IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 2, 2020, accepted July 14, 2020, date of publication July 30, 2020, date of current version August 13, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3012967*

# Computer-Aided Diagnosis and Staging of Pancreatic Cancer Based on CT Images

**MIN LI[1,2,3], XIAOHAN NIE[4], YILIDAN REHEMAN[4], PAN HUANG[5], SHUAILEI ZHANG[5], YUSHUAI YUAN[5], CHEN CHEN[5], ZIWEI YAN[5], CHENG CHEN[5], XIAOYI LV[1,2,3], AND WEI HAN[4]**

[1]College of Software, Xinjiang University, Urumqi 830046, China
[2]Key Laboratory of Software Engineering Technology, Xinjiang University, Urumqi 830046, China
[3]Key Laboratory of Signal Detection and Processing, Xinjiang University, Urumqi 830046, China
[4]The First Affiliated Hospital of Xinjiang Medical University, Urumqi 830000, China
[5]College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China

Corresponding authors: Xiaoyi Lv (xjuwawj01@163.com) and Wei Han (13999846637@139.com)

This work was supported by the Science and Technology Project on aid to Xinjiang Uygur Autonomous Region under Grant 2019E0215.

**ABSTRACT** Pancreatic cancer (PC) is a malignant tumor that seriously threatens the survival of patients. Artificial classification has practical difficulties, such as unstable classification accuracy, a heavy workload, and the classification results depend on the subjective judgment of the clinician during the diagnosis and staging of PC. In addition, accurate PC staging could better help clinicians deliver the optimal therapeutic schedule for PC patients of different stages. Therefore, this study proposes a comprehensive medical computer-aided method for preoperative diagnosis and staging of PC based on an ensemble learning-support vector machine (EL-SVM) and computed tomography (CT) images. The least absolute shrinkage and selection operator (LASSO) algorithm was chosen for feature selection. In contrast to no feature selection, the model optimization time decreased by 19.94 seconds while maintaining precision. The EL-SVM learner was used to classify 168 CT images of normal pancreas and different stages of PC. The experimental results demonstrated that the normal pancreas (normal)-pancreatic cancer early stage (early stage) classification accuracy was 86.61%, the normal-pancreatic cancer stage III (stage III) classification accuracy was 87.04%, the normal-pancreatic cancer stage IV (stage IV) classification accuracy was 91.63%, the normal-PC classification accuracy was 87.89%, the early stage-stage III classification accuracy was 75.03%, and the early stage-stage IV classification accuracy was 81.22%, and the stage III-stage IV classification accuracy was 82.48%. Our experimental results prove that our proposed method is feasible and promising for clinical applications for the preoperative diagnosis and staging of PC via CT images.

**INDEX TERMS** Pancreatic cancer, diagnosis and staging, EL-SVM, CT, LASSO.

## I. INTRODUCTION

Pancreatic cancer (PC) is a highly malignant tumor of the digestive tract that presents considerable challenges in both the early screening stage and later treatment [1]–[3]. According to statistics from the American Cancer Society, the death rates of patients with PC continue to increase, ranking fourth in the United States [4]–[7]. It is estimated that approximately 57,600 people will be diagnosed with PC, and approximately 47,050 people will die of PC in 2020, therefore PC is known as an incurable disease [7], [8]. In developing countries, PC is still widely distributed. Therefore, comprehensive preoperative diagnosis and staging of PC are particularly important, especially in the detection of PC staging, which could better help the clinicians to deliver the optimal therapeutic schedule for different stages of PC and allow the patients to receive early medical interventions before advanced PC are formed [9], [10].

At present, the imaging modalities commonly used in the diagnosis of PC are as follows: (1) Computed tomography (CT) is the common imaging for many patients due to low cost and high penetration. Several studies have shown that CT plays a significant role in the preoperative diagnosis of PC and has become the preferred approach for many

The associate editor coordinating the review of this manuscript and approving it for publication was Haiyong Zheng.

patients with PC screening [1], [5], [11]–[13]. (2) Endoscopic ultrasonography (EUS), including contrast-enhanced harmonic EUS (CH-EUS) and EUS-guided fine needle aspiration (EUS-FNA). When CH-EUS is used in combination with EUS-FNA, it can effectively enhance the sensitivity of EUS-FNA [14]. EUS has high sensitivity and spatial resolution when detecting small-sized pancreatic tumors (<= 1 cm), and it is easy to obtain tissue sections [14]–[16]. However, the accuracy of the preoperative staging of PC and diagnosis between PC and some nonneoplastic lesions are both controversial [1], [12], [16]–[19]. (3) Magnetic resonance imaging (MRI) /magnetic resonance cholangiopancreatography (MRCP). MRI is highly sensitive for the detection of PC liver metastases. However, it needs high cost, low spatial resolution and high expertise. Therefore, MRI has not developed into a widely used imaging modality compared with CT [4], [12], [16], [20], [21]. MRCP can display the pancreatic duct system with a noninvasive delineating method for expert observation, but its diagnostic accuracy is controversial [15], [22]. (4) Endoscopic retrograde cholangiopancreatography (ERCP) is at risk in the diagnosis of PC because of low diagnostic sensitivity [4], [12], [15], [23], [24]. (5) Positron emission tomography (PET) is often used to identify distant metastases and evaluate the effect of treatment, and its anatomical coverage is quite wide. However, its disadvantage is low spatial resolution [1], [22], [24]. PET is currently not recommended, because it cannot notice the information of PC staging [16]. (6) Laparoscopy (LAP) is mainly used to confirm the inspection results [21].

Many research institutions have carried out a series of studies on the diagnosis of PC based on CT images. Qiu *et al.* [25] proposed a method based on CT images to assist in the diagnosis of PC. The method of combining wavelet transform and a statistical method was used to extract texture features. After using the interactive information method for feature selection, the supported vector machine (SVM) and probabilistic neural network (PNN) classifiers were selected for classification. Through experimental verification, the accuracy of the two classifiers in classifying PC and normal pancreas reached 95.48% and 96.13%, respectively. Although this method can better help clinicians diagnose PC, it only classifies the normal pancreas and PC. Li *et al.* [26] performed a hybrid feedback-support vector machine-random forest (HFB-SVM-RF) computer-aided diagnosis model for PC. The test results of 80 cases of PET/CT images showed that the average accuracy of the model was 96.47%, and the sensitivity and specificity were both above 95%. Although the model has high classification accuracy, the problem of incomplete experimental objects still exists. Liu *et al.* [27] proposed the faster region-based convolution network model (Faster R-CNN) artificial intelligence system for PC diagnosis. CT images of 100 PC patients were used for verification, the area under the receiver operating characteristic curve (AUC) value of the model reached 0.9632, and the model could automatically classify a CT image in only 0.2 seconds. Although the subjects of this study included different stages of PC, there is a lack of further study on the normal pancreas and different stages of PC.

Some research institutions have also paid attention to the ensemble learning of SVM. Shi *et al.* [28] proposed a computer aided pulmonary nodule detection system for chest radiography. The multiple massive training supported vector machine (MTSVM) model was proposed in this system, which consists of four non-linear SVMs arranged in parallel and one neural network arranged in serial. Each SVM classifier is an expert with excellent classification performance. However, the bagging algorithm was the ensemble learning algorithm used in this paper. Our paper used a bootstrap sampling method to generate different data sets; Then, we generated T base learners for combination, which ensures that each individual basic learner is ''good and different''. In order to improve the overall classification accuracy, the absolute majority voting method was used to explore the classification problem in this paper.

Aiming at the problems in the current research on the diagnosis of PC based on CT images, this study proposed a method of computer-aided diagnosis and staging of PC based on the ensemble learning-support vector machine (EL-SVM). Firstly, 168 nonenhanced CT images of different stages of PC and normal pancreas were chosen for pancreas segmentation. After extracting 202-dimensional features using six texture analysis methods, feature selection was performed based on the least absolute shrinkage and selection operator (LASSO). Finally, the EL-SVM learner was used to classify CT images of normal pancreas and different stages of PC. By comparing with SVM, the least absolute shrinkage and selection operator-support vector machine (LASSO-SVM), the k-nearest neighbor (KNN), the back propagation (BP) neural network classifiers, Softmax classifier, VGG16 model and DenseNet121 model [29], EL-SVM model achieved better performance for normal pancreas (normal)-pancreatic cancer early stage (early stage), normal-pancreatic cancer stage III (stage III), normal-pancreatic cancer stage IV (stage IV), normal-PC, early stage-stage III, early stage-stage IV, and stage III-stage IV classification. The main contributions of this paper can be listed as follows:

- This study was the first to comprehensively explore early screening and staging of PC by machine learning algorithms, and it shows the great potential of machine learning algorithms to be used in the diagnosis of PC at different stages.
- This study proposed a PC nonenhanced CT image classification model based on the EL-SVM.
- This study has a certain guiding role for junior clinicians in the diagnosis and staging of PC based on CT images, it can be used as a reference for diagnosis and can reduce the workload of clinicians.

