JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

IDENTIFICATION OF BREAST CANCER USING CNN

[1]Mrs.N.Fathima Shrene Shifna, [2]B.Hemanth, [3]B.Sai Mani Ram, [4]B.Chandra Mohan Reddy

[1]*Assistant professor, School of Computing, Department of Computer Science and Engineering, Bharath Institute of Higher Education And Research, Chennai, India- 600073.*

[2, 3,4,] *Student, School of Computing, Department of Computer Science and Engineering, Bharath Institute of Higher Education And Research, Chennai, India- 600073.*

ABSTRACT:

Breast cancer is a global health challenge that demands early detection to improve patient outcomes. In this regard, Convolutional Neural Networks (CNNs) are deep learning techniques that have in recent times exhibited value in diverse medical image analysis applications including breast cancer detection and diagnosis. This paper presents in-depth research on breast cancer identification using CNNs to improve the accuracy and efficiency of diagnosis. In this approach, a dataset was constructed from mammography images collected from different sources which consist of several kinds and stages of breast cancer. Pre-processing methods were used to improve image quality and reduce noise, to enhance the performance of the CNN model. Afterward, a custom CNN architecture designed specifically for identifying breast cancers is created which consists of multiple convolutional & pooling layers followed by fully connected layers for classification. For robust training and evaluation purposes, the dataset is divided into training, validation, and test sets. Extensive experiments were carried out to fine-tune hyperparameters and optimize the CNN architecture to achieve better performances in terms of accuracy, sensitivity, specificity, and computational efficiency

Keywords: Pancreatic cancer; Diagnosis and staging; EL-SVM; CT; LASSO;

INTRODUCTION:

Breast cancer is among the most common and deadliest forms of cancer that affect women globally, with around 2.3 million new cases diagnosed each year and over a hundred thousand deaths reported worldwide. Early detection combined with correct diagnosis is crucial for better management as well as treatment of breast cancer which greatly enhance patient outcomes and survival rates. In breast cancer screening, mammography plays a vital role in

enabling the finding of some abnormalities such as masses, microcalcification, and architectural distortion. Nevertheless, the interpretation of mammograms can be difficult and subjective sometimes necessitating specialized expertise. In recent times, advances in artificial intelligence (AI) and deep learning have fueled significant progress toward medical image analysis thus offering promising ways to improve accuracy as well as efficiency of breast cancer diagnosis. A type of deep learning model called Convolutional Neural Networks (CNNs) has shown a remarkable ability to extract intricate features from images with high classification accuracy. CNN's ability to learn hierarchical representations automatically from data has made it very useful for different medical imaging purposes whereby breast cancer detection involving utilization of CNNs has the potential to address current challenges in conventional mammogram interpretation like inter-observer variability, false positives, and missed diagnoses. It is possible to enhance diagnostic accuracy by analyzing mammography images using CNNs, thus enabling radiologists to improve their expertise and expedite the triaging of patients so that timely interventions occur thereby leading to improved patient outcomes. This paper seeks to contribute to ongoing efforts aimed at leveraging CNNs for breast cancer identification by presenting a comprehensive study covering data preprocessing, CNN architecture design, model training, and evaluation. The main aim is to develop a powerful and effective CNN-based system that can accurately recognize and categorize breast cancer from mammography images. The proposed method starts with the collection of a diverse dataset containing mammography pictures obtained from different sources such as public repositories and healthcare institutions. To ensure consistency and reliability in further analysis, various pre-processing techniques are used on these images including enhancing their quality; correcting artifacts, and normalizing intensity variations. In addition to this, a CNN system designed specifically for breast cancer detection is discussed in the paper. The proposed architecture employs multiple convolutional layers followed by pooling layers that enable the extraction of relevant features at different scales. Other layers such as fully connected layers and dropout layers are included for classification purposes and prevention of overfitting.

The dataset is divided into training, validation, and testing sets to train and evaluate the CNN model, following best practices in machine learning model development. Hyperparameter tuning is done to optimize the performance of the model using metrics like accuracy, sensitivity, specificity and area under the receiver operating characteristic curve (AUC-ROC) depicts how good CNN performs in the identification of breast cancer.

