Prediction of clinically relevant postoperative pancreatic fistula using radiomic features and preoperative data

Nithya Bhasker[1,*,+], Fiona R. Kolbinger[2,3,*,+], Nadiia Skorobohach[4], Alex Zwanenburg[1,5], Steffen Löck[5], Jürgen Weitz[2,3], Ralf-Thorsten Hoffmann[4], Marius Distler[2], Stefanie Speidel[1], Stefan Leger[1,++], and Jens-Peter Kühn[4,*,++]

[1]National Center for Tumor Diseases (NCT/UCC) Partner Site Dresden, Germany
[2]Department of Visceral, Thoracic and Vascular Surgery, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany
[3]Else Kröner Fresenius Center for Digital Health, TU Dresden, Dresden, Germany
[4]Institute and Polyclinic for Diagnostic and Interventional Radiology, University Hospital Carl Gustav Carus Dresden, TU Dresden, Dresden, Germany
[5]OncoRay - National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany
[*]nithya.bhasker@nct-dresden.de; fiona.kolbinger@uniklinikum-dresden.de; jens-peter.kuehn@uniklinikum-dresden.de
[+]these authors contributed equally to this work
[++]shared last authorship

ABSTRACT

Clinically relevant postoperative pancreatic fistula (CR-POPF) can significantly affect the treatment course and outcome in pancreatic cancer patients. Preoperative prediction of CR-POPF can aid surgical decision-making process and lead to better perioperative management of patients. In this retrospective study, we have developed radiomic risk models that use preoperative computed tomography (CT)-based radiomic features in a cohort of 148 patients having undergone pancreaticoduodenectomy. The developed models were compared with risk models using preoperative clinical features and mesh-based volumes of the annotated structures. The model signatures were analysed in detail by visualising feature expression maps and comparing significant features to the standard CR-POPF risk measures. The best average area under receiver operating characteristic curve (AUC) of 0.81 on validation data was achieved by the radiomic signature model retrained using the most important features. The following features showed significant correlation with outcome ($p < 0.05$)- texture and morphology of healthy pancreatic segment, intensity volume histogram-based feature of pancreatic duct segment and morphology of combined segment; mesh-based volumes of healthy pancreatic segment, segment containing arteries, combined segment and pancreatic duct segment; BMI and sex. The results of the study indicate that preoperative data can be useful for an accurate risk assessment of CR-POPF.

Introduction

Despite advances in conservative and surgical treatment approaches, pancreatic cancer remains one of the deadliest malignant diseases. All curative treatment approaches necessitate surgical resection. However, 80-85% of patients are diagnosed with locally advanced or metastasised, and thus unresectable tumours[1,2]. Surgical treatment options for patients with tumours of the pancreatic head include pylorus-resecting (Whipple surgery) and pylorus-preserving pancreaticoduodenectomy. These surgical procedures are complex and associated with severe postoperative complications like pancreatic fistula, which can significantly delay or impede neoadjuvant chemotherapy, thus affecting the oncological treatment course and outcome[3]. Postoperative pancreatic fistula (POPF) is caused by pancreato-enteric anastomotic leakage[4] and affects about 10-30% of patients undergoing pancreatic head resection[5–7]. The definition of POPF published in 2016 by the International Study Group on Pancreatic Fistula Definition (ISGPF) distinguishes between asymptomatically elevated levels of drain amylase (termed biochemical leak), and clinically relevant (CR-)POPF, which necessitates invasive treatment and prolongs hospital stay[5,8].

A plethora of risk stratification methods have been introduced in the past decade based on the association of CR-POPF with parenchymal risk factors like soft pancreatic parenchyma, small pancreatic duct width, and clinical factors like sex, intraoperative blood loss and body mass index (BMI)[6,9–13]. Most of the prevalent methods heavily rely on intraoperatively

assessed factors like pancreatic duct width, texture and blood loss for risk stratification. However, an early and accurate risk estimate based on the vast amount of available preoperative data has the potential to play a key role in pre-surgical decision-making, especially in case of patients with terminal illness like the cancer patients, where postoperative complications could delay or impede neoadjuvant therapy. For instance, small pancreatic duct size and large pancreatic remnant volume (PRV), estimated from preoperative computed tomography (CT) images, was found to be significantly associated with the risk of CR-POPF after pancreatic head resection[14–17]. The preoperative risk estimate could also complement the intraoperative and postoperative findings for better patient state monitoring.

It has been established that radiomic features (a large number of quantitative imaging descriptors extracted from medical images using data characterisation algorithms) contain good prognostic power in predictive modelling[18]. Skawran et al.[19] investigated preoperative magnetic resonance imaging (MRI)-based radiomic features from the pancreatic body and tail in addition to pancreas-to-muscle T1 signal intensity ratio in a cohort of 62 patients undergoing pancreaticoduodenectomy and achieved an average area under the receiver operating characteristic curve (AUC) of 0.90 on the test data for CR-POPF prediction. Given that CT is the preoperative routine imaging modality in patients undergoing pancreatic resection, it would be preferable to predict the complication risk based on features extracted from CT imaging data rather than MRI data. A study conducted by Zhang et al.[20] considered CT-based radiomic features from the pancreatic parenchyma remaining after tumour removal (healthy parenchyma) for postoperative prediction of CR-POPF, and achieved an AUC of 0.76 on the validation cohort. Based on the above-mentioned clinically established risk factors such as small pancreatic duct size, one could deduce that considering pancreatic structures besides the healthy parenchyma for preoperative image based risk assessment might reveal interesting correlations with the surgical outcome. Therefore, one of the objectives of this study was to develop and investigate radiomic models for predicting the risk of CR-POPF using features from different intra- and peripancreatic anatomical structures in addition to the healthy parenchyma.