## II. MATERIALS AND METHODS

Figure 1 shows the overall flowchart for assisted diagnosis and staging of the PC method, which consists of following parts that are processed consecutively: 1) Pancreas region of interest (ROI) segmentation; 2) Feature selection and fusion;

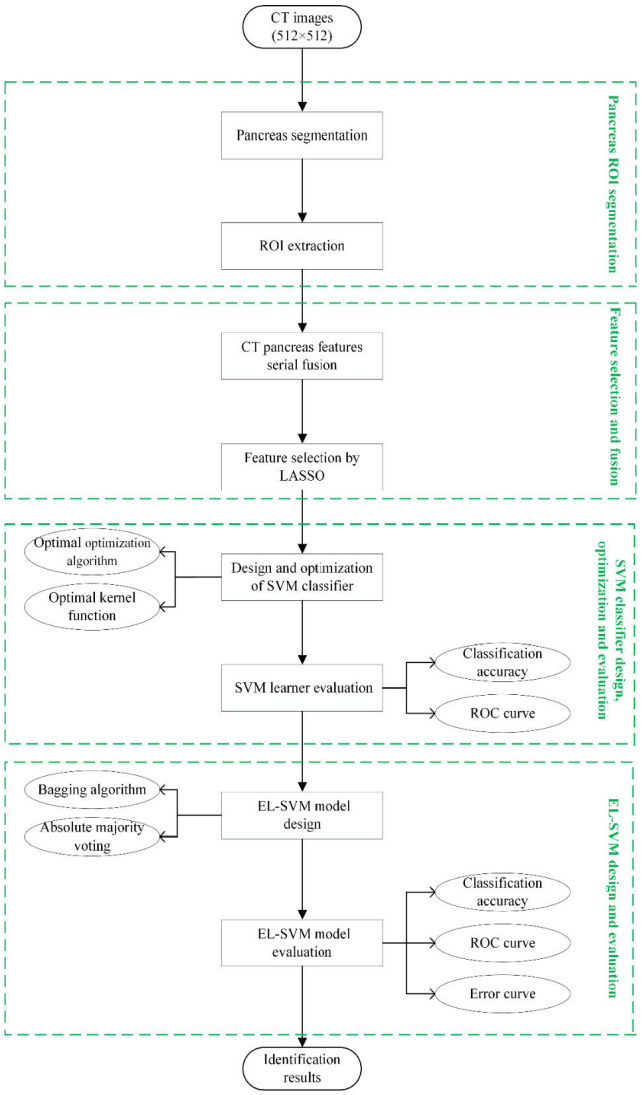



**FIGURE 1.** **Overall flowchart for assisted diagnosis and staging of the PC method. The green boxes in the figure represent the four main parts in this paper.**

3) SVM classifier design, optimization and evaluation; 4) EL-SVM design and evaluation. In the following sections, we will expand on the details.

### *A. DATA COLLECTION*

This paper selected 54 patients from the First Affiliated Hospital of Xinjiang Medical University, from November 2017 to August 2019, of which 39 cases had PC and 15 cases had normal pancreas. Patients with PC were retrieved from the hospital pathology database, and the stage of PC was referred to the American Joint Committee on Cancer (AJCC) standard [30]. And they were pathologically confirmed after pathology or surgery, as shown in Table 1. The 15 cases of normal pancreas (7 men [mean age, 57.57 ± 14.76 years; range, 36-80 years] and 8 women [mean age, 49 ± 17.20 years;

**TABLE 1.** **Patient demographics (n = 39).**

| Characteristic | Value |
|---|---|
| Sex | |
| Men | 24 |
| Women | 15 |
| Age | 59.72 ± 12.25 |
| PC stage | |
| I | 2(5.13) |
| Tumor location in the pancreas | |
| Head and neck | 1(2.56) |
| Uncinate process | 0 |
| Body | 0 |
| Tail | 1(2.56) |
| II | 13(33.33) |
| Tumor location in the pancreas | |
| Head and neck | 10(25.64) |
| Uncinate process | 2(5.13) |
| Body | 0 |
| Tail | 1(2.56) |
| III | 9(23.08) |
| Tumor location in the pancreas | |
| Head and neck | 7(17.95) |
| Uncinate process | 0 |
| Body | 0 |
| Tail | 2(5.13) |
| IV | 15(38.46) |
| Tumor location in the pancreas | |
| Head and neck | 8(20.51) |
| Uncinate process | 1(2.56) |
| Body | 0 |
| Tail | 6(15.38) |
| Tumor maximum diameter (in centimeters) | |
| Mean ± SD | 3.45 ± 2.19 |
| Range | 1.50-12.00 |
| Surgery | |
| Pancreatoduodenectomy | 21(53.85) |
| Distal pancreatectomy | 5(12.82) |
| Biliary stenting | 2(5.13) |
| Palliative surgery | 2(5.13) |
| Needle biopsy of the pancreas | 2(5.13) |
| Exploratory laparotomy | 1(2.56) |
| Removal of intraabdominal foreign body | 1(2.56) |
| Cancer antigen 19-9 level | |
| <37 U/mL | 4(10.26) |
| 37–200 U/mL | 7(17.95) |
| >200 U/mL | 28(71.79) |

Note-Except for age and tumor maximum diameter (mean ± SD), values are the number of patients with the percentage in parentheses.

**TABLE 2. The obtained experimental images.**

| Type | | Number of images |
|---|---|---|
| Total | | 168 |
| Normal | | 51 |
| PC | Stage I | 5 |
| | Stage II | 34 |
| | Stage III | 26 |
| | Stage IV | 52 |

range, 26-77 years]) were determined by a doctor as having a normal pancreas through routine abdominal CT.

### B. PANCREAS SEGMENTATION

Accurate segmentation of the pancreas is still a complex and challenging task. Considering the abdominal features of the pancreas and the variation in the volume and shape of the pancreas presented by the interpatient [31], [32], this paper manually drew the ROI of the pancreas. First, the CT sequence images in DICOM format were converted into JPG format and then imported into LabelMe software where two clinicians (Wei Han has 21 years and Yilidan Reheman has 3 years of clinical experience in abdominal CT) manually drew ROIs along the edge of the pancreas [33]. To improve model generalization ability, we selected 3 cross-sectional images of plain CT scans with significant features from each CT sequence on average [27], [34], [35]. A total of 168 CT images (as shown in Table 2) were evaluated, and a manually drawn ROI flowchart is shown in Figure 2. Pancreatic cancer stage I (stage I) or pancreatic cancer stage II (stage II) can be treated radically by surgical resection [36]; therefore, in this paper, stage I and stage II are collectively referred to as early stage [16], [37]–[40].

### C. FEATURE EXTRACTION

We extracted 202-dimensional features using six feature analysis methods, including the shape, gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level gradient co-occurrence matrix (GLGCM), gray-level difference statistics (GLDS), and wavelet transform, as shown in Table 3.

### D. FEATURE SELECTION AND FUSION

The 202-dimensional features are fused together by serial fusion. To avoid the dimensional disaster and low generalization performance of the learner caused by too many features [41], we must use a suitable feature selection algorithm to remove redundant features on the basis of retaining key factors, reduce model complexity, and reduce the waste of time and resources [42]. We compared the three methods of no feature selection, the relevant features (Relief) and LASSO, and the classification results of the comparison are recorded in Table 4.

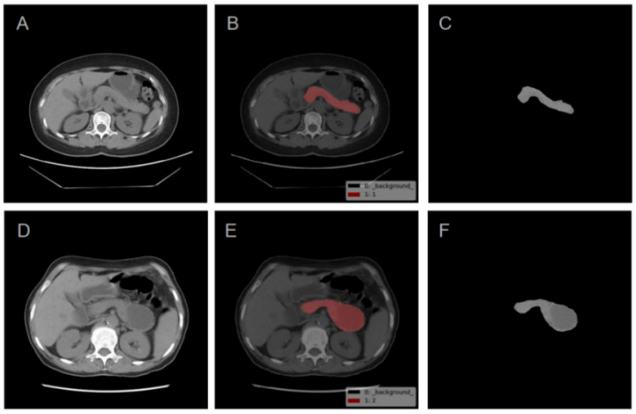



**FIGURE 2. The pancreatic ROI segmentation. A. Normal CT image. B. ROI drawn on a pancreas image with the normal (red area). C. Normal segmentation. D. Stage IV CT image. E. ROI drawn on a pancreas image with the stage IV (red area). F. Stage IV segmentation.**

**TABLE 3. Extracted features of the six methods.**

| Methods | Feature name | Total number |
|---|---|---|
| Shape | height, width, perimeter, area, complexity, rectangularity, elongation, equivalent area radius | 8 |
| GLCM | mean and standard deviation of energy, entropy, moment of inertia, and correlation | 8 |
| GLRLM | short run emphasis, long run emphasis, gray-level nonuniformity, run percentage, run-length nonuniformity, low gray-level run emphasis, high-gray level run emphasis | 7 |
| GLGCM | small grads dominance, big grads dominance, gray asymmetry, grads asymmetry, energy, gray mean, grads mean, gray variance, grads variance, correlation, gray entropy, grads entropy, entropy, inertia, differ moment | 15 |
| GLDS | mean, contrast, angular second moment, entropy | 4 |
| Wavelet transform | | 160 |
| Total | | 202 |

As shown in Table 4, the LASSO selection algorithm not only reduced the dimensions but also retained the key features. Under the premise of ensuring classification accuracy, compared with no feature selection, the optimization time of LASSO algorithm decreased by 19.94 seconds and compared with the Relief feature selection, the optimization time decreased by 7.17 seconds, which verified the feasibility of the LASSO embedded feature selection algorithm selected in this paper. The LASSO algorithm combines L1 norm regularization and a linear regression model so that some unimportant variables in the model were directly compressed to 0 to streamline the model. The LASSO algorithm is shown in Equation (1). In addition, a 10-fold cross-validation method was used to verify the stability of the LASSO algorithm,

