Mathematical Biosciences and Engineering

Research article

Computer-aided differentiates benign from malignant IPMN and MCN with a novel feature selection algorithm

Chengkang Li[1], Ran Wei[1], Yishen Mao[2], Yi Guo[1,3,*], Ji Li[2,*], Yuanyuan Wang[1,3,*]

[1] The School of Information Science and Technology of Fudan University, Shanghai 200433, China
[2] Department of Pancreatic Surgery, Pancreatic Disease Institute, Huashan Hospital, Shanghai Medical College, Fudan University, Shanghai, 200040, China
[3] Key Laboratory of Medical Imaging, Computing and Computer-Assisted Intervention, Shanghai Medical College, Fudan University, Shanghai, 200433, China

* **Correspondence:** Email: guoyi@fudan.edu.cn, liji@huashan.org.cn, yywang@fudan.edu.cn.

Abstract: In clinical practice, differentiating benign from malignant intraductal papillary mucinous neoplasm (IPMN) and mucinous cystic neoplasm (MCN) preoperatively is crucial for deciding future treating algorithm. However, it remains challenging as benign and malignant lesions usually show similarities in both imaging appearances and clinical indices. Therefore, a robust and accurate computer-aided diagnosis (CAD) system based on radiomics and clinical indices was proposed in this paper to solve this dilemma. In the proposed CAD system, 107 patients were enrolled, where 90 cases were randomly selected for the training set with 5-fold cross validation to build the diagnostic model, while 17 cases were remained for an independent testing set to validate the performance. 436 high-throughput radiomics features while 9 clinical indices were designed and extracted. A novel feature selection algorithm named BLR (Bootstrapping repeated LASSO with Random selections) was proposed to select the most effective features. Then the selected features were sent to Support Vector Machine (SVM) to differentiate the benign or malignant. In the cross-validation cohort and independent testing cohort, the area under receiver operating characteristic curve (AUC) of CAD scheme were 0.83 and 0.92, respectively. The results fully prove the proposed CAD system achieves significant effect in tumors diagnosis.

1. Introduction

In recent years, due to a higher morbidity and the malignant potential of intraductal papillary mucinous neoplasm (IPMN) and mucinous cystic neoplasm (MCN), these two pancreatic cystic neoplasms (PCNs) have drawn great attention of clinical researchers [1–3]. As there is a tremendous difference in the prognosis between benign and malignant cases, recognizing malignant IPMN/MCN preoperatively is very important for tailoring an optimal strategy of treatment for a patient. Patients with benign tumors can receive a function-preserving surgery or even can escape surgery and stick to a close surveillance plan, while patients with malignant tumors may undergo a more radical surgery. However, even for experienced clinicians, it's currently always challenging to distinguish benign lesions from malignant ones, because they show similar imaging appearance and clinical indices and the potential difference cannot be easily recognized by human in routine clinical practice.

Multi-detector row computed tomography (MDCT) is recommended for the initial diagnosis and assessment of IPMN and MCN [4,5]. And the incidence of pancreatic cancer raised dramatically, owing to the wide application of MDCT examination [6]. In clinical practice, there are a lot of high-risk imaging findings associated with cancerization. According to the international consensus guidelines published by Tanaka M et al. [7,8], the diffuse dilation of main pancreatic duct (MPD) > 5 mm is an important signal of malignant IPMN and MCN, while the large cyst > 3 cm, the thickened cystic wall, and the papillary mural nodules are also prominent features of malignant tumors. In addition to radiologic approaches, studies also have shown that certain laboratory results and clinical symptoms can also indicate the malignant transformation of pancreatic neoplasms, e.g., the raising of serum carbohydrate antigen (CA) 19-9, CA12-5 or carcinoembryonic antigen (CEA), the increasing of fasting plasma glucose, and even abdominal pain [9]. However, due to the similar imaging appearances and the subjective clinical diagnosis, there is a solid need from clinicians for an assistant system to help them make a more accurate and objective diagnosis on benign or malignant IPMN/MCN. As a result, some studies have emerged on computer-aided diagnosis (CAD).

Radiomics-based CAD method constructs an objective image evaluation system and the whole process is highly automated, time-saving and effective. At present, there are some CAD methods and radiomics researches related to pancreatic disease. Jayasree et al. [10] predicted cancer risk in branch duct (BD)-IPMN through the extraction of 135 features, then applied wilcoxon rank-sum test (WRST) based p-value feature selection and obtained the AUC of 0.77. Park et al. [11] extracted 431 3D CT radiomics features and used minimum-redundancy maximum-relevancy (mRMR) algorithm to reduce feature dimension, then realized the differential diagnosis of autoimmune pancreatitis and pancreatic ductal adenocarcinoma. Zhang et al. [12] extracted 251 radiomics features from 2D and 3D PET/CT images, and used support vector machine recursive feature elimination (SVM-RFE) to select effective features, while support vector machine (SVM) was choose as classifier to discriminate autoimmune pancreatitis (AIP) and pancreatic ductal adenocarcinoma (PDAC). Wei et al. [13] extracted 409 MDCT radiomics features and used bootstrapping repeated least absolute shrinkage selection operator (LASSO) regression to select effective features. Finally, they distinguished between serous cystic neoplasm (SCN) and other subtypes of PCN, which obtained the AUC of 0.84.

Throughout the abovementioned CAD methods, there are some urgent problems to be solved. First, the above feature extraction modules did not form a complete and comprehensive feature system optimized for PCN. Second, most of the feature selection methods only selecting once, which leads to the instability and contingency of the results. Third, the CAD methods have not made many attempts

focusing on the task of the differentiation between benign and malignant of both IPMN and MCN. Therefore, a robust and improved radiomics CAD system is highly needed to solve these problems.

In this paper, an enhanced CAD system was proposed, which combines the radiomics features with the clinical indices and forms a complete, comprehensive and PCN-optimized feature system. In addition, a fashion and high-performance feature selection algorithm named Bootstrapping repeated LASSO with Random selections (BLR) was used in our CAD system. Finally, the features selected by BLR were sent to SVM to classify benign and malignant of IPMN and MCN.

The workflow of the proposed CAD system is shown in Figure 1. Our major contributions can be summarized as follows:

1) The feature extraction modules in the existing CAD system are not systematic, nor optimized for PCN. Besides, it lacks of some important clinical symptoms or laboratory indicators. Therefore, a comprehensively, completely, and PCN-optimized feature system has been built including 436 radiomics feature and 9 clinical indices to describe the image characteristics of PCNs automatically and accurately.