The main goal of this study was to make use of preoperative data like CT-based radiomic features, mesh-based volumes of different pancreatic structures and clinical data to get an accurate and early risk estimate of CR-POPF that could complement the current intraoperative and postoperative risk measures. The contributions of this study are (i) to develop risk models using CT-based radiomic features extracted not only from the normal pancreatic parenchyma but also from other intra- and peripancreatic anatomical structures, (ii) to compare the developed radiomic risk models with risk models developed using preoperative clinical data and mesh-based volumes of the selected anatomic structures, (iii) to quantitatively and qualitatively assess the feature importance for all the models using univariate analyses and visualisation, and (iv) to analyse in detail the relation between the resultant important radiomic features and the established risk factors for CR-POPF.

Methods

Characteristics of patient cohort

This was a retrospective study. A total of 381 patients having undergone pancreatic surgery at the University Hospital Carl Gustav Carus Dresden between 2011 and 2019 were screened for eligibility. Patients with a preoperative contrast-enhanced venous-phase CT with a delineable tumour were selected for the study. The exclusion criteria were: patients with calcifying chronic pancreatitis or acute pancreatitis, presence of peripancreatic fluid collections, pancreatic pseudocysts, or a stent in the common bile duct. The final retrospective cohort consisted of 148 patients, all of whom had a clinical indication for the surgical procedure and suffered from a localised pathology of the pancreas. The characteristics of the patient cohort are summarised in Table 1.

Frequency and severity of POPF were assessed for at least 30 days after the pancreatic head resection or until discharge from hospital, whichever occurred last. Following ISGPF standards, CR-POPF was sub-divided into Grade B POPF (defined as a clinically evident fistula with mild restriction of patient condition, typically managed using antibiotic medication, transfusion, parenteral nutrition, prolonged drain maintenance or additional interventional drain placement) and Grade C POPF (a severe complication characterised by critical restriction of patient condition, typically necessitating intensive care or an operative intervention). Ethical approval for the retrospective analysis of clinical and imaging data was obtained from the Ethics Committee at the Technische Universität Dresden (BO-EK-263062020). All analyses were carried out in accordance with the relevant guidelines and regulations, in particular the Declaration of Helsinki and its later amendments. Written informed consent was waived by the Ethics Committee at the Technische Universität Dresden.

Segmentation of the pancreas and image feature computation

For all patients, different anatomical structures within and adjacent to the pancreas were manually segmented in preoperative contrast-enhanced CT images (venous phase) using the picture archiving system PACS (Agfa Impax EE R20, Agfa Healthcare) and the open-source software Slicer 3D (version 4.10.2, www.slicer.org). In particular, the following structures were delineated (refer to Figure 1): pancreatic pathology (panc_pat), healthy pancreas (panc_heal), pancreatic duct (panc_duct), portal vein (pv), arteries (art), and bile duct (bile_duct). An additional segment was created by combining these delineated structures

Clinical variable	Statistic	CR-POPF	Non CR-POPF
Patients	n	48	100
Sex:	n		
male		35	50
female		13	50
Pre-existing diabetes:	n		
yes		9	29
no		39	71
History of smoking:	n		
yes		5	18
no		43	82
History of alcohol abuse:	n		
yes		7	16
no		41	84
Age	mean ± standard deviation	68.10 ± 10.37	67.5 ± 10.78
BMI	mean ± standard deviation	27.09 ± 4.81	24.86 ± 4.29
CA 19-9 (U/mL)	mean ± standard deviation	138.61 ± 294.37	489.89 ± 1114.72
Preoperative amylase (μmol/s*L)	mean ± standard deviation	0.92 ± 1.40	0.72 ± 1.60
Preoperative lipase (μmol/s*L)	mean ± standard deviation	1.88 ± 3.04	2.22 ± 8.54
Mesh-based volumes of segmented structures (cm^3):	mean ± standard deviation		
pancreatic duct		3.05 ± 2.23	4.66 ± 4.18
healthy pancreas		92.62 ± 38.31	67.38 ± 34.59
pancreatic pathology		53.23 ± 143.63	22.69 ± 49.73
arteries		58.72 ± 23.32	46.69 ± 17.31
bile duct		11.58 ± 18.65	11.04 ± 12.46
portal vein		39.21 ± 14.68	36.18 ± 15.09
Pancreatic surgery type:	n		
pylorus preserving pancreaticoduodenectomy (PPPD)		33	72
Whipple		2	4
pancreatic left resection with splenectomy		11	16
pancreatic left resection without splenectomy		1	7
enucleation		1	1

Table 1. Patient characteristics of the retrospective cohort used in the study.