**TABLE 4. Compared feature selection algorithms before and after LASSO and RELIEF to select the optimal solution.**

| Linear kernel (Test/training) | Category | Normal-early stage | Normal-stage III | Normal-stage IV | Normal-PC | Early stage-stage III | Early stage-stage IV | Stage III-stage IV |
|---|---|---|---|---|---|---|---|---|
| No feature selection | Dimension | 202 | 202 | 202 | 202 | 202 | 202 | 202 |
| | Accuracy rate% | 73.08/96.88 | 69.57/100.00 | 73.33/98.63 | 72.55/100.00 | 42.86/88.64 | 60.71/93.65 | 68.00/96.23 |
| | Time/second | 3.06 | 2.12 | 3.77 | 8.71 | 1.58 | 2.94 | 2.15 |
| RELIEF feature selection | Dimension | 57 | 59 | 30 | 156 | 19 | 25 | 15 |
| | Accuracy rate% | 65.38/96.88 | 65.22/98.15 | 56.67/78.08 | 68.63/96.58 | 57.14/68.18 | 75/79.37 | 68.00/81.13 |
| | Time/second | 1.05 | 0.90 | 0.94 | 6.75 | 0.62 | 0.77 | 0.53 |
| LASSO feature selection | Dimension | 26 | 8 | 11 | 19 | 9 | 12 | 4 |
| | Accuracy rate% | 76.54/92.19 | 60.87/97.41 | 55.33/91.64 | 70.78/96.15 | 49.52/85.23 | 71.43/87.62 | 66.40/74.53 |
| | Time/second | 0.68 | 0.37 | 0.64 | 1.48 | 0.37 | 0.51 | 0.34 |

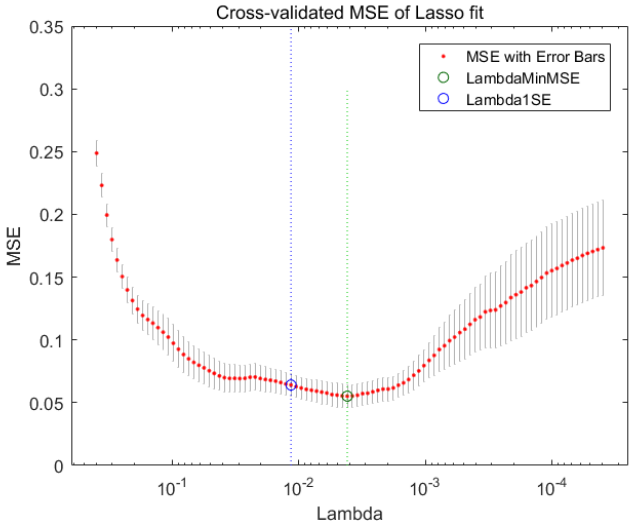



**FIGURE 3. Cross-validated MSE of LASSO fit (normal-early stage).**

as shown in Figure 3 [43].

$$min_{\mathbf{w}} \sum_{i=1}^{m} \left(y_i - w^T x_i\right)^2 + \lambda \| \mathbf{w} \|_1 \tag{1}$$

### E. SVM CLASSIFIER DESIGN AND IMPLEMENTATION

SVM was introduced by Vapnik [44], because it shows good performance on binary classification tasks, it has gradually developed and is now commonly used to solve multidomain classification matters [45]. The primary target of SVM is to find a "maximum margin" partitioning hyperplane suitable for classification samples so that the classification results are the most robust and have strong generalization capabilities [46]. Due to the existence of the kernel function, the samples of the original space reach the linearly separable high-dimensional feature space through mapping [47]. The kernel function and the feature space show a corresponding relationship, so the choice of the kernel function becomes the largest variable that determines the SVM classification performance. In addition, the selection of optimization algorithms is also quite important. We used multiple experiments and comparisons to attempt to find optimal optimization algorithms and kernel functions so that the SVM achieves the best performance and the largest increase in classification accuracy [46], [48].

### F. SVM LEARNER EVALUATION

To better evaluate the reliability of the SVM learner, this paper drew the receiver operating characteristic (ROC) curve to judge the performance of the learner and used the AUC to quantitatively verify the learner generalization ability [49].

According to its real category and the predictor category of the learner, combining the sample divided into the following four cases: true positive (TP), false positive (FP), true negative (TN), false negative (FN), which comprise the confusion matrix. Finally, the horizontal and vertical axes of the ROC curve were calculated, where the horizontal axis represents the false positive rate (FPR) and the vertical axis represents the true positive rate (TPR), defined as Equations (2) and (3), respectively.

$$FPR = \frac{FP}{TN + FP} \tag{2}$$

$$TPR = \frac{TP}{TP + FN} \tag{3}$$

### G. EL-SVM MODEL DESIGN

From Table 4, we learned that when we used the LASSO-SVM, although we demonstrated good performance in reducing time cost and dimensions, the classification accuracy was still low, of which normal-stage IV, and early stage-stage III classification accuracy were low. It can be

**TABLE 5.** Comparing three optimization algorithms (OA) and four kernel functions (KF) to choose the optimal Kernel function and optimization algorithm.

| (Test/ training(%)/time) | KF / OA | Linear kernel | Polynomial kernel | Gaussian kernel | Sigmoid kernel |
|---|---|---|---|---|---|
| Normal-early stage | GS-SVM | 76.54/92.19/0.68 | 59.23/99.06/0.41 | 77.69/96.25/0.58 | 53.85/57.81/0.60 |
| | PSO-SVM | 65.77/97.81/0.33 | 63.85/92.66/0.40 | 74.23/96.09/0.41 | 64.62/75.47/0.39 |
| | GA-SVM | 70.77/95.78/2.41 | 60.38/100.00/2.07 | 71.54/100.00/2.12 | 53.85/57.81/2.14 |
| Normal-stage III | GS-SVM | 60.87/97.41/0.37 | 65.22/100.00/0.34 | 60.87/100.00/0.32 | 60.87/68.52/0.56 |
| | PSO-SVM | 64.78/100.00/0.30 | 60.87/68.52/0.30 | 62.17/100.00/0.28 | 63.48/86.30/0.28 |
| | GA-SVM | 63.91/99.81/1.09 | 65.22/100.00/1.16 | 63.48/100.00/1.09 | 60.87/68.52/1.24 |
| Normal-stag IV | GS-SVM | 55.33/91.64/0.64 | 42.33/99.86/0.58 | 55.67/100.00/0.59 | 46.67/50.68/0.57 |
| | PSO-SVM | 58.33/91.23/0.41 | 44.00/87.80/0.42 | 59.33/89.59/0.44 | 53.67/77.40/0.38 |
| | GA-SVM | 54.67/91.64/1.94 | 44.33/100.00/1.84 | 51.33/99.04/1.92 | 46.67/50.68/1.86 |
| Normal-PC | GS-SVM | 70.78/96.15/1.48 | 71.76/99.83/2.99 | 66.86/98.46/3.06 | 72.55/68.38/2.26 |
| | PSO-SVM | 72.16/97.44/1.89 | 72.94/97.44/0.93 | 73.92/88.29/0.83 | 76.47/89.15/1.72 |
| | GA-SVM | 69.22/96.75/4.07 | 66.86/100.0/11.52 | 66.67/98.63/9.27 | 72.55/68.38/4.26 |
| Early stage-stage III | GS-SVM | 49.52/85.23/0.37 | 45.24/86.14/0.34 | 52.38/96.36/0.33 | 57.14/61.36/0.34 |
| | PSO-SVM | 48.10/93.96/0.29 | 55.24/66.14/0.31 | 53.33/86.14/0.29 | 53.33/71.82/0.29 |
| | GA-SVM | 51.90/77.05/2.31 | 43.33/91.36/2.25 | 48.10/82.95/1.21 | 57.14/61.36/2.36 |
| Early stage-stage IV | GS-SVM | 71.43/87.62/0.51 | 65.71/95.24/0.50 | 74.64/89.68/0.49 | 57.14/57.14/0.50 |
| | PSO-SVM | 68.93/89.21/0.37 | 57.14/57.14/0.39 | 76.43/85.40/0.41 | 72.86/76.51/0.38 |
| | GA-SVM | 73.93/87.46/1.89 | 67.14/93.81/1.59 | 72.86/91.11/1.95 | 59.64/61.43/1.81 |
| Stage III-stage IV | GS-SVM | 66.40/74.53/0.34 | 70.80/92.08/.0.34 | 67.60/89.25/0.36 | 64.00/67.92/0.34 |
| | PSO-SVM | 66.80/79.06/0.34 | 68.40/93.02/0.46 | 63.20/86.42/0.32 | 64.00/67.92/0.34 |
| | GA-SVM | 67.20/83.96/1.94 | 64.40/93.58/1.98 | 68.00/91.13/2.94 | 64.00/67.92/2.51 |

**TABLE 6.** The experimental parameter of GS-SVM, PSO-SVM and GA-SVM.

| Classifier | Parameter | Parameter value | Parameter | Parameter value |
|---|---|---|---|---|
| GS-SVM | Range of C | $[2^{-4}, 2^{4}]$ | Range of g | $[2^{-4}, 2^{4}]$ |
| | Step length of C | 0.3 | Step length of g | 0.3 |
| PSO-SVM | Range of C | $[10^{-1}, 10^{2}]$ | Range of g | $[10^{-2}, 10^{3}]$ |
| | Local search capability | 1.5 | Global search capability | 1.7 |
| | Maximum number of generations | 50 | Maximum number of populations | 5 |
| GA-SVM | Range of C | $[0,10^{2}]$ | Range of g | $[0,10^{3}]$ |
| | Maximum number of generations | 200 | Maximum number of populations | 20 |
| | Generation gap | 0.9 | | |

known from theory that good generalization performance can be obtained by using weak learner integration, so this paper proposed an EL-SVM to improve the classification accuracy and obtain a model with good generalization performance.

