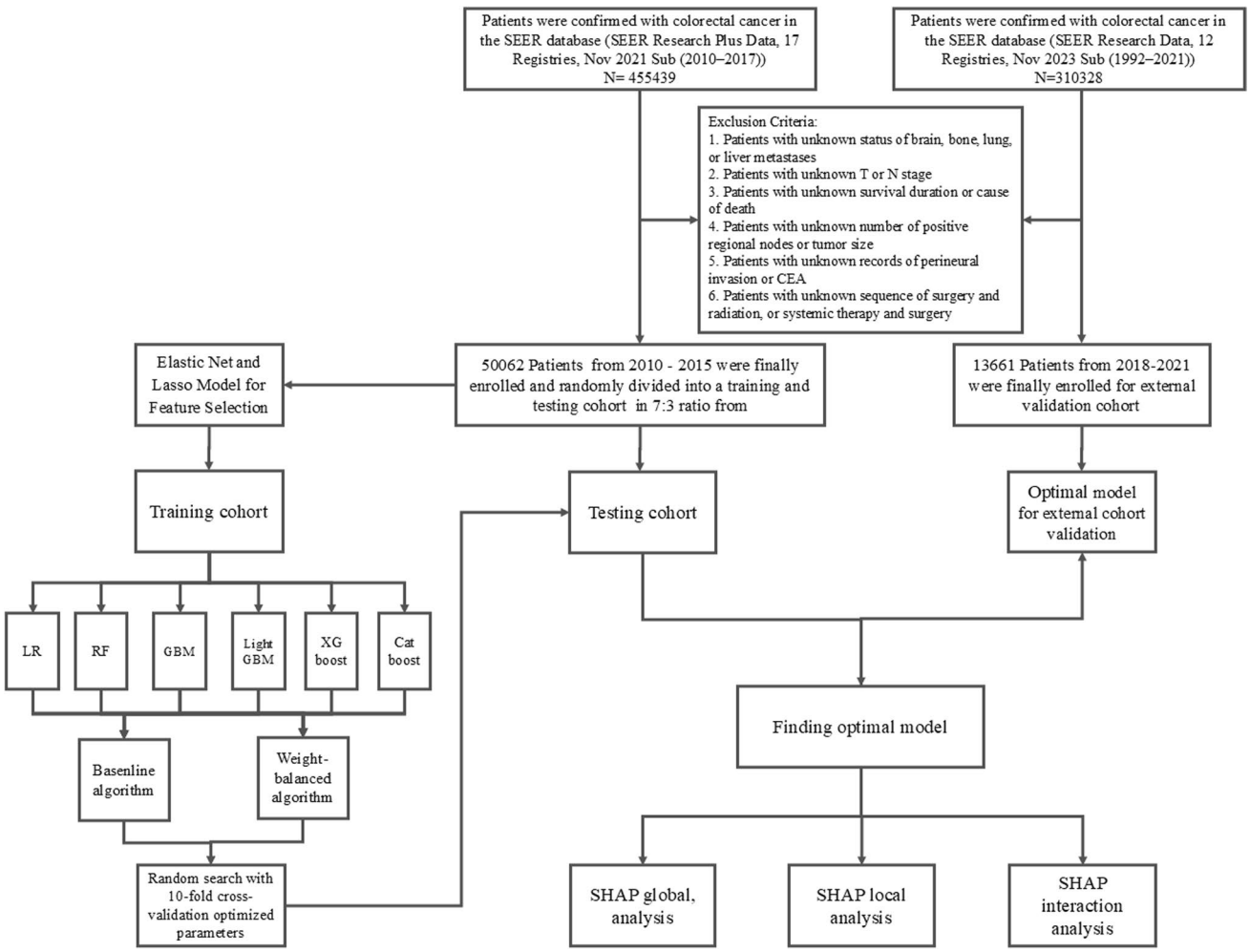

Fig. 1 Flowchart of research framework

to consolidate all surgery codes into a single category labeled "Surgery performed," while entries with the code of 0 were categorized as "No Surgery." These consolidations were necessary to facilitate cleaner analysis and interpretation of the data.

The overall research framework including data preprocessing procedure is illustrated in Fig. 1.

2.2 Statistical analyses

For predictive modeling in this study, the scikit-learn Python library, a widely recognized and robust toolkit for machine learning, was utilized in conjunction with various model packages to train the predictive models. Prior to the model-building phase, categorical variables were manually encoded to ensure they were represented in a format that could be effectively processed by machine learning algorithms. This manual encoding process was carefully performed to preserve the integrity and interpretability of the data, minimizing potential biases introduced by automated encoding methods. Subsequently, the dataset was subjected to a random division process. Specifically, 70% of the patients were strategically allocated to the training cohort, which served as the foundation for the models to learn the underlying patterns and relationships within the data. The remaining 30% of the patients were assigned to the testing cohort, which was reserved for evaluating the performance and generalizability of the trained models. This division ratio is a commonly adopted approach in machine learning, as it strikes a balance between providing sufficient data for model training and retaining a substantial portion for unbiased model assessment.

To ensure randomness and reproducibility during dataset splitting, stratified sampling was employed by specifying "stratify = y" in the function used to divide the data into training and testing cohorts. This approach maintained the original proportion of positive and negative classes (i.e., presence or absence of lung metastasis) within each subset, minimizing sampling bias and ensuring that both cohorts accurately reflected the overall class distribution in the dataset. By effectively addressing class imbalance, this method enhanced the reliability and robustness of the experimental results.

Feature selection was performed using two approaches—Lasso and ElasticNet regression—to identify the most relevant features for model development and eliminate those with limited predictive value. This study employed a range of predictive models, spanning from classical regression algorithms to advance nonlinear machine learning techniques, to comprehensively evaluate predictive performance and enhance the credibility of the results. Logistic Regression (LR), a classical linear model, was included as a baseline due to its simplicity, efficiency, and interpretability, making it a widely used benchmark for assessing the performance of more complex models. Random Forest (RF), an ensemble learning algorithm based on decision trees, improves robustness and generalization by aggregating multiple decision trees, effectively mitigating overfitting. Gradient Boosting Machines (GBM), LightGBM, and XGBoost utilize gradient boosting frameworks to iteratively minimize prediction errors, with LightGBM optimized for high-dimensional and sparse data, and XGBoost offering advanced regularization and parallel computing capabilities for enhanced performance. CatBoost was selected as a sophisticated gradient boosting technique designed to handle categorical features and deliver high performance [8].

During the model training stage, grid search and ten-fold cross-validation were employed to optimize hyperparameters and evaluate performance. Specifically, the dataset was divided into ten equally sized subsets. In each iteration, nine subsets were used as the training set, and the remaining one served as the validation set. This process was repeated ten times, cycling through all subsets. Within this framework, grid search systematically explored combinations of hyperparameters under the ten-fold cross-validation setup to identify the optimal configuration, ensuring the best model performance on the validation set. This approach effectively minimized overfitting risks while improving the robustness and generalizability of the performance evaluation.

The predictive performance of the six models was evaluated using ROC-AUC curves, precision-recall curves, and calibration plots. In addition, their precision, recall, and F1 scores were assessed. The dataset used in this study was highly imbalanced, with a total of 50,062 colorectal cancer patients, of whom only 5604 had liver metastases. If not properly addressed, this imbalance could lead the models to favor the majority class and neglect accurate classification of the minority class, which, in this study, is the subset of patients with liver metastases—our primary research focus. To address this issue, the study implemented a cost-sensitive learning approach by applying a weighted cross-entropy loss function to optimize the models. The formula for the proposed model improvement method is:

$$L_{weighted}(y,p) = -\frac{N_0}{N_1} \cdot ylog(p) - w_0 \cdot (1-y)log(1-p)$$

where y represents the actual class of the sample, and p is the predicted probability of the sample belonging to the positive class., the terms w_0 and $\frac{N_0}{N_1}$ are the weights assigned to the negative and positive classes, respectively. This weighted

loss function is designed to address class imbalance by assigning a higher weight to the minority class (liver metastasis class), and a lower weight to the majority class (negative class). The weight $\frac{N_0}{N_1}$ scales the contribution of the positive class in proportion to the imbalance ratio, ensuring that the model gives more attention to the minority class during training.

During training process, the higher weight for minority class samples increases their contribution to the total loss, resulting in larger gradients and more significant adjustments to reduce misclassification of these samples. Conversely, the lower weight for majority class samples decreases their penalty for misclassification, preventing the model from overfitting to the majority class and ignoring the minority class. Eventually, this weighting mechanism enhances the model's sensitivity to all classes, improving its ability to learn from imbalanced data and boosting performance in practical applications.

SHAP is a versatile method used for both individual predictions and global interpretation of machine learning models [9]. Rooted in game theory, SHAP applies Shapley values to optimally distribute credit among features, quantifying their impact on model predictions. SHAP force plots offer a clear and intuitive visualization of how individual features influence a specific prediction, facilitating a deeper understanding of the model's decision-making process. For global interpretation, SHAP not only highlights the importance of each feature but also elucidates their interactions and relationships with the model's output. Furthermore, SHAP ensures that predictions are fairly distributed across different feature values, thereby enhancing the model's interpretability and reliability. In this study, SHAP feature importance assessment was employed for the global interpretation of the optimal model, illustrating how individual predictions can be explained both locally and interactively.

