BMC Cancer

RESEARCH Open Access

Development of a radiomics-3D deep learning fusion model for prognostic prediction in pancreatic cancer

Zheng Dou[1†], Chenghao Lu[3†], Xinan Shen[5†], Chenqi Gu[4], Yuntian Shen[5], Wei Xu[1], Songbing Qin[2], Jinzhou Zhu[3], Chao Xu[2*] and Jinli Li[2*]

Abstract

Objective With pancreatic cancer's dismal prognosis, developing accurate predictive tools is crucial for personalized treatment. This study aims to develop and evaluate a radiomics-3D deep learning fusion model to enhance survival prediction accuracy and explore its potential for clinical risk stratification in pancreatic cancer patients.

Methods This study retrospectively analyzed from pancreatic cancer patients treated at two hospitals between 2013 and 2023. Patients were split into training and test cohorts (7:3). Baseline clinical data and portal venous phase contrast-enhanced CT images were collected. Two physicians independently delineated tumor regions of interest (ROIs), and 1,037 radiomic features were extracted. After dimensionality reduction via Principal component analysis (PCA) and feature selection with LASSO regression, a radiomics model was developed using the random survival forest (RSF) algorithm to predict overall survival, accounting for censored data. A separate 3D-DenseNet model was trained using ROI-based image inputs to extract deep features. For fusion models, we adopted a binary classification approach to predict survival status at 1-, 2-, and 3-year time points. Radiomics features, 3D-DenseNet outputs, and clinical variables were integrated using logistic regression, random forest, support vector machine, and decision tree classifiers. Model performance was evaluated using receiver operating characteristic (ROC) curves, area under the curve (AUC), and accuracy. The best-performing fusion model was selected for clinical risk stratification. Kaplan-Meier curves and Log-rank tests were used to assess survival differences between risk groups.

Result A total of 880 eligible patients were included in this study. In the test cohort, the performance of each model in predicting 1-year, 2-year, and 3-year survival was evaluated. The radiomics model achieved AUC values of 0.78, 0.85, and 0.91, with corresponding accuracies of 0.75, 0.77, and 0.77. The 3D-DenseNet model demonstrated AUC values of 0.81, 0.79, and 0.75, with accuracies of 0.72, 0.76, and 0.77. The fusion model, developed using logistic regression, exhibited superior predictive performance with AUC values of 0.87, 0.92, and 0.94, and accuracies of 0.84, 0.86, and

†Zheng Dou, Chenghao Lu and Xinan Shen contributed equally to this work.

*Correspondence:
Chao Xu
xuchao87@suda.edu.cn
Jinli Li
ljl2049@163.com

Full list of author information is available at the end of the article

0.89, outperforming the individual unimodal models. Risk stratification based on the fusion model categorized patients into high-risk and low-risk groups, revealing a statistically significant difference in OS between the two groups ($P < 0.001$). Feature contribution analysis indicated that the 3D-DenseNet model had the greatest influence on the predictions of the fusion model, followed by the radiomics model.

Conclusion This study developed a fusion model incorporating radiomics features, deep learning-derived features, and clinical data, which outperformed unimodal models in predicting survival outcomes in pancreatic cancer and demonstrated potential utility in patient risk stratification.

Keywords Pancreatic cancer, Radiomics, Deep learning, 3D-DenseNet, Prognostic prediction

Introduction

Pancreatic cancer is one of the most aggressive and lethal malignancies globally, with a 5-year relative survival rate of only 12.8% [1]. In 2022, it ranked 9th in cancer-related mortality in China, with the 7th highest number of deaths [2]. The asymptomatic nature of the disease in its early stages often leads to late diagnosis, typically when the tumor is already in advanced stages. Therefore, timely prognostic prediction after diagnosis is crucial for improving survival rates in pancreatic cancer patients. Stratifying patients based on mortality risk before treatment can assist physicians in developing personalized treatment plans and provide important guidance for subsequent relapse monitoring and follow-up. Developing accurate and reliable prognostic prediction methods for pancreatic cancer remains an urgent challenge.

Tumor staging and differentiation play a critical role in determining the prognosis of pancreatic cancer patients. However, the prognostic factors for pancreatic cancer are multifactorial, and even among patients with the same stage of disease, there can be significant variation in recurrence rates and survival outcomes [3]. Recent studies have identified numerous histopathological features, clinical factors, and biological fluid biomarkers that could serve as prognostic predictors for pancreatic cancer [4, 5]. However, many of these factors require invasive procedures or are expensive, which limits their clinical utility. Moreover, common prognostic factors may be affected by interobserver variability, and their sensitivity and specificity can vary significantly. In this context, radiomics, as a technique for extracting quantitative features from medical imaging [6], has shown potential in prognostic prediction of tumors [7, 8].

The concept of radiomics was first formally introduced by Lambin et al. in 2012, and later further refined and expanded in 2017 [9, 10]. Radiomics utilizes computational algorithms to extract a wide range of high-dimensional, quantifiable features from medical imaging, such as shape, texture, intensity distribution, and contrast, which can reflect the tumor's microstructure and biological behavior, and subsequently assist in disease diagnosis and treatment [11].

In recent years, the development of 3D deep learning technology has further advanced this field. Traditional radiomics methods typically rely on two-dimensional imaging data and predefined feature extraction techniques, which may overlook the full spatial heterogeneity of the tumor. In contrast, 3D deep learning models can automatically learn complex spatial features from the entire tumor volume in an end-to-end manner, eliminating the need for handcrafted features and enhancing consistency, scalability, and predictive accuracy [12, 13]. These models are typically built upon architectures such as DenseNet, ResNet, or hybrid frameworks incorporating attention mechanisms or transformers [14]. Such approaches are capable of capturing both local texture and global spatial relationships within volumetric data, providing a richer and more comprehensive representation of tumor morphology. Furthermore, 3D deep learning is applicable to a wide range of medical imaging tasks, including prognosis prediction, tumor segmentation, detection, and classification, thereby offering strong potential for integration into clinical decision support systems [15].

This study aims to compare the performance of radiomics models and 3D deep learning models in prognostic prediction for pancreatic cancer patients. We propose a fusion model that combines the predictive results of radiomics, 3D deep learning, and clinical features to further enhance the accuracy of prognostic prediction. Through this study, we hope to provide a more effective tool for pancreatic cancer risk stratification and clinical decision-making, ultimately improving patient management and outcomes.

Methods

Patients

The study was approved by the Ethics Committee of the First Affiliated Hospital of Soochow University. A total of 880 patients diagnosed with pancreatic cancer between January 1, 2013, and December 31, 2023, were retrospectively collected from two hospitals, including 520 patients from the First Affiliated Hospital of Soochow University and 360 patients from the Second Affiliated Hospital of Soochow University. All patients received

follow-up treatment in accordance with the CSCO guidelines for the diagnosis and treatment of pancreatic cancer [16]. The inclusion criteria were: (1) Histopathological diagnosis of pancreatic cancer; (2) Abdominal contrast-enhanced CT examinations performed at our hospital before definitive treatment. The exclusion criteria were: (1) Lack of complete baseline clinical data or follow-up data; (2) Combined with a history of other malignant tumors; (3) Poor quality of contrast-enhanced CT imaging, defined as scans with significant motion artifacts, insufficient contrast enhancement (i.e., poor visualization of pancreatic parenchyma and vessels during the portal venous phase), image blurring affecting tumor boundary visibility, or incomplete coverage of the pancreas within the scanning range. Patients were divided into a training set ($n=616$) and a testing set ($n=264$) in a 7:3 ratio. The specific selection process for the study cohort is shown in Fig. 1. Demographic characteristics such as gender and age were collected from electronic medical records. Tumor clinical staging was assessed according to the 8th edition of the AJCC (American Joint Committee on Cancer) staging system. Overall survival (OS) was followed up through telephone interviews and inpatient medical records, defined as the time from initial pathological diagnosis to death from any cause or the last known date of survival (Figure S 4). The follow-up was censored on June 1, 2024. OS was binarized as "survival" or "death" for machine learning classification. Death within 1-year, 2-year, and 3-year post-diagnosis was labeled as "death," while survival at these time points was labeled as "survival." Patients without follow-up information were already excluded according to the predefined exclusion criteria.

Imaging protocol and preprocessing

All patients underwent contrast-enhanced abdominal CT scans before initiating treatment. Imaging was performed using a 256-slice CT scanner, the GE Revolution CT from GE, USA, at the First Affiliated Hospital of Soochow University, and a 64-slice helical CT scanner, the GE Discovery CT 750 HD, at the Second Affiliated Hospital of Soochow University. Scan parameters were similar at both hospitals, with a tube voltage of 120 kV and a slice thickness and interslice gap of 5 mm. The Second Affiliated Hospital used automatic milliampere modulation, while the First Affiliated Hospital set the tube current between 200 and 450 mAs. Both hospitals used iodixanol 320 as the contrast agent, administered at a dose of