Ensemble learning improves performance by constructing and combining multiple base learners. The key is that the integration individual needs to have certain accuracy and diversity. The bagging algorithm was the ensemble learning algorithm used in this paper [50], [51]. The bagging algorithm is a significant representative of the parallel ensemble learning method; it uses a bootstrap sampling method to generate different initial data sets. T-times bootstrap sampling is performed, and m samples are returned each time; approximately 63.2% of the samples appeared in the training set, and the remaining 36.8% of the uncollected samples were used as the test set to perform an ''out-of-bag estimate'' of the generalization performance, which was determined from Equation (4). Finally, we generated T differential base learners for combination. The basic flow of the EL-SVM

**TABLE 7.** Comparison of the results of the EL-SVM model and classifier.

| Test/training (%) | Normal-early stage | Normal-stage III | Normal-stage IV | Normal-PC | Early stage-stage III | Early stage-stage IV | Stage III-stage IV |
|---|---|---|---|---|---|---|---|
| BP | 59.23/95.78 | 62.18/90.37 | 77.14/97.26 | 72.75/88.80 | 50.95/87.50 | 57.50/96.51 | 57.61/90.76 |
| KNN | 80.77/91.57 | 72.18/94.80 | 81.00/81.10 | 83.14/86.25 | 46.67/70.47 | 65.00/74.77 | 78.00/87.34 |
| SVM | 73.08/96.88 | 69.57/100.00 | 73.33/98.63 | 72.55/100.00 | 42.86/88.64 | 60.71/93.65 | 68.00/96.23 |
| LASSO-SVM | 76.54/92.19 | 60.87/97.41 | 55.33/91.64 | 70.78/96.15 | 49.52/85.23 | 71.43/87.62 | 66.40/74.53 |
| EL-SVM | 86.61/92.78 | 87.04/94.67 | 91.63/100.0 | 87.89/99.28 | 75.03/95.69 | 81.22/96.70 | 82.48/99.23 |

**TABLE 8.** The experimental parameter of BP and KNN.

| Classifier | Parameter | Parameter value | Parameter | Parameter value |
|---|---|---|---|---|
| BP | Maximal iteration times | 2000 | Target error | 0.00001 |
| | Display interval | 500 | Learning rate | 0.05 |
| KNN | Number of neighbors | 5 | Distance metric | Cosine |
| | Distance weight | Equal | Standardize data | True |

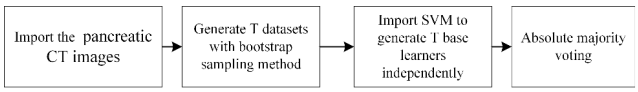



**FIGURE 4.** EL-SVM algorithm flow.

algorithm is shown in Figure 4.

$$lim_{m\to\infty}(1-\frac{1}{m})^m=\frac{1}{e}\approx 0.368 \tag{4}$$

where m is the number of samples of each base learner.

To complete the classification task of EL-SVM, this paper used the absolute majority voting method to select the class with more than half of the predicted votes, otherwise rejecting predictions to improve overall classification accuracy.

### H. EL-SVM MODEL EVALUATION

The classification error curve was used to further verify the reliability of the EL-SVM classifier in this paper. The number of base learners was used on the horizontal axis, and the classification error that still existed after ensemble learning was used on the vertical axis.

### I. HORIZONTAL COMPARISON WITH OTHER MODELS

We compared SVM, LASSO-SVM, KNN, BP neural network classifier, Softmax classifier, VGG16 model, DenseNet121 model and EL-SVM model [52]. We used the no data augmentation, data augmentation 3 times, data augmentation 6 times, data augmentation 12 times, and data augmentation 24 times to expand the training set, and sequentially import two deep learning models for experimentation. The rest of the models use the no data augmentation data set for classification.

## III. RESULTS

The experiments in this paper were based on the MATLAB 2016a platform and PyCharm platform. The LIBSVM toolbox was used for the discrimination and analysis of the SVM algorithm [53].

### A. SVM MODEL PARAMETER OPTIMIZATION

Three optimization algorithms were used to build a diagnostic model, including grid search-support vector machine (GS-SVM), particle swarm optimization-support vector machine (PSO-SVM), and genetic algorithms based on support vector machine (GA-SVM), to choose the optimal optimization algorithm and the optimal kernel function, as shown in Table 5 [54]. The primary experimental parameters of GS-SVM, PSO-SVM and GA-SVM are listed in Table 6. To avoid the risk of overfitting, we adopted 5-fold cross-validation methods.

It can be seen from Table 5 that under the premise of using the grid optimization algorithm, the linear kernel and the Gaussian kernel have little difference in accuracy and optimization time. However, the linear kernel function had the largest growth rate in the later integrated learning classifier and the highest optimization improvement. Therefore, this paper chose a grid search optimization algorithm and linear kernel function to classify pancreatic CT images.

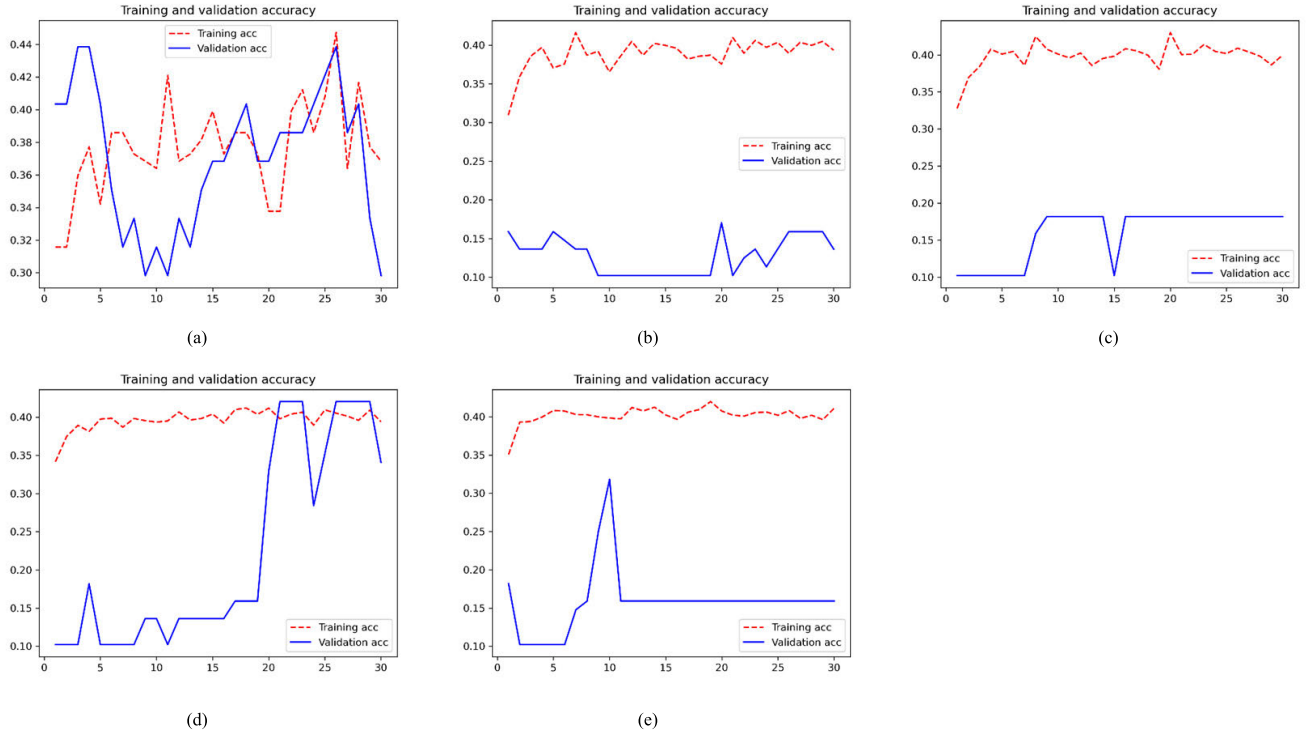
### B. EVALUATION OF EL-SVM

To reduce the experimental error and ensure the validity of the experimental results, the results in Table 7 all take the average of 10 experiments in each category [55].

**TABLE 9.** The experimental parameter of Softmax classifier.

| Classifier | Parameter | Parameter value | Parameter | Parameter value |
|---|---|---|---|---|
| Softmax | Simulation environment | 64-bit Windows 10 operating system | Software | PyCharm |
| | Learning rate | 0.001 | Epoch value | 100 |
| | Loss type | Categorical Cross-entropy | | |

**TABLE 10.** Visual picture of the VGG16 model accuracy.




Note-The visual picture of the VGG16 model accuracy. (a) no data augmentation. (b) data augmentation 3 times. (c) data augmentation 6 times. (d) data augmentation 12 times. (e) data augmentation 24 times.