In terms of improving the generalization ability of the CNN model, such techniques include data augmentation as well as transfer learning especially on limited annotated datasets as used in this research. By utilizing prior knowledge and pre-trained models, transfer learning makes it easier for a CNN to adjust itself accordingly to new data sets and new tasks with higher efficiency thereby leading to better outcomes anticipated with the use of reduced training time frames.

In this paper, we will review the proposed CNN-based method that demonstrates how effectively it can detect breast cancer from mammography images with computational efficiency. In this way, by comparing with existing methods, the author shows an improved performance of the suggested method in terms of its accuracy, robustness, and scalability which indicates a possibility of introducing it into real medical settings.

To sum up, the main aim of this article is to develop advanced CNNs for breast cancer diagnosis by highlighting critical issues in traditional mammogram reading and offering effective suggestions for early detection and prevention. Thus, using medical imaging together with deep learning, the recommended solution is expected to become a key milestone in improving health outcomes globally once adopted worldwide.

OBJECTIVE:

This research project aims to create a Convolutional Neural Network (CNN) model to detect breast cancer from images, and mammograms accurately. Given the impact of breast cancer, on mortality rates, there is a demand for precise and effective diagnostic tools. By harnessing the image recognition and classification abilities of CNNs this study aims to improve the accuracy, speed, and scalability of detection. Key objectives include developing and fine-tuning a CNN structure for identifying breast cancer enhancing feature extraction through preprocessing and image augmentation techniques assessing the models' performance using metrics comparing it with existing approaches and delving into interpretability for potential clinical insights. The ultimate goal of this endeavor is to contribute towards diagnosis and improved patient outcomes, in breast cancer detection. **Materials and Methods:**

Figure 1 shows the overall flowchart for assisted diagnosis and staging of the PC method, which consists of the following parts that are processed consecutively: 1) Pancreas region of interest (ROI) segmentation; 2) Feature selection and fusion; 3) SVM classifier design, optimization, and evaluation; 4) EL-SVM design and evaluation. In the following sections, we will expand on the details.

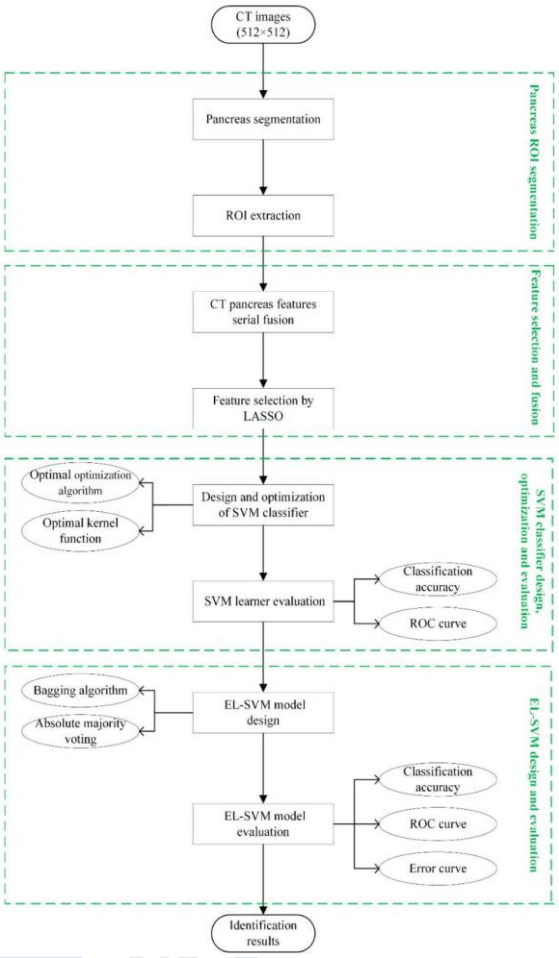


Fig 1. Overall flowchart for assisted diagnosis and staging of the PC method. The green boxes in the figure represent the four main parts of this paper.

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

.**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 21 years and Yilidan Rehman 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 168CT 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].

