Development and validation of a machine learning model for predicting early death in metastatic pancreatic ductal adenocarcinoma: a study based on the SEER database

Leiming Zhang[1,2], Jikai He[1]

[1]Department of Hepatobiliary Surgery, Zhoushan Hospital of Zhejiang Province, Zhoushan, China; [2]The Affiliated Li Huili Hospital, Health Science Center, Ningbo University, Ningbo, China

Contributions: (I) Conception and design: L Zhang; (II) Administrative support: J He; (III) Provision of study materials or patients: Both authors; (IV) Collection and assembly of data: L Zhang; (V) Data analysis and interpretation: Both authors; (VI) Manuscript writing: Both authors; (VII) Final approval of manuscript: Both authors.

Correspondence to: Jikai He, MS. Department of Hepatobiliary Surgery, Zhoushan Hospital of Zhejiang Province, No. 739 Dingshen Road, Zhoushan 316000, China. Email: hejikaizsyy@163.com.

Background: Metastatic pancreatic ductal adenocarcinoma (mPDAC) has a poor prognosis, with a significant number of patients experiencing early death. Identifying these high-risk patients at diagnosis is critical for personalizing treatment intensity, facilitating timely palliative care discussions, and improving clinical trial stratification. Therefore, this study aimed to develop and validate a machine learning (ML)-based algorithm to estimate the probability of early death in patients with mPDAC.

Methods: We recruited a total of 14,820 patients diagnosed with mPDAC from the Surveillance, Epidemiology, and End Results (SEER) databases. Key exclusion criteria were missing data on survival time or essential variables. The cohort was randomly split into a training set (70%) and an internal test set (30%). For external validation, we retrospectively enrolled patients with mPDAC from a Chinese medical center (2017–2019), representing a distinct geographic and healthcare population. The primary outcome was early death, defined as all-cause mortality within three months of diagnosis. Baseline clinical predictors included demographic, tumor, and treatment characteristics. Four ML models were constructed based on clinical and pathological features. The effectiveness of these models was assessed through various metrics such as the area under the curve (AUC), calibration plots, and decision curve analysis (DCA). The optimal model was selected based on 10-fold cross-validation and its generalizability was internally and externally validated. Additionally, Shapley values for relevant features were calculated using the SHapley Additive exPlanations (SHAP) method.

Results: The extreme gradient boosting classifier (XGBoost) model demonstrated the best performance (AUC =0.757). Crucially, it maintained strong generalizability in the independent external Chinese cohort (AUC =0.780), demonstrating robust cross-population applicability. According to the feature importance ranking plot generated, chemotherapy stood out as the most crucial feature, followed by age, and marital status.

Conclusions: We developed and validated an interpretable ML model that accurately predicts the risk of early death in mPDAC patients. The model's robust performance across US and Chinese populations underscores its broad clinical utility. This tool can assist clinicians in identifying high-risk individuals at diagnosis, thereby informing personalized treatment strategies, prioritizing palliative care, and optimizing resource allocation in diverse healthcare settings.

Keywords: Machine learning (ML); metastatic pancreatic cancer; model prediction; prognostic model; early death

Submitted Jun 14, 2025. Accepted for publication Nov 07, 2025. Published online Jan 27, 2026. This article was updated on Apr 20, 2026.

The original version is available at: https://dx.doi.org/10.21037/tcr-2025-1276.

doi: 10.21037/tcr-2025-1276

Introduction

Pancreatic ductal adenocarcinoma (PDAC) is a group of malignant tumors that exhibit a rising mortality trend, characterized by a notably low 5-year survival rate of 6%. The mortality rates closely align with the incidence rates, highlighting the severity of this disease. It is classified among the most invasive and resilient solid tumors, consistently exhibiting resilience to both traditional chemotherapy agents and targeted therapeutics (1). Projections indicate that by 2030, PDAC will become the second leading cause of cancer-related deaths in the US and Europe (2). However, metastatic PDAC (mPDAC) presents an even grimmer prognosis. In excess of 50% of patients are diagnosed with mPDAC, attributable to the absence of typical clinical symptoms and the inadequacy of effective screening methods (3,4). The estimated 5-year overall survival (OS) for mPDAC is approximately 3%, with a median OS frequently below 1 year (5). Generally, current therapeutic options for mPDAC provide a slight increase in survival above palliative care alone, while simultaneously presenting substantial toxicity (5).

Early death is particularly prominent in patients with mPDAC, posing not only significant psychological and social burdens on patients and their families but also presenting substantial challenges for clinicians. Due to the highly aggressive nature and rapid progression of mPDAC, many patients face life-threatening conditions shortly after diagnosis, with a median survival potentially shorter than three months. Such an extremely short survival window often renders traditional treatment strategies ineffective, requiring clinicians to make critical therapeutic decisions within a very limited timeframe (6). Therefore, accurate prediction of early death not only aids in identifying patients who are in urgent need of intervention but also optimizes resource allocation, avoids unnecessary treatments, and enhances overall healthcare efficiency (7). Furthermore, the ability to predict early death provides patients and their families with clearer treatment expectations, enabling them to make more informed medical decisions that align with their personal preferences within the limited time available.

Several studies have sought to develop prognostic models for PDAC using clinical variables, imaging features, and serum biomarkers. Commonly identified predictors include CA19-9 levels, tumor stage, performance status, neutrophil-to-lymphocyte ratio (NLR), and presence of liver metastases (8-10). While existing models—often based on Cox regression or logistic regression (LR)—provide valuable insights, they are primarily designed for general survival estimation rather than specifically predicting early death (e.g., within 3 months of diagnosis). Many of these models also face limitations such as modest accuracy, reliance on linear assumptions, and limited ability to capture complex, non-linear interactions among predictors. Furthermore, most were developed using traditional statistical approaches that may not fully leverage high-dimensional clinical data.

Machine learning (ML) algorithms offer several advantages in this context, including the capacity to model complex, non-linear relationships, handle missing data, and integrate diverse types of predictors without strong prior assumptions (11). These characteristics make ML particularly suited for prognostication in heterogeneous conditions like mPDAC. However, to date, no dedicated ML-based tool has been developed specifically for screening mPDAC patients at high risk of early death.

Therefore, the primary objective of this study is to develop and validate a ML-based algorithm to estimate the probability of early death in mPDAC. By leveraging ML

Highlight box

Key findings

- The XGBoost model accurately predicted early death (within three months of diagnosis) in metastatic pancreatic ductal adenocarcinoma (mPDAC) patients, achieving an area under the curve (AUC) of 0.757. The model demonstrated strong generalizability in an external Chinese cohort (AUC =0.780). Chemotherapy was the most important predictor, followed by age and marital status.

What is known and what is new?

- mPDAC carries a poor prognosis with high rates of early death. Current methods lack accuracy in identifying high-risk patients at diagnosis, limiting opportunities for treatment personalization and timely palliative care.
- This study developed and validated a machine learning model that effectively predicts early death risk in mPDAC patients at diagnosis. Our model represents the first externally validated tool demonstrating robust performance across diverse populations (US and Chinese cohorts), providing a practical solution for clinical risk stratification.

What is the implication, and what should change now?

- This validated tool enables clinicians to stratify mPDAC patients by early death risk at diagnosis. High-risk patients should receive immediate palliative care discussions and avoid aggressive treatments, while lower-risk patients can proceed with standard therapy. Clinical practice should incorporate such data-driven tools to personalize treatment decisions and optimize healthcare resource allocation.

techniques, we aim to improve the accuracy and clinical relevance of prognostic assessment for mPDAC, ultimately supporting more personalized and timely clinical decision-making. We present this article in accordance with the TRIPOD reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-2025-1276/rc).