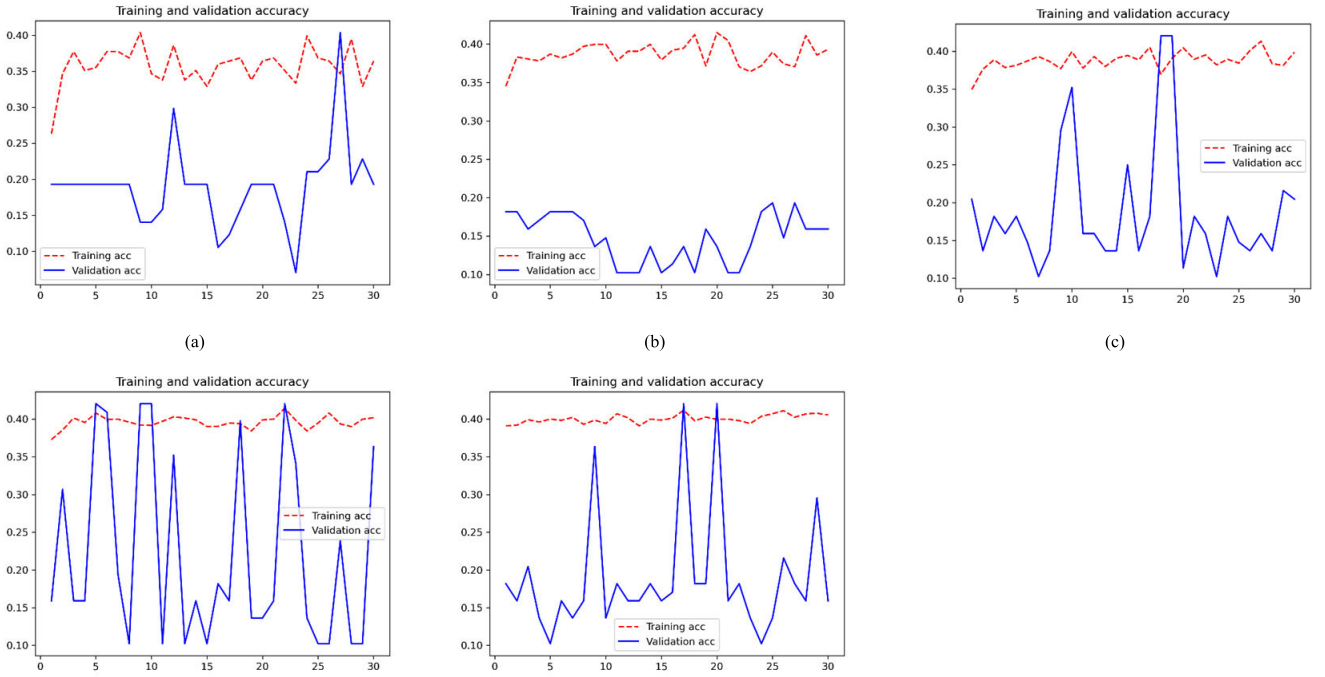
### 1) COMPARISON OF THE EL-SVM MODEL AND OTHER MODELS

The mainstream classification algorithms the KNN and the BP neural network were compared horizontally with the EL-SVM model in this paper, and the comparison results are shown in Table 7 [56]. The primary experimental parameters of KNN and BP are listed in Table 8. Table 7 shows that after using the ensemble learning algorithm, the classification accuracy of normal pancreas and PC in different stages greatly improved, and the generalization ability was enhanced. The classification accuracy rate of normal-stage IV reached 91.63%, and the classification accuracy of EL-SVM in normal-stage IV was improved by 36.30%, compared with LASSO-SVM. The average classification accuracy of BP was lower than EL-SVM 22.08%, and KNN was lower than EL-SVM 12.17%, which verified the rationality of selecting the EL-SVM classifier in this paper.

This paper also used Softmax classifier for comparison [57]. The 202-dimensional features extracted by the above six texture analysis methods are passed to the Softmax classifier to classify normal pancreas and different stages of PC. Experimental results showed that the average classification accuracy of the Softmax classifier was lower than EL-SVM 14.45%. As shown in Figure 5, the overall performance of EL-SVM is better than Softmax classifier. The primary experimental parameters of Softmax classifier are listed in Table 9.

This work attempted to train two different network structures (VGG16, DenseNet121) and built two models [58]–[61]. The visual images of VGG16 model and DenseNet121 model are placed in Table 10 and Table 11, respectively. It can be seen from the figure that the effects of the two models are not ideal, which also verifies that our proposed EL-SVM is our appropriate choice.

**TABLE 11.** Visual picture of the DenseNet121 model accuracy.




Note-The visual picture of the DenseNet121 model accuracy. (a) no data augmentation. (b) data augmentation 3 times. (c) data augmentation 6 times. (d) data augmentation 12 times. (e) data augmentation 24 times.

**TABLE 12.** The primary experimental parameter of VGG16 and DenseNet121.

| Architecture model | Parameter | Parameter value | Parameter | Parameter value |
|---|---|---|---|---|
| VGG16 | Simulation environment | 64-bit Windows 10 operating system | Software | PyCharm |
| | Image dimensions | 224×224 | Loss function | Categorical Cross-entropy |
| | Batch size | 32 | Epoch | 30 |
| | Learning rate | 0.001 | Optimization algorithm | Stochastic gradient descent |
| DenseNet121 | Image dimensions | 224×224 | Loss function | Categorical Cross-entropy |
| | Batch size | 32 | Epoch | 30 |
| | Learning rate | 0.001 | Optimization algorithm | Stochastic gradient descent |

**TABLE 13.** The AUC of five classifiers.

| AUC | Normal-early stage | Normal-stage III | Normal-stage IV | Normal-PC | Early stage-stage III | Early stage-stage IV | Stage III -stage IV |
|---|---|---|---|---|---|---|---|
| BP | 0.6310 | 0.6151 | 0.7098 | 0.7046 | 0.5000 | 0.5833 | 0.6042 |
| KNN | 0.8095 | 0.7262 | 0.7991 | 0.8475 | 0.4722 | 0.6250 | 0.7222 |
| SVM | 0.7262 | 0.7381 | 0.8393 | 0.7471 | 0.3796 | 0.6302 | 0.6181 |
| LASSO-SVM | 0.7440 | 0.6905 | 0.5848 | 0.7548 | 0.4259 | 0.7656 | 0.7014 |
| EL-SVM | 0.8750 | 0.9444 | 0.9583 | 0.9500 | 0.7500 | 0.7917 | 0.8750 |

**TABLE 14.** The classification error convergence value.

| Classification | Normal-early stage | Normal-stage III | Normal-stage IV | Normal-PC | Early stage-stage III | Early stage-stage IV | Stage III-stage IV |
|---|---|---|---|---|---|---|---|
| Convergence value | 0.1333 | 0.1418 | 0.0971 | 0.1307 | 0.2991 | 0.2053 | 0.1954 |

Note-The convergence value is the classification error convergence value.

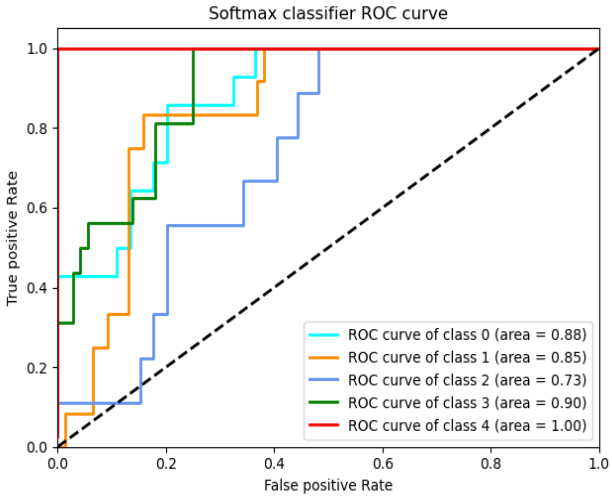



**FIGURE 5.** Softmax classifier ROC curve. Class 0 (normal), class 1(early stage), class 2 (stage III), class 3(stage IV), class 4(PC).