3 Results

3.1 Patient basic characteristics

A total of 50,062 CRC patients from the SEER database were included in this study. The baseline characteristics of these patients are presented in Table 1. Among the patients, 24,115 (48.2%) were female, and the majority (77.6%) were white. A total of 5604 patients developed liver metastases, representing 11.3% of the overall cohort. Additional parameters, including marital status, rural–urban continuum, TNM stage, tumor size, CEA level, perineural invasion, tumor deposits, positive regional lymph nodes, lymph node positivity rate, and occurrences of lung, bone, and brain metastases, as well as therapeutic strategies, were also collected for further analysis. With the exception of the rural–urban continuum, all other demographic and clinical features showed significant differences between the liver metastasis group and the no liver metastasis group (Table 1).

3.2 Features selection in models

In this study, Lasso and ElasticNet regression techniques were employed as feature selection algorithms to identify the most significant variables for constructing the predictive model. Notably, as depict in Fig. 2, the outcomes of these two methods demonstrated a striking degree of similarity.

This convergence of results underscored the stability and validity of the features that were ultimately selected for further analysis. It is noteworthy that certain variables, namely Rural–Urban Continuum, Surgery Performed, and Brain Metastasis, displayed only tenuous correlations with the target variable. This observation implied that these variables possessed a relatively low predictive capacity and thus were deemed to be of limited utility in the context of the model.

Consequently, a decision was made to exclude these aforementioned variables from the subsequent model training phase. This exclusion was predicated on the rationale that incorporating variables with weak predictive associations could dilute the overall performance and accuracy of the model. By omitting such variables, the focus could be more sharply directed towards those features that exhibited stronger and more meaningful relationships with the target variable, thereby enhancing the model's ability to generate accurate and reliable predictions.

3.3 Model comparison

Six ML algorithms (LR, RF, XGBoost, GBM, LightGBM and CatBoost) were utilized in the development of prediction models. Initially, these models were trained using basic algorithm. As shown in Supplementary Table 1, while the models demonstrated strong performance in terms of AUCs, with values of 0.8641, 0.8395, 0.8740, 0.8828, 0.8821 and 0.8840 for LR, RF, XGBoost, GBM, LightGBM and CatBoost respectively, their recall rates were relatively low at 0.5930, 0.6361, 0.6306, 0.6228, 0.6313 and 0.6205. This indicates that the models' ability to accurately identify true cases of

Table 1 Baseline characteristics of CRC patients with and without liver metastasis from the SEER database (n = 50,062)

Variables	Liver metastasis (n = 5604)	No liver metastasis (n = 44,458)	P value
Age, years			< 0.001
< 60	2522(45.0%)	15,899(2.4%)	
60–69	1526(27.2%)	11,503(25.9%)	
70–79	973(17.4%)	9371(21.1%)	
80–89	520(9.3%)	6613(14.9%)	
> 90	63(1.1%)	1072(2.4%)	
Sex			< 0.001
Male	3083(55.0%)	22,864(51.4%)	
Female	2521(45.0%)	21,594(48.6%)	
Race			< 0.001
White	4224(75.4%)	34,610(77.8%)	
Black	846(15.1%)	4844(10.9%)	
Other (American Indian/AK Native, Asian/Pacific Islander)	534(9.5%)	5004(11.3%)	
Marital status			< 0.001
Married	3136(56.0%)	23,998(54.0%)	
Single	1056(18.8%)	7112(16.0%)	
Divorced/Separated/Unknown	1412(25.2%)	13,348(30.0%)	
Rural–Urban continuum			0.6555
ge 1 million	3156(56.3%)	25,137(56.5%)	
250,000 to 1 million	1242(22.2%)	9913(22.3%)	
lt 250 thousand	427(7.6%)	3484(7.8%)	
Unknown/others	779(13.9%)	5924(13.3%)	
T stage			< 0.001
T1	49(0.9%)	3970(8.9%)	
T2	147(2.6%)	7062(15.9%)	
T3	3349(59.8%)	26,198(58.9%)	
T4	2059(36.7%)	7228(16.3%)	
N stage			< 0.001
N0	914(16.3%)	24,658(55.5%)	
N1	2143(38.2%)	12,912(29.0%)	
N2	2547(45.4%)	6888(15.5%)	
Grade			< 0.001
I	197(3.5%)	3214(7.2%)	
II	3862(68.9%)	32,194(72.4%)	
III	1163(20.8%)	6683(15.0%)	
IV	262(4.7%)	1336(3.0%)	
Unknown	120(2.1%)	1031(2.3%)	
Tumor size, cm			< 0.001
< 5	2913(52.0%)	28,639(64.4%)	
> 5	2691(48.0%)	15,819(35.6%)	
CEA			< 0.001
Negative	1009(18.0%)	27,334(61.5%)	
Positive	4578(81.7%)	16,867(37.9%)	
Borderline	17(0.3%)	257(0.6%)	
Perineural invasion recode			< 0.001
No	3744(66.8%)	39,208(88.2%)	
Yes	1860(33.2%)	5250(11.8%)	
Tumor deposits			< 0.001
Negative	3379(60.3%)	37,941(85.3%)	
Positive	1990(35.5%)	5307(11.9%)	

Table 1 (continued)

Variables	Liver metastasis (n = 5604)	No liver metastasis (n = 44,458)	P value
Unknown	235(4.2%)	1210(2.7%)	
Positive regional nodes			< 0.001
0 nodes	1122(20.0%)	26,713(60.1%)	
1–5 nodes	2666(47.6%)	13,653(30.7%)	
6–10 nodes	1111(19.8%)	2655(6.0%)	
> 10 nodes	705(12.6%)	1437(3.2%)	
Lymph node positivity rate, %			< 0.001
0	1122(20.0%)	26,713(60.1%)	
> 0 and < 0.5	3233(57.7%)	15,442(34.7%)	
> 0.5 and < 1	1105(19.7%)	2045(4.6%)	
1	144(2.6%)	258(0.6%)	
Lung metastasis			< 0.001
No	4813(85.9%)	43,986(98.9%)	
Yes	791(14.1%)	472(1.1%)	
Bone metastasis			< 0.001
No	5478(97.8%)	44,397(99.9%)	
Yes	126(2.2%)	61(0.1%)	
Brain metastasis			< 0.001
No	5582(99.6%)	44,417(99.9%)	
Yes	22(0.4%)	41(0.1%)	
Chemotherapy			< 0.001
No	1341(23.9%)	24,624(55.4%)	
Yes	4263(76.1%)	19,834(44.6%)	
Surgery performed			< 0.001
No/unknown	13(0.2%)	29(0.1%)	
Yes	5591(99.8%)	44,429(99.9%)	
Surgery/radiance sequence			
No radiation and/or cancer-directed surgery	5145(91.8%)	37,669(84.7%)	
Radiation prior to surgery	247(4.4%)	4832(10.9%)	< 0.001
Radiation after surgery	193(3.4%)	1779(4.0%)	
Radiation before and after surgery	19(0.3%)	178(0.4%)	
Systemic therapy/surgery sequence			
No systemic therapy and/or surgical procedures	1341(23.9%)	24,610(55.4%)	< 0.001
Systemic therapy after surgery	3336(59.5%)	14,675(33.0%)	
Systemic therapy before surgery	491(8.8%)	2995(6.7%)	
Systemic therapy both before and after surgery	436(7.8%)	2178(4.9%)	

metastasis is relatively poor. In real clinical settings, it is more important to identify as many CRC patients who will actually develop liver metastases as possible, even if this may compromise precision, which means that some patients may be incorrectly predicted to develop metastasis. Therefore, given the actual higher incidence of liver metastasis rate than our data [10] and the necessity to identify more true positive (liver metastases) cases, we subsequently adjusted our weight-balanced algorithm (seen in methods) for each model. As shown in Table 2 and Fig. 3A, following these adjustments, the six ML models (LR, RF, XGBoost, GBM, LightGBM and CatBoost) showed a much better performance in identifying CRC patients who will actually develop liver metastases with recalls of 0.7857, 0.6508, 0.7862, 0.8049, 0.8039 and 8060, respectively and still demonstrated strong discriminative capacity for detecting liver metastasis in CRC patients with AUCs of 0.8638, 0.8198, 0.8685, 0.8820 0.8820 and 0.8844, respectively. Of note, the CatBoost model achieved the highest F1-score of 0.6736, the highest recall of 0.8060, and the highest AUC of 0.8844 among these ML models, indicates that CatBoost overall outperformed the other ML models. Figure 3C presents the confusion matrix of the CatBoost model.