2) A novel algorithm named BLR was proposed, which increases the stability of feature selection through a large number of bootstrapping repetitions while reduces the bias and overfitting, and maximizes the diagnostic potential of features by means of random selections.

3) A robust and precise preoperative CAD system was established aimed to classify the benign and malignant of IPMN/MCN, which forms a reliable and objective image evaluation and diagnostic method.
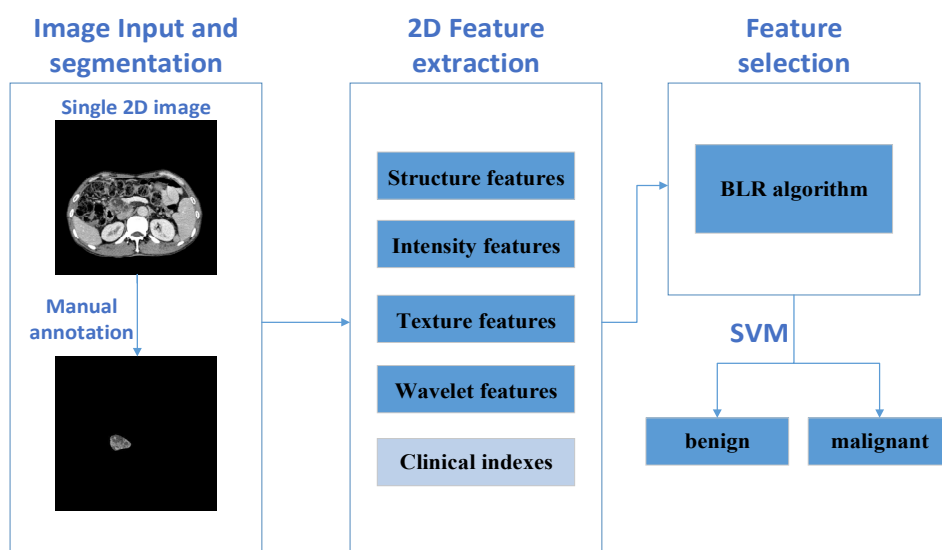

Figure 1. The workflow of the proposed CAD system.

2. Materials and methods

2.1. Dataset and segmentation

This study is a retrospective study. The 107 patients were enrolled at Department of Pancreatic Surgery, Huashan Hospital of Fudan University, Shanghai, China, from December 2007 to August 2016. There are 73 IPMN and 34 MCN cases, including 71 benign and 36 malignant. All patients have signed

informed consent, and our study has been approved by the ethics committee of the Huashan Hospital. The characteristics of enrolled patients are shown in Table 1.

All patients underwent preoperative abdominal enhance MDCT scans with the thickness of 1.5 mm and already received surgery, the postoperative pathological examination result was taken as the ground-truth and final diagnosis of the patient. Due to the higher image quality of portal venous phase, the single 2D image with the largest tumor cross section in portal venous phase of abdominal enhance MDCT sequence was chosen as the input. The portal venous phase is the images collected in 60–65 s after intravenous injection of nonionic-iodinated contrast agent with the reagents concentration of 370 mg I/mL and the injection rate of 4 mL/s. Preoperative data used in this study of these patients is complete and available.

The size of all images was 512×512 and the tumors were outlined by experienced radiologists manually. The 90 cases were randomly selected as the training cohort with 5-fold cross-validation, while the remaining 17 cases formed the independent testing cohort. The data processing was implemented in Matlab R2018b (Mathworks, Inc, Natick, Massachusetts).

Table 1. Patient characteristics.

Category	Training cohort		Independent testing cohort	
	malignant	benign	malignant	benign
IPMN	25	36	7	5
MCN	2	27	2	3
total	27	63	9	8

2.2. High-throughput feature design and extraction

A PCN-optimized feature system was designed according to the clinical guidelines published by Sahani et al. [14] in 2013, which practically covers all the features included in the clinical guidelines.

Specifically, 436 radiomics features were designed and extracted, including 21 structure features, 16 intensity features, 67 texture features, and 332 wavelet features. Besides, 9 clinical indices were added to our feature system to promote the effects of radiomics-based CAD scheme and make it closer to the real clinical diagnosis model. The extracted features are summarized in Table 2, while the detailed features are shown in Appendix. The correspondences between the extracted features and the clinical guidelines features are shown in Table 3.

Structure features are used for describe the shape, edge, size and internal structure of the tumor [13,15]. Shape features comprehensively reflect the shape information of tumors, such as tumor size, shape, edge roughness and so on. The internal structure features focus on PCN such as internal cyst, calcification, wall thickness and central scar.

Intensity features reflect the intensity and histogram information of the tumors.

Texture features calculate the texture information of tumor region, including Inner regional comparison features and Other texture features.

2.2.1. Inner regional comparison features

To better describe the gray level and texture distribution disparity between the different inner

regions of the tumor, 6 texture features were designed, named inner regional comparison features. The greatest advantage of inner regional comparison features is that it can calculate the inner-regional disparity of the tumor with any shape, in the meanwhile, without the serious interference of the other tissues outside of the mass or compared areas. Figure 2 reflects the process of obtaining the internal regions of tumor. Figure 2(a) is the original image with the size of R × C. The R and C are both 512 in this study; (b) is the mask of the tumor; (c) is the tumor region; (d) and (e) represent $ROI1_{ij}$ and $ROI2_{ij}$, which are the upper left corner and the lower right corner of (c) respectively and with the same size of $R - i \times C - j$; (f) shows the intersection of $ROI1_{ij}$ and $ROI2_{ij}$; (g) is the mask of intersection region; (h) and (i) are $Region1_{ij}$ and $Region2_{ij}$, which are the point multiplication of (g) and (d) or (e), respectively. In simple terms, $Region1_{ij}$ and $Region2_{ij}$ are the different inner regions of the tumor to be compared, all of the other subpictures are the intermediate processes to obtain $Region1_{ij}$ and $Region2_{ij}$. Notice that the size of $Region1_{ij}$ and $Region2_{ij}$ vary with the value of i and j.

Table 2. Summary of the extracted high-throughput radiomics features and clinical indices. The detailed features are shown in Appendix.