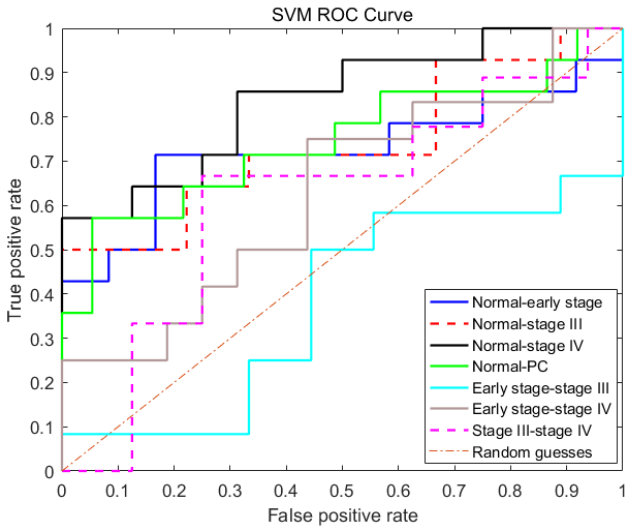



**FIGURE 6.** SVM classifier ROC curve.

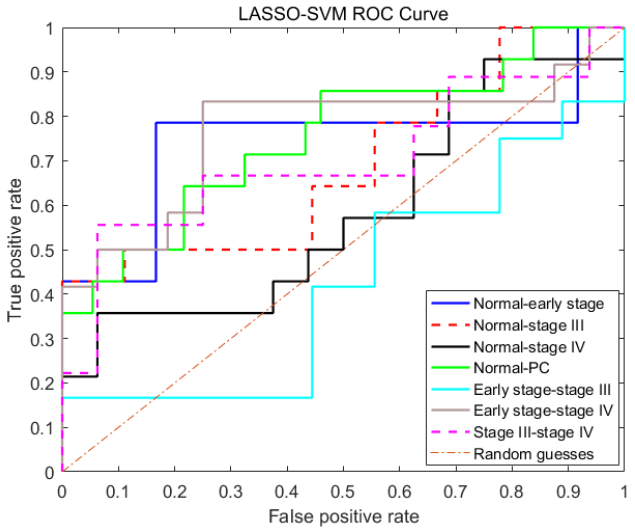



**FIGURE 7.** LASSO-SVM ROC curve.

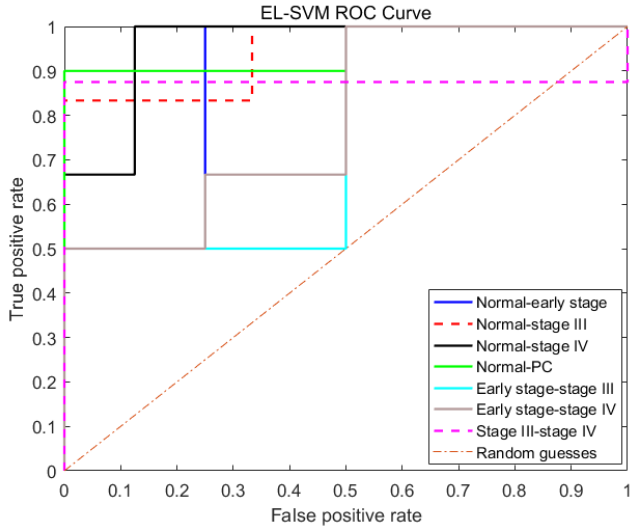



**FIGURE 8.** EL-SVM ROC curve.

The primary experimental parameters of VGG16 model and DenseNet121 model are listed in Table 12.

### 2) ROC CURVE AND AUC

We drew the ROC curves of SVM, LASSO-SVM, EL-SVM, BP, and KNN, as shown in Figures 6, 7, 8, 9 and 10, and the corresponding AUC values are given in Table 13.

It can be seen from Table 13, Figure 6, Figure 7, and Figure 8 that the AUC value of EL-SVM for normal-stage IV classification increased by 0.3735 compared to LASSO-SVM and the AUC value of EL-SVM for early stage-stage III classification increased by 0.3241 compared to LASSO-SVM. The average AUC of EL-SVM reached 0.8778, and the AUC of normal-stage IV reached 0.9583, which greatly improved

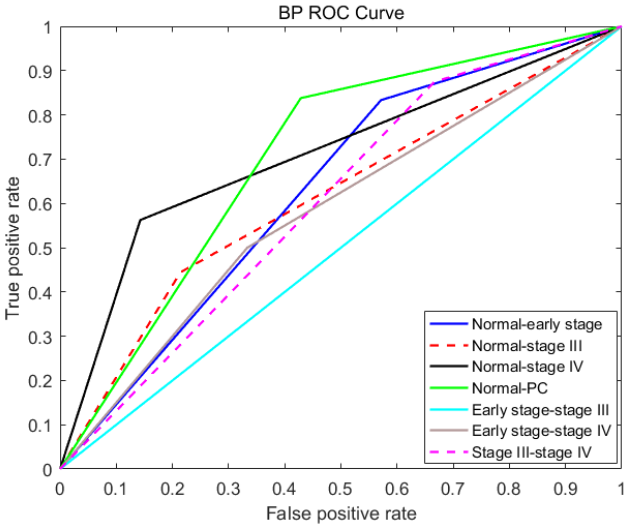



**FIGURE 9.** BP classifier ROC curve.

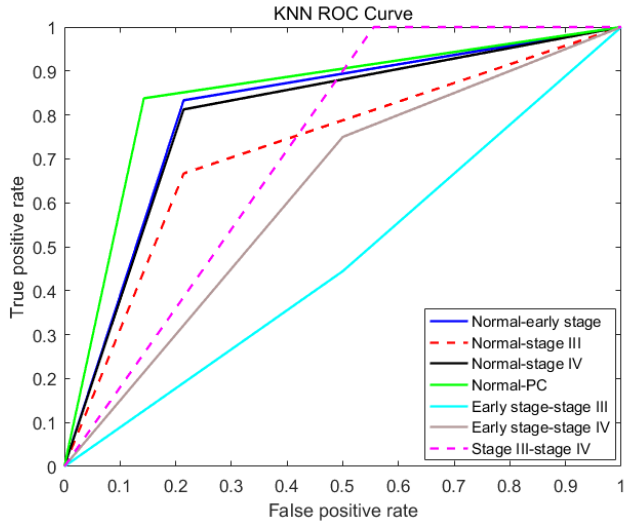



**FIGURE 10.** KNN classifier ROC curve.

compared with LASSO-SVM and SVM. The results show that EL-SVM has good feasibility.

From Table 13 and Figures 8, 9 and 10, it can be seen that the AUC values of BP and KNN were lower than those of EL-SVM. The average AUC of BP was lower than that of EL-SVM (0.2567) and that of KNN was lower than that of EL-SVM (0.1633). Therefore, the performance of the KNN and BP neural networks on pancreatic CT images was inferior to that of the EL-SVM model.

### 3) EL-SVM ERROR CURVE

This paper used 400 base learners for ensemble learning and then drew the classification error curve of EL-SVM as shown in Figure 11. The classification error convergence values are shown in Table 14, which shows that the classification error convergence value of normal-stage IV was the lowest and the classification performance was the best.

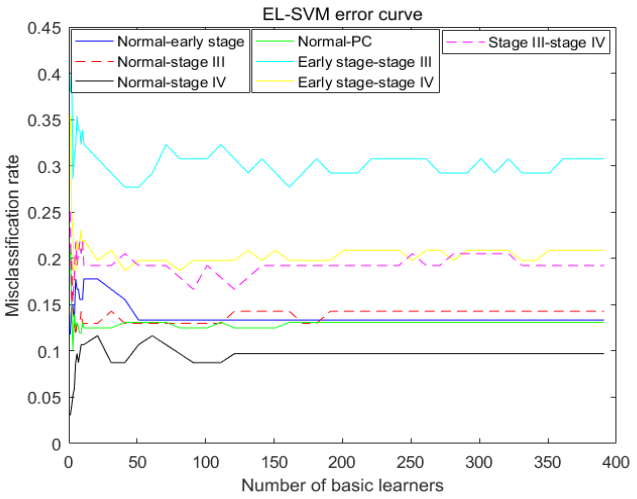



**FIGURE 11.** EL-SVM error curve.

## IV. CONCLUSION

This paper was the first to proposed a classification method of pancreatic CT images based on EL-SVM. We performed experiments on 168 CT images of normal pancreas and different stages of PC, and the results demonstrated that the EL-SVM method obtained the best classification performance. Our study could help solve the problems existing in preoperative PC diagnosis to a certain extent and deliver treatment options for different stages of PC patients, which have certain feasibility and practicability. In the future, we will continue to collect more samples from different central institutions, and establish a classification model with better performance; therefore, it has greater clinical significance for the preoperative diagnosis and staging of PC via CT images.

## ACKNOWLEDGMENT

*(Min Li and Xiaohan Nie are co-first authors.)*

## CONFLICTS OF INTEREST

The authors have no relevant financial interests in this article and no potential conflicts of interest to disclose.

## REFERENCES

[1] J. R. Treadwell, H. M. Zafar, M. D. Mitchell, K. Tipton, U. Teitelbaum, and J. Jue, "Imaging tests for the diagnosis and staging of pancreatic adenocarcinoma: A meta-analysis," *Pancreas*, vol. 45, no. 6, pp. 789–795, Jul. 2016.

[2] S. H. Wang, Y. F. Sun, Y. Liu, Y. Zhou, and Y. Liu, "CT contrast enhancement correlates with pathological grade and microvessel density of pancreatic cancer tissues," *Int. J. Clin. Experim. Pathol.*, vol. 8, no. 5, pp. 5443–5449, 2015.

[3] Z. Q. Wang, J. S. Li, G. M. Lu, X. H. Zhang, Z. Q. Chen, and K. Meng, "Correlation of CT enhancement, tumor angiogenesis and pathologic grading of pancreatic carcinoma," *World J. Gastroenterology*, vol. 9, no. 9, pp. 2100–2104, Sep. 2003.

[4] M. S. Bhutani, P. Koduru, V. Joshi, P. Saxena, R. Suzuki, A. Irisawa, and K. Yamao, "The role of endoscopic ultrasound in pancreatic cancer screening," *Endoscopic Ultrasound*, vol. 5, no. 1, pp. 8–16, Jan./Feb. 2016.

[5] A. Dallongeville, L. Corno, S. Silvera, I. Boulay-Coletta, and M. Zins, "Initial diagnosis and staging of pancreatic cancer including main differentials," *Seminars Ultrasound, CT MRI*, vol. 40, no. 6, pp. 436–468, Dec. 2019.