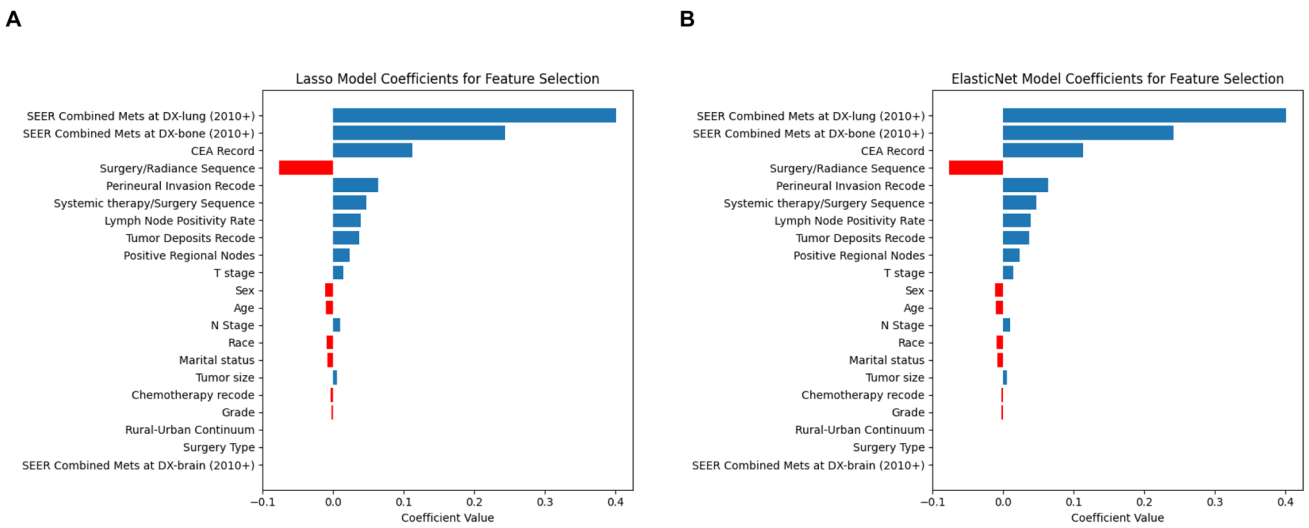

Fig. 2 **A** Feature selection using lasso coefficients. **B** Feature selection using ElasticNet coefficients

Furthermore, the performance of the improved CatBoost model was evaluated using the Precision–Recall (P–R) curve and calibration performance metrics to further validate its predictive capabilities. As illustrated in Fig. 4A, the CatBoost model demonstrated superior performance on the P–R curve, achieving the highest Average Precision (AP) score among all models, which highlights its enhanced ability to identify true positive cases of liver metastasis while maintaining acceptable precision. Figure 4C presents the calibration performance of the CatBoost model, revealing that it not only provides accurate probability estimates but also maintains excellent calibration across different probability thresholds. As a result, CatBoost was chosen as the optimal model for predicting liver metastasis.

3.4 Interpretability analysis

After determining the optimal model for predicting CRC patients with liver metastases, we employed SHAP to interpret the output of the optimal model (CatBoost) by quantifying the contribution of each variable to the prediction. Firstly, the global explanation describes the overall functionality of the model. As shown in the SHAP important plot (Fig. 5B), the mean SHAP values were used to assess the contribution of features to the model, with the results presented in descending order: CEA, Systemic/Surgery Sequence, T stage, N stage, Systemic/Radiation Sequence, Chemotherapy record and Positive Regional Nodes and are the top seven most important variables in the model. Additionally, Lymph Node Positivity Rate and Tumor deposits also show the strength of the influence on the prediction model.

Furthermore, the SHAP summary plot (Fig. 5A) visually represents the direction and magnitude of each feature's impact on the model's predictions. The plot reveals that higher CEA levels, as well as advanced T and N stages, are associated with an increased predicted risk of liver metastasis, while a lower value for the Surgery/Radiation sequence feature corresponds to a reduced predicted risk of liver metastasis.

After establishing a global understanding of feature importance and their overall impact on the model's predictions, we now turn to the SHAP local analysis to explore the contributions of specific features to individual predictions in greater

Table 2 Comparison of the performance of six ML models in the developing cohort

Performance	LR	RF	XGBoost	GBM	LightGBM	CatBoost
Precision	0.6406	**0.6592**	0.6475	0.6486	0.6523	0.6495
Recall	0.7857	0.6508	0.7862	0.8049	0.8039	**0.8060**
F1-score	0.6574	0.6548	0.6693	0.6667	0.6682	**0.6736**
AUC	0.8638	0.8198	0.8685	0.8820	0.8820	**0.8844**

Bold values indicate better results than other methods

LR, logistic regression; RF, random forest; XGBoost, extreme gradient boosting; GBM, gradient boosting machine; AUC, area under curve

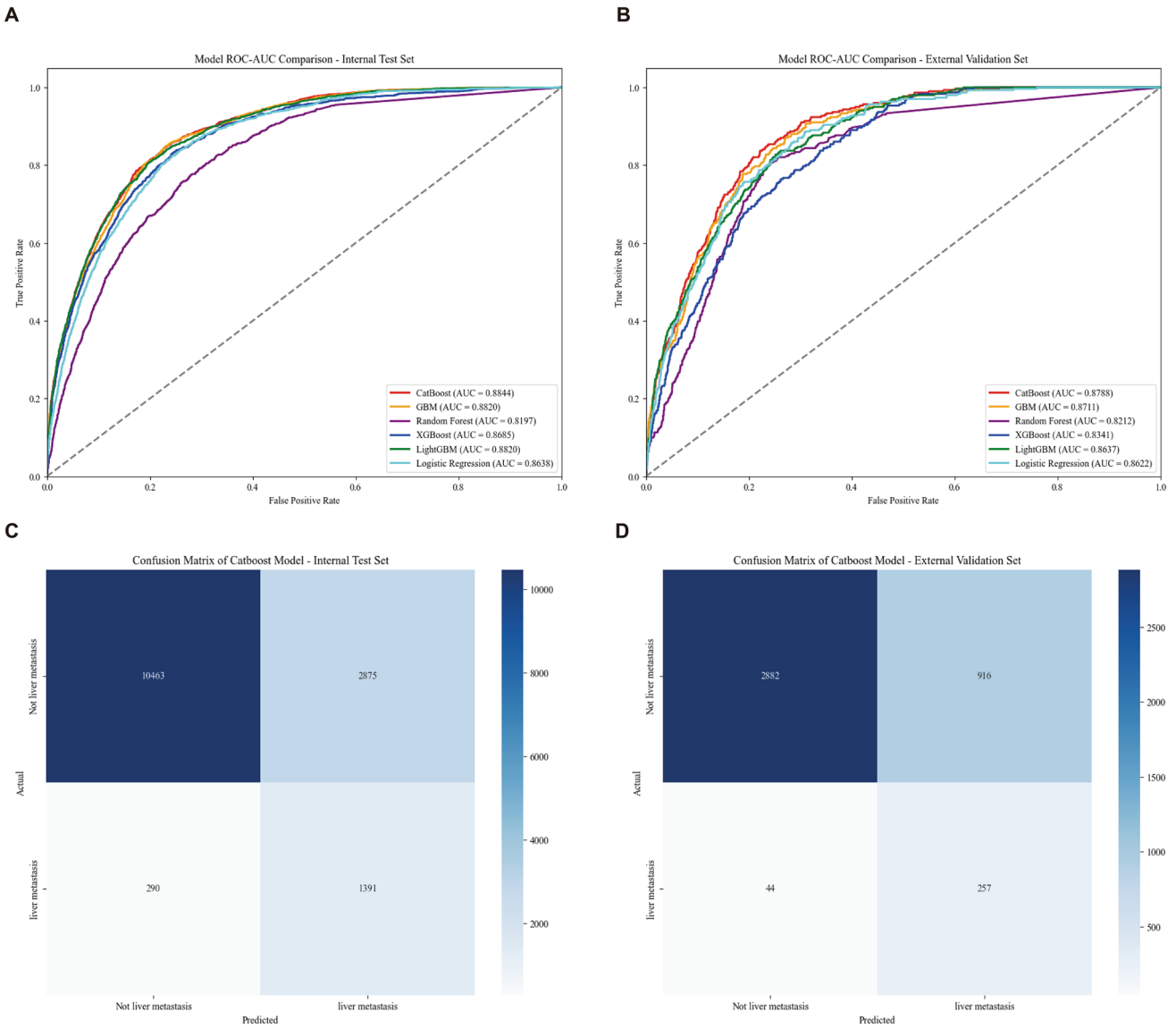

Fig. 3 AUROCs and confusion matrix plots for models in the internal and external validation cohorts. **A** AUROC curves comparing the performance of CatBoost, GBM, Random Forest, XGBoost, LightGBM, and Logistic Regression models in the internal test set, with the corresponding AUC values provided in the legend. **B** AUROC curves comparing the performance of the same models in the external validation set, with AUC values detailed in the legend. **C** Confusion matrix for the CatBoost model in the internal set, presenting the counts of true positives, false positives, true negatives, and false negatives. **D** Confusion matrix for the CatBoost model in the external validation test set, showing the classification results for actual and predicted outcomes, AUROCs and confusion matrix plots for models in the internal and external cohort. AUROC, area under the receiver operating characteristic curve; AUC, area under curve

detail. Based on Fig. 6, Elevated CEA levels (Fig. 6A), systemic therapy both before and after surgery (Fig. 6B), advanced T (Fig. 6C) or N stages (Fig. 6D) and radiation not performed (Fig. 6E), chemotherapy performed (Fig. 6F) and 10 more positive regional nodes (Fig. 6G) were associated with higher predicted risks for liver metastasis. In addition, factors such as the number of positive regional nodes, absence of radiation therapy, increased tumor deposits, and lung metastasis were also positively correlated with the risk of liver metastasis. Notably, radiation therapy prior to surgery appeared to be more effective in reducing the risk of liver metastasis compared to radiation administrated post-surgery.