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 $\pm$ 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	
$< 37U/mL$	4(10.26)
$37 - 200U/mL$	7(17.95)
$> 200U/mL$	28(71.79)

Normal	51
Stage I	5
Stage II	34
Stage III	26
Stage IV	52

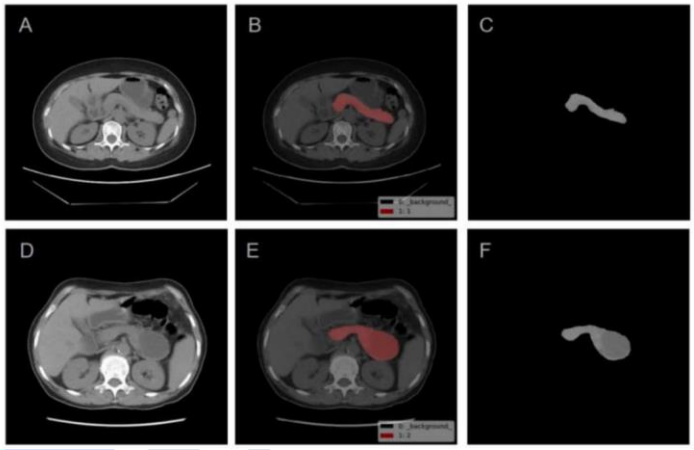


Fig 2. The pancreatic ROI segmentation. A. Normal CT image. B. ROI drawn on a pancreas image with the normal (red area). C. Normal segmentation. D. Stage IV CT image. E. ROI drawn on a pancreas image with the stage IV (red area). F. Stage IV segmentation.

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

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.

Table 2. The obtained experimental images

Type	Number of images
Total	168

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, different moment	15
GLDS	mean, contrast, angular second moment, entropy	4
Wavelet transform		160
Total		202

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.

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, as shown in Figure 3 [43].

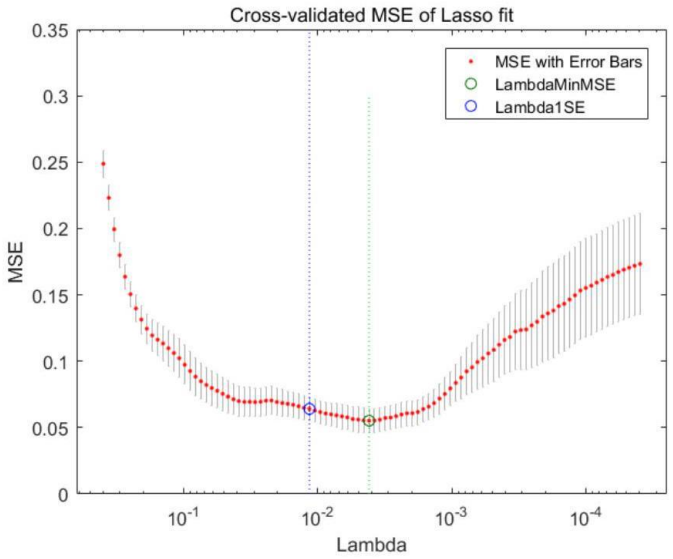


Fig 3. Cross-validated MSE of LASSO fit (normal-early stage)

$$\text{필필필}\sum_{\text{필1}}^{\text{필}}(\text{픓}-\text{픓}\text{픹}\text{픓})^2+\text{휆}\|\text{퓐}\|_1$$

SVM Classifier Design And Implementation:

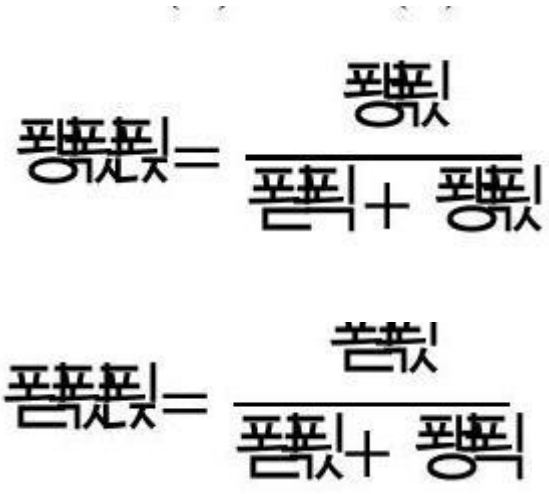
SVM was introduced by Vapnik et al. [44], because it shows good performance on binary classification tasks, it has gradually developed and is now commonly used to solve multidomain classification matters [45]. The primary target of SVM is to find a "maximum margin" partitioning hyperplane suitable for 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].