Type	Name		Number
Structure	Shape features (13), Inner-structure features (8)		21
Intensity			16
Texture	Inner regional comparison features (6)	mean of inner-regional dissimilarity, SD of inner-regional dissimilarity, mean of inner-regional contrast, SD of inner-regional contrast, mean of inner-regional covariance, SD of inner-regional covariance	67
	ROI-based features (7), High-order matrices texture features (GLCM (23), GLRLM (13), GLSZM (13), NGTDM (5))		
Wavelet	LL HL LH HH decomposition		332
Clinical indices	Sex, age, preoperative fasting plasma glucose, CA199, CA125, CEA, tumor size, tumor location, clinical symptoms		9

Abbreviations: ROI is region of interest; IMC is informational measure of correlation; SD is standard deviation.

Table 3. Correspondence between the extracted features and the clinical guidelines features.

| Clinical guidelines features	Extracted features
Age	Clinical indices-Age
Sex	Clinical indices-Sex
Location	Clinical indices-Location
Shape	Shape features
Size	Clinical indices-tumor size, General shape features-diameter of equivalent circle
Wall	Average wall thickness
Internal cysts	number of cysts, cyst size, SD of cysts area
Central scar	central scar density
Calcification	number of calcifications, SD of calcification area, calcification area location
Intensity	Intensity features, Texture features, Wavelet features
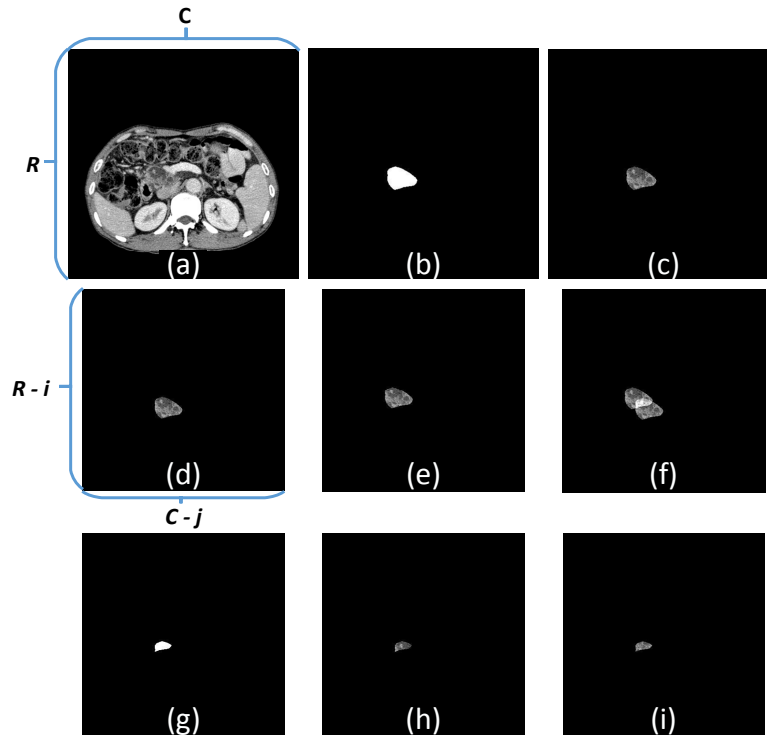
Figure 2. The process of obtaining the internal regions of tumor. (a)–(g) are the intermediate processes to get (h) and (i); (h) and (i) are the two obtained tumor's internal regions.

So, the calculation method and formula of each feature are as follows:

1) Mean of inner-regional dissimilarity ($Diss_m$):

$$Diss_m = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{K-1} \frac{sum\left|Region1_{ij} - Region2_{ij}\right|}{N}}{M \times K} \quad (1)$$

where the sum means to sum all pixels of an image; N is the number of pixels of foreground in (g); M and K are the positive integers, which can be regarded as the maximum offset when using Figure 2(c) to get (d) and (e). Their values should be appropriate, not too large or too small. In this study, M and K are both equal to 5. The dissimilarity of different inner regions of the tumor is proportional to the value of this feature.

2) Standard deviation of inner-regional dissimilarity ($Diss_{std}$):

$$Diss_{m_{ij}} = \frac{sum\left|Region1_{ij} - Region2_{ij}\right|}{N} \quad (2)$$

$$Diss_{std} = \sqrt{\frac{\sum_{i=0}^{M-1} \sum_{j=0}^{K-1} \left(Diss_{m_{ij}} - Diss_m\right)^2}{M \times K}} \quad (3)$$

This feature reflects the stability of inner-regional dissimilarity.

3) Mean of inner-regional contrast (Con_m):

$$Con_m = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{K-1} \frac{sum(ROI1n_{ij}*ROI2n_{ij})}{N}}{M \times K} \tag{4}$$

$$Tumor_n = \left(\frac{Tumor - min(Tumor)}{max(Tumor) - min(Tumor)}\right) * BW \tag{5}$$

$$ROI1\_n_{ij} = Tumor_n(1:R-i, 1:C-j) \tag{6}$$

$$ROI2\_n_{ij} = Tumor_n(1+i:R, 1+j:C) \tag{7}$$

where the $*$ represents point multiplication of two images; BW and $Tumor$ represent the Figure 2(b),(c), respectively. min/max are the minimum and maximum value of the tumor region in (c); $Tumor_n$ is the normalized image of the tumor region in (c). $ROI1\_n_{ij}$ and $ROI2\_n_{ij}$ are the upper left and the lower right corner of $Tumor_n$ respectively, and with the same size of $R-i \times C-j$. This feature reflects the contrast inside the tumor.

4) Standard deviation of inner-regional contrast (Con_{std}):

$$Con_{m_{ij}} = \frac{sum(ROI1\_n_{ij}*ROI2\_n_{ij})}{N} \tag{8}$$

$$Con_{std} = \sqrt{\frac{\sum_{i=0}^{M-1} \sum_{j=0}^{K-1}(Con_{m_{ij}} - Con_m)^2}{M \times K}} \tag{9}$$

This feature reflects the stability of inner-regional contrast.

5) Mean of inner-regional covariance ($Cova_m$):

$$I_m = \frac{sum(Tumor_n)}{P} \tag{10}$$

$$Cova_{m_{ij}} = \frac{sum[(ROI1\_n_{ij} - I_m)*(ROI2\_n_{ij} - I_m)*\alpha]}{N} \tag{11}$$

$$Cova_m = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{K-1} Cova_{m_{ij}}}{M \times K} \tag{12}$$

P is the number of pixels with the value of 1 in the Figure 2(b); α represents the Figure 2(g), which is the mask of intersection region of $ROI1\_n_{ij}$ and $ROI2\_n_{ij}$. This feature calculates the covariance of the tumor's different inner-regions.