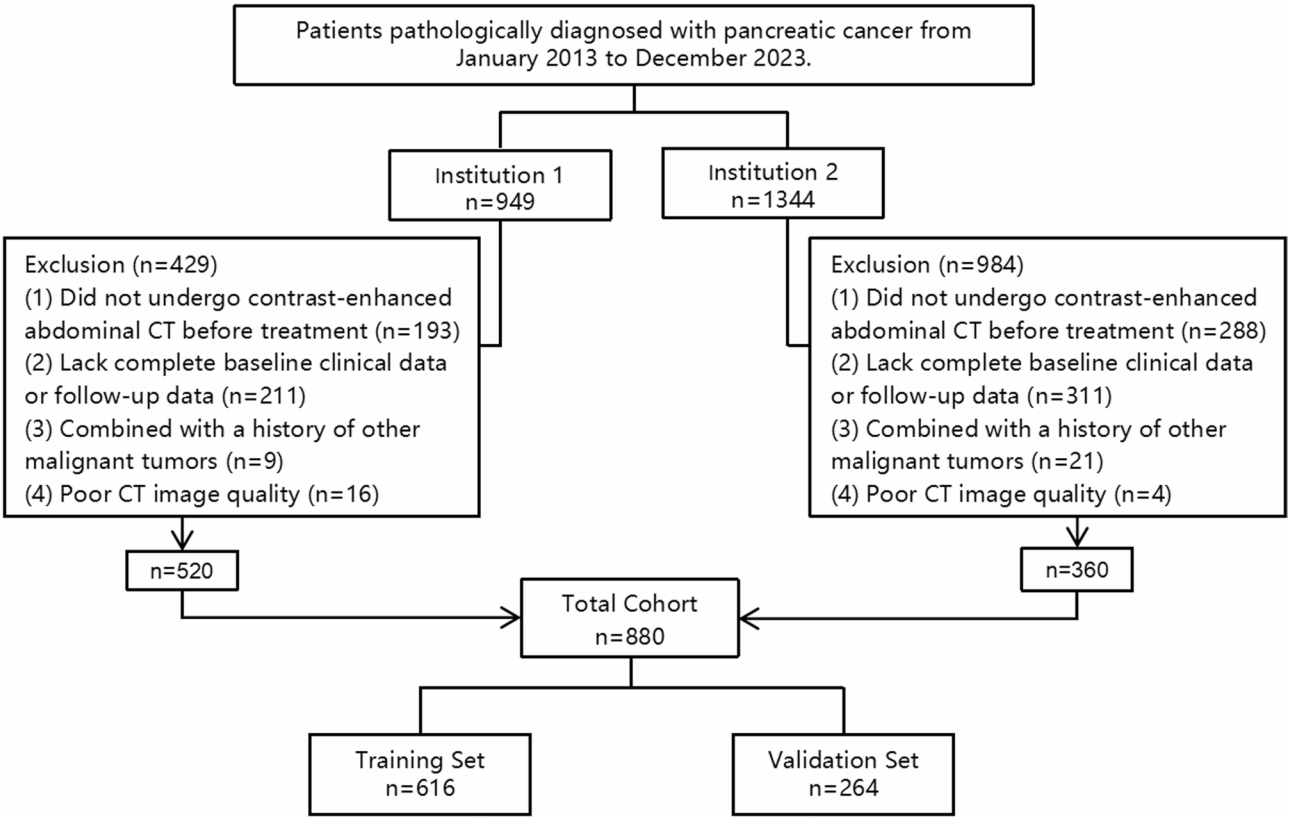

Fig. 1 Flowchart diagram shows the patient selection process from two medical centers. A total of 880 patients were included in this study and randomly divided into a training set ($n=616$) and a test set ($n=264$) at a ratio of 7:3. *Institution 1: The First Affiliated Hospital of Soochow University; Institution 2: The Second Affiliated Hospital of Soochow University

approximately 1.4 mL per kilogram of body weight at an injection rate of 3 mL per second. Arterial phase images were acquired using an automatic triggering method, with a threshold of 100 Hounsfield units. Venous phase images were obtained after a 60-second delay.

Portal venous phase contrast-enhanced CT images in DICOM format were imported into 3D Slicer software (version 5.6.1). Guided by a senior radiotherapy oncologist (20 years of experience), a radiotherapy oncologist and a gastroenterologist (each with 3 years of experience) independently delineated tumor boundaries to create regions of interest (ROIs), carefully avoiding adjacent vessels, dilated bile ducts, and pancreatic ducts. Both physicians were blinded to the clinical outcomes prior to the ROI segmentation. If the two physicians produced significantly different ROIs for the same patient, a third radiologist (20 years of experience) reviewed the case and made the final decision in accordance with NCCN (National Comprehensive Cancer Network) guidelines. To assess the reliability and consistency of the ROI delineation, the intraclass correlation coefficient (ICC) was used for validation.

Radiomic feature extraction

Before feature extraction, the images were resampled using 3D Slicer software to a voxel size of $3 \times 3 \times 3$ mm^3 to ensure isotropy, thereby minimizing variability due to differences in scanning equipment and protocols. Radiomic features were then extracted using the Radiomics plugin within the 3D Slicer software. A total of 1,037 features were extracted, including 108 original features and 929 filtered features (LoG and wavelet), which were classified into seven categories: shape features, first-order (FO) features, features based on gray-level co-occurrence matrices (GLCM), gray-level dependence matrix (GLDM) features, features based on run-length matrices (RLM), features based on size-zone matrices (SZM), and features based on neighborhood gray-tone difference matrices (NGTDM).

The development and evaluation of models

Radiomics based model

To reduce the dimensionality of features and address the issue of multicollinearity, principal component analysis (PCA) was applied to the extracted radiomic features. We selected principal components that together explained 80% of the cumulative variance to minimize redundancy while retaining the most informative aspects of the original feature set. Subsequently, feature selection was further refined using the Least Absolute Shrinkage and Selection Operator (LASSO) regression, with the optimal λ value determined via 10-fold cross-validation. The final input features were determined by retaining principal components with non-zero coefficients.

The Random Survival Forest (RSF) algorithm, an ensemble tree-based method, is designed to analyze survival time data and predict individual survival probabilities or risks. RSF constructs models by integrating multiple survival trees, providing high stability and robustness while effectively reducing the risk of overfitting. Additionally, it captures complex non-linear relationships and automatically identifies key features closely associated with survival outcomes, thereby enhancing the interpretability of the model [17, 18]. In recent years, RSF has been widely applied to survival analysis problems [19]. In this study, the RSF algorithm was utilized to construct models in the training set for predicting 1-, 2-, and 3-year survival probabilities of patients. The RSF model was implemented using the randomForestSRC package in R. The number of trees (ntree) was set to 345, and the minimum terminal node size (nodesize) was optimized using the tune.nodesize() function, which selects the optimal value based on the out-of-bag (OOB) prediction error. The number of random splits (nsplit) was set to 11. The discriminatory ability of the models was evaluated using the receiver operating characteristic (ROC) curve and its area under the curve (AUC), while prediction consistency was assessed through accuracy measurements.

3D-Densenet based model

Given the importance of capturing three-dimensional structural information in medical imaging, we developed a deep learning model based on a three-dimensional densely connected convolutional network (3D-DenseNet) to predict 1-, 2-, and 3-year survival outcomes in pancreatic cancer patients. The 3D-DenseNet model was implemented using the MONAI (Medical Open Network for AI) framework, which is specifically designed for medical imaging analysis and provides robust deep learning tools optimized for healthcare applications.

The previously segmented tumor regions (ROI) were standardized to a fixed size of $96 \times 96 \times 96$ voxels using linear interpolation to ensure uniform input dimensions and reduce computational complexity. To address the limitations of small sample datasets and enhance model robustness, image preprocessing was performed, including pixel intensity normalization, explicit channel dimension addition, and data type standardization. During the training phase, data augmentation techniques such as random 90-degree rotations, axis flipping, and Gaussian noise injection were applied to improve the model's generalization capability.

The 3D-DenseNet architecture consisted of four densely connected blocks containing 6, 12, 24, and 16 convolutional layers, respectively. By leveraging dense connectivity, the model reused features from previous layers, improving feature learning efficiency and

reducing the number of parameters. To extract both global and local features, a dual-input design (DualInputDenseNet) was adopted, integrating a whole-image branch and an ROI branch. The whole-image branch utilized 3D-DenseNet to extract global features, while the ROI branch employed two 3D convolutional layers, each followed by a ReLU activation function and max pooling operation, to progressively downsample and extract high-level ROI features. The outputs of both branches were concatenated along the channel dimension and passed through fully connected layers for feature fusion and final classification. The model was optimized using the Adam optimizer and trained for 100 epochs with a learning rate of 1×10^{-4}. In each epoch, a batch loader was used to retrieve training data in batches. Cross-entropy loss was employed as the optimization objective, and model performance was monitored using test accuracy computed on the hold-out test set at each epoch. The model checkpoint that achieved the highest test set accuracy was saved for final evaluation. During the testing phase, data augmentation steps were disabled, and only normalization and resizing were applied to ensure stability and consistency in the evaluation process.

After training, the optimal 3D-DenseNet model was used to predict 1-, 2-, and 3-year survival probabilities for patients in both the training and test sets. AUC was calculated to assess the model's discriminative ability. Additionally, accuracy was evaluated to measure classification consistency.

Fusion model

To further enhance the performance of the prognostic prediction model, this study built a fusion model by performing decision-level fusion of the radiomics model and the 3D-DenseNet model. The predictions from both models were combined with clinical features of patients (gender, age, T stage, N stage, and M stage) to predict the 1-, 2-, and 3-year survival outcomes. This approach aimed to fully leverage the different types of data to enhance the model's predictive ability, thereby improving its generalizability and accuracy. To identify the optimal fusion model, four different machine learning algorithms were employed, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and Decision Tree. The model performance was evaluated using accuracy and AUC.

After determining the optimal fusion model, patients were stratified based on mortality risk scores predicted by the model. Specifically, the optimal risk cutoff value was identified using X-tile software (version 3.6.1), and all patients were classified into low-risk and high-risk groups accordingly. Kaplan-Meier (KM) survival curves for the high-risk and low-risk groups were then plotted separately in both the training and test cohorts. A log-rank test was used to evaluate the statistical significance of survival differences between the two groups. Figure 2 provide an overview of the overall study workflow.