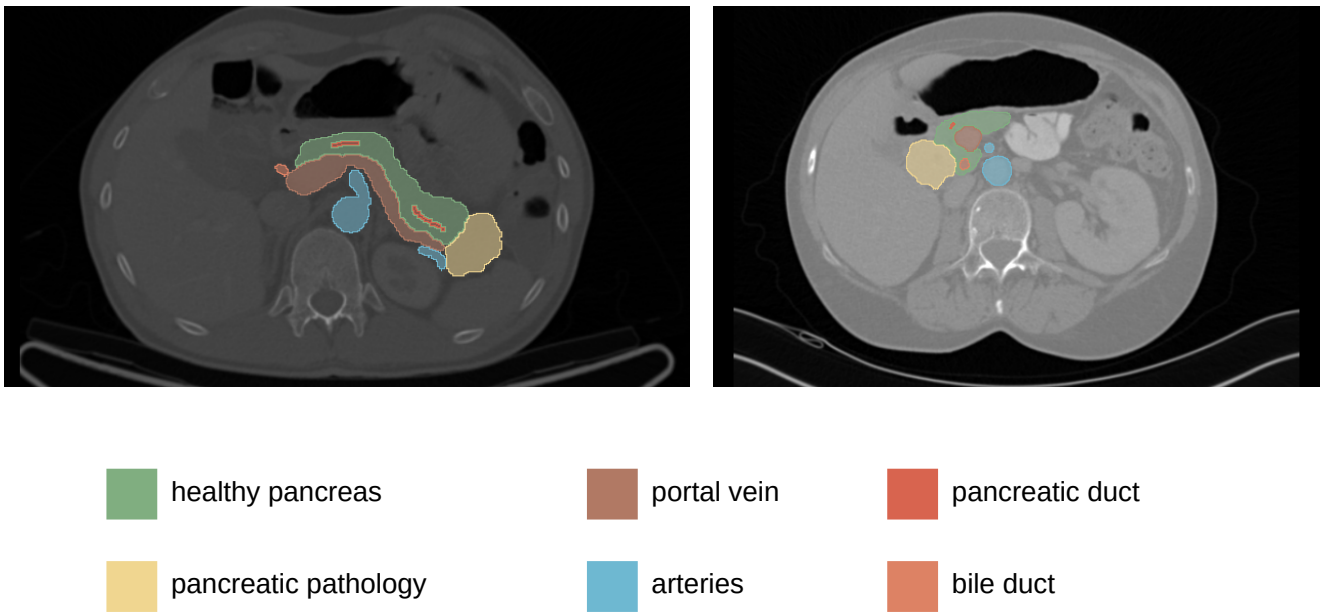

Figure 1. Examples of delineated structures: healthy pancreas, pancreatic pathology, portal vein, arteries, pancreatic duct, bile duct.

(combined). The annotation was performed by a radiologist (NS) with 5 years of experience in abdominal CT imaging, and was independently reviewed by a second radiologist (JPK) with 20 years of experience in pancreatic imaging. Subsequently, the imaging features were computed and extracted in compliance with the Image Biomarker Standardisation Initiative using the medical image radiomics processor (MIRP - www.github.com/oncoray/mirp)[21].

In brief, to correct differences in voxel spacing and slice thickness the CT images were resampled using linear image interpolation to a voxel size of $1.0 \times 1.0 \times 1.0$ mm^3. Subsequently, to emphasise image characteristics such as edges and blobs, nine additional images were created by applying spatial filtering to the base image. In particular, eight images were created using stationary coiflet-1 wavelet high-/low-pass filter along each of the three spatial directions and one image was created by applying a Laplacian of Gaussian (LoG) filter consisting of different filter kernel widths. Finally, for each segment statistical, histogram-based and texture features were calculated using the base and the nine transformed images. Furthermore, morphological features were computed for each of the pancreatic segments on the base image only. The configuration settings of the MIRP pipeline are summarised in Supplementary Table 1.

Development of risk models

The familiar R package (version 0.0.0.54)[22] was used for the development and validation of the radiomic models to predict the risk of CR-POPF. The framework consisted of the following steps: feature pre-processing, feature selection, hyper-parameter optimisation, model development and model validation. The radiomic risk models were developed as previously described[23,24] within a 10-fold cross validation scheme, repeated three times. Based on the entire retrospective cohort, the computed imaging features were transformed and normalised using Yeo-Johnson and z-normalisation methods, respectively. Subsequently, highly correlated features were clustered using Spearman correlation coefficient ($\rho > 0.9$). Feature selection was performed using 10 bootstrap samples (with replacement) of the training folds. The risk models were trained on these 10 bootstrap samples of the training folds, using the most important features and the optimised hyper-parameters[25]. Finally, an ensemble prediction was made by averaging the prediction scores of each model for the training and validation folds separately.

Different feature selection methods and learning algorithms were used for the model development to reduce the risk of incidental findings[26]. In particular, the following four feature selection methods were used: mutual information maximisation (MIM), minimum redundancy maximum relevance (MRMR)[27], lasso regression- and elastic net regression-based feature selection methods. For the learning algorithms, the following three approaches were considered: logistic regression without penalty, with lasso penalty and elastic-net penalty. The lasso and elastic-net regression based approaches prevent the model from overfitting by using a regularisation term[28].

Risk models were developed based on (a) the computed radiomic features (radiomic model), (b) the clinical variables (clinical model), (c) mesh-based volumes of the annotated pancreatic segments (volume model). The clinical model used the following patient data: age, sex, BMI, preoperative presence of diabetes mellitus, smoking history, history of alcohol abuse,

and preoperative lab values like CA 19-9 level and preoperative digestive enzyme levels (amylase and lipase). The continuous values for CA 19-9, amylase and lipase levels were transformed into binary values (0 - normal, 1 - abnormal) by considering their corresponding reference intervals used as a standard in the University Hospital Carl Gustav Carus Dresden. The normal reference intervals were: < 24 U/mL for CA 19-9, < 0.88 μmol/s*L for amylase, and < 1 μmol/s*L for lipase. The frequency, mean and standard deviation values of these variables for the used patient cohort are summarised in Table 1.

Performance assessments

The performance of the models was assessed using the average area under the receiver operating characteristic curve (AUC) and F1-score on the ensemble predictions. The median AUC over all combinations of feature selection methods and machine learning algorithms were determined on the training and validation folds, and one representative model combination (out of the 12 combinations) was selected for each risk model (a)-(c). The heatmaps representing the performances of all the model combinations for different risk models are illustrated in Supplementary Fig. 1, 2 and 3. To ensure robustness, the representative model was chosen based on the following scheme[24]: (i) median performance of each feature selection method over all learning algorithms and vice versa was determined on the exploratory data; (ii) the feature selection and machine learning algorithms with AUCs closest to the respective median performances were chosen as the representative model combination.

The choice of features for the volume and clinical model was based on prior knowledge and existing literature[14,15,17,29]. However, for the radiomic model, the aim of feature extraction was to characterise the region of interest quantitatively and explore potential correlations between the extracted radiomic features and the outcome. As a result, the number of features provided as input to the radiomic model after feature clustering (1706) was orders of magnitude greater than the number of features fed to the clinical (9) or the volume (7) models. The machine learning algorithms used as feature selection methods in this study are influenced by the size of the input features and are known to perform comparatively well when a smaller number of relevant features are provided as input[30].