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.



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 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 algorithm is shown in Figure 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.

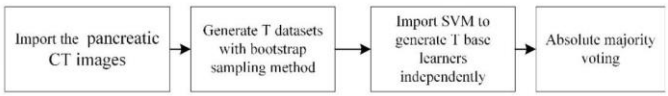


Fig 4. EL-SVM algorithm flow

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

Horizontal comparison with other models:

We compared the SVM, LASSO-SVM, KNN, BP neural network classifier, Soft-max 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.

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].

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.

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]

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 Soft-max classifier for comparison [57]. The 202dimensional features extracted by the above six texture analysis methods are passed to the Soft-max classifier to classify normal pancreas and different stages of PC. Experimental results showed that the average

classification accuracy of the Soft-max classifier was lower than EL-SVM 14.45%. As shown in Figure 5, the overall performance of EL-SVM is better than Soft-max classifier. The primary experimental parameters of Soft-max classifier are listed in Table 9

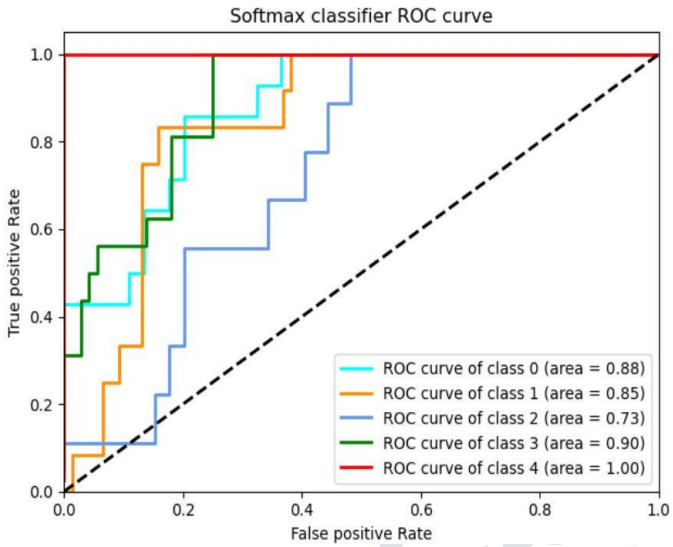


Fig 5. Soft-max classifier ROC curve. Class 0 (normal), class 1(early stage), class 2 (stage III), class 3(stage IV), class 4(PC).