While the local analysis offers important insights into the contributions of individual features to specific predictions, we now advance our interpretation by examining feature interactions through the SHAP-based interaction interpretation plot, which elucidates the joint influence of feature pairs on the model's predictions. Notably, while systemic therapy was associated with a higher SHAP value, indicating an increased risk of liver metastasis (Fig. 5A), SHAP interaction plots revealed that administrating systemic therapy either before or after the surgery had a protective

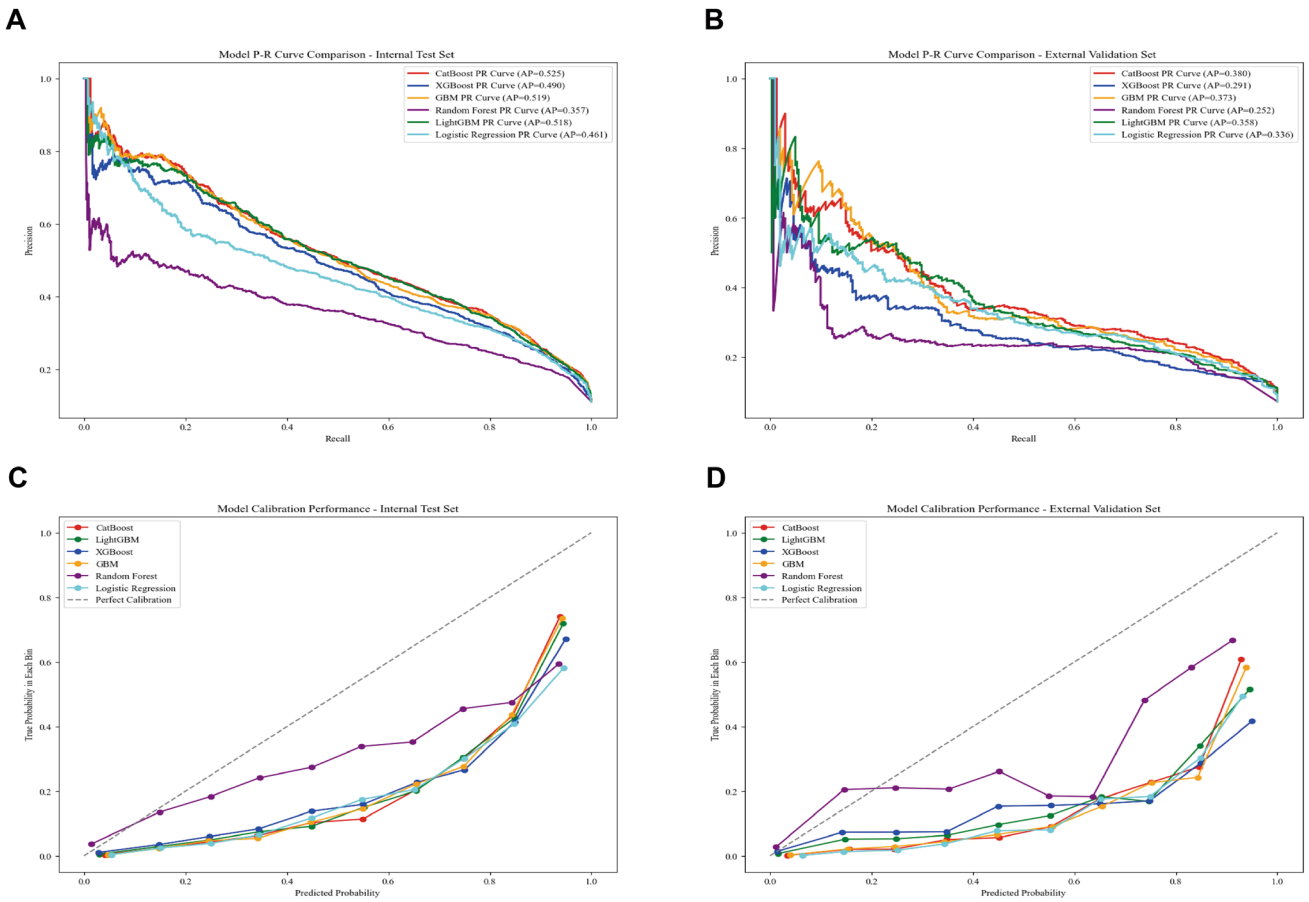

Fig. 4 Validation of six prediction models. **A** P–R curve of six models in internal test set. The CatBoost achieved better than the other models; **B** P–R curve of six models in external validation set; **C** Calibration performance in internal test set; **D** Calibration performance in external validation set

effect against liver metastasis in CRC patients in N1 and N2 stage or with positive regional nodes (Fig. 7A–D). Interestingly, in N0 patients (Fig. 7C), systemic therapy especially exerted both before and after surgery amplified the liver metastasis conversely. This suggests that systemic therapy may be particularly beneficial for patients with lymph node metastasis. Similarly, applying systemic therapy to CRC patients at T3 or T4 stage reduced the predicted risk of liver metastasis, as evidenced by the diminished SHAP value in Fig. 7E. And increased risk was observed in T1 and T2 patients receiving systemic therapy, consistent with our prior investigation that patients at early N stage may not be suitable to receive supplementary systemic therapy. Meanwhile, chemotherapy interacting with T, N stage showed the similar performance (Fig. 7F, G). Furthermore, compared with the single use of either systemic therapy or radiation therapy, combining systemic therapy with radiation therapy appeared to further decrease the liver metastasis risk (Fig. 7H, I).

To further elucidate the interactions between key clinical features, SHAP boxplots were generated to visually depict the distribution of SHAP values across different subgroups, as shown in Fig. 8A–I. These visualizations correspond to the interaction effects described in Fig. 7A–I, providing a clear representation of how combinations of these features influence the predicted risk of liver metastasis.

To complement the visual analysis, a one-way analysis of variance (ANOVA) was performed to evaluate differences between groups if the normality assumption was satisfied. For groups that did not meet the normality assumption, the Kruskal-Walli's test was conducted as a non-parametric alternative. The P-values displayed alongside the boxplots indicate whether the observed differences in SHAP values between subgroups are statistically significant. A P-value < 0.05 denotes a significant interaction effect, suggesting that the influence of two clinical features on liver metastasis risk is interdependent rather than independent. These findings underscore the presence of interactive effects, wherein the impact of one feature—such as systemic therapy—may vary depending on another feature, such as tumor stage or

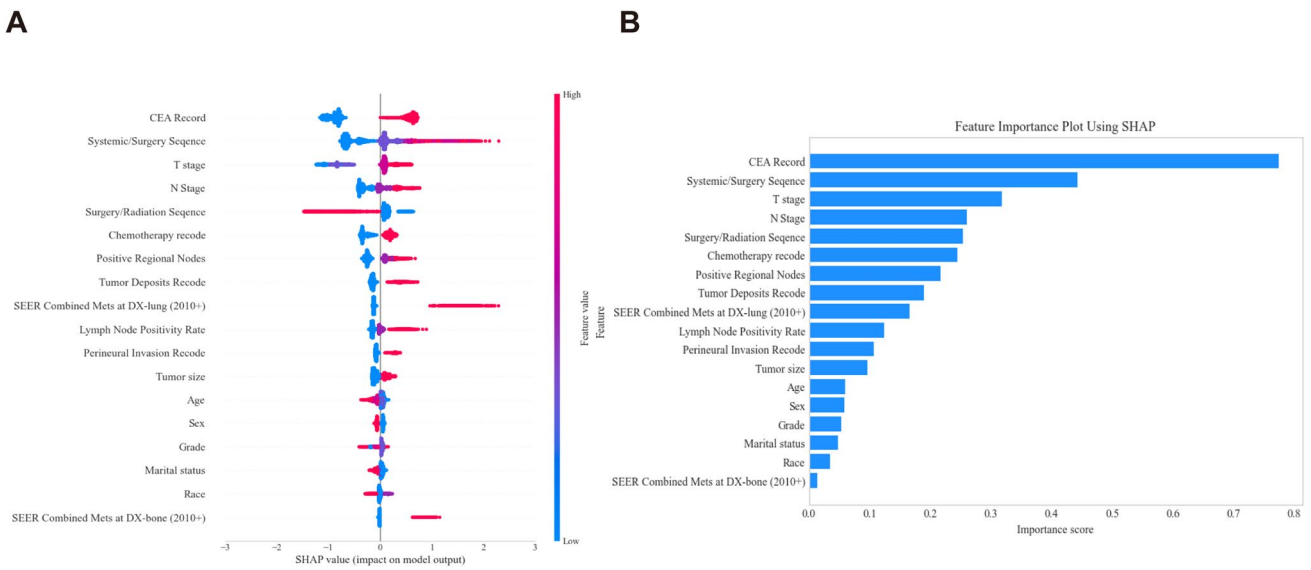

Fig. 5 SHAP-based global interpretation of the CatBoost model. **A** The SHAP Summary Plot. Positive SHAP values indicate that the feature increases the predicted risk, while negative SHAP values suggest a reduction in the predicted risk. Each dot represents a SHAP value for a particular feature in an individual patient. The position along the x-axis shows the SHAP value's magnitude and direction (positive or negative), while the y-axis groups the dots by feature. The color of the dots encodes the feature's value. For continuous features, red corresponds to higher values and blue to lower values. For binary features, red represents "yes" or "true", while blue indicates "no" or "false". **B** Feature Importance Plot. Features are ranked by their importance scores

positive lymph node status. This analysis provides additional insights into the nonlinear relationships captured by the model, enhancing its interpretability and highlighting its relevance for personalized clinical decision-making.