Therefore, the radiomic model was retrained within the same cross validation setting using the four most important features. The importance of the features was assessed by ranking the features according to their occurrence (in the top 'n' position based on feature selection and signature size parameter from hyperparameter optimisation) over the 300 trained radiomic models (10 training folds $\times$ 10 bootstrap samples with replacement of each training fold $\times$ 3 repetitions) of the representative model combination. The resulting model is referred to as the radiomic signature model (d). To assess the relative importance of radiomic features in the prediction of CR-POPF, the radiomic signature model was combined: (i) with mesh-based volumes of annotated pancreatic segments (referred to as (e) combined radiomic signature and volume model), and (ii) with clinical features (referred to as (f) combined radiomic signature and clinical model).

The model signatures were quantitatively and qualitatively examined to understand the decision rationale of the various risk models. The significance of the association between individual features and CR-POPF outcome was assessed by performing univariate analyses using the logistic regression model with Benjamini-Hochberg adjustment for multiple testing[31]. The features with significant association with the CR-POPF outcome ($p < 0.05$) were standardised, and heatmaps of the resulting standardised features were visualised for the entire patient cohort in order to qualitatively assess the trends and correlations with the outcome. Additionally, the features included in the radiomic signature were compared with established risk factors[16] such as intraoperative pancreatic texture and pancreatic duct width, and preoperatively estimated pancreatic remnant volume (PRV). The standardised values of these established risk factors were plotted against the corresponding standardised values of the radiomic features.

Results

Prognostic performance

The median AUC, F1-score and their 95% confidence intervals (CI) for the risk models (a)-(f) are shown in Table 2. The representative model combinations (a)-(c) (one best feature selection and machine learning algorithm for each risk model) were chosen according to the scheme described in the previous section. The models (a)-(c) showed similar performance with an F1-score of 0.80 on the validation data. The best AUC of 0.77 on the validation data was achieved by the volume model (c). The median performance of the models over all combinations of feature selection methods and machine learning algorithms for the models (a)-(c) are presented in Supplementary Table 3. The models (d)-(f) were retrained based on the results and hyperparameters of the representative radiomic model (a). Among the retrained models (d)-(f), the radiomic signature model (d) and the combined radiomic signature and clinical model (f) showed the best AUC of 0.81 on the validation data.

Signature analyses

The feature expression maps for combined radiomic signature and volume model (e), and for combined radiomic signature and clinical model (f) are illustrated in Figures 2 and 3 respectively. The radiomic features significantly associated ($p < 0.05$) with the outcome included two morphological features related to the panc_heal and combined segments, a feature cluster

Model	Exploratory data		Validation data		FS method	ML method
	AUC	F1-score	AUC	F1-score		
(a) radiomic	0.89 [0.82-0.95]	0.84 [0.79-0.89]	0.70 [0.61-0.79]	0.80 [0.74-0.85]	lasso	elastic-net
(b) clinical	0.74 [0.65-0.82]	0.80 [0.74-0.85]	0.68 [0.59-0.76]	0.80 [0.74-0.84]	lasso	elastic-net
(c) volume	0.83 [0.77-0.90]	0.85 [0.80-0.89]	0.77 [0.67-0.83]	0.80 [0.75-0.86]	MIM	lasso
(d) radiomic signature	0.83 [0.76-0.89]	0.83 [0.77-0.88]	0.81 [0.73-0.87]	0.82 [0.77-0.87]	-	elastic-net
(e) combined radiomic signature and volume	0.84 [0.76-0.90]	0.86 [0.81-0.91]	0.79 [0.70-0.86]	0.83 [0.77-0.88]	-	lasso
(f) combined radiomic signature and clinical	0.84 [0.77-0.90]	0.84 [0.79-0.89]	0.81 [0.73-0.88]	0.84 [0.78-0.89]	-	elastic-net

Table 2. Median average area under the receiver operating characteristic curve (AUC), F1-score and their 95% confidence intervals (CI) for the representative model combinations of (a)-(c), and retrained models (d)-(f). The best feature selection methods and machine learning algorithms are listed. Abbreviations: FS - feature selection, ML - machine learning, MIM - mutual information maximisation.

represented by a texture feature related to the panc_heal segment and one intensity volume histogram (IVH) based feature related to the panc_duct segment. The details about these features can be found in Supplementary Table 2. The other features, which were significantly associated with outcome included the mesh-based volumes of the healthy pancreatic segment, the combined segment, the arteries segment, and the pancreatic duct segment, and BMI and sex. The feature expression maps for risk models (b)-(d) are illustrated in Supplementary Fig. 4, 5 and 6.

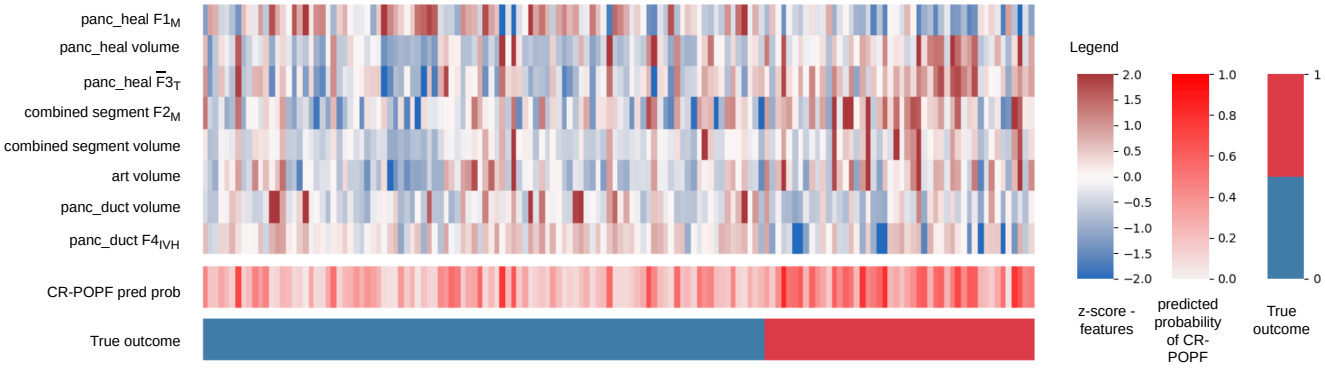

Figure 2. Feature expression maps for the combined radiomic signature and volume model. The values indicated in the heatmaps are z-scores of standardised features, probability of developing CR-POPF as predicted by the model and the true outcome for all the patients included in the study. True outcome of 1 indicates the presence of CR-POPF. Abbreviations: pred prob – prediction probability, F_M – morphological feature, $\overline{F}$ – representative feature for a cluster, F_T – texture feature, , F_{IVH} - intensity volume histogram based feature.