Methods

Data source and extraction

Information regarding patients diagnosed with mPDAC was obtained from the Surveillance, Epidemiology, and End Results (SEER) database, covering the timeframe from 1st January 2010 till 31st December 2015. The cancer sites analyzed in this research were categorized based on the 3rd edition of the International Classification of Diseases in Oncology (ICD-O-3) as follows: C25.0-Head of the pancreas, C25.1-Body of the pancreas, C25.2-Tail of the pancreas, C25.3-Pancreatic duct, C25.7-Other specified parts of the pancreas, C25.8-Overlapping lesions of the pancreas and C25.9-Pancreas. The study identified patients presenting with pathologies that correspond to ICD-O-3 histology/behavior codes 8140/3 (adenocarcinoma) and 8500/3 (infiltrating duct carcinoma).

The prerequisite for patient inclusion in the study was the initial diagnosis of metastatic PC (pancreatic cancer), confirmed through histopathological examination of the tissue. Patients were not eligible for this study if they met any of the following exclusion criteria: (I) survival time remains unspecified, died due to non-cancer-related factors; (II) alive but survival ≤3 months at the follow-up cut-off date [early death refers to patients who passed away at or within 3 months following the initial diagnosis of the disease (12)]; (III) patients <18 years of age at diagnosis; (IV) information about tumor site, grade, tumor size, metastasis or therapy unavailable; (V) patients with non-pathological diagnosis.

Baseline demographic and clinical characteristics were extracted from the SEER database, including age, sex, race, primary tumor site, tumor grade, tumor size, metastatic sites (e.g., liver, lung, bone, brain), and first-course treatment (surgery, chemotherapy, radiation). Follow-up procedures within SEER are standardized, with active follow-up for vital status conducted through linkage to state vital records, the National Death Index, and other registries. OS was defined as the time from diagnosis to death from any cause. For patients still alive, data were censored at the last follow-up date (December 31, 2015).

To assess the generalizability of the model, an external validation cohort was obtained from the Ningbo University Affiliated Li Huili Hospital in China. This cohort included patients diagnosed with mPDAC between October 1, 2017, and October 31, 2019. The same inclusion and exclusion criteria applied to the SEER cohort were used for patient selection in this external dataset. The use of a Chinese cohort for external validation allows for testing the model's performance across different healthcare systems, genetic backgrounds, and environmental exposures, which is critical for evaluating its robustness and potential for broad clinical application.

Ethics approval

The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments. The study was approved by the Ethics Committee of Ningbo University Affiliated Li Huili Hospital (No. KY2023SL029-01) and informed consent was obtained from all individual participants. Subsequently, a data use agreement was signed for the SEER 1975-2019 study data file, granting us permission to extract and utilize the data.

Potential risk variables and outcome measurement

The analysis encompassed a comprehensive array of 18 potential risk factors, encompassing patient demographics, metastatic status, tumor status, and various treatment interventions. These variables were methodically examined to evaluate their potential influence on the outcomes of the study. Upon confirmed diagnosis of mPDAC, all patients were systematically registered in the database. Early death was defined as patients who succumbed to causes directly associated with their cancer within a period of 3 months after the diagnosis of mPDAC.

Statistical analysis

A two-sided P value <0.05 was considered statistically significant in all analyses. Based on common standards in oncological prediction models, an area under the curve (AUC) of ≥0.75 was considered indicative of good discrimination. Additionally, sensitivity and specificity values both exceeding 70% were targeted to ensure a balanced identification of both high-risk and low-risk patients, though the optimal threshold would be determined based

on clinical priorities (e.g., maximizing sensitivity to avoid missing high-risk cases).

Patients in the SEER cohort initially diagnosed with mPDAC were randomly assigned to the training and internal testing sets, maintaining a 7:3 ratio. We performed stepwise regression (forward-backward selection) first to screen out obviously irrelevant variables from the initial 18 candidate factors, obtaining an initial variable subset. To further refine this subset and mitigate potential biases from stepwise regression, we then applied Least Absolute Shrinkage and Selection Operator (LASSO) regression—using 10-fold cross-validation to determine the optimal regularization parameter (λ, selected as "Lambda Min" to minimize mean squared error). Notably, the final variable set selected by LASSO (with Lambda Min) was identical to the initial subset from stepwise regression, confirming the stability of our selected variables. The final filtered clinical features were used as input features for training and optimizing the model. The training dataset is utilized to train four ML algorithms aimed at predicting early mortality in mPDAC, comprising extreme gradient boosting classifier (XGBoost), Gaussian Naive Bayes (GNB), k-nearest neighbor (KNN) algorithm, and LR. These algorithms were selected based on their proven effectiveness in handling classification tasks and their widespread use in medical research (13-16). Table S1 provides detailed information on the advantages, disadvantages, and applicable scopes of each ML model, while the hyperparameters adjusted for each model are summarized in Table S2.

In order to enhance the robustness of our models, we implemented k-fold cross-validation with 5 folds as the resampling technique. Additionally, we fine-tuned the hyperparameters through the utilization of grid search. The selected methodology involved the utilization of AUC as the optimization objective for hyperparameter tuning. AUC is a well-recognized metric in the context of classification tasks.

In the validation set, the predictive performance of the model was assessed using 7 metrics: AUC, accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1 score. Moreover, subsequent to computing net benefits across diverse threshold probabilities, the assessment of the models' clinical efficacy involved the application of decision curve analysis (DCA). Based on the discriminative power of the models, as measured by AUC, the model demonstrating the best performance was chosen for internal validation. Furthermore, the most optimal ML model was employed for external validation.

In the context of medical research, beyond the consideration of the model itself, a critical focus should be directed towards discerning which variables exert an influence on the manifestation of the disease. SHapley Additive exPlanations (SHAP), is a post hoc explanatory method for models that introduces the concept of Shapley value, which measures the contribution of each feature to an ensemble or coalition, into ML (17). After identifying the optimal ML algorithm, to further elucidate the relationship between various indicators and the target variable, Shapley values for relevant features were calculated using the SHAP method. A detailed flowchart is presented in *Figure 1*.

Results

Baseline characteristics of patients

In our study, a total of 14,820 eligible patients participated. Among these patients, 8,274 cases (55.8%) were over the age of 65, 5,593 cases (37.7%) were between the ages of 50 and 64, and 953 cases (6.4%) were under the age of 50. There were 7,942 male patients (53.6%) and 6,878 female patients (46.4%). According to the 8th edition of the AJCC cancer staging manual, all patients were in the metastatic stage, with 48.7% (7,219/14,820) in stages T3 and T4, and 27.0% (3,998/14,820) in stages T1 and T2. Patients in stage N1 accounted for 33.4% (4,949/14,820), and those in stage N0 accounted for 48.9% (7,253/14,820). A significant portion of patients (48.7%, 7,219/14,820) were found to have extrapancreatic organ invasion at the initial diagnosis. The majority of patients had liver metastases (74.2%, 11,000/14,820), followed by lung metastases (19.6%, 2,911/14,820), and bone metastases (6.8%, 1,012/14,820). Only a very small fraction underwent surgical treatment for the primary tumor (2.9%, 425/14,820) or for distant metastatic lesions (4.6%, 688/14,820). *Table 1* provides a summary of demographic characteristics, therapeutic interventions, and organ metastases. Additionally, it presents the allocation of clinical features within the training and internal test cohorts, demonstrating the comparability of these two groups.