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

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

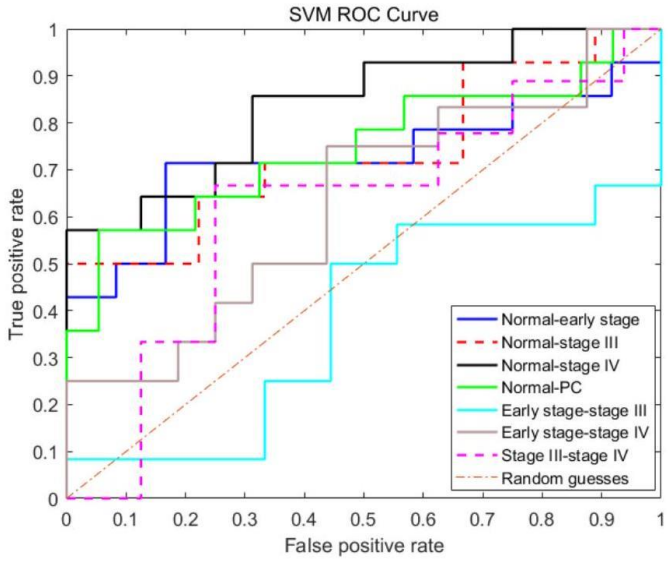


Fig 6. SVM classifier ROC Curve

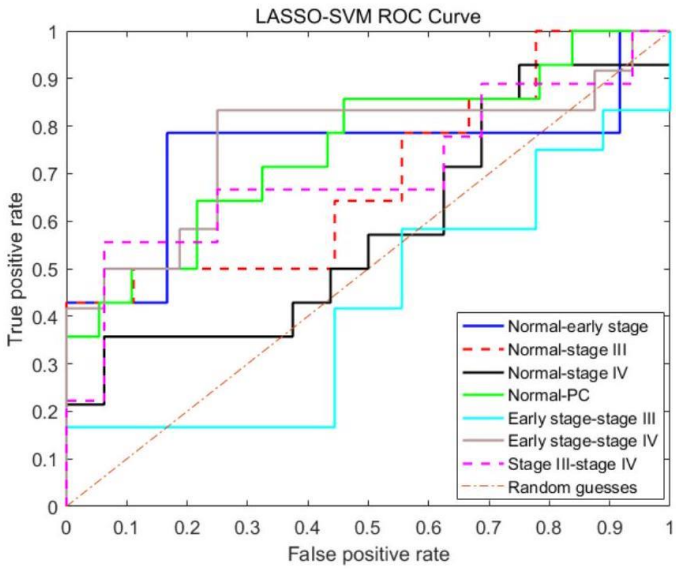


Fig 7. LASSO-SVM ROC Curve

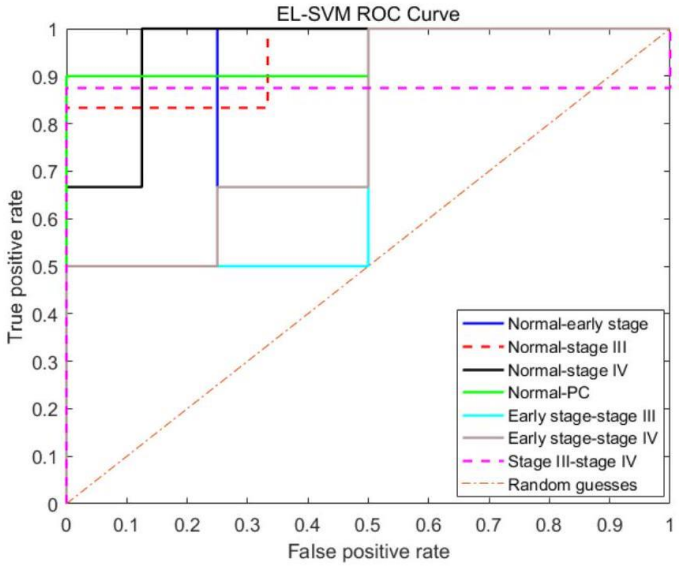


Fig 8. EL-SVM ROC Curve

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.