6) Standard deviation of inner-regional covariance ($Cova_{std}$):

$$Cova_{std} = \sqrt{\frac{\sum_{i=0}^{M-1} \sum_{j=0}^{K-1}(Cova_{m_{ij}} - Cova_m)^2}{M \times K}} \tag{13}$$

This feature reflects the stability of inner-regional covariance.

2.2.2. Other texture features

ROI-based features calculate the relevant texture characteristics in the bounding-box of the tumor [13,16]. Besides, gray-level co-occurrence matrix (GLCM) [17], gray-level run-length matrix

(GLRLM) [18], gray-level size zone matrix (GLSZM) [19], and neighborhood gray-tone difference matrix (NGTDM) [20] are common high-order matrices to describe image texture.

In order to get the deeper intensity and texture information of the image, wavelet transform was applied to all two-dimensional images and four components were obtained, which are Low pass/Low pass (LL), Low pass/High pass (LH), High pass/Low pass (HL), and High pass/High pass (HH). The 83 intensity and texture features were extracted from four component respectively, and obtained 332 wavelet features in total.

Clinical indices record the patients' demographic information, preoperative laboratory indicators, and clinical symptoms comprehensively. The quantization methods of each feature are described below:
1) Sex: 0 and 1 represent male and female, respectively.
2) Age: 1–4 represent age range from 1 to 19, 20 to 39, 40 to 59 and above 60.
3) Preoperative fasting plasma glucose (mmol/L): range from 4.1 to 11.3.
4) Serum cancer antigen 199 (CA199, u/mL): range from 0.6 to 999.
5) Serum cancer antigen 125 (CA125, u/mL): range from 2.0 to 116.9.
6) Carcinoembryonic antigen (CEA, ug/mL): range from 0.2 to 6.1.
7) tumor size (cm): range from 0.2 to 15.
8) Location: 1, 1.5, 2, 2.5, 3, 3.5, 4, in which 1, 2, 3, 4 represent the head, neck, body, tail of pancreas, respectively. While 1.5 represents between the head and the neck of the pancreas, 2.5, 3.5 are similar.
9) Clinical symptoms: 0–6 represent asymptomatic, abdominal pain, abdominal distension, weight loss, jaundice, mass, and pancreatitis, respectively.

To sum up, a complete, comprehensive and PCN-optimized feature system with 445 high-throughput features has been established in this study.

2.3. BLR feature selection algorithm

LASSO regression has been widely used in the feature extraction because of its efficiency and high performance [21,22]. It can shrink the coefficients of the redundant features to zero. But selecting features only once can be extremely occasional and cause overfitting. Therefore, in our previous works [13,16], LASSO regression was combined with a large number of bootstrapping repetitions to increase the reliability of feature selection meanwhile prevent overfitting. To be exact, all of the input features were ranked according to their occurrence number in the multiple bootstrapping repeated LASSO regression, and the features with top 10% reproducibility were selected as the final feature subset.

However, it is far from sufficient if only combines features with top reproducibility mechanically, because it ignores the potential ability of features with non-top but high reproducibility and will reduces the classification accuracy. Therefore, in this study, to obtain better performances, an extremely huge number of random selections were implemented among the features with high reproducibility, then chose the best feature combination according to the comprehensive classification effect. This algorithm named BLR.

The steps of proposed BLR algorithm are as follows:
1) Use the LASSO regression model on the training cohort with 5-fold cross-validation to select the effective features, and bootstrapping 300 times.
2) Sort the features according to the reproducibility in above LASSO selections, and the top M features were selected to form the feature subset1.

3) Select N (N < M) features from the feature subset1 randomly K times (K is extremely large) to obtain K feature combinations, feed them to SVM (use 5-fold cross-validation, besides, bootstrapping repeat 100 times and take the average) to get their classification ability.

4) The threshold method was used for select L (L << K) feature combinations with excellent classification performance from K feature combinations obtained in step 3).

5) Select the best one from L feature combinations obtained in step 4) manually according to the comprehensive classification performance, and get the corresponding final feature subset.

BLR increases the reliability of LASSO regression through a lot of bootstrapping repetitions. Besides, it avoids simple feature combination, which picks up the features with the non-top reproducibility and maximizes their diagnostic potential through a large number of random extractions. Meanwhile, it explores various possibilities of feature combinations. Therefore, the BLR algorithm has a great probability to obtain better classification effect.

2.4. Diagnosis of benign and malignant

The final feature subset selected by BLR was sent to SVM with linear kernel to complete the differentiation between benign and malignant of IPMN and MCN. All selected features were normalized to [−1, +1].

Some important details of SVM were configured as follows: The maximal number of optimization iterations was 1,000,000. In order to solve the slight sample imbalance in the training cohort, we doubled the cost of the misjudged malignant tumors in the training stage. In the K-fold training set, the classification results were averaged as the final classification scores.

3. Experiments

In this section, a series of experiments were conducted to illustrate the performance of our method.

3.1. Compared methods

Our proposed method was compared with some state-of-the-art methods in four aspects: the comprehensive feature sets, the feature selection methods, the performance of various classifiers, and the overall pancreas related CAD methods.

As for the feature sets, the effect of inclusion and exclusion of clinical indices on the proposed CAD system was contrasted.

As for the feature selection methods, our BLR algorithm was compared with three most commonly feature selection algorithms, which were p-value, relief and logistic regression, respectively. In addition, the influences of two important factors of BLR were observed, which are bootstrapping repetitions and random selections. For a fair comparison, the feature sets fed into various selection methods were the same. The condition of each method is as follows:

1) P-value algorithm: The WRST based p-value feature selection method with $p < 0.01$ was applied.

2) Relief and logistic regression algorithm: The features with top 20 weights were selected.

3) LASSO: The lasso fit was constructed by 10-fold cross-validation, and the features corresponding to the minimum cross-validated mean squared error (MSE) were combined as final feature subset.

4) BL: The effect of bootstrapping repetitions on the LASSO model was verified and repeated this

step 300 times. Then the features with top 10% reproducibility were selected as final feature subset.
5) BLR: N features were selected randomly from the top M features in BL. Repeated this step K times and obtained K feature combinations, fed them to SVM. Threshold method was used for selecting L combinations with excellent classification effect, then the best combination was selected manually as the final feature subset. The hyperparameters like N, M, K, and thresholds were optimized according to the experimental conditions.