Comparison of the radiomic signature to established risk factors

Figures 4 and 5 show the association between the radiomic signature features and the established risk factors for CR-POPF such as pancreatic texture, estimated PRV and pancreatic duct width[16]. The data for these risk factors was available for 70 out of the 148 patients analyzed in this study. Out of the 70 patients, 26 had developed CR-POPF. The morphological feature $F1_M$ of the healthy pancreatic segment was inversely associated with the estimated PRV with Spearman correlation co-efficient of $\rho = -0.63$, while the morphological feature $F2_M$ showed modest positive association with $\rho = 0.54$. The

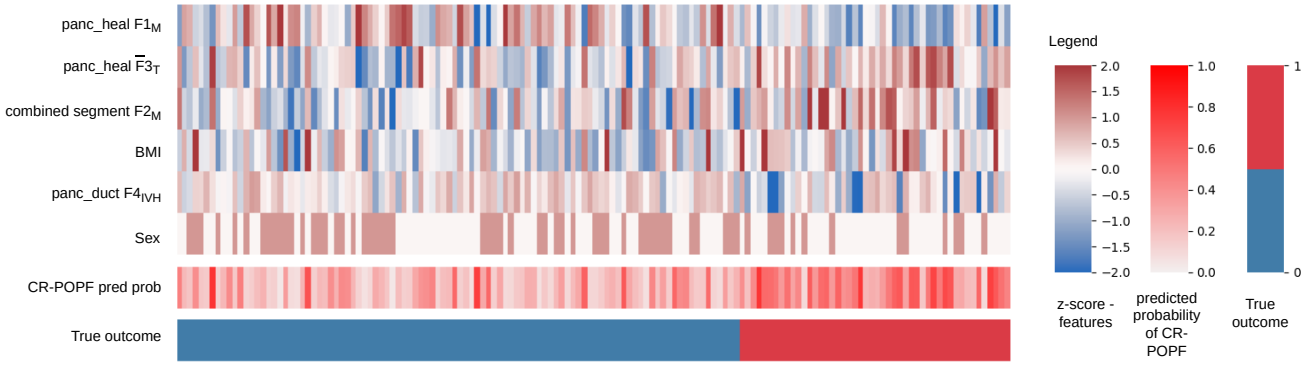

Figure 3. Feature expression maps for the combined radiomic signature and clinical model. The values indicated in the heatmaps are z-scores of standardised features, probability of developing CR-POPF as predicted by the model and the true outcome for all the patients included in the study. True outcome of 1 indicates the presence of CR-POPF. Abbreviations: pred prob - prediction probability, F_M - morphological feature, $\overline{F}$ - representative feature for a cluster, F_T - texture feature, , F_{IVH} - intensity volume histogram based feature.

intensity volume histogram based feature $F4_{IVH}$ of the pancreatic duct segment also showed moderate association with the intraoperatively assessed pancreatic duct width. Figure 5 displays that the frequency of CR-POPF was higher in patients with soft intraoperatively determined pancreatic texture, and above average values of the texture feature $\overline{F3}_T$ of the healthy pancreatic segment were associated with soft texture of the pancreas.