Model development

The variables were systematically screened through stepwise regression to minimize the Akaike Information Criterion (AIC). LASSO regression was applied to further refine the variable selection. Ultimately, demographic factors (age,

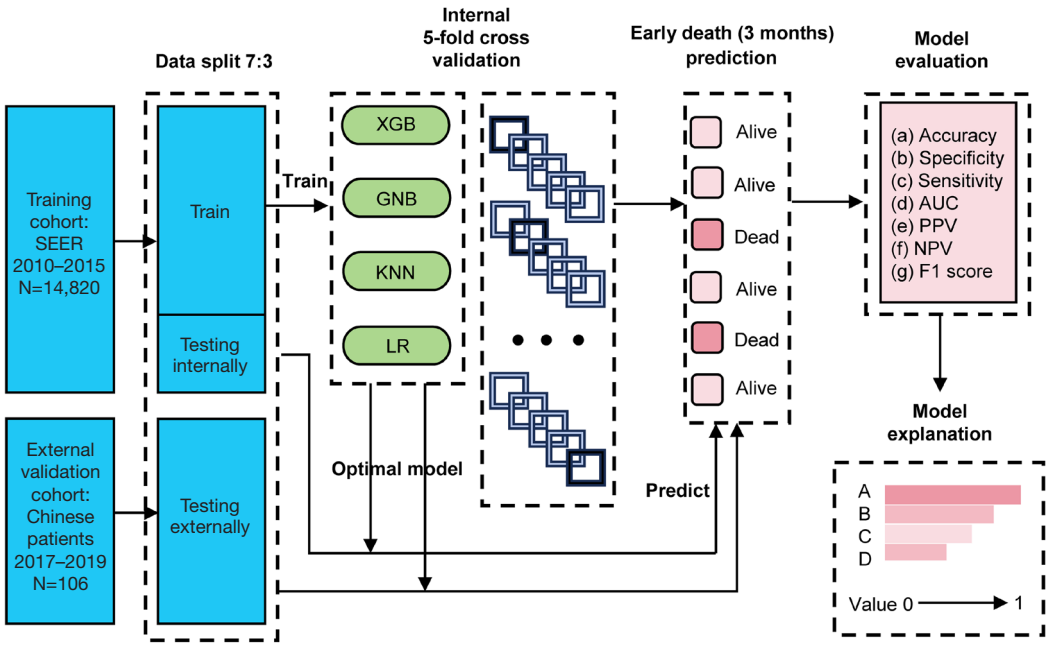

Figure 1 Steps involved in developing the models and a flowchart outlining the study procedure. AUC, area under the curve; GNB, Gaussian Naive Bayes; KNN, k-nearest neighbor; LR, logistic regression; NPV, negative predictive value; PPV, positive predictive value; SEER, Surveillance Epidemiology, and End Results; XGB, extreme gradient boosting.

gender, marital status, income), tumor characteristics (site, differentiation grade, T stage), metastasis status (liver, lung, bone metastasis), and treatment modalities (primary surgery, non-primary surgery, adjuvant radiotherapy/chemotherapy) were observed to exhibit a statistically significant association with early death (*Table 2*, *Figure 2*) and were included as input features for model training and optimization. To assess the potential multicollinearity among the prognostic variables, we generated a correlation heatmap (*Figure 3*). The heatmap illustrates the pairwise correlation coefficients between the key prognostic features included in our analysis. As shown in the heatmap, most variables exhibited weak to moderate correlations, with no strong multicollinearity observed.

Model validation

To ensure the accuracy of the model, we employed a 10-fold cross-validation. The 10-fold cross-validation randomly divided the patients in the training set (n=10,374) into training and validation sets at a ratio of 9:1, and the process was repeated 10 times to compute validation results. In the training set, XGBoost achieved an AUC of 0.764 (*Figure 4A*). Based on the 10-fold cross-validation results in the training cohort, XGBoost demonstrated the best overall performance, achieving the highest AUC (0.757) (*Figure 4B*) and maintaining a balance between sensitivity (0.611) and specificity (0.803). Its ability to capture non-linear relationships and robustness to overfitting, combined with lower variability in performance metrics, makes it particularly suitable for our dataset. In contrast, LR, while competitive in AUC and accuracy, showed slightly lower specificity, and GNB and KNN exhibited lower overall performance and higher variability. *Table 3* presents comprehensive metrics for the validation set, covering accuracy, sensitivity, specificity, PPV, NPV, and F1 score. In addition, to quantitatively assess the performance of our predictive models on an independent validation set, we employed calibration plots and DCA. The results from the calibration plots reveal a high degree of concordance between the predicted probabilities and the actual observed probabilities across the four ML models (*Figure 4C*). The DCA indicates that, compared to strategies of either no intervention (no treatment) or extreme intervention (full treatment), all models demonstrated significant clinical net benefits (*Figure 4D*). When the AUC index on the validation set is lower than the training set and the ratio is less than 10%, it can be considered that the model fitting is

Table 1 Baseline clinical characteristics

Characteristics	Overall (N=14,820)	Training group (N=10,374)	Test group (N=4,446)	P value
Age (years)				0.61
<50	953 (6.4)	673 (6.5)	280 (6.3)	
50–64	5,593 (37.7)	3,889 (37.5)	1,704 (38.3)	
65+	8,274 (55.8)	5,812 (56.0)	2,462 (55.4)	
Sex				0.65
Male	7,942 (53.6)	5,572 (53.7)	2,370 (53.3)	
Female	6,878 (46.4)	4,802 (46.3)	2,076 (46.7)	
Marital status				0.05
Married	8,317 (56.1)	5,759 (55.5)	2,558 (57.5)	
Other	6,503 (43.9)	4,615 (44.5)	1,888 (42.5)	
Race				0.45
White	11,727 (79.1)	8,226 (79.3)	3,501 (78.7)	
Black	1,822 (12.3)	1,278 (12.3)	544 (12.2)	
Other	1,271 (8.6)	870 (8.4)	401 (9.0)	
Primary site				0.56
Head	5,463 (36.9)	3,831 (36.9)	1,632 (36.7)	
Body/tail	5,419 (36.6)	3,812 (36.7)	1,607 (36.1)	
Other	3,938 (26.6)	2,731 (26.3)	1,207 (27.1)	
Grade				0.92
G1/G2	1,542 (10.4)	1,078 (10.4)	464 (10.4)	
G3/G4	1,623 (11.0)	1,143 (11.0)	480 (10.8)	
Unknown	11,655 (78.6)	8,153 (78.6)	3,502 (78.8)	
T stage				0.51
T1	371 (2.5)	250 (2.4)	121 (2.7)	
T2	3,627 (24.5)	2,545 (24.5)	1,082 (24.3)	
T3	4,281 (28.9)	2,968 (28.6)	1,313 (29.5)	
T4	2,938 (19.8)	2,082 (20.1)	856 (19.3)	
TX	3,603 (24.3)	2,529 (24.4)	1,074 (24.2)	
N stage				0.45
N0	7,253 (48.9)	5,043 (48.6)	2,210 (49.7)	
N1	4,949 (33.4)	3,492 (33.7)	1,457 (32.8)	
NX	2,618 (17.7)	1,839 (17.7)	779 (17.5)	
Organ involvement				0.91
Yes	7,219 (48.7)	5,050 (48.7)	2,169 (48.8)	
No	4,459 (30.1)	3,115 (30.0)	1,344 (30.2)	
Unknown	3,142 (21.2)	2,209 (21.3)	933 (21.0)	

Table 1 (*continued*)

Table 1 (*continued*)