[6] S. Schneeweiß, M. Horger, A. Grözinger, K. Nikolaou, D. Ketelsen, R. Syha, and G. Grözinger, "CT-perfusion measurements in pancreatic carcinoma with different kinetic models: Is there a chance for tumour grading based on functional parameters?" *Cancer Imag.*, vol. 16, no. 1, pp. 1–8, Dec. 2016.

[7] R. L. Siegel, K. D. Miller, and A. G. Sauer, S. A. Fedewa, L. F. Butterly, J. C. Anderson, A. Cercek, R. A. Smith, and and A. Jemal, "Cancer statistics, 2020," *Ca-a Cancer J. Clinicians*, vol. 70, no. 1, pp. 7–30, Jan. 2020.

[8] K. Holub and C. Conill, "Unveiling the mechanisms of immune evasion in pancreatic cancer: May it be a systemic inflammation responsible for dismal survival?" *Clin. Transl. Oncol.*, vol. 22, no. 1, pp. 81–90, Jan. 2020.

[9] M. Ilic and I. Ilic, "Epidemiology of pancreatic cancer," *World J. Gastroenterology*, vol. 22, no. 44, pp. 9694–9705, Nov. 2016.

[10] Q. J. Lin, F. Yang, C. Jin, and D. L. Fu, "Current status and progress of pancreatic cancer in China," *World J. Gastroenterology*, vol. 21, no. 26, pp. 7988–8003, Jul. 2015.

[11] Y. N. Shen, X. L. Bai, G. G. Li, and T. B. Liang, "Review of radiological classifications of pancreatic cancer with peripancreatic vessel invasion: Are new grading criteria required?" *Cancer Imag.*, vol. 17, no. 1, p. 14, May 2017.

[12] A. Singh and A. L. Faulx, "Endoscopic evaluation in the workup of pancreatic cancer," *Surgical Clinics North Amer.*, vol. 96, no. 6, p. 1257, Dec. 2016.

[13] H. Taoka, E. Hauptmann, L. W. Traverso, M. J. Barnett, M. G. Sarr, and H. A. Reber, "How accurate is helical computed tomography for clinical staging of pancreatic cancer?" *Amer. J. Surg.*, vol. 177, no. 5, pp. 428–432, May 1999.

[14] M. Kitano, M. Kudo, K. Yamao, T. Takagi, H. Sakamoto, T. Komaki, K. Kamata, H. Imai, Y. Chiba, M. Okada, T. Murakami, and Y. Takeyama, "Characterization of small solid tumors in the pancreas: The value of contrast-enhanced harmonic endoscopic ultrasonography," *Amer. J. Gastroenterology*, vol. 107, no. 2, pp. 303–310, Feb. 2012.

[15] M. I. Canto, F. Harinck, R. H. Hruban, G. J. Offerhaus, J. W. Poley, I. Kamel, Y. Nio, R. S. Schulick, C. Bassi, I. Kluijt, and M. J. Levy, "International cancer of the pancreas screening (CAPS) consortium summit on the management of patients with increased risk for familial pancreatic cancer," *Gut*, vol. 62, no. 3, pp. 339–347, Dec. 2013.

[16] P. Moutinho-Ribeiro, J. Iglesias-Garcia, R. Gaspar, and G. Macedo, "Early pancreatic cancer—The role of endoscopic ultrasound with or without tissue acquisition in diagnosis and staging," *Digestive Liver Disease*, vol. 51, no. 1, pp. 4–9, Jan. 2019.

[17] L. Helmstaedter and J. F. Riemann, "Pancreatic cancer—EUS and early diagnosis," *Langenbecks Arch. Surg.*, vol. 393, no. 6, pp. 923–927, Nov. 2008.

[18] J. Iglesias-Garcia and J. Domínguez-Muñoz, "Early detection of pancreatic cancer in patients with intraductal papillary mucinous neoplasms: The pivotal role of endoscopic ultrasound," *Endoscopy*, vol. 46, no. 1, pp. 30–31, Dec. 2013.

[19] T. Yoshida, Y. Yamashita, and M. Kitano, "Endoscopic ultrasound for early diagnosis of pancreatic cancer," *Diagnostics*, vol. 9, no. 3, p. 81, Jul. 2019.

[20] M. Alabousi, M. D. McInnes, J. Salameh, J. Satkunasingham, Y. K. Kagoma, L. Ruo, B. M. Meyers, T. Aziz, and C. B. van der Pol, "MRI vs. CT for the detection of liver metastases in patients with pancreatic carcinoma: A comparative diagnostic test accuracy systematic review and meta-analysis," *J. Magn. Reson. Imag.*, to be published, doi: 10.1002/jmri.27056.

[21] I. Borbath, B. E. Van Beers, M. Lonneux, D. Schoonbroodt, A. Geubel, J.-F. Gigot, and P. H. Deprez, "Preoperative assessment of pancreatic tumors using magnetic resonance imaging, endoscopic ultrasonography, positron emission tomography and laparoscopy," *Pancreatology*, vol. 5, no. 6, pp. 553–561, Jan. 2005.

[22] E. S. Lee and J. M. Lee, "Imaging diagnosis of pancreatic cancer: A state-of-the-art review," *World J. Gastroenterology*, vol. 20, no. 24, pp. 7864–7877, Jun. 2014.

[23] S. P. Kauhanen, G. Komar, M. P. Seppänen, K. I. Dean, H. R. Minn, S. A. Kajander, I. Rinta-Kiikka, K. Alanen, R. J. Borra, P. A. Puolakkainen, P. Nuutila, and J. T. Ovaska, "A prospective diagnostic accuracy study of 18F-fluorodeoxyglucose positron emission tomography/Computed tomography, multidetector row computed tomography, and magnetic resonance imaging in primary diagnosis and staging of pancreatic cancer," *Ann. Surg.*, vol. 250, no. 6, pp. 957–963, Dec. 2009.

[24] H. Saisho and T. Yamaguchi, "Diagnostic imaging for pancreatic cancer: Computed tomography, magnetic resonance imaging, and positron emission tomography," *Pancreas*, vol. 28, no. 3, pp. 273–278, Apr. 2004.

[25] J.-J. Qiu, Y. Wu, B. Hui, Z.-X. Huang, and J. Chen, "Texture analysis of computed tomography images in the classification of pancreatic cancer and normal pancreas: A feasibility study," *J. Med. Imag. Health Informat.*, vol. 8, no. 8, pp. 1539–1545, Oct. 2018.

[26] S. Li, H. Jiang, Z. Wang, G. Zhang, and Y.-D. Yao, "An effective computer aided diagnosis model for pancreas cancer on PET/CT images," *Comput. Methods Programs Biomed.*, vol. 165, pp. 205–214, Oct. 2018.

[27] S.-L. Liu, S. Li, Y.-T. Guo, Y.-P. Zhou, Z.-D. Zhang, S. Li, and Y. Lu, "Establishment and application of an artificial intelligence diagnosis system for pancreatic cancer with a faster region-based convolutional neural network," *Chin. Med. J.*, vol. 132, no. 23, pp. 2795–2803, Dec. 2019.

[28] Z. Shi, M. Zhao, L. He, Y. Wang, M. Zhang, and K. Suzuki, "A computer aided pulmonary nodule detection system using multiple massive training SVMs," *Appl. Math. Inf. Sci.*, vol. 7, no. 3, pp. 1165–1172, May 2013.

[29] T. Fujioka, L. Katsuta, K. Kubota, M. Mori, Y. Kikuchi, A. Kato, G. Oda, T. Nakagawa, Y. Kitazume, and U. Tateishi, "Classification of breast masses on ultrasound shear wave elastography using convolutional neural networks," *Ultrason. Imag.*, Jun. 2020, doi: 10.1177/0161734620932.

[30] S. van Roessel *et al.*, "International validation of the eighth edition of the American joint committee on cancer (AJCC) TNM staging system in patients with resected pancreatic cancer," *Jama Surg.*, vol. 153, no. 12, Dec. 2018, Art. no. e183617.

[31] S. Liu, X. Yuan, R. Hu, S. Liang, S. Feng, Y. Ai, and Y. Zhang, "Automatic pancreas segmentation via coarse location and ensemble learning," *IEEE Access*, vol. 8, pp. 2906–2914, 2020.

[32] H. R. Roth, L. Lu, N. Lay, A. P. Harrison, A. Farag, A. Sohn, and R. M. Summers, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Med. Image Anal.*, vol. 45, pp. 94–107, Apr. 2018.

[33] C. Guo, X. Zhuge, Z. Wang, Q. Wang, K. Sun, Z. Feng, and X. Chen, "Textural analysis on contrast-enhanced CT in pancreatic neuroendocrine neoplasms: Association with WHO grade," *Abdominal Radiol.*, vol. 44, no. 2, pp. 576–585, Feb. 2019.

[34] S. Pandey, P. R. Singh, and J. Tian, "An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101782.

[35] X. Wang, K. Wang, and S. Lian, "A survey on face data augmentation for the training of deep neural networks," in *Neural Computing and Applications*. London, U.K.: Springer-Verlag, Mar. 2020, pp. 1–29, doi: 10.1007/s00521-020-04748-3.

[36] H. Zhu, T. Li, Y. Du, and M. Li, "Pancreatic cancer: Challenges and opportunities," *BMC Med.*, vol. 16, no. 1, Nov. 2018.