The SHAP method also provides explanations for model predictions at an individual level (Fig. 9), which is valuable for personalized treatment in clinical practice and aids in decision-making for clinicians and patients. Three representative cases were selected in the figure. The plots display the individual predicted value for the risk of liver metastasis and the contributions of each characteristic (represented by the arrows). Red arrows denote positive contributions to the predicted values, while blue indicate negative contributions. The length of the arrows reflects the importance of predictors, with longer arrows signifying more important predictors. As shown in Fig. 9A and B, for the first and second patients, the absence of systemic therapy, negative CEA levels, earlier N stages, no chemotherapy and no positive regional node predict a lower risk of liver metastasis (Fig. 9A, B). Conversely, in Fig. 9C, advanced age, elevated CEA level, higher on stage and T stage, systemic therapy prior to surgery and absence of radiation therapy are key predictors of liver metastasis for the first patient. By accumulating SHAP values, the force plot visually demonstrates the formation process of the prediction result for a specific patient, helping us to understand in depth the decision-making mechanism of the model.

3.5 External validation

We validated the external cohort using data obtained from the SEER Research Data, 12 Registries (November 2023 submission) for the years 2018 to 2021. The data for these years were preprocessed using the same methodology as previously applied, with the exception that the T and N stages were classified based on the AJCC 8th edition criteria.

In the external validation cohort, we first evaluated the baseline models shown in Supplementary Table 2, which were previously trained without applying weight-balanced techniques. The AUC values were 0.8689 for LR, 0.8395 for RF, 0.8246 for XGBoost, 0.8774 for GBM, 0.8607 for LightGBM, and 0.8710 for CatBoost. Although the AUC were slightly lower than their performance in the internal test set, they remained relatively high, indicating good discriminative ability. However, consistent with the internal training cohort, the recall rates, which assess the ability to accurately identify true cases of metastasis, remained relatively low, with values of 0.5688 for LR, 0.6361 for RF, 0.5745 for XGBoost, 0.5746 for GBM, 0.5839 for LightGBM, and 0.5661 for CatBoost.

This study then evaluated the trained weight-balanced techniques enhanced models using external validation cohort. As presented in Supplementary Table 3 and Fig. 3B, the AUC values exhibited remain stable, reaching 0.8622 for LR, 0.8212 for RF, 0.8341 for XGBoost, 0.8711 for GBM, 0.8637 for LightGBM, and 0.8788 for CatBoost. In addition, the recall rates

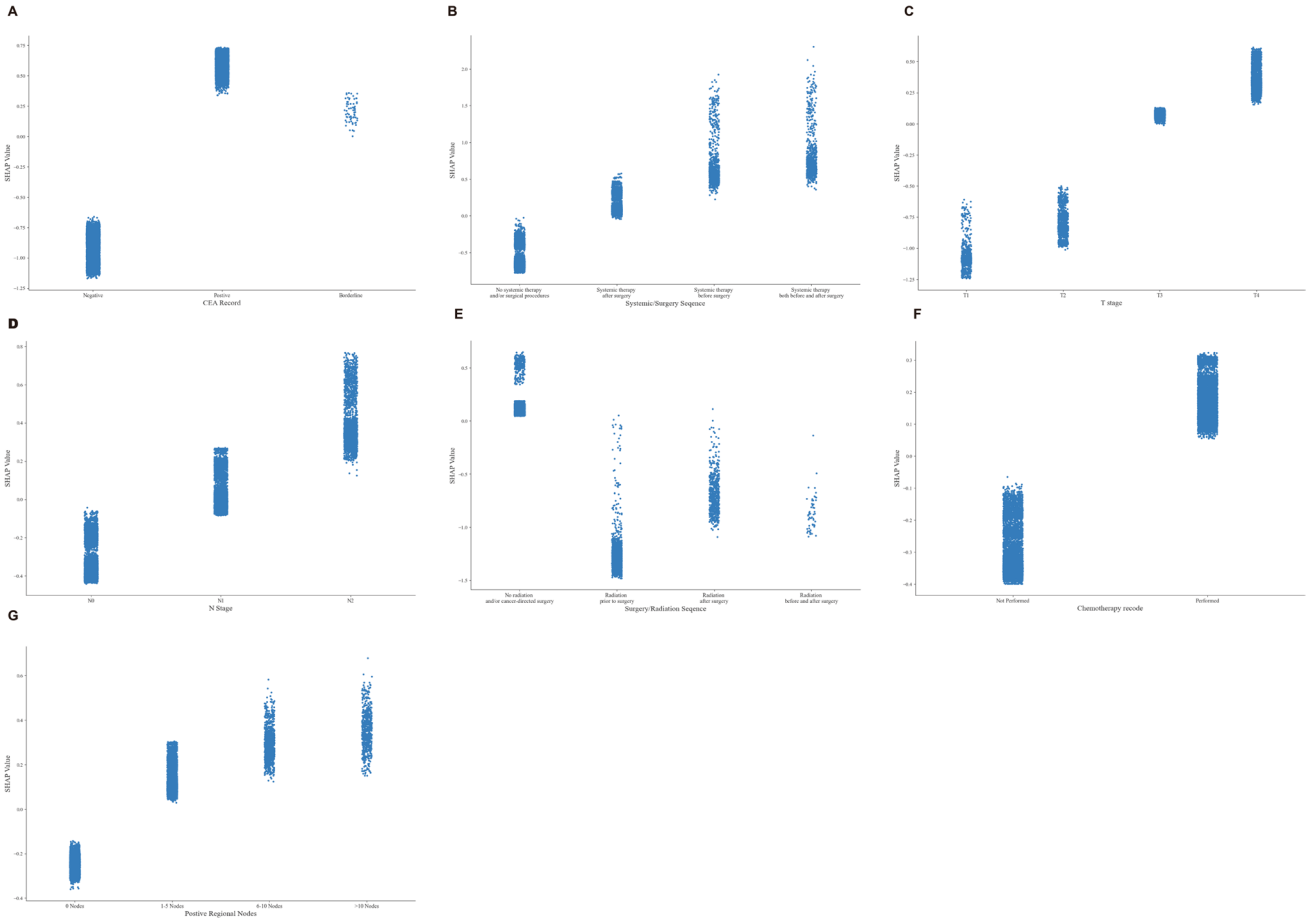

Fig. 6 SHAP-based local interpretation plot of the CatBoost model. SHAP local plots for the top 7 most important features. Higher SHAP values indicate that the feature contributes to increasing the predicted risk, while negative SHAP values suggest a contribution to reducing the predicted risk. Each dot represents an individual patient in the test set, showing the SHAP value for a specific feature in that prediction. **A** SHAP values for the feature "CEA Record" showing its impact on risk prediction. **B** SHAP values for the feature "Systemic/Surgery Sequence" illustrating its influence on predicted outcomes. **C** SHAP values for the feature "T Stage" highlighting its contribution to the model's risk predictions. **D** SHAP values for the feature "N Stage" and its effect on the predicted risk. **E** SHAP values for the feature "Surgery/Radiation Sequence" showing its impact on prediction. **F** SHAP values for the feature "Chemotherapy Recode" demonstrating its role in the model's predictions. **G** SHAP values for the feature "Positive Regional Nodes" and its contribution to predicted risk. SHAP, Shapley additive explanations

showed more notable improvements, consistent with the internal test set, reaching 0.7813 for LR, 0.5610 for RF, 0.7156 for XGBoost, 0.7967 for GBM, 0.7715 for LightGBM, and 0.8063 for CatBoost. Similarly, the F1-scores were enhanced, with values of 0.5877 for LR, 0.5735 for RF, 0.6102 for XGBoost, 0.5941 for GBM, 0.6169 for LightGBM, and 0.6030 for CatBoost. Figure 3D presents the confusion matrix of the CatBoost model in the external validation set, and the precision-recall curve comparisons and calibration plots are shown in Fig. 4B and D. These results suggest that the weight-balancing techniques improved the models' ability to identify true cases of metastasis while also enhancing overall predictive performance, as demonstrated in the external validation set.