Characteristics	Overall (N=14,820)	Training group (N=10,374)	Test group (N=4,446)	P value
Liver metastasis				0.83
Yes	11,000 (74.2)	7,695 (74.2)	3,305 (74.3)	
No/unknown	3,820 (25.8)	2,679 (25.8)	1,141 (25.7)	
Lung metastasis				0.45
Yes	2,911 (19.6)	2,021 (19.5)	890 (20.0)	
No/unknown	11,909 (80.4)	8,353 (80.5)	3,556 (80.0)	
Bone metastasis				0.92
Yes	1,012 (6.8)	707 (6.8)	305 (6.9)	
No/unknown	13,808 (93.2)	9,667 (93.2)	4,141 (93.1)	
Primary surgery				0.95
Yes	425 (2.9)	297 (2.9)	128 (2.9)	
No	14,395 (97.1)	10,077 (97.1)	4,318 (97.1)	
Non-primary surgery				0.58
Yes	688 (4.6)	488 (4.7)	200 (4.5)	
No/unknown	14,132 (95.4)	9,886 (95.3)	4,246 (95.5)	
Chemotherapy				0.43
Yes	9,364 (63.2)	6,534 (63.0)	2,830 (63.7)	
No/unknown	5,456 (36.8)	3,840 (37.0)	1,616 (36.3)	
Radiation				0.65
Yes	866 (5.8)	612 (5.9)	254 (5.7)	
No/unknown	13,954 (94.2)	9,762 (94.1)	4,192 (94.3)	
Metropolitan counties				0.50
Yes	13,093 (88.3)	9,177 (88.5)	3,916 (88.1)	
No	1,727 (11.7)	1,197 (11.5)	530 (11.9)	
Mean household income				0.32
<$60,000 USD	4,776 (32.2)	3,369 (32.5)	1,407 (31.6)	
$60,000 USD+	10,044 (67.8)	7,005 (67.5)	3,039 (68.4)	

Data are presented as number (%). USD, United States dollar.

successful. The learning curve shows that both the training set and validation set possess strong fitting ability and high stability (*Figure 4E*).

The effectiveness of XGBoost in the internal test set was assessed based on the same metrics as above: AUC =0.759 (*Figure 5A*), accuracy =0.710, sensitivity =0.630, specificity =0.782, PPV =0.686, NPV =0.727 and F1 score =0.657. The DCA curve reveals the most advantageous net benefit of the model at 0.32, within a probability threshold range from 0.21 to 0.79. This highlights the model's potential value, as it can provide valuable information for clinical decision-making within this range (*Figure 5B*). The calibration curve is constructed by plotting the predicted probabilities against the ratio of actual occurrences in the internal dataset (*Figure 5C*).

To evaluate the universality and reliability of XGBoost on various datasets, we performed an external validation employing data from 106 Chinese patients (*Figure 5D-5G*). The XGBoost model proposed in this study achieved

Table 2 Univariate and multivariate logistic regression analysis of prognostic factors for early death in mPDAC

Characteristics	Univariate analysis			Multivariate analysis		
	OR	95% CI	P value	OR	95% CI	P value
Age (years)						
<50	Ref			Ref		
50–64	1.48	1.24–1.76	<0.001	1.37	1.13–1.67	<0.001
65+	2.18	1.84–2.58	<0.001	1.77	1.47–2.14	<0.001
Sex						
Male	Ref			Ref		
Female	0.95	0.88–1.02	0.16	0.83	0.76–0.91	<0.001
Marital status						
Married	Ref			Ref		
Other	1.50	1.39–1.62	<0.001	1.32	1.21–1.45	<0.001
Race						
White	Ref			–		
Black	1.16	1.03–1.30	0.01	–		
Other	1.05	0.91–1.21	0.48	–		
Primary site						
Head	Ref			Ref		
Body/tail	1.22	1.12–1.34	<0.001	1.31	1.19–1.46	<0.001
Other	1.56	1.42–1.73	<0.001	1.46	1.30–1.64	<0.001
Grade						
G1/G2	Ref			Ref		
G3/G4	1.63	1.38–1.93	<0.001	1.68	1.38–2.04	<0.001
Unknown	1.57	1.38–1.80	<0.001	1.34	1.15–1.55	<0.001
T stage						
T1	Ref			Ref		
T2	1.40	1.08–1.84	0.01	1.53	1.13–2.06	0.006
T3	1.14	0.88–1.49	0.31	1.42	1.05–1.91	0.02
T4	1.07	0.82–1.41	0.60	1.31	0.97–1.77	0.08
TX	1.73	1.33–2.26	<0.001	1.60	0.97–1.77	0.08
N stage						
N0	Ref			–		
N1	0.91	0.84–1.00	0.04	–		
NX	1.32	1.19–1.47	<0.001	–		

Table 2 (*continued*)

Table 2 (*continued*)

Characteristics	Univariate analysis			Multivariate analysis		
	OR	95% CI	P value	OR	95% CI	P value
Organ involvement						
Yes	Ref			–		
No	1.26	1.15–1.38	<0.001	–		
Unknown	1.53	1.39–1.70	<0.001	–		
Liver metastasis						
Yes	Ref			Ref		
No/unknown	0.64	0.59–0.70	<0.001	0.55	0.50–0.62	<0.001
Lung metastasis						
Yes	Ref			Ref		
No/unknown	0.82	0.74–0.90	<0.001	0.78	0.70–0.88	<0.001
Bone metastasis						
Yes	Ref			Ref		
No/unknown	0.66	0.57–0.77	<0.001	0.57	0.47–0.68	<0.001
Primary surgery						
Yes	Ref			Ref		
No	3.46	2.61–4.65	<0.001	2.91	2.10–4.08	<0.001
Non-primary surgery						
Yes	Ref			Ref		
No/unknown	1.70	1.40–2.06	<0.001	1.26	1.01–1.58	0.04
Chemotherapy						
Yes	Ref			Ref		
No/unknown	6.33	5.79–6.91	<0.001	6.25	5.70–6.86	<0.001
Radiation						
Yes	Ref			Ref		
No/unknown	1.65	1.39–1.96	<0.001	1.42	1.16–1.75	<0.001
Metropolitan counties						
Yes	Ref			–		
No	1.11	0.98–1.25	0.10	–		
Mean household income						
<$60,000 USD	Ref			Ref		
$60,000 USD+	0.91	0.84–0.99	0.02	0.92	0.83–1.01	0.06

CI, confidence interval; mPDAC, metastatic pancreatic ductal adenocarcinoma; N, node; OR, odds ratio; T, tumor; USD, United States dollar.

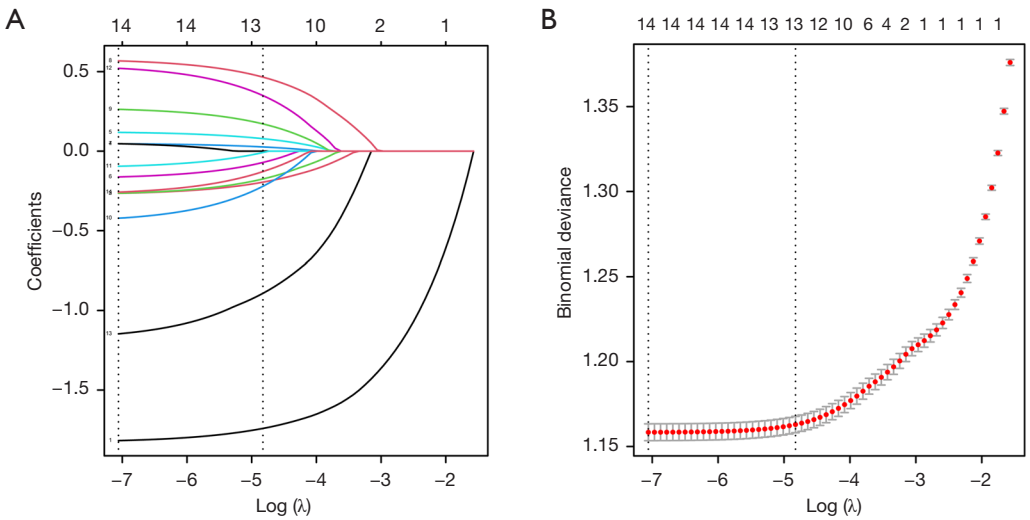

Figure 2 LASSO regression for variable selection: (A) coefficient profiles and (B) cross-validation curve. LASSO, Least Absolute Shrinkage and Selection Operator.