Statistical analysis

All statistical analyses were performed using SPSS 26.0, R software (version 4.3.2), and Python 2.7.5. The demographic and clinical characteristics between the training and test sets were compared using SPSS 26.0. For continuous variables, an independent sample t-test was applied if the data followed a normal distribution, while the Mann-Whitney U test was used for non-normally distributed variables. Categorical variables were compared using the chi-square test. All statistical tests were two-sided, and a p-value < 0.05 was considered statistically significant. Feature selection, model construction, and evaluation of radiomics-based predictive models were performed using R (version 4.3.2). The "glmnet" package was employed for LASSO regression to select significant radiomic features. The "randomForestSRC" and "survival" packages were used to develop survival prediction models, while the "ggplot2" package was utilized for data visualization. The 3D deep learning model was constructed and evaluated using Python 2.7.5. The remaining general statistical analyses were conducted using R software.

Result

Patient characteristics

A total of 880 patients were included in this study, with 616 patients in the training cohort and 264 patients in the test cohort. The baseline characteristics of both cohorts are summarized in Table 1. The median follow-up time was 36 months for the training cohort and 29 months for the test cohort. During the follow-up period, 448 patients (72.7%) in the training cohort and 203 patients (76.9%) in the test cohort were confirmed deceased. In the training cohort, there were 354 male and 262 female patients, while the test cohort included 147 male and 117 female patients. The mean age of patients in the training cohort was 68.68 years, whereas the mean age in the test cohort was 66.39 years. There were no statistically significant differences in baseline characteristics between the two cohorts ($p > 0.05$).

Consistency analysis

Contrast-enhanced CT images from 50 randomly selected pancreatic cancer patients were independently segmented by two researchers. The interobserver consistency of tumor boundary delineation was quantitatively assessed using the intraclass correlation coefficient (ICC). The results showed an ICC value of 0.81 between the two researchers, indicating good reliability and consistency in tumor boundary segmentation. As shown in Figure S 1.

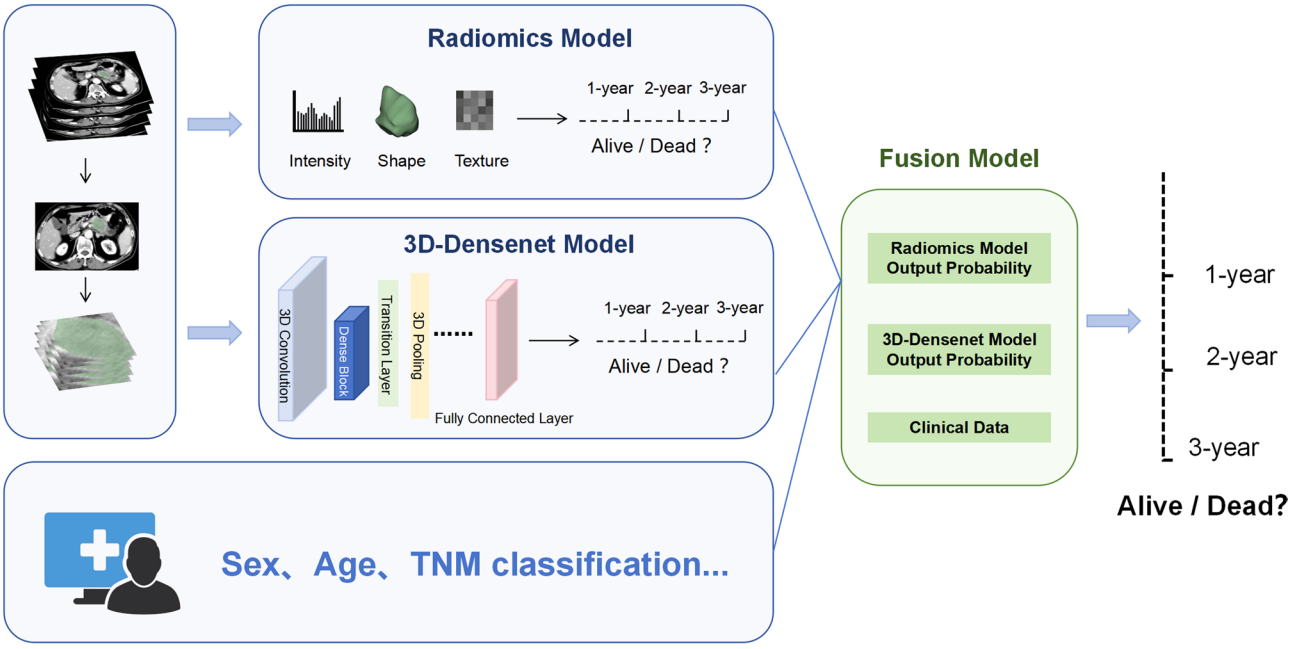

Fig. 2 Workflow diagram for developing a predictive model. Tumor segmentation and region of interest (ROI) delineation were performed by two physicians. Quantitative features, including intensity, shape, and texture, were extracted from the ROI to train the radiomics model, while a deep learning model was trained based on the 3D-DenseNet framework. On this basis, a fusion model was constructed by integrating the predicted 1-year, 2-year, and 3-year survival probabilities from the radiomics model and 3D-DenseNet models along with baseline clinical characteristics of the patients. *3D-DenseNet, three-dimensional densely connected convolutional networks

Table 1 Baseline characteristics of study sets

	Group	Training set ($n=616$)	test set($n=264$)	*p*-value
		Mean (Median) or Count (Percentage)	Mean (Median) or Count (Percentage)	
Age		68.68 (68)	66.39 (67)	0.42
Gender				
	Male	354 (57.50%)	147 (55.70%)	0.66
	Female	262 (42.50%)	117 (44.30%)	
T Stage				0.29
	1	53 (8.6%)	22 (8.3%)	
	2	290 (47.10%)	127 (48.10%)	
	3	143 (23.20%)	48 (18.20%)	
	4	130 (21.1%)	67 (25.40%)	
N Stage				0.08
	0	267 (43.30%)	93 (35.20%)	
	1	235 (38.10%)	118 (44.70%)	
	2	114 (18.50%)	53 (20.10%)	
M Stage				0.08
	0	456 (75.50%)	184 (69.70%)	
	1	151 (24.50%)	80 (30.30%)	
Follow-up Time		46.73 (36)	44.71 (29)	0.19

$P<0.05$ indicates statistical significance

Radiomics feature selection

A total of 1,037 radiomic features were extracted for each patient. Dimensionality reduction was performed using PCA, retaining the top 21 principal components that cumulatively explained 80% of the total variance (Figure S 2). The 80% threshold was chosen as it is a widely recognized standard in previous literature [10, 20], ensuring maximal information retention while effectively eliminating redundant features. Subsequently, LASSO regression was applied, identifying nine significant features with nonzero coefficients. The optimal regularization parameter (λ) was determined using 10-fold cross-validation within the training set, selecting the value corresponding to the minimum mean cross-validated error. Figure 3 illustrates the coefficient path and cross-validation curve for feature selection using LASSO.

Performance of the radiomics model and 3D deep learning model

Using the nine principal components selected through PCA and LASSO regression, we developed a survival prediction model for 1-, 2-, and 3-year survival outcomes using RSF. Figure 4 presents the ROC curves for predicting 1-, 2-, and 3-year survival, with AUC values of 0.78, 0.85, and 0.91, respectively. The corresponding accuracy rates were 0.73, 0.77, and 0.77 (Table 2).

In the test cohort, the 3D-DenseNet deep learning model achieved AUC values of 0.81, 0.79, and 0.75

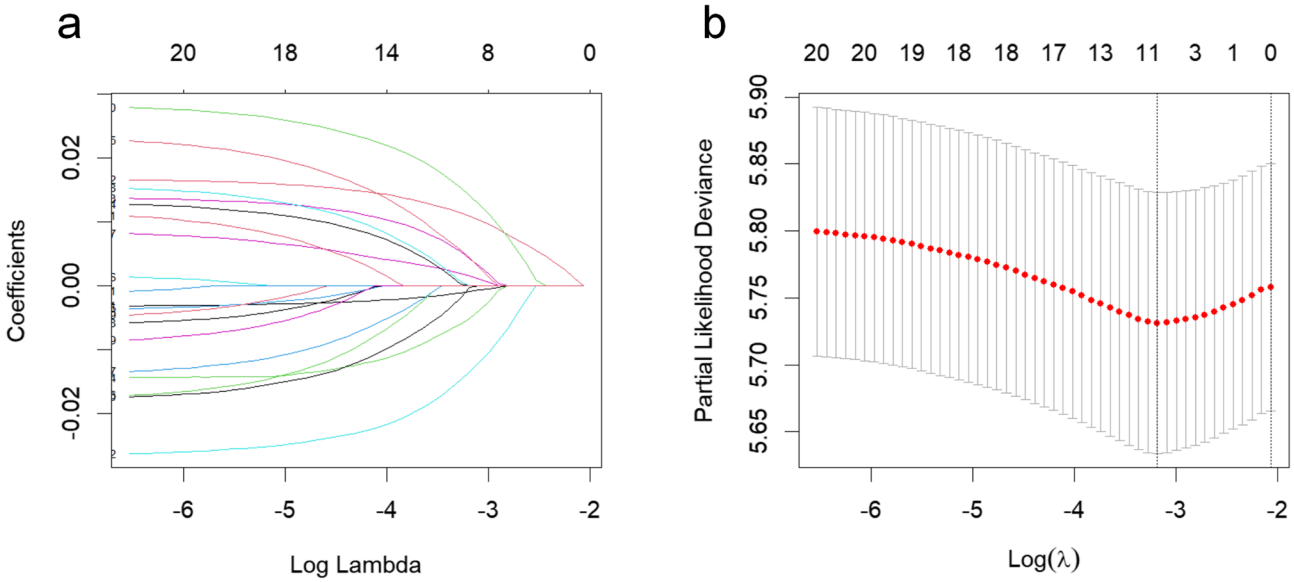

Fig. 3 The results of LASSO regression analysis. **A** LASSO coefficient profiles of the expression of 21 variables. **B** Selection of the λ in the regression analysis via 10-fold cross-validation. The dotted vertical lines are plotted at the optimal values following the minimum criteria (right) and "one standard error" criteria (left).

for predicting 1-, 2-, and 3-year survival, respectively. The corresponding accuracy rates were 0.72, 0.76, and 0.77. Its performance was slightly inferior to that of the radiomics model (Table 2). Figure 5 presents the ROC curves of the 3D-DenseNet model for predicting 1-, 2-, and 3-year survival in both the training and test cohorts.