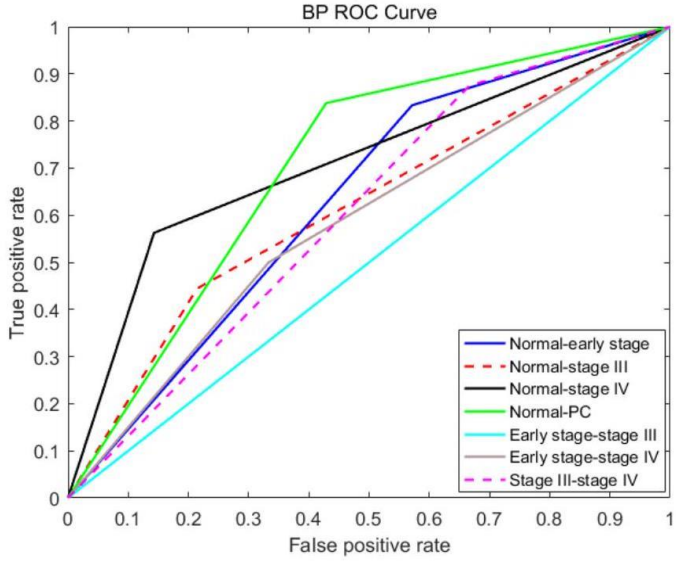


Fig 9. BP classifier ROC curve

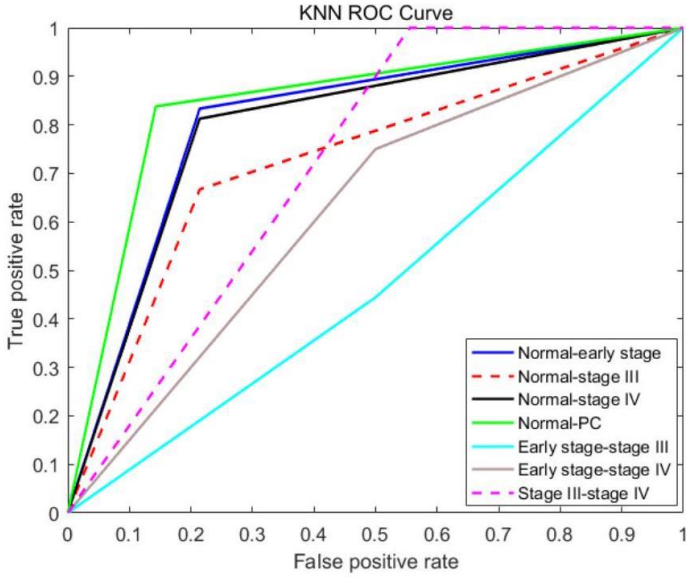


Fig 10. KNN classifier ROC curve

Classifier	Parameter	Parameter value	Parameter	Parameter value
GS-SVM	Range of C	$[2^{-4}, 2^{5}]$	Range of g	$[2^{-4}, 2^{5}]$
	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		

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

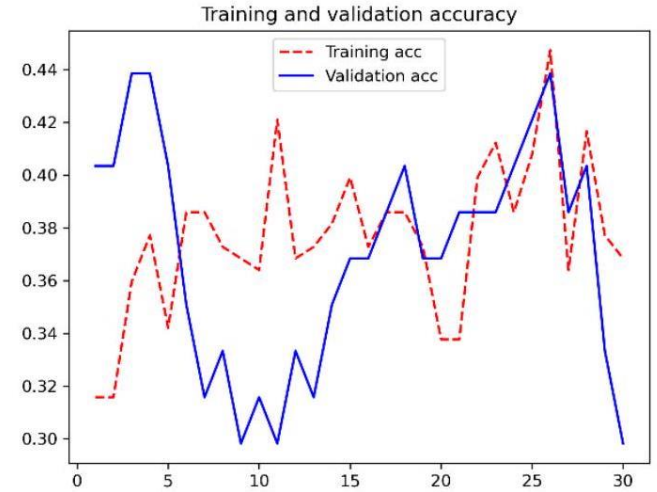
Table 8. The experimental parameter of BP and KNN

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

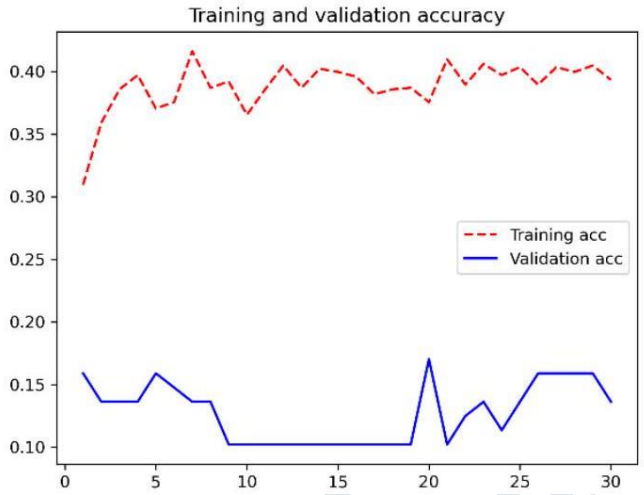
Table 9. The experimental parameter of Softmax classifier

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		

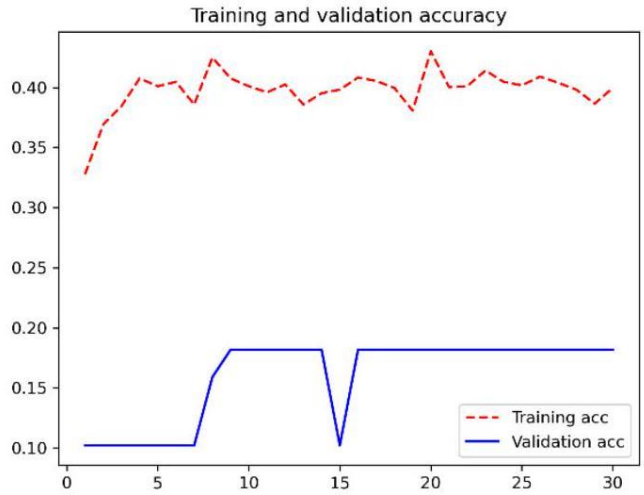
Table 4. Visual picture of the VGG16 model accuracy