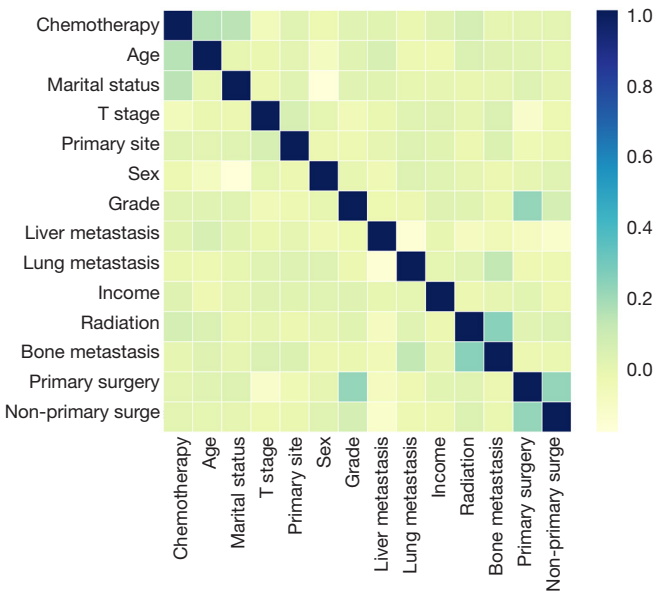

Figure 3 Correlation heatmap of prognostic variables. Colors represent the strength and direction of correlations.

an AUC of 0.78 and an accuracy of 0.71 on the external validation dataset. Based on the risk scores predicted by the XGBoost model, we stratified the external validation cohort of mPDAC patients into high-risk and low-risk groups. Kaplan-Meier survival curves were generated to compare the OS between the two groups (*Figure 6*). Patients in the low-risk group had a median survival time of 5 months [95% confidence interval (CI): 4–6 months], whereas those in the high-risk group had a median survival time of 2 months (95% CI: 2–3 months). These results demonstrate the strong prognostic value of the XGBoost model in stratifying mPDAC patients based on their risk of early death.

Model interpretation and feature importance

Figure 7A presents a SHAP feature summary chart, which

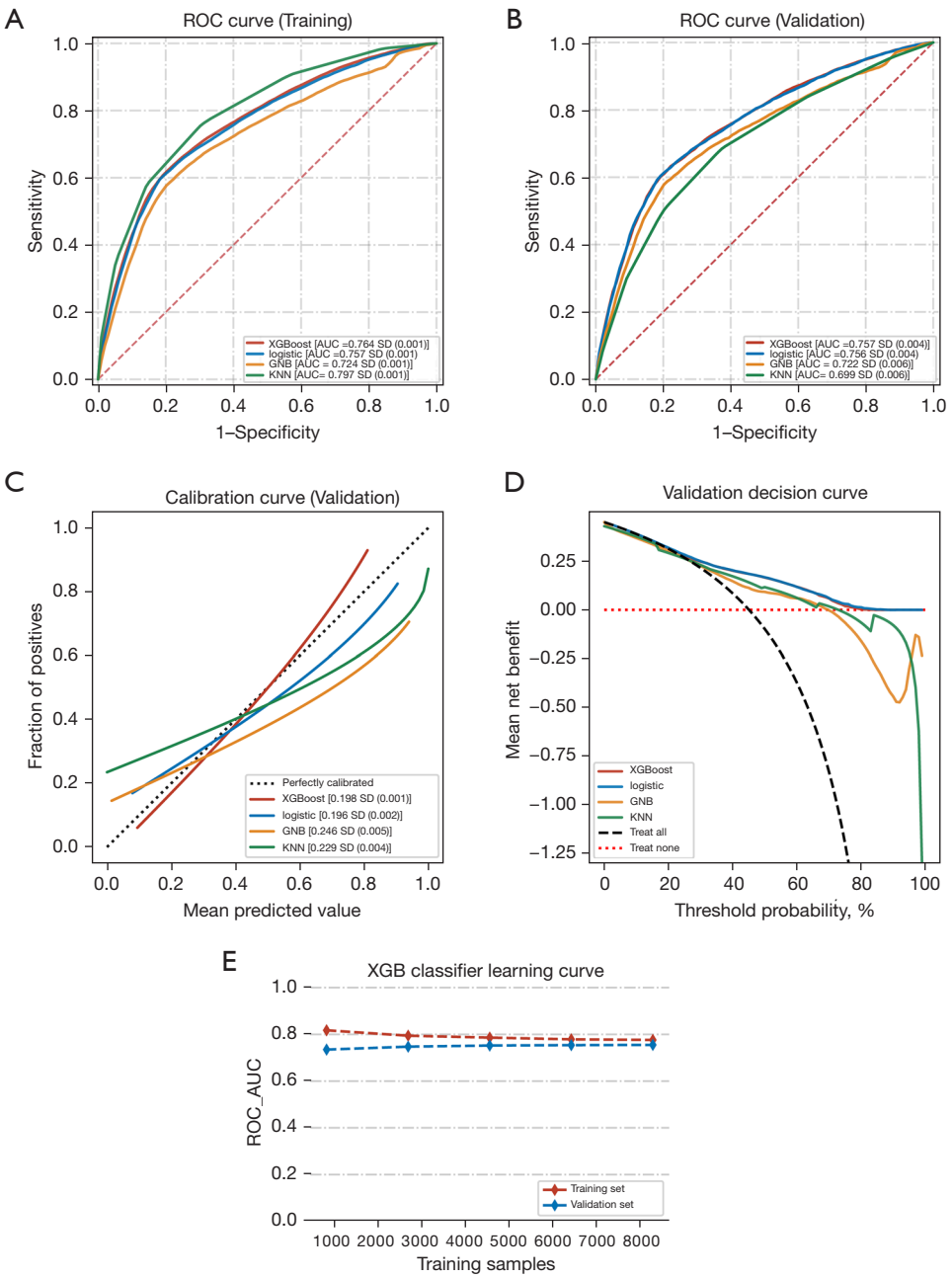

Figure 4 Results of tenfold cross-validation of four ML models in the training cohort: (A) the ROC curves for four models in the training set; (B) the ROC curves for four models in the validation set; (C) the calibration plots for four models; (D) the DCA plots for four models; (E) the learning curve of XGBoost. AUC, area under the curve; DCA, decision curve analysis; GNB, Gaussian Naive Bayes; KNN, k-nearest neighbor; ML, machine learning; ROC, receiver operating characteristic; SD, standard deviation; XGB, extreme gradient boosting; XGBoost, extreme gradient boosting classifier.

Table 3 Performance metrics for four models computed in the validation set

Model	AUC (SD)	Accuracy (SD)	Sensitivity (SD)	Specificity (SD)	PPV (SD)	NPV (SD)	F1 score (SD)
XGB	0.757 (0.004)	0.712 (0.004)	0.611 (0.011)	0.803 (0.014)	0.708 (0.008)	0.716 (0.003)	0.656 (0.009)
LR	0.756 (0.004)	0.715 (0.004)	0.622 (0.018)	0.794 (0.023)	0.720 (0.010)	0.712 (0.005)	0.667 (0.012)
GNB	0.722 (0.006)	0.696 (0.007)	0.585 (0.027)	0.794 (0.017)	0.694 (0.016)	0.697 (0.005)	0.635 (0.020)
KNN	0.699 (0.006)	0.666 (0.006)	0.653 (0.073)	0.662 (0.068)	0.674 (0.012)	0.662 (0.004)	0.661 (0.042)

AUC, area under the curve; GNB, Gaussian Naive Bayes; KNN, k-nearest neighbor; LR, logistic regression; NPV, negative predictive value; PPV, positive predictive value; SD, standard deviation; XGB, extreme gradient boosting.