Performance of the fusion model and feature contribution

Figure 6 presents the ROC curves of fusion models constructed using different algorithms, with detailed performance metrics summarized in Table 2. Among them, the fusion model based on logistic regression demonstrated the best performance in the test cohort, achieving AUC values of 0.87, 0.92, and 0.94 for predicting 1-, 2-, and 3-year survival, respectively. The corresponding accuracy rates were 0.84, 0.86, and 0.89, making it the best-performing fusion model.

Finally, Figure 7 compares the ROC curves of the optimal fusion model based on logistic regression, the radiomics model, and the 3D-DenseNet model in the test cohort. The results indicate that the fusion model demonstrated superior predictive performance compared to the radiomics model and the 3D-DenseNet model for 1-, 2-, and 3-year survival predictions.

By analyzing the contribution of each feature in the fusion model, we obtained the feature importance ranking (Figure 8). The results showed that the prediction output from the 3D-DenseNet model contributed the most to the predictive performance of the fusion model, with a significantly higher Mean Decrease Gini value than other features. This was followed by the prediction output from the radiomics model, which also played a key role in the fusion model's risk assessment. Patient age and T stage ranked third and fourth, respectively, indicating that clinical characteristics also had a certain impact on the risk assessment capability of the fusion model.

Patient risk stratification using the optimal fusion model

To assess the clinical applicability of the optimal fusion model, patients were stratified based on the mortality risk predicted by the model. Using X-tile software, a risk score cutoff of 0.8 was determined to maximize the intergroup difference, classifying all patients into low-risk and high-risk groups (Figure S 3). KM survival curves were plotted for both groups in the training and test cohorts, showing clear separation (Figure 9), indicating a significant difference in survival outcomes between the two groups. The Log-rank test confirmed that the survival differences were statistically significant in both the training and test cohorts ($p < 0.0001$).

Discussion

This study developed a multimodal fusion model integrating radiomics and 3D deep learning for prognosis prediction in pancreatic cancer patients. The fusion model demonstrated significantly better performance compared to the standalone radiomics model and the 3D-DenseNet model. Risk stratification analysis showed that the model effectively distinguished patients into high-risk and low-risk groups, with a significant survival difference between the two.

Radiomics, which extracts high-throughput features from medical images, provides a novel tool for prognosis prediction, treatment response evaluation, and biological

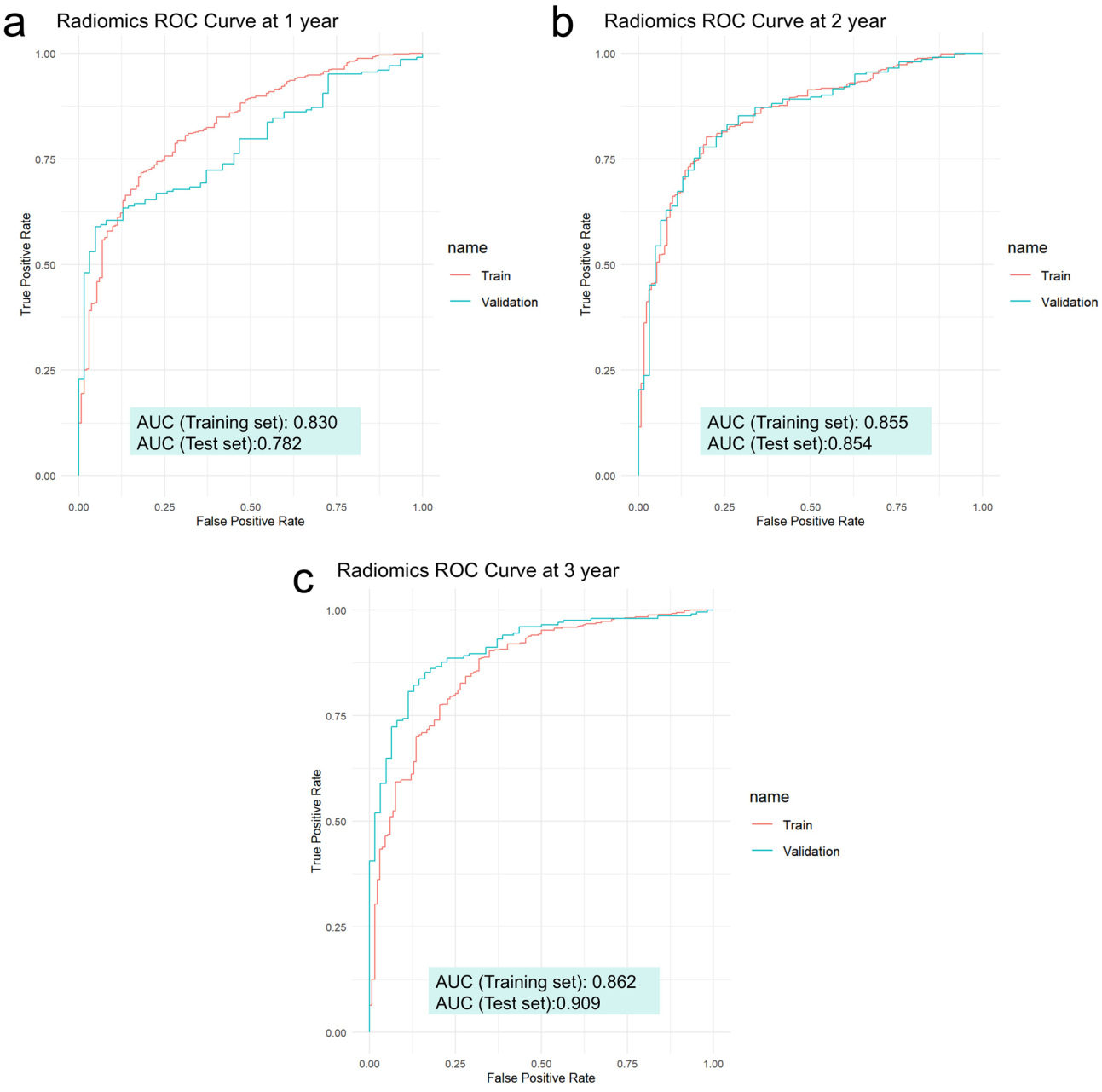

Fig. 4 Performance of survival prediction. The ROC curves of the radiomics model in the training and test sets. **a** Model predicting 1-year survival; **b** Model predicting 2-year survival; **c** Model predicting 3-year survival. * AUC, the Area Under the Receiver Operating Characteristic Curve

behavior analysis of malignant tumors. In recent years, radiomics has gained increasing attention in pancreatic cancer research. Numerous studies [21–23] have demonstrated the significant advantages of radiomics in the diagnosis and differential diagnosis of pancreatic cancer. Additionally, Borhani et al. [24] found that texture features from CT images could predict early responses to neoadjuvant chemotherapy in pancreatic cancer patients. Nasief et al. [25] reported that combining CA19-9 with radiomic features improved the prediction of treatment outcomes in pancreatic cancer, highlighting the potential of radiomics in evaluating treatment response. Parr et al. [26] demonstrated that radiomics-based models outperformed traditional clinical models in predicting overall survival and recurrence. Similarly, Cen et al. [27] developed a nomogram model based on radiomic features from contrast-enhanced CT, achieving high performance in preoperative staging and survival prediction of pancreatic ductal adenocarcinoma (PDAC). Our study highlights the clinical utility of radiomics in pancreatic cancer prognosis, with an RSF-based model accurately

Table 2 The AUC and accuracy of different models in predicting 1-year, 2-year, and 3-year survival outcomes in the training and test sets

		1 Year		2 Year		3 Year	
		AUC	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**
Training Set	Radiomics Model	0.831	0.739	0.855	0.805	0.862	0.786
(*n* = 616)	3D-DenseNet Model	0.821	0.729	0.837	0.774	0.789	0.779
	Fusion Model						
	Logistic Regression	0.916	0.862	0.941	0.877	0.941	0.89
	Random Forest	0.965	0.904	0.973	0.907	0.979	0.932
	SVM	0.922	0.867	0.949	0.888	0.953	0.903
	Decision Tree	0.901	0.877	0.888	0.898	0.92	0.906
Test Set	Radiomics Model	0.782	0.754	0.854	0.765	0.909	0.765
(*n* = 264)	3D-DenseNet Model	0.808	0.716	0.785	0.758	0.749	0.773
	Fusion Model						
	Logistic Regression	0.874	0.841	0.919	0.864	0.944	0.886
	Random Forest	0.843	0.822	0.925	0.864	0.938	0.886
	SVM	0.852	0.845	0.927	0.867	0.932	0.867
	Decision Tree	0.763	0.803	0.791	0.852	0.912	0.860

**AUC* The Area Under the Receiver Operating Characteristic Curve, *SVM* Support Vector Machine

predicting 1-, 2-, and 3-year survival, reinforcing its role in guiding personalized treatment.