[37] X. Li, S. Li, L. Liu, J. Hong, T. Zhao, and C. Gao, "Effect of perioperative CEA and CA24-2 on prognosis of early resectable pancreatic ductal adenocarcinoma," *J. Cancer*, vol. 11, no. 1, pp. 9–15, 2020.

[38] H. Matsubayashi, H. Ishiwatari, K. Sasaki, K. Uesaka, and H. Ono, "Detecting early pancreatic cancer: Current problems and future prospects," *Gut Liver*, vol. 14, no. 1, pp. 30–36, Jan. 2020.

[39] Z.-Y. Wang, X.-Q. Ding, H. Zhu, R.-X. Wang, X.-R. Pan, and J.-H. Tong, "KRAS mutant allele fraction in circulating cell-free DNA correlates with clinical stage in pancreatic cancer patients," *Frontiers Oncol.*, vol. 9, p. 1295, Nov. 2019.

[40] S. Yamamoto, Y. Tomita, Y. Hoshida, H. Nagano, K. Dono, K. Umeshita, M. Sakon, O. Ishikawa, H. Ohigashi, S. Nakamori, M. Monden, and K. Aozasa, "Increased expression of valosin-containing protein (p97) is associated with lymph node metastasis and prognosis of pancreatic ductal adenocarcinoma," *Ann. Surgical Oncol.*, vol. 11, no. 2, pp. 165–172, Feb. 2004.

[41] P. Wang, Y. Li, B. Chen, X. Hu, J. Yan, Y. Xia, and J. Yang, "Proportional hybrid mechanism for population based feature selection algorithm," *Int. J. Inf. Technol. Decis. Making*, vol. 16, no. 05, pp. 1309–1338, Sep. 2017.

[42] S. Zhu, D. Wang, K. Yu, T. Li, and Y. Gong, "Feature selection for gene expression using model-based entropy," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 7, no. 1, pp. 25–36, Jan. 2010.

[43] P. Huang, S. Zhang, M. Li, J. Wang, C. Ma, B. Wang, and X. Lv, "Classification of cervical biopsy images based on LASSO and EL-SVM," *IEEE Access*, vol. 8, pp. 24219–24228, 2020.

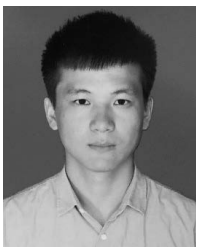
[44] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, Sep. 1999.

[45] Q. Zhang, H. Wang, and S. W. Yoon, "A 1-norm regularized linear programming nonparallel hyperplane support vector machine for binary classification problems," *Neurocomputing*, vol. 376, pp. 141–152, Feb. 2020.
[46] Y. Huo, L. Xin, C. Kang, M. Wang, Q. Ma, and B. Yu, "SGL-SVM: A novel method for tumor classification via support vector machine with sparse group lasso," *J. Theor. Biol.*, vol. 486, Feb. 2020, Art. no. 110098.
[47] R. Ullah, S. Khan, S. Javaid, H. Ali, M. Bilal, and M. Saleem, "Raman spectroscopy combined with a support vector machine for differentiating between feeding male and female infants mother's milk," *Biomed. Opt. Express*, vol. 9, no. 2, pp. 844–851, Feb. 2018.
[48] J. Liu, J. Wang, Z. Tang, B. Hu, F.-X. Wu, and Y. Pan, "Improving Alzheimer's disease classification by combining multiple measures," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 5, pp. 1649–1659, Sep. 2018.
[49] J. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, and J. Wang, "Classification of Alzheimer's disease using whole brain hierarchical network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 2, pp. 624–632, Mar. 2018.
[50] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization," *Mach. Learn.*, vol. 40, no. 2, pp. 139–157, 2000.
[51] S. E. Roshan and S. Asadi, "Improvement of bagging performance for classification of imbalanced datasets using evolutionary multi-objective optimization," *Eng. Appl. Artif. Intell.*, vol. 87, Jan. 2020, Art. no. 103319.
[52] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery," *Remote Sens.*, vol. 10, no. 7, p. 1119, Jul. 2018.
[53] J. Liu, M. Li, Y. Pan, F.-X. Wu, X. Chen, and J. Wang, "Classification of schizophrenia based on individual hierarchical brain networks constructed from structural MRI images," *IEEE Trans. Nanobiosci.*, vol. 16, no. 7, pp. 600–608, Oct. 2017.
[54] C. Chen, G. Du, D. Tong, G. Lv, X. Lv, R. Si, J. Tang, H. Li, H. Ma, and J. Mo, "Exploration research on the fusion of multimodal spectrum technology to improve performance of rapid diagnosis scheme for thyroid dysfunction," *J. Biophotonics*, vol. 13, no. 2, Feb. 2020, Art. no. e201900099.
[55] Y. Li and X. Zeng, "Sequential multi-criteria feature selection algorithm based on agent genetic algorithm," *Int. J. Speech Technol.*, vol. 33, no. 2, pp. 117–131, Oct. 2010.
[56] Y. Li, S. Zhang, and X. Zeng, "Research of multi-population agent genetic algorithm for feature selection," *Expert Syst. Appl.*, vol. 36, no. 9, pp. 11570–11581, Nov. 2009.
[57] M. Toğaçar, B. Ergen, and Z. Cömert, "Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks," *Biocybern. Biomed. Eng.*, vol. 40, no. 1, pp. 23–39, Jan. 2020.
[58] S. Hussein, P. Kandel, C. W. Bolan, M. B. Wallace, and U. Bagci, "Lung and pancreatic tumor characterization in the deep learning era: Novel supervised and unsupervised learning approaches," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1777–1787, Aug. 2019.
[59] Z. Shi, H. Hao, M. Zhao, Y. Feng, L. He, Y. Wang, and K. Suzuki, "A deep CNN based transfer learning method for false positive reduction," *Multimedia Tools Appl.*, vol. 78, no. 1, pp. 1017–1033, Jan. 2019.
[60] P. Wu, X. Sun, Z. Zhao, H. Wang, S. Pan, and B. Schuller, "Classification of lung nodules based on deep residual networks and migration learning," *Comput. Intell. Neurosci.*, vol. 2020, pp. 1–10, Mar. 2020.
[61] H. Zhu, H. Jiang, S. Li, H. Li, and Y. Pei, "A novel multispace image reconstruction method for pathological image classification based on structural information," *BioMed Res. Int.*, vol. 2019, pp. 1–9, Apr. 2019.

**MIN LI** received the bachelor's degree in computer science and technology from Shandong Women's University, China, in 2019. She is currently pursuing the master's degree in software engineering with Xinjiang University. Her current research interest includes medical image processing.

**XIAOHAN NIE** is currently pursuing the Master of Medicine with the First Affiliated Hospital of Xinjiang Medical University. He is also a Surgical Resident with the First Affiliated Hospital of Xinjiang Medical University. He makes a speciality of pancreatic surgery, having made a series of achievements. His major research interest includes pancreas-related diseases.

**YILIDAN REHEMAN** received the bachelor's degree in clinical medicine from Xinjiang Medical University. He is currently pursuing the master's degree in surgery with the First Affiliated Hospital of Xinjiang Medical University. His current research interest includes general surgery.



**PAN HUANG** studied at Xinjiang University, China, majoring in communication engineering. He will pursue the Ph.D. degree in information and communication engineering. He was admitted to Chongqing University, China, by recommendation, in September 2020. His current research interests are medical artificial intelligence, medical image processing and recognition, and biological signal processing.

**SHUAILEI ZHANG** received the bachelor's degree in communications engineering from Xinjiang University, China, in June 2019, where she is currently pursuing the master's degree in information and communications engineering. Her main research interests include image processing and bioinformatics.

**YUSHUAI YUAN** received the bachelor's degree from the Tianjin University of Technology, China, in June 2018. He is currently pursuing the master's degree with Xinjiang University. His main research interest includes image processing.

**CHEN CHEN** received the bachelor's degree from Xinjiang University, China, in June 2018, where he is currently pursuing the Ph.D. degree. His main research interest includes medical signal processing.

**ZIWEI YAN** received the bachelor's degree from Xinjiang University, China, in June 2018, where she is currently pursuing the master's degree. Her main research interest includes medical signal processing.

**CHENG CHEN** received the bachelor's degree from Xinjiang University, China, in June 2018, where he is currently pursuing the master's degree. His main research interest includes bioinformatics.

**XIAOYI LV** received the M.S. degree in information and communication engineering from Xinjiang University, China, in 2006, and the Ph.D. degree in electronic and information engineering from Xi'an Jiaotong University, China, in 2010. He is currently a Professor with the School of Software, Xinjiang University. His current research interests include bioinformatics and artificial intelligence in medical diagnosis.

**WEI HAN** is currently pursuing the Doctor of Medicine degree with the First Affiliated Hospital of Xinjiang Medical University. He is also a Chief Physician, an Associate Professor, a Master's Supervisor, and the Director of the Pancreatic Surgery Department, The First Affiliated Hospital of Xinjiang Medical University. He is also the Deputy Director of Pancreatic Surgery, and the Secretary of Pancreatology Group of General Surgery Committee of Xinjiang Medical Association. He is expert in the diagnosis and treatment of hepatobiliary and pancreatic diseases, especially the minimally invasive treatment of pancreatic diseases and biliary diseases. He has participated in four provincial and ministerial level projects. He has published more than 20 core journal articles. He is an Editorial Board Member of the *Journal of Laparoscopic Surgery*.