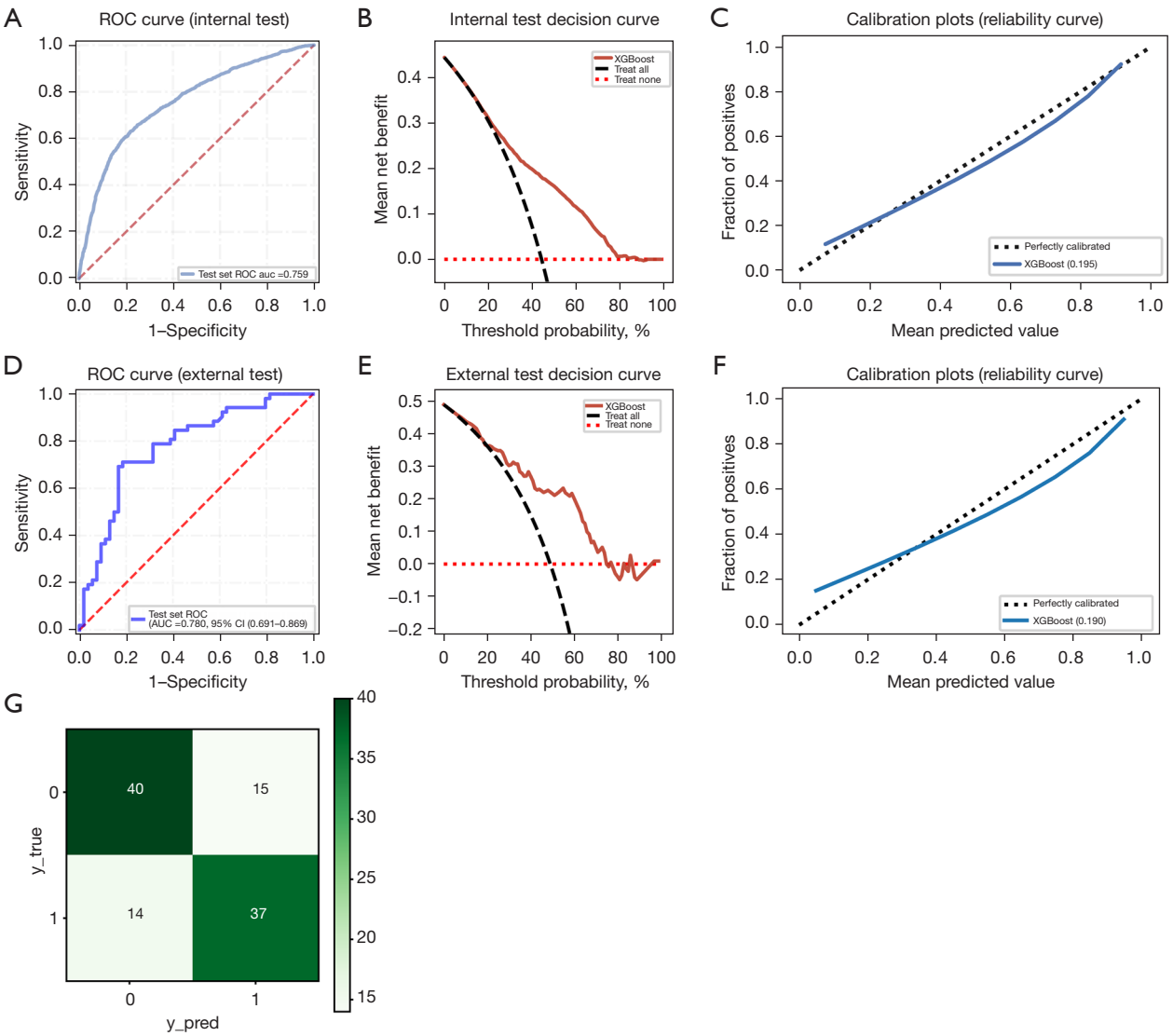

Figure 5 XGBoost's performance in the test sets: (A) the ROC curve of XGBoost for the internal test set; (B) the DCA plot of XGBoost for the internal test set; (C) the calibration plot of XGBoost for the internal test set; (D) the ROC curve of XGBoost for the external test set; (E) the DCA plot of XGBoost for the external test set; (F) the calibration plot of XGBoost for the external test set; (G) the confusion matrix for the external test set. AUC, area under the curve; CI, confidence interval; DCA, decision curve analysis; ROC, receiver operating characteristic; XGBoost, extreme gradient boosting classifier.

analyzes factors affecting early death in patients with mPDAC, organized by the importance of these features. Each point on the feature summary plot represents the Shapley value for a feature in a specific instance. The y-axis of the plot denotes the features, while the position on the x-axis is determined by the Shapley value. The color of the points indicates the magnitude of the feature values. Due to the variation in Shapley values across different samples, overlapping points occur, causing fluctuations along the y-axis. Ranking chart of feature importance of the SHAP method is mainly obtained by averaging the SHAP values of all sample points for each feature and arranging them in descending order from top to bottom. As shown in *Figure 7B*, chemotherapy was the most important feature, followed by age, marital status and T stage.

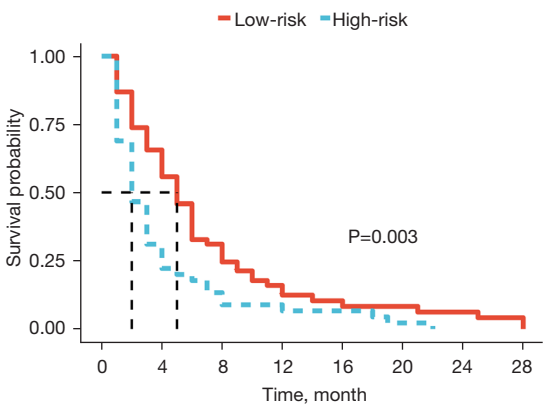

Figure 6 Kaplan-Meier survival curves for mPDAC patients stratified into high-risk and low-risk groups based on XGBoost model risk scores. mPDAC, metastatic pancreatic ductal adenocarcinoma; XGBoost, extreme gradient boosting classifier.

Discussion

Principal findings

The establishment of a reliable method for predicting early death in patients with mPDAC is crucial for timely treatment planning. In this study, we assessed the prognosis of mPDAC patients using clinicopathological features and developed a precise model to evaluate the risk of early death. We systematically evaluated multiple algorithms and found that the XGBoost model demonstrated the highest predictive performance for early death in metastatic patients.