Compared with traditional 2D methods, 3D models can more comprehensively analyze the overall morphology of tumors, boundary characteristics, and their spatial relationships with surrounding tissues. Additionally, they better align with the inherent three-dimensional structure of common medical imaging modalities such as CT and Magnetic Resonance Imaging (MRI). By processing multiple consecutive slices simultaneously, 3D models help prevent information loss [28]. Deep learning enables the automatic extraction of high-level features directly from raw images, eliminating the subjectivity and limitations associated with manually designed features [29]. As a result, 3D deep learning has become a research hotspot in medical image analysis and precision medicine. It has shown great potential in various areas, including tumor diagnosis, prognosis prediction, and treatment response assessment [30, 31]. A retrospective study [32] proposed a multi-modal deep learning method based on 3D convolutional neural network (3D-CNN), which utilizes multi-channel MRI images combined with support vector machine (SVM) to predict the survival of high-grade glioma patients. The method achieved an accuracy of 90.66%, outperforming traditional radiomics-based approaches. Similarly, Joo S et al. [33] designed a deep neural network that combined ResNet-50 with 3D-CNN for MRI analysis, significantly improving the prediction of pathological complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients (AUC = 0.888). To our knowledge, systematic research on the application of 3D deep learning in pancreatic cancer remains limited. In our study, we developed a prognostic prediction model based on 3D-DenseNet using contrast-enhanced abdominal CT images of pancreatic cancer patients. While the model demonstrated good predictive performance, it showed slightly lower AUCs than the radiomics model for 2- and 3-year survival predictions, though it performed better for 1-year prediction. Overall, its average performance was slightly inferior to the radiomics-based model. A study by Kim H et al. [34] reported similar findings, employing a 2D-DenseNet to analyze axial, coronal, and sagittal sections of lung CT scans. They developed a model termed the "2.5D" model, which outperformed the 3D-DenseNet in lung cancer classification. Several factors may explain this finding. First, 3D deep learning models require large datasets for optimal training, and the limited sample size in this study may have constrained the model's learning capacity. Second, radiomics models leverage handcrafted features based on prior knowledge and clinical experience, enhancing interpretability and specificity. In contrast, deep learning models extract features automatically, which, while capturing complex patterns, may introduce redundancy or overlook key prognostic factors, potentially affecting predictive accuracy. To address these limitations, we integrated both models' predictive outputs, leveraging the interpretability of radiomics with the automated feature extraction of deep learning. The resulting fusion model outperformed single-modal models, highlighting its potential for improving pancreatic cancer prognosis. Feature contribution analysis based on the Gini importance ranking showed that predictions from the 3D-DenseNet and radiomics models had the greatest influence on the fusion model's output. This suggests that imaging-derived features may provide greater prognostic value for pancreatic cancer survival than traditional clinical variables such as gender, age, or TNM staging. In

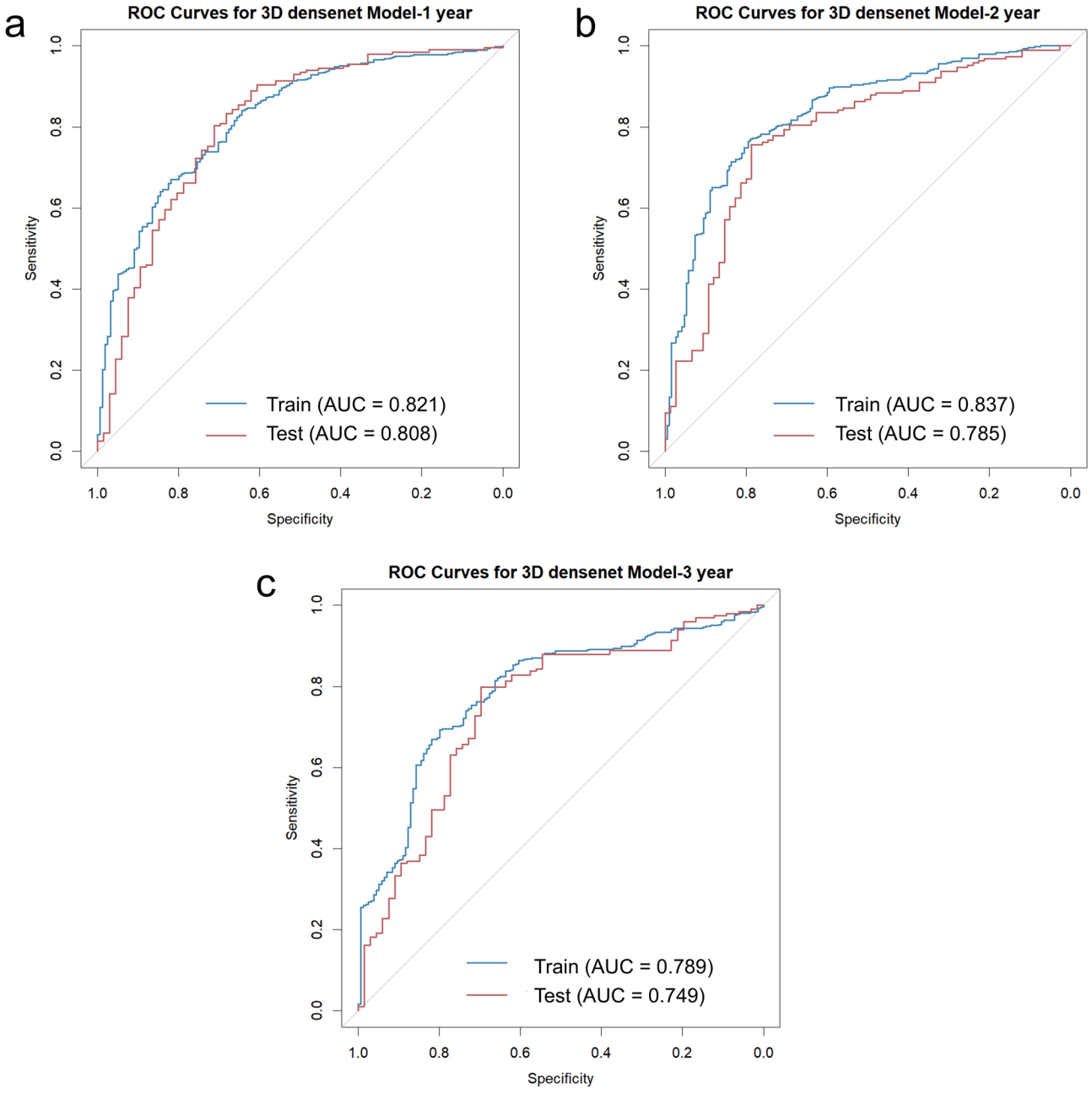

Fig. 5 Performance of survival prediction. The ROC curves of the 3D-DenseNet model in the training and test sets. **a** Model predicting 1-year survival; **b** Model predicting 2-year survival; **c** Model predicting 3-year survival. *AUC, the area under the receiver operating characteristic curve

future work, we will explore multimodal fusion strategies incorporating additional imaging modalities (e.g., MRI or PET-CT), particularly when such modalities offer complementary anatomical or functional information. However, we acknowledge that incorporating more data types does not automatically guarantee better performance and should be supported by specific clinical rationale and validated through comparative studies.

A systematic review compared the performance of radiomics models, deep learning (DL) models, and multimodal fusion models in medical imaging. The results demonstrated that fusion models exhibited superior performance in 63% of the studies, underperformed in 25% of the studies, and showed comparable performance to other models in 13% of the studies [35]. These variations may be influenced by the different fusion strategies employed. Feature-level fusion (early fusion) and decision-level fusion (late fusion) are two common strategies in multimodal data analysis. Feature-level fusion integrates multimodal features (such