4 Discussion

In this study, we have developed and validated machine learning models using a carefully selected set of clinically relevant features. These features include CEA levels, the sequence of systemic therapy and surgery, T stage, N stage, the sequence of systemic therapy and radiation, chemotherapy records, and the presence of positive regional nodes. These models aim to predict the risk of liver metastasis in colorectal cancer patients. Importantly, these features are readily accessible during hospitalization, which facilitates the practical application of the models in clinical settings. This

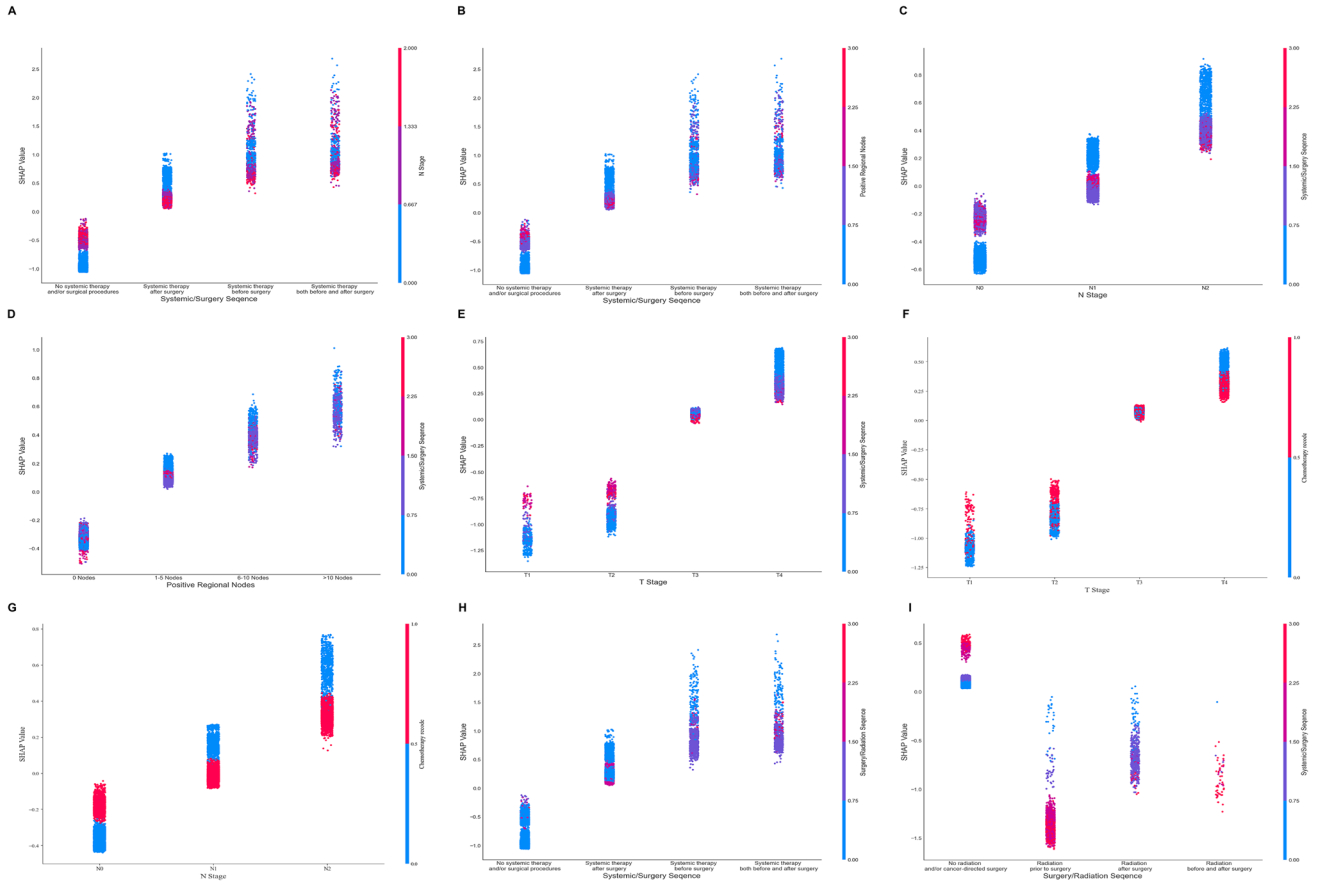

Fig. 7 SHAP-based interaction plot of the CatBoost model. The values of the first feature are represented by the x-axis, and the values of the second feature are represented by the color gradient. SHAP values are represented by the y-axis. All displayed features are categorical variables. **A** Interaction between "Systemic/Surgery Sequence" and "N Stage". **B** Interaction between "Systemic/Surgery Sequence" and "Positive Regional Nodes". **C** Interaction between "N Stage" and "Systemic/Surgery Sequence". **D** Interaction between "Positive Regional Nodes" and "Systemic/Surgery Sequence". **E** Interaction between "T Stage" and "Systemic/Surgery Sequence". **F** Interaction between "T Stage" and "Chemotherapy Recode". **G** Interaction between "N Stage" and "Chemotherapy Recode". **H** Interaction between "Systemic/Surgery Sequence" and "Surgery/Radiation Sequence". **I** Interaction between "Surgery/Radiation Sequence" and "Systemic/Surgery Sequence". For systemic/surgery sequence, 0 represents no systemic therapy and/or surgical procedures, 1 represents systemic therapy after surgery, 2 represents systemic therapy before surgery, and 3 represents systemic therapy both before and after surgery. For N stage, 0 represents N0, 1 represents N1, and 2 represents N2. For positive regional nodes, 0 represents 0 nodes, 1 represents <5 nodes, and 2 represents 6-10 nodes. For T stage, 0 represents T1, 1 represents T2, 2 represents T3, and 3 represents T4. For surgery/radiation sequence, 0 represents no radiation and/or cancer-directed surgery, 1 represents radiation prior to surgery, 2 represents radiation after surgery, and 3 represents radiation before and after surgery. For chemotherapy record, 0 represents no chemotherapy and 1 represents chemotherapy applied. SHAP, Shapley additive explanations

accessibility allows for seamless integration into clinical workflows, enabling clinicians to make more informed decisions regarding the risk of liver metastasis in CRC patients.

The performance of each model was rigorously assessed using a comprehensive set of evaluation metrics, including the AUROC, PR curves, calibration plots, recall, precision, and F1-scores. Among the models tested, CatBoost exhibited superior predictive performance, consistently outperforming the others across these metrics.

Furthermore, SHAP analysis was employed to interpret the predictions generated by the CatBoost model. This provided valuable insights into the decision-making process of the model, enhancing its interpretability. By elucidating the contributions of individual features to the model's predictions, SHAP analysis helps clinicians and researchers better understand the underlying factors driving the model's behavior. This interpretability is crucial for gaining confidence in the model's predictions and for integrating the model into clinical decision-making processes.

Considering the complexity of CRC and the challenges associated with early-stage detection of CRLM, it is essential to integrate medical knowledge with technological advancements. By customizing artificial intelligence and machine learning for medical diagnosis, these diagnostic challenges could be potentially bridged. Currently, machine learning