To our knowledge, only a few studies have modelled the prognosis of mPDAC. Compared to existing prognostic tools, our model offers several advantages. First, it is derived from a large, diverse cohort of 14,820 mPDAC patients, ensuring broad generalizability. In contrast, many existing tools, such as the MPS score [The Memorial Sloan Kettering (MSK) Prognostic Score], are based on single-center datasets, limiting their external validity (18). Second, our model incorporates a wide range of clinical

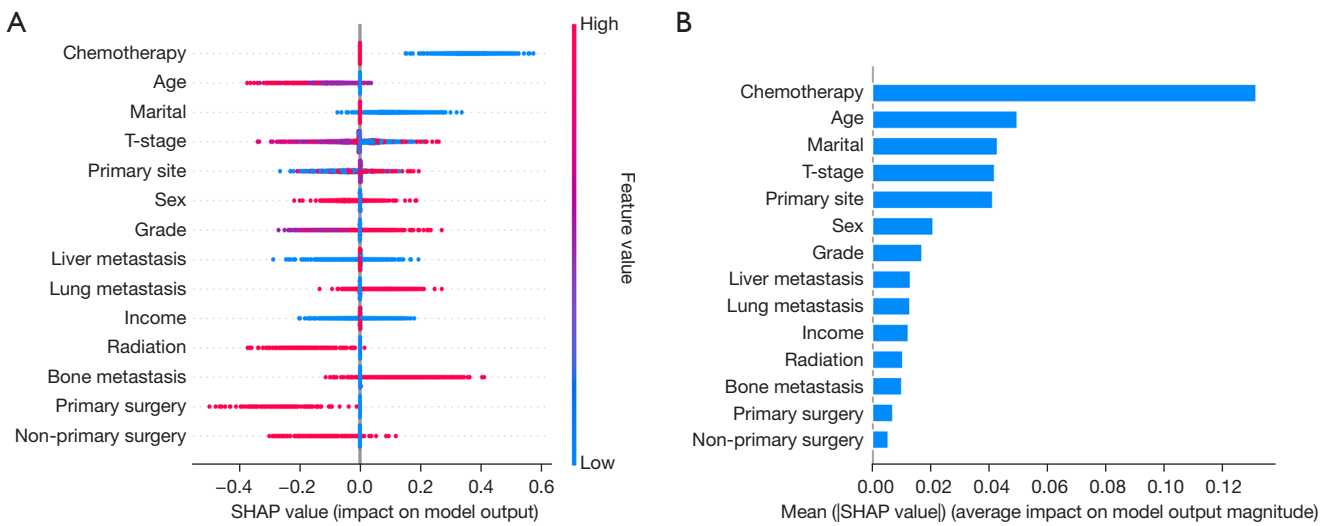

Figure 7 The interpretation and feature importance of XGBoost. (A) The SHAP feature summary chart visually demonstrates the impact of each feature in XGBoost; (B) ranking chart of feature importance. SHAP, SHapley Additive exPlanations; XGBoost, extreme gradient boosting classifier.

and pathological features, making it applicable to a broader patient population regardless of treatment regimen, unlike models that focus on specific therapies (19). Third, our use of ML and SHAP provides interpretable insights into feature contributions, a significant advancement over traditional prognostic tools. Finally, our model's performance was rigorously validated using both internal and external datasets, along with 10-fold cross-validation, ensuring robust and reliable predictions. While our model requires a broader set of features compared to some biomarker-based tools, its comprehensive approach offers superior predictive accuracy and clinical utility, as demonstrated by DCA.

Consistent with previous research, traditional clinical factors such as age, tumour size, staging, surgical resection and adjuvant radiotherapy and chemotherapy will influence patient survival (20,21). According to the results of interpretation from the XGBoost model, the implementation of chemotherapy emerges as the most crucial variable in this study, highlighting the pivotal role of chemotherapy in predicting early death in mPDAC. Chemotherapy remains one of the primary treatment modalities for the majority of patients with mPDAC. A substantial body of literature recommends that patients with mPDAC and an Eastern Cooperative Oncology Group (ECOG) performance status of 0–2 should undergo systemic treatment primarily centered around chemotherapy (22,23). This recommendation is based on the potential of chemotherapy to improve both OS and quality of life. Adjuvant chemotherapy promises to reduce the need for analgesia, delay weight loss and extend the time before quality of life deteriorates (24). The ranking chart of feature importance of the SHAP method indicates that surgical resection ranks as the least influential independent factor. A retrospective cohort study involving 85 patients with oligometastatic PDAC patients suggested that, compared to palliative therapy, cancer-directed surgery was associated with a lower incidence of surgical complications/death risk and is correlated with improved long-term survival rates (25). However, these cancer-directed surgeries for oligometastatic PDAC are often performed incidentally or in carefully selected patients. Furthermore, it is a single-center study with a limited sample size, potentially restricting the accuracy and reliability of these results. Currently, there is still a lack of high-quality and reliable randomized controlled trials (RCT) to provide evidence for the development and updating of guidelines (26). Additionally, we observed that marital status is also a significant predictor of early death in mPDAC, with its importance ranking just below chemotherapy and age. Similarly, Baine *et al.* (27) confirmed that marital status is an independent prognostic factor influencing the perioperative and long-term survival of PC patients. Li *et al.* (28) reported that bereaved patients may experience higher mortality rates due to the loss of social support and the failure of stress response. Marital status significantly influences the survival outcomes of patients with mPDAC, with its effects mediated through multiple interconnected pathways. A key factor is the psychosocial benefits associated with marriage, particularly spousal support, which has been shown to exert a substantial positive impact on long-term health outcomes, as evidenced by studies from Holt-Lunstad *et al.* (29). Conversely, unmarried individuals are more prone to experiencing loneliness and social isolation, factors that have been linked to the adoption of adverse health behaviors, including smoking, as highlighted in research by Shankar *et al.* (30). These behaviors may further deteriorate prognosis in mPDAC patients. Moreover, emerging evidence indicates that psychological stress can drive tumor progression through the induction of specific biological mediators, ultimately compromising the efficacy of cancer therapies (31).

In terms of modeling approaches, this study employed four ML algorithms—XGBoost, LR, GNB, KNN—to train and optimize the model. In comparison to traditional LR models, these sophisticated ML algorithms can leverage large datasets and deliver improved performance (32). Especially when dealing with large datasets, such as the one in our study with tens of thousands of samples, ML algorithms possess the advantageous qualities of flexibility and scalability, rendering them highly appropriate for a diverse range of tasks, including risk stratification and survival estimation (11). In addition to internal and external validation confirming the generalisability of the model, we used the SHAP method to provide a rational explanation for the key variables in the model. The model was developed to evaluate the likelihood of premature mortality in patients diagnosed with mPDAC. Its applicability extends to various time points, depending on the clinical context and the patient's treatment history. For instance, it can be employed to assess the probability of early death in newly diagnosed mPDAC patients, thereby facilitating treatment decision-making. Similarly, for patients who have undergone systemic treatment or are considering treatment options, the model can provide supplementary information regarding the likelihood of premature mortality, aiding in

the selection of the most suitable treatment strategy.

Limitations

Herein, we acknowledge several limitations of our study and outline corresponding directions for future research. First, owing to constraints of the SEER database, our analysis lacks specific biochemical parameters (e.g., CA199), detailed chemotherapy regimen information, and specific palliative surgical data. To address these gaps, we plan to integrate multi-center electronic medical records and surgical databases to supplement relevant biomarker, treatment, and operative details, thereby refining prognostic accuracy and validating model applicability in surgical subgroups. Second, we aim to enhance the model by incorporating multi-omics data—such as plasma CA199 and KRAS mutation status—along with comprehensive electronic health records (e.g., Charlson Comorbidity Index), and to perform external validation in larger, multi-institutional cohorts, including non-Chinese populations, to minimize sampling bias and improve AUC performance. Third, the absence of variables related to social support, treatment adherence, and healthcare access in the SEER database precludes definitive investigation into the proxy effect of marital status; future studies should explicitly include these factors to elucidate underlying mechanisms. Fourth, as the current model was trained on data from 2010–2015, prior to the widespread adoption of novel therapies such as immunotherapy and targeted agents, its predictive utility is limited to patients receiving conventional chemotherapy. We intend to update the model by incorporating data on contemporary treatment regimens and associated biomarkers. Furthermore, the omission of patient-reported outcomes—such as quality of life and symptom burden—limits the model's ability to support patient-centered decision-making where palliative and survival goals intersect. Subsequent iterations will seek to integrate these dimensions to better align predictive outputs with holistic clinical priorities. Finally, the retrospective nature of the study precluded the collection of dynamic in-treatment data (e.g., real-time treatment response, disease progression status), limiting full validation of the model's dynamic predictive performance. Future research will launch a prospective observational study enrolling newly diagnosed mPDAC patients, systematically collecting data at two key time points (at diagnosis and 1 month after treatment initiation) to evaluate: (I) the accuracy of preliminary predictions using pre-treatment features; (II) performance improvements after supplementing post-treatment data; (III) feasibility of real-time clinical application (e.g., prediction time, ease of operation for clinicians); and (IV) the impact of model-guided decision-making (e.g., reducing unnecessary treatment, improving follow-up compliance).