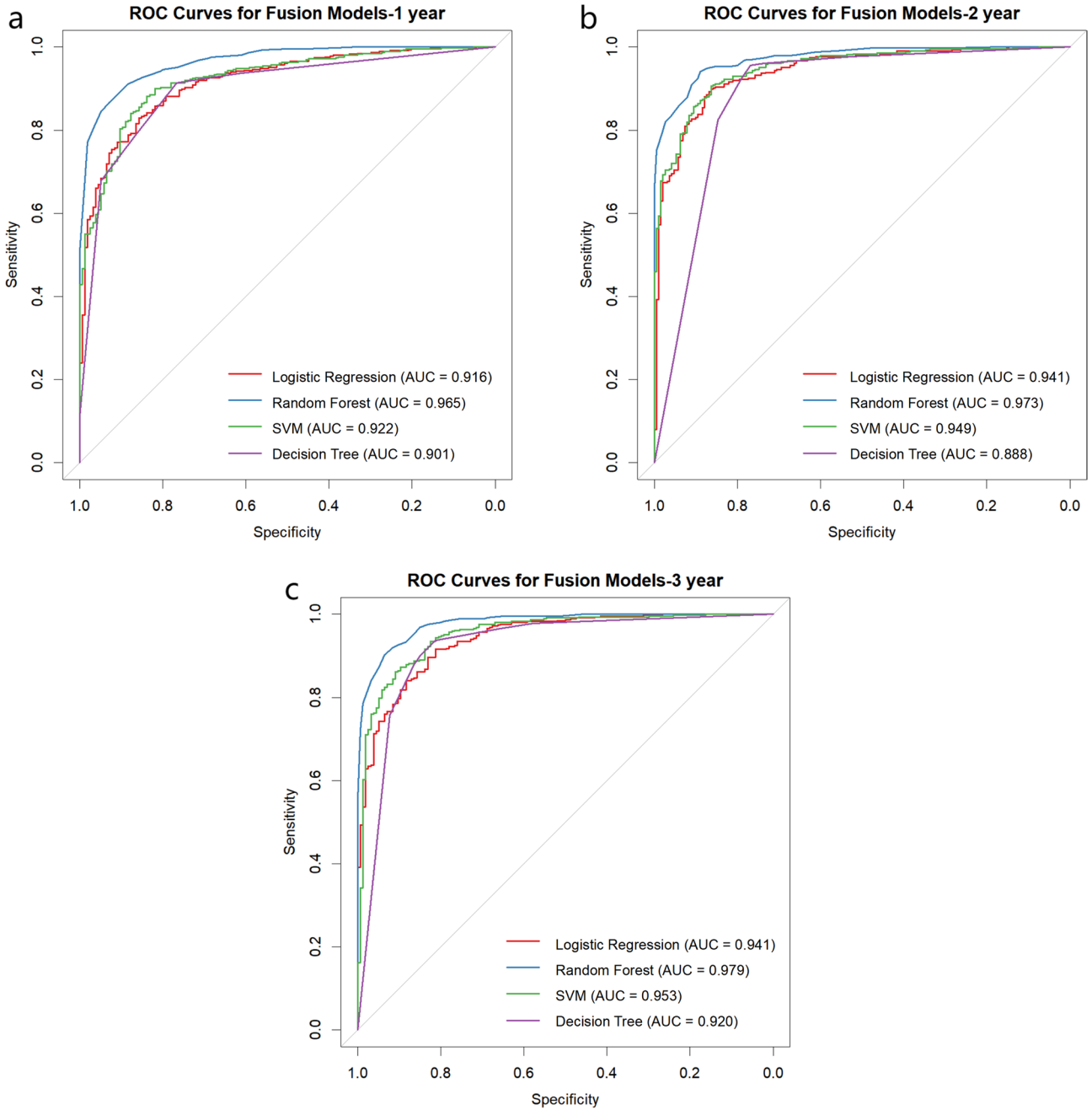

Fig. 6 Performance of survival prediction. The ROC curves of the fusion model based on different algorithms in the training and test sets. **a** Model predicting 1-year survival; **b** Model predicting 2-year survival; **c** Model predicting 3-year survival. *AUC, the area under the receiver operating characteristic curve; svm: support vector machine.

as radiomics, deep learning, and clinical data) into a single feature vector for classification, capturing complex interactions but facing challenges like high dimensionality and sensitivity to noise. Decision-level fusion, on the other hand, combines predictions from different models, offering flexibility and reducing overfitting risk, though it may lead to a loss of feature interactions [36]. Given that our study primarily relies on imaging-based features (radiomics and 3D deep learning), which differ significantly in nature and dimensionality, we adopted a decision-level fusion strategy to avoid introducing feature-level noise and to maintain model interpretability. Moreover, since the radiomics and deep learning models were already optimized independently, their outputs were well-suited for decision-level integration. This strategy is particularly advantageous when the feature distributions across modalities are heterogeneous, the feature spaces are high-dimensional, and early fusion

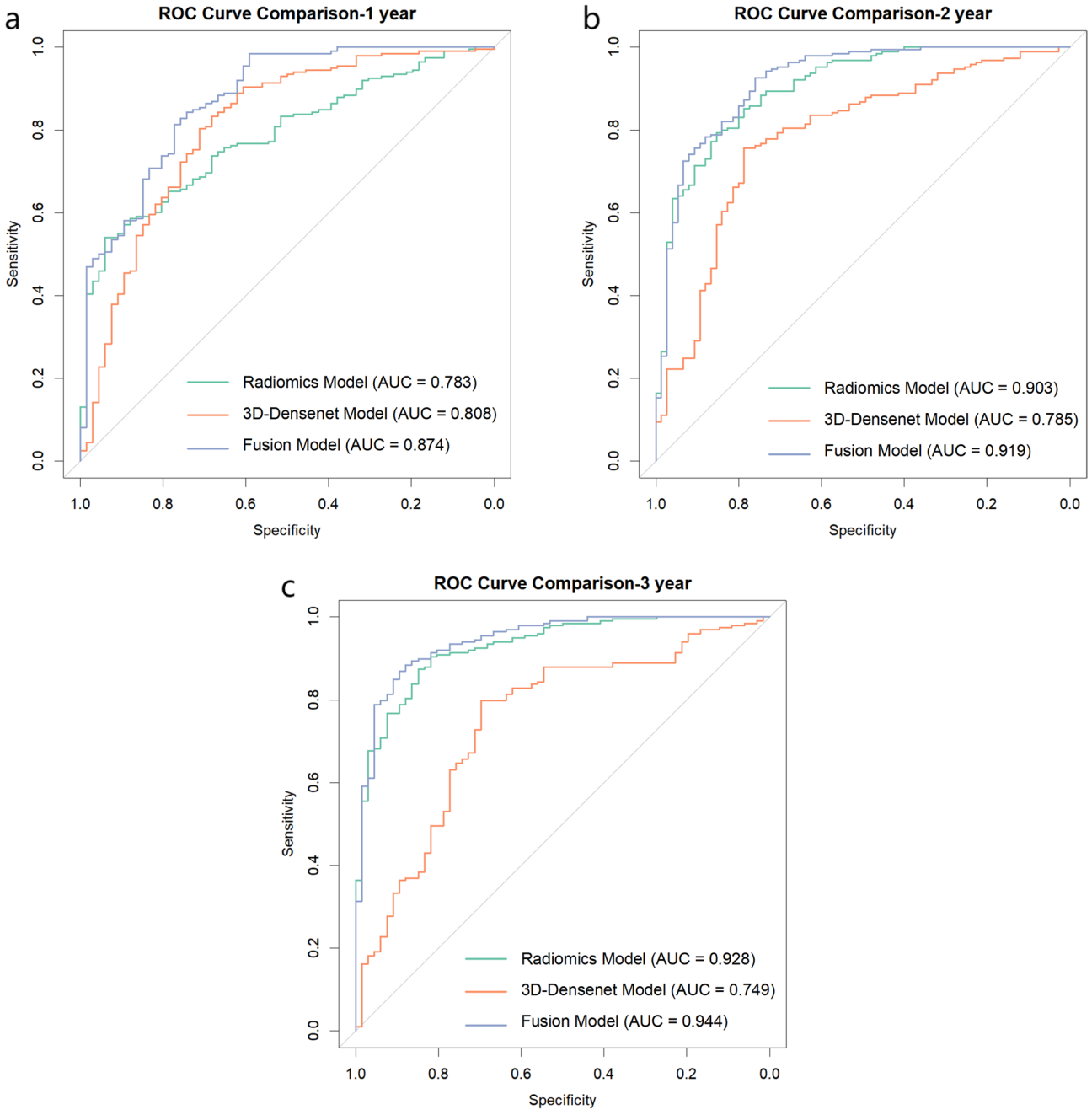

Fig. 7 Performance of survival prediction. The ROC curves of the radiomics model, 3D-DenseNet model, and fusion model in the test set. **a** Model predicting one-year survival; **b** Model predicting two-year survival; **c** Model predicting three-year survival. *AUC, the area under the receiver operating characteristic curve

may lead to overfitting due to increased model complexity. Decision-level fusion also allows for more flexible model designs and enables each modality to contribute without being constrained by strict alignment or normalization requirements. Although previous studies in other disease domains have demonstrated the effectiveness of decision-level fusion [37, 38], the choice of fusion strategy remains an open research question in machine learning. Our decision was therefore based on a combination of theoretical considerations—such as mitigating overfitting risk and preserving modular model structure—and empirical evidence supporting its feasibility in medical imaging tasks.

Our study has certain limitations. First, as a retrospective study, it may be subject to selection bias and difficulty in fully controlling confounding factors. Second, the sample size is relatively small, particularly the test cohort, which may affect the model's generalizability

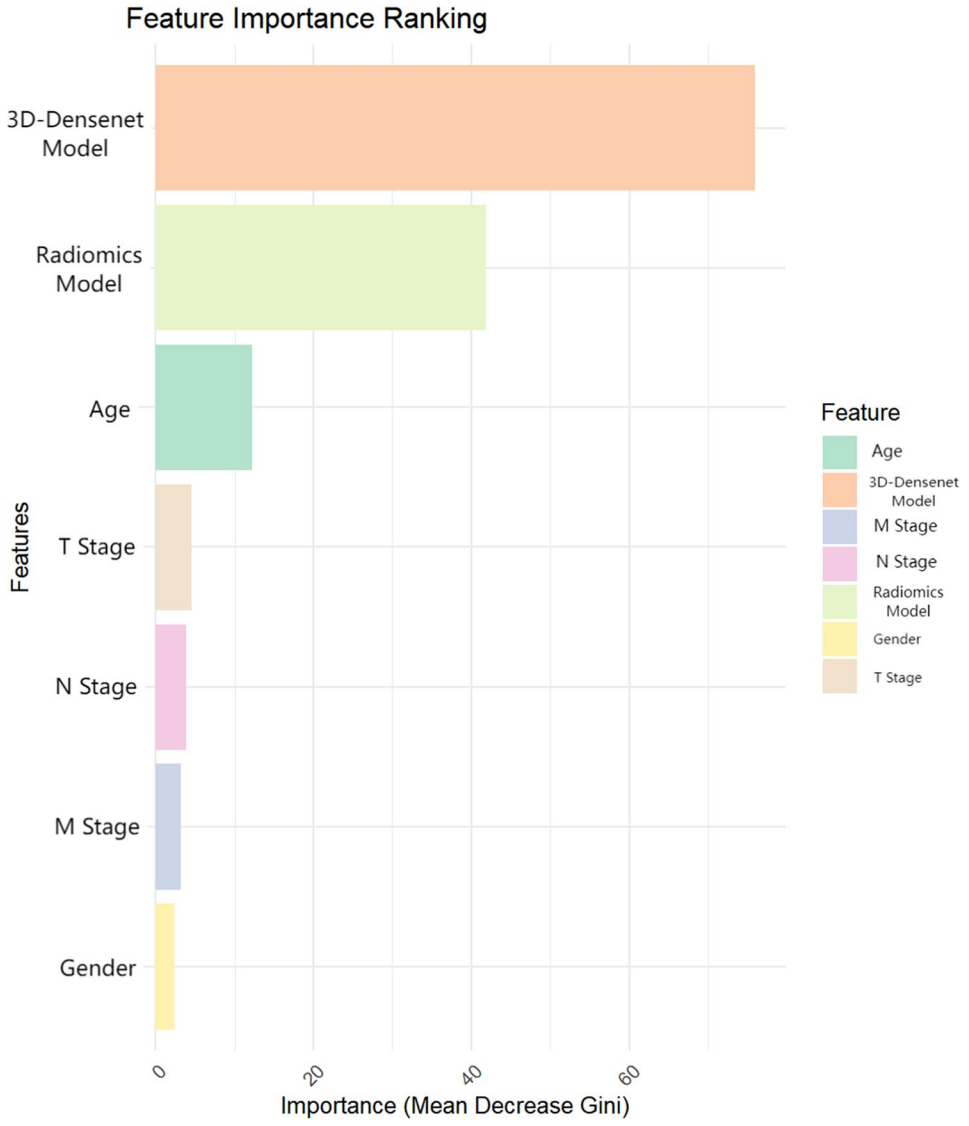

Fig. 8 Feature importance ranking in the fusion model. The x-axis represents feature importance, while the y-axis lists the features analyzed. The color legend distinguishes different feature categories. Features with higher values contribute more to the model's predictive performance