As for the classifiers, the effect of different classifiers was compared on the 5-fold cross-validation cohort. The image preprocessing was exactly the same. The proposed feature extraction module without clinical indices was used. And the feature selection algorithm was the baseline LASSO regression. The selected features were sent to the compared classifiers, including Decision Tree (DT), K-Nearest Neighbor (KNN), Back Propagation Neural Network (BPNN), Naive Bayes (NB), and SVM.

As for the overall pancreas related CAD methods, we compared our CAD system with other pancreas related radiomics-based method, including Park et al. [11], Zhang et al. [12], and Wei et al. [13]. For Park et al., the features of volume were replaced with those of area to adapt our data dimensions. And mRMR algorithm was used to feature selection, then the random forest was used to classification. For Zhang et al., we extracted the 2D CT radiomics features mentioned in their article and used SVM-RFE for feature selection, then the selected features were sent to SVM. For Wei et al., we extracted 409 features and used their LASSO-based feature selection algorithm, then the SVM was used to diagnosis.

3.2. Evaluation metrics

Four indices were used for evaluate the performance of established model, there are the area under the ROC curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE), respectively.

4. Results and discussion

4.1. Efficiency of different classifiers

The classification performances of 5 classifiers were compared. We extracted the proposed feature system without clinical indices and sent them to the baseline LASSO algorithm to select features. The selected features were fed to the compared classifiers, and the related experimental results are listed in Table 4.

From Table 4, in the 5-fold cross-validation cohort, the SVM obtained AUC, ACC, SEN, SPE = 0.80, 0.74, 0.68, and 0.77, respectively, which achieves the best classification effects among the compared classifiers. So, we choose SVM as the final classifier of the following experiments.

4.2. Efficiency of feature selection algorithms

The classification efficiencies of 6 feature selection algorithms were compared in this part. The feature system without clinical indices was fed to the above 6 algorithms, and the selected feature subsets were sent to SVM to analyze their classification performance.

In a cross-validation cohort, for p-value, 24 features with $p < 0.01$ were fed to SVM and obtained

AUC, ACC, SEN, SPE = 0.72, 0.72, 0.65, and 0.75, respectively. For relief and logistic regression, 20 features with top weights were selected as final feature subsets and obtained AUC, ACC, SEN, SPE = 0.63 versus 0.66, 0.63 versus 0.65, 0.50 versus 0.54, and 0.69 versus 0.70, respectively. For LASSO, 11 features were selected and obtained AUC, ACC, SEN, SPE = 0.80, 0.74, 0.68, and 0.77, respectively. And for BL, 13 features with top 10% reproducibility were selected and obtained AUC, ACC, SEN, SPE = 0.80, 0.74, 0.67, and 0.77, respectively. For BLR, M, N, and K were set to 23, 13, and 20,000, the thresholds were set to AUC = 0.81, ACC = 0.75, SEN = 0.70, SPE = 0.76, and selected L = 100 feature combinations. Finally, the best feature subset with 13 features was fed to SVM, obtained AUC, ACC, SEN, SPE = 0.82, 0.76, 0.71, 0.79. The above results are summarized in Table 5.

Table 4. The classification results of compared feature selection algorithms without clinical indices in cross-validation cohort.

Classifier	5-fold cross-validation cohort			
	AUC	ACC	SEN	SPE
DT	0.63	0.66	0.51	0.72
KNN	0.67	0.67	0.60	0.70
BPNN	0.70	0.66	0.62	0.70
NB	0.78	0.71	0.65	0.76
SVM	**0.80**	**0.74**	**0.68**	**0.77**

Table 5. The classification results of compared feature selection algorithms without clinical indices in cross-validation cohort.

Algorithm	num	5-fold cross-validation cohort			
		AUC	ACC	SEN	SPE
P-value	24	0.72	0.72	0.65	0.75
relief	20	0.63	0.63	0.50	0.69
logistic	20	0.66	0.65	0.54	0.70
LASSO	11	0.80	0.74	0.68	0.77
BL	13	0.80	0.74	0.67	0.77
BLR	13	**0.82**	**0.76**	**0.71**	**0.79**

Hence, the BLR algorithm has the best performance. By a large number of bootstrapping repetitions and random selections, the BLR algorithm avoids the bias and overfitting of single LASSO regression while maximizes the potential diagnosis abilities of the features with non-top reproducibility. It should be noted that the value of N is approximately equal to the number of features selected by BL algorithm, and M should not be too large, because the excessive value of M may lead to dramatic number of combinations while consider many useless features. At the same time, the value of K should be very large to make the experiment more general.

4.3. Analysis of the selected features

Some typical characteristics selected by BLR and their relevant information are shown in Table 6.

And Figure 3 shows the boxplots of some representative features.

We found that the roundness, mean of inner-regional covariance, the run-length variance and so on are of outstanding significance for the classification of benign and malignant IPMN or MCN. Now, we will briefly observe the imaging differences between benign and malignant tumors from these important selected features.

As we can see in Figure 4, the malignant tumors are more likely to have irregular shapes, while the benign tumors tend to be rounder. This finding is in accordance with the different growth pattern of benign and malignant tumors: benign tumors have an expansible growth pattern, while malignant ones have an infiltrative pattern and tend to invade adjacent structures.

Figure 5 shows the tumors with different H-variance, this feature reflects the intensity and histogram distribution inside the tumor, indicating that the intensity of the malignant tumor is more uneven. This might be caused by the solid components inside the cyst of the tumor. These solid nodules or components grow irregularly and are long regarded as the origin of malignant formation in the tumor.

Table 6. Representative features selected by BLR. The corresponding wavelet features are no longer listed separately.