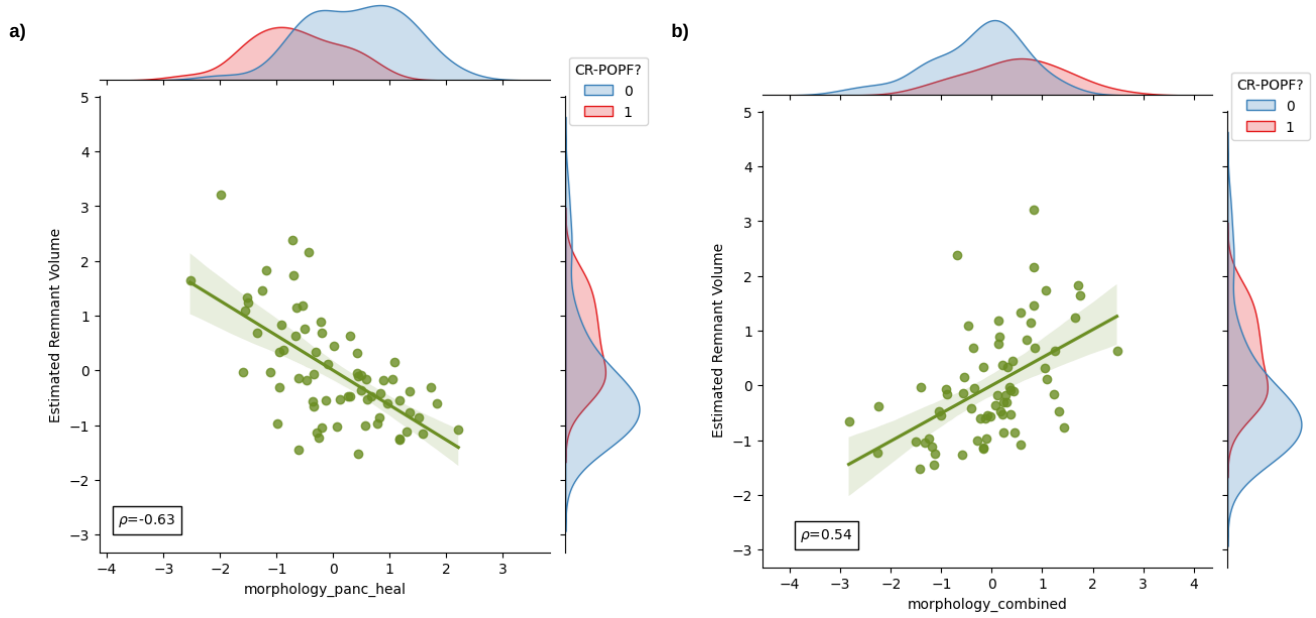

Figure 4. Comparison of radiomic features with standard CR-POPF risk factors. Comparison of estimated remnant pancreatic volume with - **a**): the morphological feature $F1_M$ of the healthy pancreatic segment; **b**): the morphological feature $F2_M$ of the combined segment. The marginal distributions for both axes categorised according to the CR-POPF outcome (0 - no CR-POPF, 1 - CR-POPF) are also illustrated. The value ρ indicates the Spearman correlation co-efficient between the two variables.

Discussion

The current gold standard of assessing CR-POPF risk intraoperatively are based on factors such as parenchymal texture and probe-measured pancreatic duct diameter. An accurate and early risk stratification based on the available preoperative data could complement the current standard and allow for better surgical planning and more transparent patient information. To this

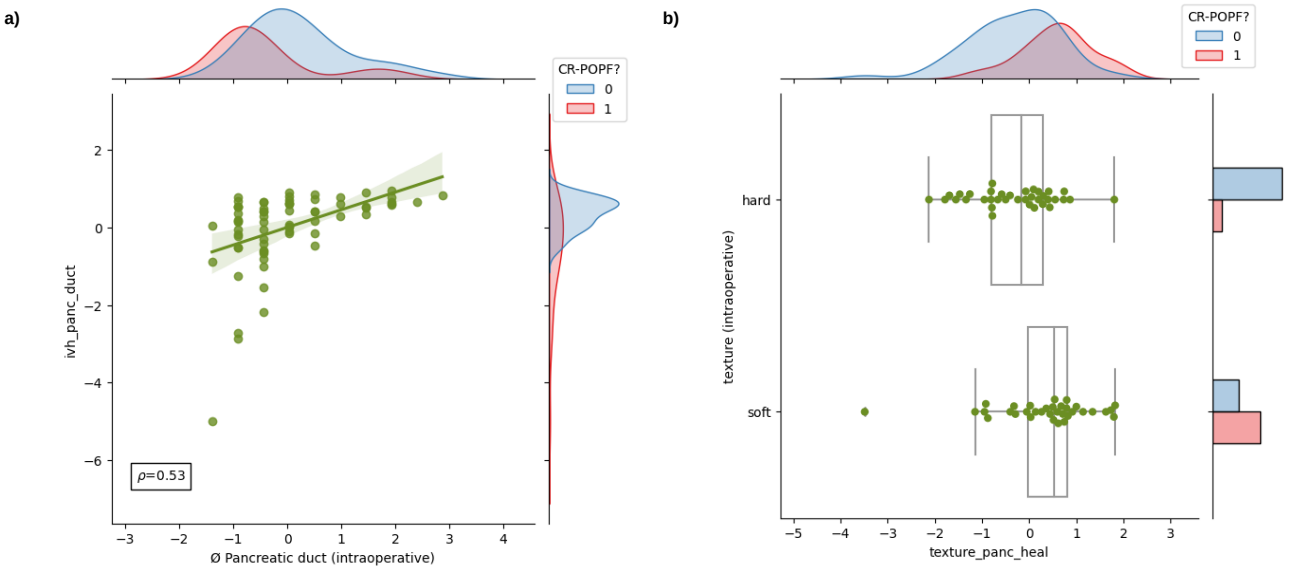

Figure 5. Comparison of radiomic features with standard CR-POPF risk factors. **a)**: Comparison of intraoperatively determined pancreatic duct width with the intensity volume histogram based feature $F4_{IVH}$ of the pancreatic duct segment. The marginal distributions for both axes categorised according to the CR-POPF outcome (0 - no CR-POPF, 1 - CR-POPF) are also illustrated. The value ρ indicates the Spearman correlation co-efficient between the two variables; **b)**: Comparison of intraoperatively determined pancreatic texture with the texture feature $\overline{F3}_T$ of the healthy pancreatic segment.

end, we combined preoperatively available clinical and laboratory data with annotated preoperative CT imaging to generate and compare radiomics-based models for CR-POPF prediction in patients undergoing pancreatic surgery. The radiomic model (a), clinical model (b) and volume model (c) showed similar performance on the validation data. Among the retrained models (d)-(f), the radiomic signature model (d) and the combined radiomic signature and clinical model (e) showed the best performance with an AUC of 0.81 on the validation data.

The radiomic signature model was trained with four important features (with no feature selection) selected based on the experiments with the radiomic model (a). On univariate analysis, all of these four radiomic features were found to be significantly associated with the outcome. These features were: (i) morphological ($F1_M$) and (ii) texture ($\overline{F3}_T$) features of the healthy pancreatic segment; (iii) morphological feature ($F2_M$) of the combined segment; and (iv) intensity volume histogram based feature ($F4_{IVH}$) of the pancreatic duct. The radiomic texture feature ($\overline{F3}_T$) describes the correlation between the voxels in close proximity[21]. From the feature expression map shown in Figure 2, it is evident that the model tends to predict a higher risk of CR-POPF for patients with above average values of this texture feature. The radiomic texture feature $\overline{F3}_T$ was also found to be associated with soft intraoperative pancreatic texture (Figure 5), which was found to be an indication of CR-POPF occurrence[6,16]. The assessment of pancreatic texture is one of the most commonly applied intraoperative methods of assessing CR-POPF risk.