and clinical applicability. Third, although data were collected from two independent medical centers, we did not designate one center as an external test set. Preliminary experiments showed that using one center solely for testing led to a significant performance drop, likely due to inter-center differences in imaging protocols and population characteristics. As a result, we chose to combine both datasets and perform stratified random sampling to construct the training and test sets. While this approach improves stability and robustness, it limits our ability to assess generalizability across truly independent settings. Future research should incorporate external datasets from additional institutions to further validate the model's clinical utility. This study integrated CT-based radiomics features with selected clinical data but did not include hematological tests, pathological characteristics, or genomic data. This decision was primarily driven by the aim of developing an imaging-centered risk stratification tool to minimize invasive procedures and additional laboratory tests, thereby reducing patient burden and healthcare costs. By focusing on imaging data and basic clinical characteristics, we aimed to construct a more streamlined and widely applicable model for clinical practice.

Conclusion

The fusion model based on the logistic regression algorithm, integrating radiomics, 3D deep learning, and clinical features, demonstrated the best performance in predicting the prognosis of pancreatic cancer patients, with the highest prediction accuracy and strong discriminative ability, effectively stratifying patients into low-risk and high-risk groups.

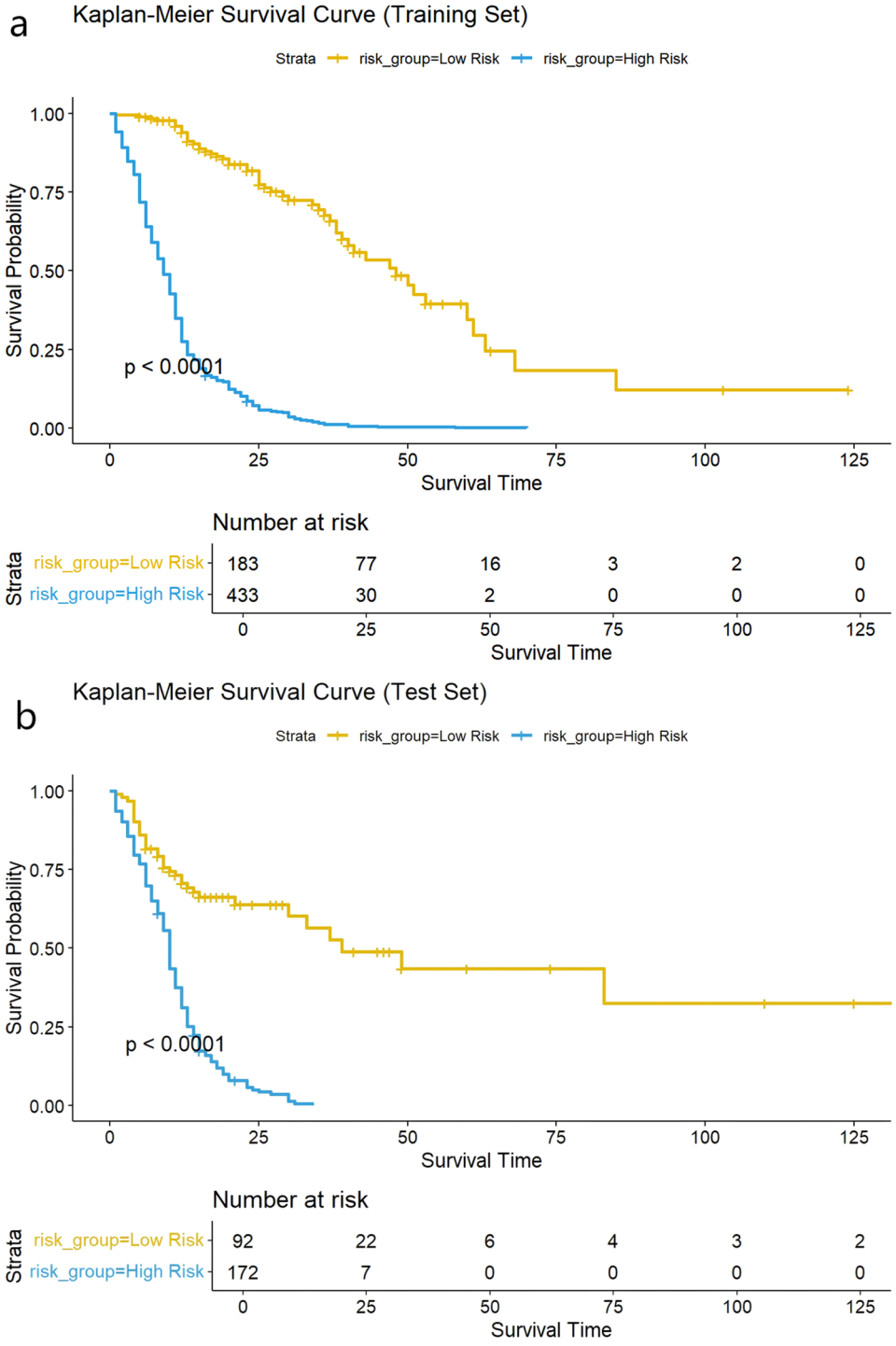

Fig. 9 The Kaplan-Meier survival curves for patients in the high-risk and low-risk groups. **a** Training set; **b** Test set

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-14889-0.

Supplementary Material 1. Figure S 1. Scatter Plot of ICC Values for Each Feature. The scatter plot illustrates the ICC values for radiomic features extracted from ROIs delineated by two independent researchers. The ICC values range from 0 to 1, with a mean value of 0.81. ICC values generally interpreted as follows: ICC < 0.5 indicates poor agreement, 0.5 ≤ ICC < 0.75 represents moderate agreement, 0.75 ≤ ICC < 0.9 reflects good agreement, and ICC ≥ 0.9 signifies excellent agreement. Figure S 2. Selection of Principal Components Based on Cumulative Variance Explained. The x-axis represents the principal components, while the y-axis shows the cumulative proportion of variance explained. A red dashed line indicates the 80% threshold, marking the number of PCs required to retain at least 80% of the total variance. This selection ensures a balance between dimensionality reduction and information preservation. Figure S 3. Optimal Cutoff Determination Using X-tile Analysis. (a) The heatmap generated by X-tile analysis illustrates different grouping strategies, with the x-axis and y-axis representing various stratifications. The color gradient from black to red indicates increasing statistical significance, ultimately identifying 0.8 as the optimal cutoff value. (b) The plot displays the statistical values computed by X-tile, where the x-axis represents different cutoff values and the y-axis indicates the corresponding statistical values. The peak of the curve corresponds to the optimal cutoff. (c) The histogram shows the distribution of predicted probabilities among patients, with the x-axis representing predicted probability and the y-axis representing the number of patients. Gray and blue bars correspond to different patient groups, and the cutoff of 0.8 effectively separates the two distributions. (d) The Kaplan-Meier survival curves illustrate the survival differences between the two groups stratified by the cutoff of 0.8. The blue curve, representing the low-risk group, is distinctly higher than the gray curve (high-risk group). Figure S 4.Patient Survival Follow-up Workflow. This flowchart outlines the methodology for collecting patient survival data. Enrolled participants underwent systematic telephone follow-up with patients or their family members to verify survival status (alive/deceased). Patients were classified as lost to follow-up after three unsuccessful contact attempts. Electronic medical record (EMR) review was performed in parallel to examine hospital records and confirm survival status, with telephone and EMR data cross-validated for accuracy. Final survival status determinations were exported as anonymized data for analysis. EMR, electronic medical record.

Acknowledgements

Not applicable.

Authors' contributions

All authors contributed to the study conception and design. Conceptualization, C.X and JZ.Z; methodology, CH.L; software, CH.L; validation, XN.S and CQ.G; formal analysis, W.X and CH.L; investigation, Z.D; data curation, XN.S; writing—original draft preparation, Z.D; writing—review and editing, W.X, C.X and JL.L; visualization, Z.D; supervision, YT.S, SB.Q, and JL.L project administration, SB.Q; funding acquisition, C.X, JL.L and SB.Q. All the authors have read and agreed to the published version of the manuscript.

Funding

This research was supported by the Jiangsu Provincial Medical Key Discipline (grant number: ZDXK202235).

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Ethics Committee of the Affiliated Hospital of Soochow University (Approval No. 176/2024). The Ethics Committee waived the requirement for individual patient informed consent for this retrospective study, as all data were anonymized and no identifiable information was included.