Category	Feature	Malignant (Mean [SD])	Benign (Mean [SD])	P Value
Structure	Roundness	0.712 (0.027)	0.786 (0.014)	0.017
	SD of normalized radius	0.109 (0.0024)	0.087 (0.0014)	0.021
Intensity	H-variance	2457 (1.3×10^6)	1653 (1.0×10^6)	< 0.01
Texture	mean of inner-regional covariance	0.0105 (2.92×10^{-5})	0.0056 (2.53×10^{-5})	< 0.01
	SD of inner-regional contrast	0.011 (4.23×10^{-5})	0.008 (2.78×10^{-5})	0.011
	SD of inner-regional covariance	0.008 (1.99×10^{-5})	0.006 (1.48×10^{-5})	< 0.01
	grey-level variance	0.054 (0.0004)	0.063 (0.0004)	0.044
	run-length variance	2.6×10^{-4} (3.3×10^{-9})	3.1×10^{-4} (3.7×10^{-9})	< 0.01
Clinical indices	Clinical symptoms	–	–	0.016

Note: Did not calculate the mean and SD of clinical symptoms because there are meaningless. And the H-variance represents histogram variance.
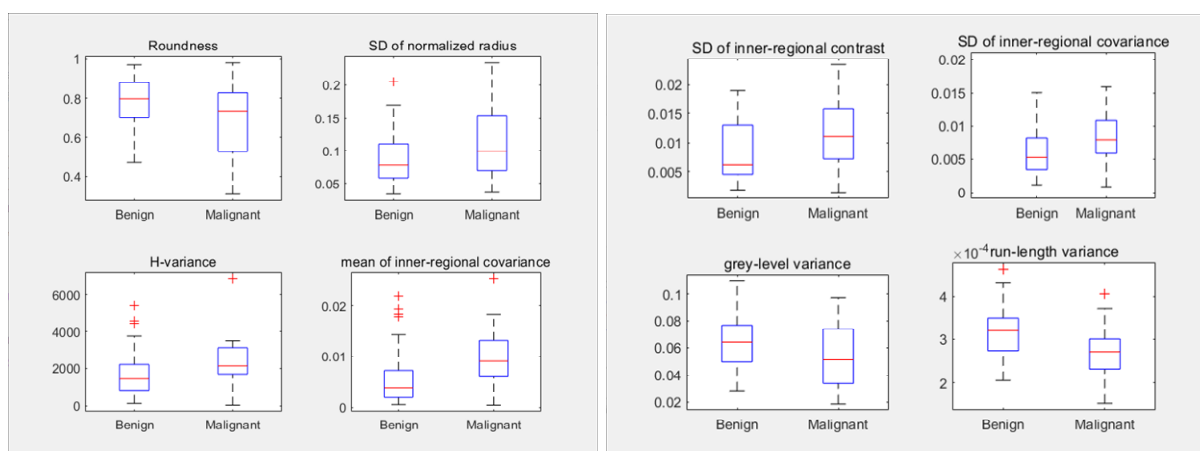

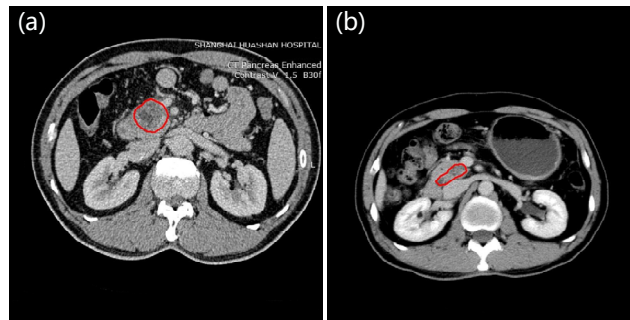**Figure 3.** The boxplots of some representative features.

Figure 4. The comparison of benign and malignant tumors from shape. (a) is the benign tumor with higher Roundness and SD of normalized radius, while (b) is the malignant tumor with lower values.
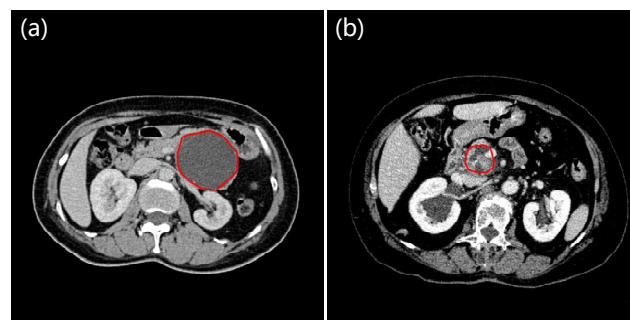

Figure 5. The comparison of benign and malignant tumors from intensity. (a) is the benign tumor with lower H-variance, while (b) is the malignant tumor with higher H-variance.
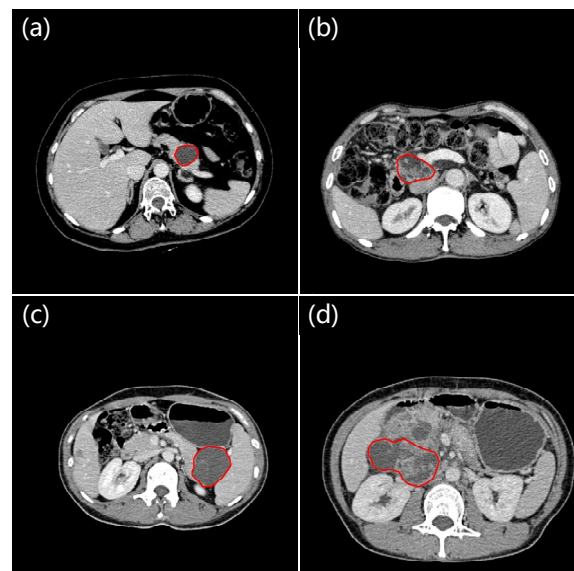

Figure 6. The comparison of benign and malignant tumors from texture. (a),(b) are the benign/malignant tumor with lower/higher mean of inner-regional covariance. While (c),(d) are the benign/malignant tumor with higher/lower run-length variance.

Figure 6 reflects the image differences of the benign and malignant tumors from the perspective of texture. In the malignant tumors, the texture and density distributions are more complex and inhomogeneous, the gray level ranges are wider, and the disparities between the different inner regions of the tumor are quite distinct. This phenomenon identified reflects the fact that malignant tumors usually have a more complex composition than benign tumors. Beside cancerous cells, there are also small nourish vessels resulted from neovascularization, the characteristic fibrotic stroma of pancreatic malignancy, and so on. All of these components above have distinct difference on CT images, which make the texture of malignant tumors shows more complex on the image.

A great difference between benign and malignant tumors in shape, density, and texture are illustrated. Compared with the benign tumors, the malignant tumors usually have more irregular shape, intensity, and texture distribution. However, the thickened cystic wall, the large cyst and papillary mural nodules may lead to the above problem. It also reflects that these imaging features are very important to the classification of benign/malignant IPMN or MCN.

4.4. Diagnostic performances

In this part, the feature system with clinical indices was fed to 6 feature selection algorithms to contrast the influence of clinical indices in the classification performance.