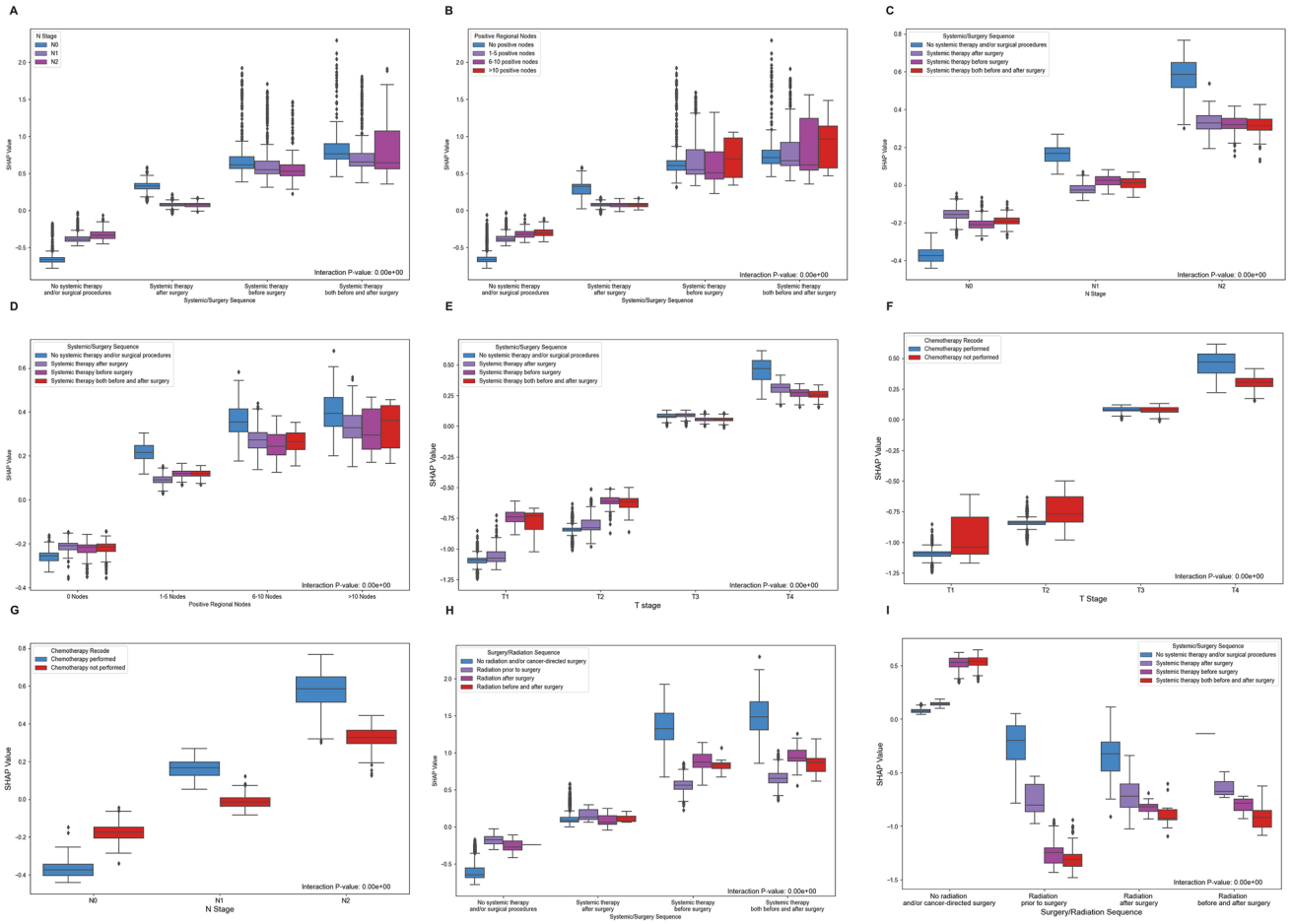

Fig. 8 SHAP-based box plots depicting interaction effects between clinical features. The box plots show the distribution of SHAP values across different subgroups for the key clinical features, highlighting their combined influence on the predicted risk of liver metastasis. Each box plot corresponds to the interaction effects visualized in Figure 7 (**A**–**I**), providing a complementary perspective to better understand the model's predictions. **A** SHAP box plot showing the interaction between "Systemic/Surgery Sequence" and "N Stage". **B** SHAP box plot depicting the interaction between "Systemic/Surgery Sequence" and "Positive Regional Nodes". **C** SHAP box plot illustrating the interaction between "N Stage" and "Systemic/Surgery Sequence". **D** SHAP box plot showing the interaction between "Positive Regional Nodes" and "Systemic/Surgery Sequence". **E** SHAP box plot highlighting the interaction between "T Stage" and "Systemic/Surgery Sequence". **F** SHAP box plot depicting the interaction between "T Stage" and "Chemotherapy Recode". **G** SHAP box plot showing the interaction between "N Stage" and "Chemotherapy Recode". **H** SHAP box plot illustrating the interaction between "Systemic/Surgery Sequence" and "Surgery/Radiation Sequence". **I** SHAP box plot highlighting the interaction between "Surgery/Radiation Sequence" and "Systemic/Surgery Sequence". SHAP, Shapley additive explanations

models combining easily accessible demographic and clinical data has be constructed to improve the predictive accuracy and utility of distant lymph node metastases in hepatocellular carcinoma patients [11]. To date, studies using machine learning on predicting the risk of liver metastasis in CRC have been rarely reported. It's urgent to use machine learning algorithms to predict liver metastasis from colorectal cancer. In this study, we selected six different models to construct a prediction model for the better performance with less feature engineering compared to logistic regression. And these machine learning models perform well in capturing complex, non-linear relationships, works well with large datasets, is less sensitive to outliers and can automatically capture interactions between features. Our analysis illustrated the significance of various variables, highlighting their distinct contributions to the model's predictive performance. Furthermore, in this developed Catboost model, we constructed top seven feature's partial dependence plots including interacting with other features to analyze further the correlation between each variable and the liver metastasis risk.

In terms of importance, pretreatment CEA emerged as the most significant feature. CEA is a widely recognized tumor marker with considerable clinical value in the differential diagnosis of malignant tumors, disease monitoring, and the evaluation of treatment efficacy. Numerous studies have established a correlation between pre-treatment CEA levels and the likelihood of metastasis [12–14]. For instance, in 2010, Wu et al. observed that CEA levels are often significantly higher

Fig. 9 Representative SHAP force plots for individual model predictions. **A**, **B**, and **C** represent three randomly selected patients from the test set, illustrating the contribution of each feature to the prediction outcomes. The base value represents the average of all prediction results. Feature values were listed at the plot bottom, with red indicating positive contributions to the model's prediction values and blue representing negative effects

in patients with liver metastasis than in those without liver metastasis [15]. Additionally, an increasing number of studies have suggested that CEA actively promotes the metastatic spread of colorectal cancer cells to the liver, representing a promising therapeutic target for managing liver metastasis. Elevated CEA expression in tumor cells may destabilize intercellular adhesion, disrupt cellular organization, inhibit immune responses, and suppress immune surveillance, thereby facilitating cell migration and invasion [16]. Therefore, colorectal cancer patients with elevated preoperative CEA levels require more vigilant postoperative monitoring to ensure timely detection of liver metastases.

In our analysis, the Systemic/Surgery Sequence ranked second in significance after CEA. Radical surgical resection of colorectal cancer lesions remains the most effective curative approach to date and is crucial for preventing tumor progression. Preoperative neoadjuvant therapy aims to eliminate micro metastases that may not be detected by imaging studies, thereby reducing the risk of distant metastasis following radical surgery [17]. Our study indicates that systemic therapy, regardless of other clinical characteristics, whether administered preoperatively or postoperatively, is positively correlated with liver metastasis, which contradicts general expectations. This unexpected relationship may be attributed to the activation of specific gene expression patterns that facilitate the metastasis process, as well as the increased production of exosomes, which are known to promote the spread of cancer cells [18, 19]. Furthermore, we speculate that patients with advanced-stage colorectal cancer, particularly those who have already developed lymph node metastasis or distant metastasis, are more likely to undergo systemic treatment and subsequently experience liver metastasis. Notably, systemic therapy reduces the incidence of liver metastasis in CRC patients with higher N stages or more positive regional nodes, while not in N0 patients. This suggests that systemic therapy may be useful to prevent patients with positive regional nodes from liver metastasis, not suitable to the N0 patients. Similar to a recent study, neoadjuvant therapy is not required in N0 and M0 patients despite of the differences in surgical treatment [20]. The results of the SHAP value analysis indicate that patients with regional lymph node invasion are at a significantly higher risk of liver metastasis. This elevated risk may be attributed to the fact that invaded regional lymph nodes can serve as metastatic stations that facilitate the proliferation of tumor cells [21]. In total, our study indicates that effective systemic therapy either before if detected preoperatively or after the surgery should be considered for patients in N1 and N2 stages to provide protective effects against liver metastasis, these results indicate that immediate surgical intervention alone may not be the most appropriate strategy for this subset of patients.

Former studies have shown that T stage is highly correlated with distant metastasis [22, 23]. Stages T3 and T4 have been identified as independent risk factors for distant metastasis [22, 23]. As the depth of tumor invasion increases, the risk of tumor cells detaching and spreading through vessels heightens, ultimately leading to settlement in the liver. Our present study also identified T stage as an important risk factor for liver metastasis in colorectal cancer. Besides, the depence plot between T stage and Systemic/Surgery Sequence suggests that adding systemic therapy espcially exerting both prior and after surgery is benefical to the patients in T3 and T4 stage against liver metastasis. In addition, our study indicates that radiation therapy reduces the incidence of liver metastasis in patients with colorectal cancer. Several potential mechanisms may explain the protective effect of radiotherapy, including the

eradication and reduction of primary tumor cells, the neutralization of micrometastases originating from colorectal cancer, and the modulation of the immune response [24, 25].

Tumor deposits (TDs) denote the sporadic dissemination of cancer cells within the mesocolon or mesorectum. Lord et al. found that TDs in combination with metastatic lymph nodes are the strongest predictor of peritoneal metastasis [26]. Similarly, Yamano et al. found that TDs are independent poor prognostic factors for recurrence-free survival in patients with lymph node metastasis, and colorectal cancer patients with positive iTDs (iTDs: cancer cell aggregates with lymphatic or perineural infiltration or cancer cell clusters) are frequently associated with a higher incidence of liver metastasis. Other researchers also reported a strong correlation between the presence of tumor deposits and heightened risks of local recurrence and distant metastasis in individuals with colorectal cancer, which consequently affects their overall survival rates [27–29]. Our model supports the notion that TDs are significantly associated with an increased risk of liver metastasis, suggesting more intensive follow-up are needed if tumor deposits are detected in surgical pathology.

Colorectal cancer is a prevalent malignancy of the digestive tract, characterized by a high propensity for distant metastasis, which substantially contributes to cancer-related mortality. The liver represents the most common metastatic site, highlighting the necessity of developing predictive models to identify patients at elevated risk of liver metastases and to facilitate early intervention strategies. Machine learning techniques have emerged as powerful tools for predictive modeling in clinical research. However, their inherent complexity and opacity often lead to their characterization as "black boxes," posing challenges for clinical adoption due to limited interpretability. Such concerns may undermine clinicians' confidence in incorporating ML-based predictions into decision-making processes. A major strength of this study lies in its application of the SHAP method to address these interpretability challenges. By providing both global and local explanations, as well as insights into feature interactions, the SHAP method elucidates the decision-making processes of the ML models. This enhanced transparency improves the interpretability and reliability of the predictions, promoting their acceptance in clinical practice.