Consent for publication

Not Applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Oncology, Wuxi No.2 People's Hospital, Jiangnan University Medical Center, Wuxi 214002, China
[2]Department of Radiation Oncology, The First Affiliated Hospital of Soochow University, Suzhou 215000, China
[3]Department of Gastroenterology, The First Affiliated Hospital of Soochow University, Suzhou 215000, China
[4]Department of Radiology, The First Affiliated Hospital of Soochow University, Suzhou 215000, China
[5]Department of Radiation Oncology, The Second Affiliated Hospital of Soochow University, Suzhou 215000, China

Received: 22 March 2025 / Accepted: 19 August 2025
Published online: 20 October 2025

References

1. Cancer Stat Facts. Pancreatic Cancer. Available online: https://seer.cancer.gov/statfacts/html/pancreas.html (accessed on.
2. Han B, Zheng R, Zeng H, Wang S, Sun K, Chen R, et al. Cancer incidence and mortality in china, 2022. J Natl Cancer Cent. 2024;4:47–53. https://doi.org/10.1016/j.jncc.2024.01.006.
3. Hlavsa J, Cecka F, Zaruba P, Zajak J, Gurlich R, Strnad R, et al. Tumor grade as significant prognostic factor in pancreatic cancer: validation of a novel TNMG staging system. Neoplasma. 2018;65:637–43. https://doi.org/10.4149/neo_2018_171012N650
4. Dell'Aquila E, Fulgenzi CAM, Minelli A, Citarella F, Stellato M, Pantano F, et al. Prognostic and predictive factors in pancreatic cancer. Oncotarget. 2020;11:924–41. https://doi.org/10.18632/oncotarget.27518.
5. Karamitopoulou E. Emerging prognostic and predictive factors in pancreatic cancer. Mod Pathol. 2023;36:100328. https://doi.org/10.1016/j.modpat.2023.100328.
6. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278:563–77. https://doi.org/10.1148/radiol.2015151169.
7. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. Nat Rev Clin Oncol. 2022;19:132–46. https://doi.org/10.1038/s41571-021-00560-7.
8. Zhu N, Meng X, Wang Z, Hu Y, Zhao T, Fan H, et al. Radiomics in diagnosis, grading, and treatment response assessment of soft tissue sarcomas: A systematic review and Meta-analysis. Acad Radiol. 2024;31:3982–92. https://doi.org/10.1016/j.acra.2024.03.029.
9. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–6. https://doi.org/10.1016/j.ejca.2011.11.036.
10. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the Bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14:749–62. https://doi.org/10.1038/nrclinonc.2017.141.
11. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to radiomics. J Nucl Med. 2020;61:488–95. https://doi.org/10.2967/jnumed.118.222893.
12. Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás. B. 3D deep learning on medical images: A review. Sens (Basel). 2020;20. https://doi.org/10.3390/s20185097.
13. Yao L, Zhang Z, Keles E, Yazici C, Tirkes T, Bagci U. A review of deep learning and radiomics approaches for pancreatic cancer diagnosis from medical imaging. Curr Opin Gastroenterol. 2023;39:436–47. https://doi.org/10.1097/mog.0000000000000966.

14. Chen Y, Lu X, Xie Q. Collaborative networks of Transformers and convolutional neural networks are powerful and versatile learners for accurate 3D medical image segmentation. Comput Biol Med. 2023;164:107228. https://doi.org/10.1016/j.compbiomed.2023.107228.
15. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, et al. Deep Learning-Based radiomics model for prediction of survival in glioblastoma multiforme. Sci Rep. 2017;7:10353. https://doi.org/10.1038/s41598-017-10649-8.
16. Cui J, Jiao F, Li Q, Wang Z, Fu D, Liang J, et al. Chinese society of clinical oncology (CSCO): clinical guidelines for the diagnosis and treatment of pancreatic cancer. J Natl Cancer Cent. 2022;2:205–15. https://doi.org/10.1016/j.jncc.2022.08.006.
17. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. RANDOM SURVIVAL FORESTS. Annals Appl Stat. 2008;2:841–60. https://doi.org/10.1214/08-aoas169.
18. Taylor JM. Random survival forests. J Thorac Oncol. 2011;6:1974–5. https://doi.org/10.1097/JTO.0b013e318233d835.
19. Ning C, Ouyang H, Shen D, Sun Z, Liu B, Hong X, et al. Prediction of survival in patients with infected pancreatic necrosis: a prospective cohort study. Int J Surg. 2024;110:777–87. https://doi.org/10.1097/js9.0000000000000844.
20. Zwanenburg A, Vallières M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for High-Throughput image-based phenotyping. Radiology. 2020;295:328–38. https://doi.org/10.1148/radiol.2020191145.
21. Mukherjee S, Patra A, Khasawneh H, Korfiatis P, Rajamohan N, Suman G, et al. Radiomics-based Machine-learning models can detect pancreatic cancer on prediagnostic computed tomography scans at a substantial lead time before clinical diagnosis. Gastroenterology. 2022;163:1435–e14461433. https://doi.org/10.1053/j.gastro.2022.06.066.
22. Liu J, Hu L, Zhou B, Wu C, Cheng Y. Development and validation of a novel model incorporating MRI-based radiomics signature with clinical biomarkers for distinguishing pancreatic carcinoma from mass-forming chronic pancreatitis. Transl Oncol. 2022;18:101357. https://doi.org/10.1016/j.tranon.2022.101357.
23. Li J, Liu F, Fang X, Cao K, Meng Y, Zhang H, et al. CT radiomics features in differentiation of Focal-Type autoimmune pancreatitis from pancreatic ductal adenocarcinoma: A propensity score analysis. Acad Radiol. 2022;29:358–66. https://doi.org/10.1016/j.acra.2021.04.014.
24. Borhani AA, Dewan R, Furlan A, Seiser N, Zureikat AH, Singhi AD, et al. Assessment of response to neoadjuvant therapy using CT texture analysis in patients with resectable and borderline resectable pancreatic ductal adenocarcinoma. AJR Am J Roentgenol. 2020;214:362–9. https://doi.org/10.2214/ajr.19.21152.
25. Nasief H, Hall W, Zheng C, Tsai S, Wang L, Erickson B, et al. Improving treatment response prediction for chemoradiation therapy of pancreatic cancer using a combination of Delta-Radiomics and the clinical biomarker CA19-9. Front Oncol. 2019;9:1464. https://doi.org/10.3389/fonc.2019.01464.
26. Parr E, Du Q, Zhang C, Lin C, Kamal A, McAlister J, et al. Radiomics-Based outcome prediction for pancreatic cancer following stereotactic body radiotherapy. Cancers (Basel). 2020;12. https://doi.org/10.3390/cancers12041051.
27. Cen C, Wang C, Wang S, Wen K, Liu L, Li X, et al. Clinical-radiomics nomogram using contrast-enhanced CT to predict histological grade and survival in pancreatic ductal adenocarcinoma. Front Oncol. 2023;13:1218128. https://doi.org/10.3389/fonc.2023.1218128.
28. Gruetzemacher R, Gupta A, Paradice. D. 3D deep learning for detecting pulmonary nodules in CT scans. J Am Med Inf Assoc. 2018;25:1301–10. https://doi.org/10.1093/jamia/ocy098.
29. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221–48. https://doi.org/10.1146/annurev-bioeng-071516-044442.
30. Nie D, Zhang H, Adeli E, Liu L, Shen D. 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In: Unal G, Ourselin S, Joskowicz L, Sabuncu MR, Wells W, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Proceedings. Lecture Notes in Computer Science, vol. 9901. Cham: Springer Verlag; 2016. pp. 212–20. https://doi.org/10.1007/978-3-319-46723-8_25.
31. Gruetzemacher R, Gupta A, Paradice. D. 3D deep learning for detecting pulmonary nodules in CT scans. J Am Med Inform Assoc. 2018;25:1301–10. https://doi.org/10.1093/jamia/ocy098.
32. Nie D, Lu J, Zhang H, Adeli E, Wang J, Yu Z, et al. Multi-Channel 3D deep feature learning for survival time prediction of brain tumor patients using Multi-Modal neuroimages. Sci Rep. 2019;9:1103. https://doi.org/10.1038/s41598-018-37387-9.
33. Joo S, Ko ES, Kwon S, Jeon E, Jung H, Kim JY, et al. Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. Sci Rep. 2021;11:18800. https://doi.org/10.1038/s41598-021-98408-8.
34. Kim H, Lee D, Cho WS, Lee JC, Goo JM, Kim HC, et al. CT-based deep learning model to differentiate invasive pulmonary adenocarcinomas appearing as subsolid nodules among surgical candidates: comparison of the diagnostic performance with a size-based logistic model and radiologists. Eur Radiol. 2020;30:3295–305. https://doi.org/10.1007/s00330-019-06628-4.
35. Demircioğlu A. Are deep models in radiomics performing better than generic models? A systematic review. Eur Radiol Exp. 2023;7. https://doi.org/10.1186/s41747-023-00325-0.
36. Mohsen F, Ali H, El Hajj N, Shah Z. Artificial intelligence-based methods for fusion of electronic health records and imaging data. Sci Rep. 2022;12:17981. https://doi.org/10.1038/s41598-022-22514-4.
37. Wang W, Liang H, Zhang Z, Xu C, Wei D, Li W, et al. Comparing three-dimensional and two-dimensional deep-learning, radiomics, and fusion models for predicting occult lymph node metastasis in laryngeal squamous cell carcinoma based on CT imaging: a multicentre, retrospective, diagnostic study. EClinicalMedicine. 2024;67:102385. https://doi.org/10.1016/j.eclinm.2023.102385.
38. Barata C, Emre Celebi M, Marques JS. Melanoma detection algorithm based on feature fusion. *Annu Int Conf IEEE Eng Med Biol Soc* 2015, *2015*, 2653–2656. https://doi.org/10.1109/embc.2015.7318937.

Publisher's Note