Conclusions

In this study, we developed and validated a ML-based model for accurately assessing the likelihood of early death in mPDAC. The model demonstrated robust performance and interpretability, offering a reliable tool for predicting early mortality in mPDAC patients. Its integration into clinical practice could aid in treatment decision-making and risk stratification, potentially improving patient outcomes. While the model shows promise, further validation in prospective, multi-center studies and exploration of additional biomarkers are needed to enhance its clinical utility. This work represents a significant step toward personalized prognosis prediction and management in mPDAC.

Acknowledgments

We extend our profound gratitude to the SEER program for granting approval for registration and providing access to the SEER database.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-2025-1276/rc

Data Sharing Statement: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-2025-1276/dss

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-2025-1276/prf

Funding: None.

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-2025-1276/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all

aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments. The study was approved by the Ethics Committee of Ningbo University Affiliated Li Huili Hospital (No. KY2023SL029-01) and informed consent was obtained from all individual participants.

References

1. Vaccaro V, Sperduti I, Vari S, et al. Metastatic pancreatic cancer: Is there a light at the end of the tunnel? World J Gastroenterol 2015;21:4788-801.
2. Rahib L, Wehner MR, Matrisian LM, et al. Estimated Projection of US Cancer Incidence and Death to 2040. JAMA Netw Open 2021;4:e214708.
3. Antoniou E, Margonis GA, Sasaki K, et al. Is resection of pancreatic adenocarcinoma with synchronous hepatic metastasis justified? A review of current literature. ANZ J Surg 2016;86:973-7.
4. Strobel O, Neoptolemos J, Jäger D, et al. Optimizing the outcomes of pancreatic cancer surgery. Nat Rev Clin Oncol 2019;16:11-26.
5. Yoo HK, Patel N, Joo S, et al. Health-Related Quality of Life of Patients with Metastatic Pancreatic Cancer: A Systematic Literature Review. Cancer Manag Res 2022;14:3383-403.
6. Han PK, Dieckmann NF, Holt C, et al. Factors Affecting Physicians' Intentions to Communicate Personalized Prognostic Information to Cancer Patients at the End of Life: An Experimental Vignette Study. Med Decis Making 2016;36:703-13.
7. Hui D. Prognostication of Survival in Patients With Advanced Cancer: Predicting the Unpredictable? Cancer Control 2015;22:489-97.
8. Chang S, Liu Y, Liang Y, et al. Biological risk based on preoperative serum CA19-9 and histological grade predicts prognosis and improves accuracy of classification in patients with pancreatic ductal adenocarcinoma. Cancer Rep (Hoboken) 2023;6:e1911.
9. Luan H, He Y, Zhang T, et al. The identification of liver metastasis- and prognosis-associated genes in pancreatic ductal adenocarcinoma. BMC Cancer 2022;22:463.
10. Newhook TE, Vreeland TJ, Griffin JF, et al. Prognosis Associated With CA19-9 Response Dynamics and Normalization During Neoadjuvant Therapy in Resected Pancreatic Adenocarcinoma. Ann Surg 2023;277:484-90.
11. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019;20:e262-73.
12. Wang X, Mao M, Xu G, et al. The incidence, associated factors, and predictive nomogram for early death in stage IV colorectal cancer. Int J Colorectal Dis 2019;34:1189-201.
13. Kleinbaum DG, Klein M. Logistic Regression: a Self-Learning Text. Springer-Verlag Inc., New York; 1994.
14. Zhang H. The Optimality of Naive Bayes. Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004); 2004.
15. Wu X, Kumar V, Quinlan JR, et al. Top 10 algorithms in data mining. Knowl Inf Syst 2008;14:1-37.
16. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016:785-94.
17. Nohara Y, Matsumoto K, Soejima H, et al. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput Methods Programs Biomed 2022;214:106584.
18. Lebenthal JM, Zheng J, Glare PA, et al. Prognostic value of the Memorial Sloan Kettering Prognostic Score in metastatic pancreatic adenocarcinoma. Cancer 2021;127:1568-75.
19. Yu KH, Ozer M, Cockrum P, et al. Real-world prognostic factors for survival among treated patients with metastatic pancreatic ductal adenocarcinoma. Cancer Med 2021;10:8934-43.
20. Fukuda Y, Asaoka T, Maeda S, et al. Prognostic impact of nodal statuses in patients with pancreatic ductal adenocarcinoma. Pancreatology 2017;17:279-84.
21. Vernerey D, Huguet F, Vienot A, et al. Prognostic nomogram and score to predict overall survival in locally advanced untreated pancreatic cancer (PROLAP). Br J Cancer 2016;115:281-9.
22. NICE guideline. Pancreatic cancer in adults: diagnosis

and management. London: National Institute for Health and Care Excellence (NICE) Copyright; 2018.
23. Tempero MA, Malafa MP, Al-Hawary M, et al. Pancreatic Adenocarcinoma, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2021;19:439-57.
24. Gourgou-Bourgade S, Bascoul-Mollevi C, Desseigne F, et al. Impact of FOLFIRINOX compared with gemcitabine on quality of life in patients with metastatic pancreatic cancer: results from the PRODIGE 4/ACCORD 11 randomized trial. J Clin Oncol 2013;31:23-9.
25. Hackert T, Niesen W, Hinz U, et al. Radical surgery of oligometastatic pancreatic cancer. Eur J Surg Oncol 2017;43:358-63.
26. Gebauer F, Damanakis A, Quaas A, et al. Non-Randomised, Open Phase II Trial, Investigating Liposomal Irinotecan and 5-Fluorouraci Followed by Surgical Resection in Patients with Hepatic Oligometastatic Pancreatic Cancer (HOLIPANC). European Journal of Surgical Oncology 2020;46:e6.
27. Baine M, Sahak F, Lin C, et al. Marital status and survival in pancreatic cancer patients: a SEER based analysis. PLoS One 2011;6:e21052.
28. Li Q, Gan L, Liang L, et al. The influence of marital status on stage at diagnosis and survival of patients with colorectal cancer. Oncotarget 2015;6:7339-47.
29. Holt-Lunstad J, Birmingham W, Jones BQ. Is there something unique about marriage? The relative impact of marital status, relationship quality, and network social support on ambulatory blood pressure and mental health. Ann Behav Med 2008;35:239-44.
30. Shankar A, McMunn A, Banks J, et al. Loneliness, social isolation, and behavioral and biological health indicators in older adults. Health Psychol 2011;30:377-85.
31. Shin KJ, Lee YJ, Yang YR, et al. Molecular Mechanisms Underlying Psychological Stress and Cancer. Curr Pharm Des 2016;22:2389-402.
32. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. Crit Care 2019;23:112.

Cite this article as: Zhang L, He J. Development and validation of a machine learning model for predicting early death in metastatic pancreatic ductal adenocarcinoma: a study based on the SEER database. Transl Cancer Res 2026;15(1):53. doi: 10.21037/tcr-2025-1276