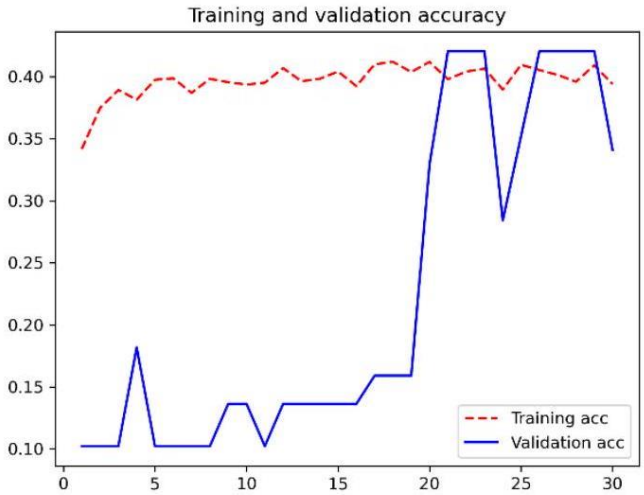
(a)



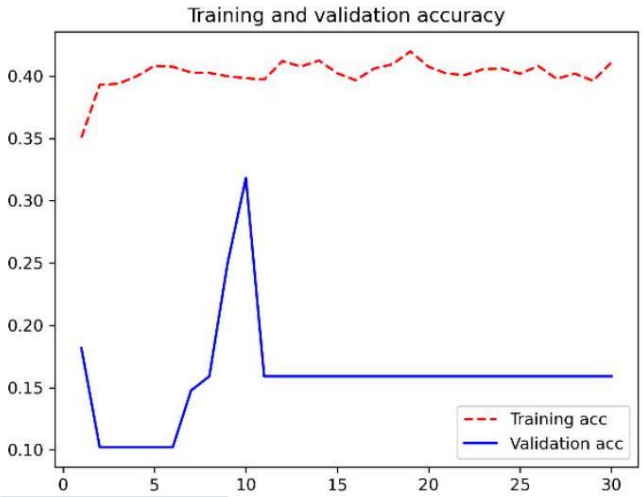
(b)



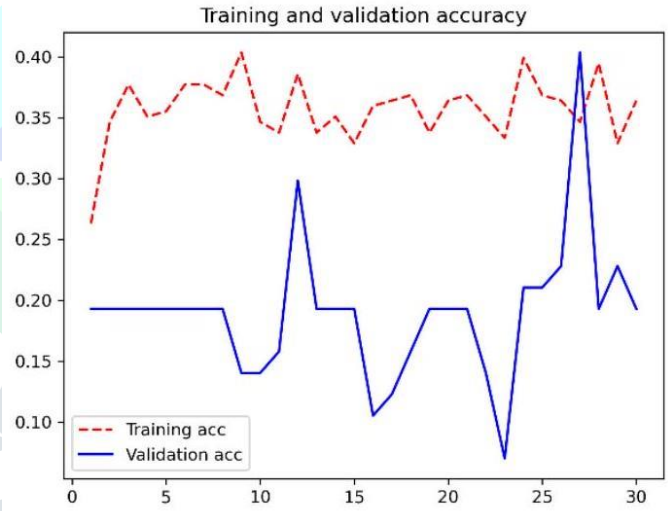
(c)



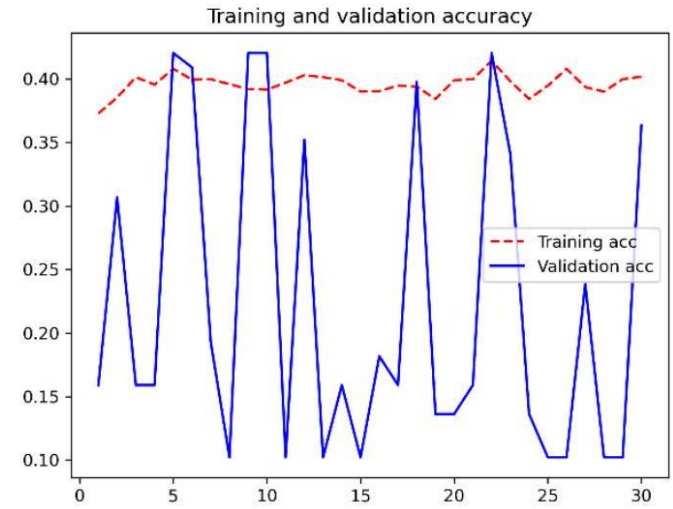
(d)