The value of $F4_{IVH}$ represents the difference between the largest volume fractions (V_{10} and V_{90}) containing intensity fractions of at least 10% and 90%[21] in the pancreatic duct segment. The patients with below average values of $F4_{IVH}$ had an increased risk of CR-POPF as predicted by the radiomic signature model (Figure 2). Since this feature showed a direct association with the intraoperative pancreatic duct width with a Spearman correlation co-efficient of $\rho = 0.53$ (Figure 5), the results are in agreement with recent studies[16,17], in which a narrow pancreatic duct was associated with higher risk of CR-POPF. A narrow pancreatic duct is also one of the most relevant intraoperatively determined CR-POPF risk factors[6,13].

The value of the $F1_M$ describes the surface to volume ratio of the healthy pancreatic segment, which decreases with increasing volume. The morphological feature $F2_M$ is a measure of the diameter along the major axis of the combined segment, and is directly influenced by either the size of the healthy pancreas or the pancreatic pathology in most of the cases. The feature expression map (Figure 2) revealed an inverse relationship between $F1_M$ and the the risk of CR-POPF as predicted by the radiomic signature model, and a direct relationship between $F2_M$ and the the risk of CR-POPF. A similar trend was observed between the features $F1_M$ and $F2_M$, and the estimated PRV (Figure 4). Therefore, the results of the radiomic signature model are in line with previous studies[16,17] that indicated an increased risk of CR-POPF for patients with higher estimated PRV.

The volume model (c) considered mesh-based volumes of all the annotated structures and exhibited an AUC of 0.77 on the

validation data. Out of the seven considered features, higher volumes of healthy pancreas, arteries and the combined segment, and lower volume of pancreatic duct were found to be significantly associated with the CR-POPF outcome according to univariate analysis and feature expression map (Figure 2). Since the combined segment is influenced by estimated PRV (Figure 5), these findings are in agreement with prior literature[16,20]. The volume model (c) and the combined radiomic signature and volume model (e) trained with mesh-based volumes of the annotated structures in addition to the radiomic signature showed similar performances on the validation data. This could be due to the increased redundancy ($|\rho|$>0.6) among the input features fed to the combined model (e) as three out of the four features in the radiomic signature ($F1_M$ of the healthy pancreas, $F2_M$ of the combined segment and $F4_{IVH}$ of pancreatic duct) depend on the volumes of the annotated structures (Supplementary Fig. 7).

The clinical model (b) trained with only the clinical features exhibited an AUC of 0.68 on the validation data. However, the combined radiomic signature and clinical model (f) trained with the clinical features in combination with the radiomic signature resulted in an AUC of 0.81 on the validation data, similar to the performance of radiomic signature model (d). While this suggests that analysis of preoperative CT data can alone accurately estimate the risk for CR-POPF, the potential benefit of combining it with clinical features should be investigated in larger patient cohorts in future studies. According to the univariate analysis and the trend seen on the feature expression map (Figure 3), an increased BMI and male sex were significantly associated with the clinical outcome in this study. This is on par with the findings of a meta-analysis conducted by Kamarajah et al.[29].

The observations from this study indicate that preoperative data like CT-based radiomic features, mesh-based volumes of different pancreatic structures and relevant clinical data can be useful in the preoperative risk assessment of CR-POPF, and analysis of preoperative CT imaging alone can provide reliable information about the CR-POPF risk, with little additional value from clinical risk factors. Moreover, on analysing the relationship of the radiomic signature with some of the well established risk factors for CR-POPF (refer to Figures 4 and 5), it was observed that the identified radiomic features were comparable to the intraoperative risk factors like parenchymal texture and pancreatic duct width, which are the current gold standard. This implies that there is great potential for obtaining an early risk estimate for CR-POPF, which could complement the current risk measures. However, this study has limitations. The radiomic signature model (d) was retrained on the same patient cohort with most suitable features from the radiomic model (a) selected based on the results and hyperparameters of the radiomic model (a). Therefore, the results of the models (d)-(f) on validation data cannot be considered as validation results in true sense, and a validation of these models on an external validation cohort would be necessary in the future. One of the other shortcomings of the work was the low occurrence of CR-POPF (48 out of 148) events in the selected patient cohort. Additionally, this was a retrospective study based on a patient cohort from a single centre. To overcome these shortcomings, a multi-centre study on a well balanced and sufficiently large patient cohort to validate the risk models is being prepared.

Data Availability

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

References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA: a cancer journal for clinicians* (2022).
2. Michalski, C. W., Weitz, J. & Büchler, M. W. Surgery insight: surgical management of pancreatic cancer. *Nat. clinical practice oncology* **4**, 526–535 (2007).
3. Mintziras, I. *et al.* Postoperative morbidity following pancreatic cancer surgery is significantly associated with worse overall patient survival; systematic review and meta-analysis. *Surg. Oncol.* **38**, 101573 (2021).
4. Ansorge, C. *et al.* Structured intraoperative assessment of pancreatic gland characteristics in predicting complications after pancreaticoduodenectomy. *J. Br. Surg.* **99**, 1076–1082 (2012).
5. Bassi, C. *et al.* The 2016 update of the international study group (isgps) definition and grading of postoperative pancreatic fistula: 11 years after. *Surgery* **161**, 584–591 (2017).
6. Callery, M. P., Pratt, W. B., Kent, T. S., Chaikof, E. L. & Vollmer Jr, C. M. A prospectively validated clinical risk score accurately predicts pancreatic fistula after pancreatoduodenectomy. *J. Am. Coll. Surg.* **216**, 1–14 (2013).
7. Nahm, C. B., Connor, S. J., Samra, J. S. & Mittal, A. Postoperative pancreatic fistula: a review of traditional and emerging concepts. *Clin. experimental gastroenterology* **11**, 105 (2018).
8. Bassi, C. *et al.* Postoperative pancreatic fistula: an international study group (isgpf) definition. *Surgery* **138**, 8–13 (2005).
9. Grendar, J. *et al.* Validation of fistula risk score calculator in diverse north american hpb practices. *HPB* **19**, 508–514 (2017).