The parameters of p-value, relief, and logistic regression, LASSO, BL algorithm were configured same as that in section 4.1. The only difference was the final feature subsets selected by LASSO and BL contain 12 and 10 features, respectively. For the BLR algorithm, the hyperparameters were as follows: M = 21, N = 10, K = 20,000, the thresholds of classifiers were AUC, ACC, SEN, SPE = 0.82, 0.75, 0.73, and 0.76, respectively. The output parameter was L = 108, and the best subset with 10 features were selected.

Table 7. The comparison results of different feature selection algorithms with or without clinical indices.

	Algorithm	num	5-fold cross-validation cohort				independent validation cohort			
			AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE
	P-value	24	0.72	0.72	0.65	0.75	0.61	0.59	0.67	0.50
	relief	20	0.63	0.63	0.50	0.69	0.68	0.59	0.56	0.63
Without	logistic	20	0.66	0.65	0.54	0.70	0.75	0.65	0.67	0.63
Clinical	LASSO	11	0.80	0.74	0.68	0.77	0.69	0.53	0.56	0.50
indices	BL	13	0.80	0.74	0.67	0.77	0.82	0.76	0.67	0.88
	BLR	13	**0.82**	**0.76**	**0.71**	**0.79**	**0.86**	**0.76**	**0.67**	**0.88**
	P-value	24	0.73	0.72	0.66	0.75	0.61	0.59	0.67	0.50
	relief	20	0.62	0.62	0.49	0.68	0.63	0.59	0.56	0.63
With	logistic	20	0.67	0.65	0.55	0.70	0.60	0.65	0.67	0.63
Clinical	LASSO	12	0.80	0.74	0.68	0.77	0.71	0.53	0.44	0.63
indices	BL	10	0.80	0.74	0.67	0.77	0.88	0.82	0.78	0.88
	BLR	10	**0.83**	**0.76**	**0.73**	**0.77**	**0.92**	**0.88**	**0.89**	**0.88**

The final feature subsets of the various algorithms were fed to SVM to realize the identification

of benign and malignant. Table 7 compares the classification results of proposed CAD scheme using different feature selection methods with or without clinical indices, and records the number of features selected by different algorithms. The ROC curves are shown in Figure 7.

As can be seen, the BLR is the state-of-the-art method. Beyond that, the clinical indices are also very important, which improved diagnosis accuracy as well as made the CAD system closer to the real clinical diagnosis. This suggests that, from a statistical point of view, the patients with malignant tumors may have some abnormal symptoms such as abdominal pain, weight loss, and so on. However, different people may exhibit different characteristics.
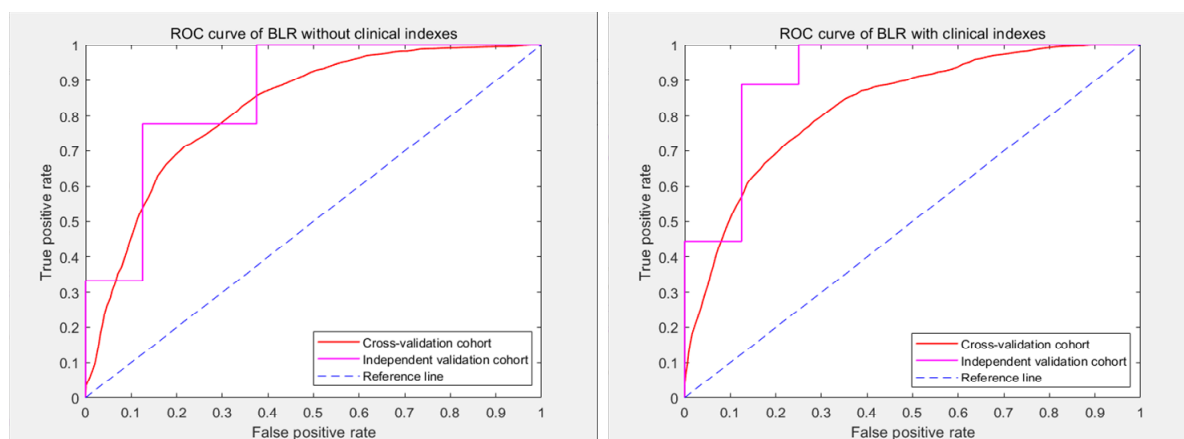

Figure 7. ROC curves of BLR feature selection algorithms with (right picture) or without (left picture) clinical indices.

4.5. Overall comparison with other CAD methods

The overall diagnosis performances of different pancreas related CAD methods were compared. The results are compared as follows:

Table 8. The comparison with other CAD methods.

Algorithm	5-fold cross-validation cohort				independent validation cohort			
	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE
Park et al. [11]	0.68	0.66	0.52	0.73	0.79	0.59	0.44	0.75
Zhang et al. [12]	0.72	0.68	0.63	0.70	0.78	0.71	0.67	0.75
Wei et al. [13]	0.79	0.74	0.64	0.77	0.71	0.71	0.56	0.88
Our	**0.83**	**0.76**	**0.73**	**0.77**	**0.92**	**0.88**	**0.89**	**0.88**

As can be seen from Table 8, our method achieves the state-of-the-art effects, proving our superiority. Especially for the sensitivity of diagnosis, which is usually low in clinical artificial diagnosis and other methods, our approach has made great progress. Compared to Park et al., the SEN of our method is increased from 0.52 to 0.73 in cross-validation cohort, and from 0.44 to 0.89 in independent validation cohort. The superiority of our method is due to the comprehensively, completely, and PCN-optimized feature system and the excellent feature selection algorithm.

5. Conclusions

In conclusion, this study builds a CAD system based on radiomics, and realizes the differentiation malignant from benign of IPMN and MCN. A complete, comprehensive, and PCN-optimized feature system is formed in our research, meanwhile 9 clinical indices are added to improve the diagnostic accuracy. Besides, a novel and wonderful feature selection algorithm is proposed and achieved remarkable effects. Our CAD system forms an objective tumor evaluation method which can reduce the probability of misdiagnosis of malignant tumors, so as to provide crucial suggestions for the treatment planning and delivery for patients. In addition, the proposed CAD system provides a time-saving diagnosis scheme with low computation cost. The calculation part of our CAD system only takes less than 2 s per patient.

However, we still have a lot of room for improvement. Firstly, a small dataset may underestimate the effect and practicability of proposed CAD system. In the future work, the data set can be expanded to make the results more universal. Secondly, manual segmentation can ensure the diagnosis accuracy of proposed CAD system, but it increases the time cost, an automatic segmentation method is considered in the next step.