This study has several limitations that should be acknowledged. The reliance on data exclusively from the SEER database, a retrospective dataset, may introduce inherent selection bias despite the implementation of stringent exclusion criteria, as well as the use of datasets from different time periods for external validation. These limitations could skew the findings and reduce their applicability to a broader patient population. The absence of an in-hospital cohort further restricts the generalizability of the results, as they may not fully capture the diverse clinical scenarios encountered in various hospital settings. Moreover, the lack of detailed treatment information, such as specific radiotherapy and chemotherapy regimens, is a significant drawback. This limitation hampers a comprehensive evaluation of how different treatment variations impact the incidence of liver metastases. Consequently, the interpretation of the results may be affected, as the study cannot fully account for the nuances of treatment-related factors that could influence patient outcomes. Additionally, while CEA was identified as a significant predictor in this study, its sensitivity is not optimal. Elevated CEA levels are observed in only a subset of CRC patients and can be influenced by external factors, such as smoking. This introduces potential bias and suggests that CEA alone may not be sufficient for accurate risk stratification. In light of these limitations, further studies that integrate additional biomarkers or complementary.

Author contributions Qunzhe Ding and Chendong Wang conceptualized and designed the study. Chendong Wang collected the data. Qunzhe Ding analyzed the data and constructed the models. Qunzhe Ding, Chendong Wang and Chenyang Li wrote and edited the manuscript. Chenyang Li prepared the figures and tables. All authors have read and approved the final manuscript.

Funding No funding.

Data availability Data is provided within supplementary information files.

Declarations

Competing interests The authors declare no competing interests.

References

1. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. CA Cancer J Clin. 2023;73(3):233–54.
2. Padmanabhan C, Nussbaum DP, D'Angelica M. Surgical management of colorectal cancer liver metastases. Surg Oncol Clin N Am. 2021;30(1):1–25.
3. Ma ZH, Wang YP, Zheng WH, Ma J, Bai X, Zhang Y, Wang YH, Chi D, Fu XB, Hua XD. Prognostic factors and therapeutic effects of different treatment modalities for colorectal cancer liver metastases. World J Gastrointest Oncol. 2020;12(10):1177–94.
4. Rompianesi G, Pegoraro F, Ceresa CD, Montalti R, Troisi RI. Artificial intelligence in the diagnosis and management of colorectal cancer liver metastases. World J Gastroenterol. 2022;28(1):108–22.
5. van der Geest LGM, Lam-Boer J, Koopman M, Verhoef C, Elferink MAG, de Wilt JHW. Nationwide trends in incidence, treatment and survival of colorectal cancer patients with synchronous metastases. Clin Exp Metastasis. 2015;32(5):457–65.
6. Gumbs AA, Croner R, Lorenz E, Cacciaguerra AB, Tsai TJ, Starker L, Flanagan J, Yu NJ, Chouillard E, Abu Hilal M. Survival study: International multicentric minimally invasive liver resection for colorectal liver metastases (SIMMILR-2). Cancers. 2022;14(17):4190.
7. Sawatzki M, Güller U, Güsewell S, Husarik DB, Semela D, Brand S. Contrast-enhanced ultrasound can guide the therapeutic strategy by improving the detection of colorectal liver metastases. J Hepatol. 2021;74(2):419–27.
8. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. arXiv.org. 2017. https://arxiv.org/abs/1706.09516v5. Accessed 2 Oct 2024.
9. Lundberg S, Lee SI. A unified approach to interpreting model predictions. arXiv.org. 2017. https://arxiv.org/abs/1705.07874v2. Accessed 2 Oct 2024.
10. Engstrand J, Nilsson H, Strömberg C, Jonas E, Freedman J. Colorectal cancer liver metastases—a population-based study on incidence, management and survival. BMC Cancer. 2018;18(1):78.
11. Sun J, Huang L, Liu Y. Leveraging SEER data through machine learning to predict distant lymph node metastasis and prognosticate outcomes in hepatocellular carcinoma patients. J Gene Med. 2024;26(9): e3732.
12. Wu S, Gu W. Association of T stage and serum CEA levels in determining survival of rectal cancer. Front Med (Lausanne). 2019;6:270.
13. Quah HM, Chou JF, Gonen M, Shia J, Schrag D, Landmann RG, Guillem JG, Paty PB, Temple LK, Wong WD, Weiser MR. Identification of patients with high-risk stage II colon cancer for adjuvant therapy. Dis Colon Rectum. 2008;51(5):503–7.
14. Becerra AZ, Probst CP, Tejani MA, Aquina CT, González MG, Hensley BJ, Noyes K, Monson JR, Fleming FJ. Evaluating the prognostic role of elevated preoperative carcinoembryonic antigen levels in colon cancer patients: results from the National cancer database. Ann Surg Oncol. 2016;23(5):1554–61.
15. Wu XZ, Ma F, Wang XL. Serological diagnostic factors for liver metastasis in patients with colorectal cancer. World J Gastroenterol. 2010;16(32):4084–8.
16. Wang Q, Shen K, Fei B, Wei M, Xie Z. Nomogram for predicting occurrence and prognosis of liver metastasis in elderly colorectal cancer patients: a population-based study. Front Oncol. 2023;13:1295650.
17. Saad AM, Abdel-Rahman O. Initial systemic chemotherapeutic and targeted therapy strategies for the treatment of colorectal cancer patients with liver metastases. Expert Opin Pharmacother. 2019;20(14):1767–75.
18. Karagiannis GS, Pastoriza JM, Wang Y, Harney AS, Entenberg D, Pignatelli J, Sharma VP, Xue EA, Cheng E, D'Alfonso TM, Jones JG, Anampa J, Rohan TE, Sparano JA, Condeelis JS, Oktay MH. Neoadjuvant chemotherapy induces breast cancer metastasis through a TMEM-mediated mechanism. Sci Transl Med. 2017;9(397):eaan0026.
19. Keklikoglou I, Cianciaruso C, Güç E, Squadrito ML, Spring LM, Tazzyman S, Lambein L, Poissonnier A, Ferraro GB, Baer C, Cassará A, Guichard A, Iruela-Arispe ML, Lewis CE, Coussens LM, Bardia A, Jain RK, Pollard JW, De Palma M. Chemotherapy elicits pro-metastatic extracellular vesicles in breast cancer models. Nat Cell Biol. 2019;21(2):190–202.
20. National Health Comimission, Chinese Society of Oncology. Chinese protocol of diagnosis and treatment of colorectal cancer of the National Health Commission (2023 edition). Zhonghua Wei Chang Wai Ke Za Zhi. 2023;26(6):505.
21. Nagtegaal ID, Tot T, Jayne DG, McShane P, Nihlberg A, Marshall HC, Påhlman L, Brown JM, Guillou PJ, Quirke P. Lymph nodes, tumor deposits, and TNM: are we getting better? J Clin Oncol. 2011;29(18):2487–92.
22. Huang CM, Huang MY, Ma CJ, Yeh YS, Tsai HL, Huang CW, Huang CJ, Wang JY. Neoadjuvant FOLFOX chemotherapy combined with radiotherapy followed by radical resection in patients with locally advanced colon cancer. Radiat Oncol. 2017;12(1):48.
23. Qiu B, Shen Z, Wu S, Qin X, Yang D, Wang Q. A machine learning-based model for predicting distant metastasis in patients with rectal cancer. Front Oncol. 2023;13:1235121.
24. Ji D, Song C, Li Y, Xia J, Wu Y, Jia J, Cui X, Yu S, Gu J. Combination of radiotherapy and suppression of Tregs enhances abscopal antitumor effect and inhibits metastasis in rectal cancer. J Immunother Cancer. 2020;8(2): e000826.
25. van den Brink M, Stiggelbout AM, van den Hout WB, Kievit J, Klein Kranenbarg E, Marijnen CA, Nagtegaal ID, Rutten HJ, Wiggers T, van de Velde CJ. Clinical nature and prognosis of locally recurrent rectal cancer after total mesorectal excision with or without preoperative radiotherapy. J Clin Oncol. 2004;22(19):3958.
26. Lord A, Brown G, Abulafi M, Bateman A, Frankel W, Goldin R, Gopal P, Kirsch R, Loughrey MB, Märkl B, Moran B. Histopathological diagnosis of tumour deposits in colorectal cancer: a Delphi consensus study. Histopathology. 2021;79(2):168.
27. Ueno H, Hashiguchi Y, Shimazaki H, Shinto E, Kajiwara Y, Nakanishi K, Kato K, Maekawa K, Nakamura T, Yamamoto J, Hase K. Peritumoral deposits as an adverse prognostic indicator of colorectal cancer. Am J Surg. 2014;207(1):70–7.
28. Wu W, Zeng S, Zhang X, Liu P, Qiu T, Li S, Gong P. The value of tumor deposits in evaluating colorectal cancer survival and metastasis: a population-based retrospective cohort study. World J Surg Onc. 2022;20(1):41.
29. Wang Y, Zhang J, Zhou M, Yang L, Wan J, Shen L, Liang L, Yao Y, Zhang H, Zhang Z. Poor prognostic and staging value of tumor deposit in locally advanced rectal cancer with neoadjuvant chemoradiotherapy. Cancer Med. 2019;8(4):1508–20.