10. Di Martino, M. *et al.* Predictive factors of pancreatic fistula after pancreaticoduodenectomy and external validation of predictive scores. *Anticancer. research* **39**, 499–504 (2019).

11. Ryu, Y. *et al.* Validation of original and alternative fistula risk scores in postoperative pancreatic fistula. *J. Hepato-Biliary-Pancreatic Sci.* **26**, 354–359 (2019).

12. Shubert, C. R. *et al.* Clinical risk score to predict pancreatic fistula after pancreatoduodenectomy: independent external validation for open and laparoscopic approaches. *J. Am. Coll. Surg.* **221**, 689–698 (2015).

13. Mungroop, T. H. *et al.* Alternative fistula risk score for pancreatoduodenectomy (a-frs): design and international external validation. *Annals surgery* **269**, 937–943 (2019).

14. Kirihara, Y. *et al.* Prediction of pancreatic anastomotic failure after pancreatoduodenectomy: the use of preoperative, quantitative computed tomography to measure remnant pancreatic volume and body composition. *Annals surgery* **257**, 512–519 (2013).

15. Frozanpor, F. *et al.* Correlation between preoperative imaging and intraoperative risk assessment in the prediction of postoperative pancreatic fistula following pancreatoduodenectomy. *World journal surgery* **38**, 2422–2429 (2014).

16. Kolbinger, F. R. *et al.* The image-based preoperative fistula risk score (prefrs) predicts postoperative pancreatic fistula in patients undergoing pancreatic head resection. *Sci. reports* **12**, 1–11 (2022).

17. Shi, Y. *et al.* Computed tomography-adjusted fistula risk score for predicting clinically relevant postoperative pancreatic fistula after pancreatoduodenectomy: Training and external validation of model upgrade. *EBioMedicine* **62**, 103096 (2020).

18. Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. communications* **5**, 1–9 (2014).

19. Skawran, S. M. *et al.* Can magnetic resonance imaging radiomics of the pancreas predict postoperative pancreatic fistula? *Eur. J. Radiol.* **140**, 109733 (2021).

20. Zhang, W. *et al.* A radiomics-based formula for the preoperative prediction of postoperative pancreatic fistula in patients with pancreaticoduodenectomy. *Cancer Manag. Res.* **10**, 6469 (2018).

21. Zwanenburg, A. *et al.* The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**, 328–338 (2020).

22. Zwanenburg, A. & Löck, S. familiar: End-to-end automated machine learning and model evaluation (2021).

23. Leger, S. *et al.* Ct imaging during treatment improves radiomic models for patients with locally advanced head and neck cancer. *Radiother. Oncol.* **130**, 10–17 (2019).

24. Leger, S. *et al.* Comprehensive analysis of tumour sub-volumes for radiomic risk modelling in locally advanced hnscc. *Cancers* **12**, 3047 (2020).

25. Hutter, F., Hoos, H. H. & Leyton-Brown, K. Sequential model-based optimization for general algorithm configuration. In *International conference on learning and intelligent optimization*, 507–523 (Springer, 2011).

26. Leger, S. *et al.* A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. *Sci. reports* **7**, 1–11 (2017).

27. Zhao, Z., Anand, R. & Wang, M. Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. In *2019 IEEE international conference on data science and advanced analytics (DSAA)*, 442–452 (IEEE, 2019).

28. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. statistical software* **33**, 1 (2010).

29. Kamarajah, S. K. *et al.* Systematic review and meta-analysis of factors associated with post-operative pancreatic fistula following pancreatoduodenectomy. *ANZ journal surgery* **91**, 810–821 (2021).

30. Kuhn, M., Johnson, K. *et al. Applied predictive modeling*, vol. 26 (Springer, 2013).

31. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal statistical society: series B (Methodological)* **57**, 289–300 (1995).

Acknowledgements

The research was carried out within the Surgomics project funded by the German Federal Ministry of Health, University Hospital Heidelberg and Uniklinikum Carl Gustav Carus, Dresden. Partners from industry: Karl Storz GmbH and Phellow Seven GmbH.

Author contributions statement

Stefan Leger (S.L.) and Fiona R. Kolbinger (F.R.K.) conceived the study; S.L., F.R.K., Nithya Bhasker (N.B.), Alexander Zwanenburg (A.Z.) and Steffen Löck (S. Löck) designed the experiments; A.Z., S.L. and N.B. developed the software; S.L., F.R.K. and N.B. conducted the experiments; N.B. and S.L. visualised the results; S.L., F.R.K., N.B., A.Z. and S. Löck analysed the results; Stefanie Speidel (S.S.), S. Löck, Jürgen Weitz (J.W.), Jens-Peter Kühn (J.K.), Marius Distler (M.D.), and Ralf-Thorsten Hoffmann (R.H.) provided resources and supervision; F.R.K., Nadiia Skorobohach (N.S.), J.K. and M.D. curated the data; N.B., F.R.K., S.L. and S.S. drafted the manuscript; N.B., F.R.K., N.S., S.S., S.L., A.Z., S.Löck, J.W., R.H., M.D. and J.K. reviewed the manuscript draft; All authors have read and agreed to the published version of the manuscript.

Ethics declarations

Ethical approval for the retrospective analyses of clinical and imaging data was obtained from the Ethics Committee at the Technische Universität Dresden (BO-EK-263062020). All analyses were carried out in accordance with the relevant guidelines and regulations, in particular the Declaration of Helsinki and its later amendments. The local institutional review board (Ethics Committee at the Technische Universität Dresden) approved the study protocol (BO-EK-263062020) and waived written patient consent for the retrospective analysis of routinely acquired anonymised clinical data.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information: Refer to the Supplementary information online for more details about the study and additional results.