Acknowledgments

Conflict of interest

There is no conflict of interest in this work.

References

1. K. Tulla, A. Maker, Can we better predict the biologic behavior of incidental IPMN? A comprehensive analysis of molecular diagnostics and biomarkers in intraductal papillary mucinous neoplasms of the pancreas, *Langenbecks Arch. Surg.,* **403** (2018), 151–194.

2. M. Daude, F. Muscari, C. Buscail, N. Carrere, P. Otal, J. Selves, et al., Outcomes of nonresected main-duct intraductal papillary mucinous neoplasms of the pancreas, *World J. Gastroenterol.,* **21** (2015), 2658–2667.

3. J. Farrell, Prevalence, diagnosis and management of pancreatic cystic neoplasms: current status and future directions, *Gut Liver*, **9** (2015), 571–589.

4. K. Ohta, M. Tanada, Y. Sugawara, N. Teramoto, H. Iguchi, Usefulness of positron emission tomography (pet)/contrast-enhanced computed tomography (ce-ct) in discriminating between malignant and benign intraductal papillary mucinous neoplasms (ipmns), *Pancreatology*, **17** (2017), 911–919.

5. S. Choi, J. Kim, M. Yu, H. Eun, H. Lee, J. Han, Diagnostic performance and imaging features for predicting the malignant potential of intraductal papillary mucinous neoplasm of the pancreas: a comparison of eus, contrast-enhanced ct and mri, *Abdom. Radiol.*, **42** (2017), 1449–1458.

6. D. D. D. Brennan, G. A. Zamboni, V. D. Raptopoulos, J. B. Kruskal, Comprehensive preoperative assessment of pancreatic adenocarcinoma with 64-section volumetric CT, *Radiographics*, **27** (2007), 1653–1666.

7. M. Tanaka, C. F. Castillo, V. Adsay, S. Chari , M. Falconi, J. Y. Jang, et al., International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas, *Pancreatology*, **12** (2012), 183–197.

8. M. Tanaka, S. Chari, V. Adsay, F. Castillo, M. Falconi, M. Shimizu, et al., International consensus guidelines for management of intraductal papillary mucinous neoplasms and mucinous cystic neoplasms of the pancreas, *Pancreatology*, **6** (2006), 17–32.

9. Y. Gu, C. Lan, H. Pei, S. N. Yang, F. Y. Liu, L. L. Xiao, Applicative value of serum CA19-9, CEA, CA125 and CA242 in diagnosis and prognosis for patients with pancreatic cancer treated by concurrent chemoradiotherapy, *Asian Pac. J. Cancer Prev.*, **16** (2015), 6569–6573.

10. C. Jayasree, M. Abhishek, G. Lior, A. Marc, L. Liana, A. Peter, et al., CT radiomics to predict high risk intraductal papillary mucinous neoplasms of the pancreas, *Med. Phys.*, **45** (2018), 5019–5029.

11. S. Park, L. C. Chu, R. Hruban, B. Vogelstein, K. W. Kinzler, A. L. Yuille, et al., Differentiating autoimmune pancreatitis from pancreatic ductal adenocarcinoma with CT radiomics features, *Diagn. Interventional Imaging*, **101** (2020), 555–564.

12. Y. Zhang, C. Cheng, Z. Liu, L. Wang, G. Pan, G. Sun, et al., Radiomics analysis for the differentiation of autoimmune pancreatitis and pancreatic ductal adenocarcinoma in 18F-FDG PET/CT, *Med. Phys.*, **46** (2019), 4520–4530.

13. R. Wei, K. Lin, W. Yan, Y. Guo, Y. Wang, J. Li, et al., Computer-aided diagnosis of pancreas serous cystic neoplasms: a radiomics method on preoperative MDCT images, *Technol. Cancer Res. Treat.*, **18** (2019), 1–9.

14. D. Sahani, A. Kambadakone, M. Macari, N. Takahashi, S. Chari, F. Castillo, Diagnosis and management of cystic pancreatic lesions, *Am. J. Roentgenol.*, **200** (2013), 343–354.

15. Y. Chou, C. Tiu, G. Hung, S. Wu, T. Chang, H. Chiang, Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis, *Ultrasound Med. Biol.*, **27** (2001), 1493–1498.

16. Y. Guo, Y. Hu, M. Qiao, Y. Wang, J. Yu, J. Li, et al., Radiomics analysis on ultrasound for prediction of biologic behavior in breast invasive ductal carcinoma, *Clin. Breast Cancer*, **18** (2018), e335–e344.

17. R. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, *IEEE Trans. Syst. Man Cybern.*, **3** (1973), 610–621.

18. M. Galloway, Texture analysis using gray level run lengths, *NASA STI/Recon Tech. Rep. N*, **4** (1975), 172–179.

19. G. Thibault, B. Fertil, C. Navarro, S. Pereira, P. Cau, N. Levy, et al., Shape and texture indices application to cell nuclei classification, *Int. J. Pattern Recognit. Artif. Intell.*, **27** (2013), 1357002.

20. M. Amadasun, R. King, Textural features corresponding to textural properties, *IEEE Trans. Syst. Man Cybern.*, **19** (1989), 1264–1274.

21. A. Dalalyan., M. Hebiri, J. Lederer, On the prediction performance of the Lasso, *Bernoulli,* **23** (2017), 552–581.

22. H. Richard, R. Patricia, R. Vincent, Lasso and probabilistic inequalities for multivariate point processes, *Bernoulli*, **21** (2015), 83–143.

Appendix

Table A1. The details of omitted features.

Type	Name
Shape features (13)	compactness, extreme point number, roundness, area circumference ratio, diameter of equivalent circle, speculation, edge roughness, SD of normalized radius, convexity, solidity, moment difference, rectangle-fitting factor, entropy of normalized radius histogram
Inner-structure features (8)	number of cysts, cyst size, SD of cysts area, central scar density, average wall thickness, number of calcifications, SD of calcification area, calcification area location
Intensity	energy, entropy, kurtosis, mean absolute deviation, mean, median, range, root mean square, skewness, standard deviation, uniformity, variance, histogram kurtosis, histogram variance, histogram skewness, histogram mean
ROI-based features (7)	mean of ROI contrast, SD of ROI contrast, mean of ROI covariance, SD of ROI covariance, mean of ROI dissimilarity, SD of ROI dissimilarity, SD of tumor area