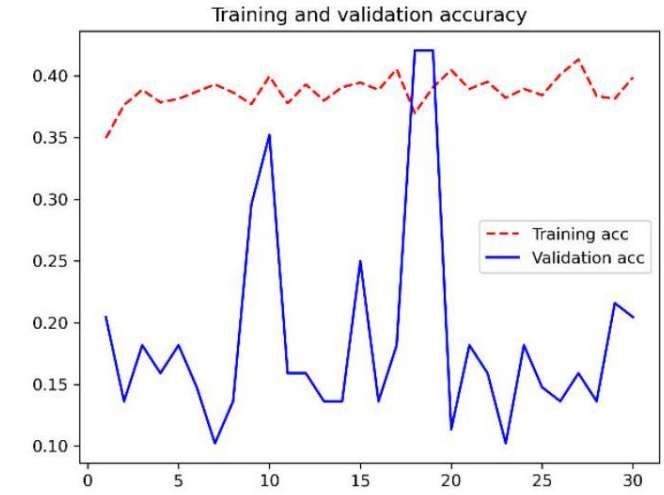
Note-The visual picture of the VGG16 model accuracy. (a) no data augmentation. (b) data augmentation 3 times. (c) data augmentation 6 times. (d) data augmentation 12 times. (e) data augmentation 24 times.



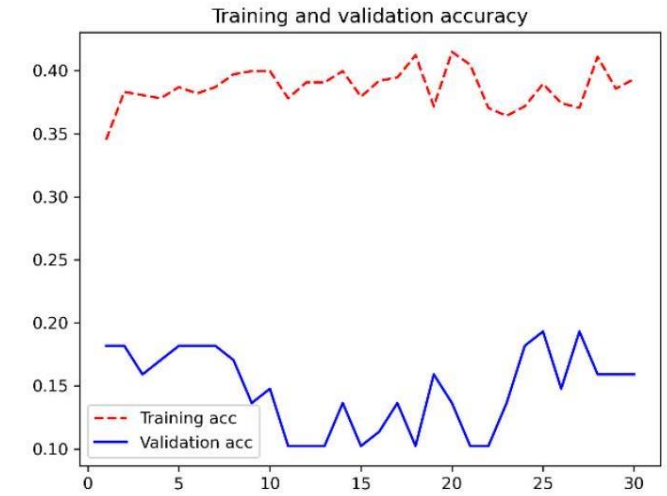
(a)



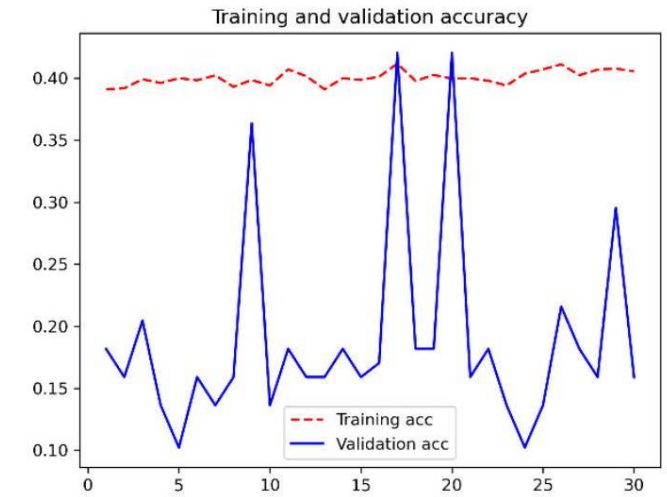
(d)



(c)



(b)



(e)

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

The convergence value is the classification error convergence value.

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.

4. Conclusion

This paper was the first to propose 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.

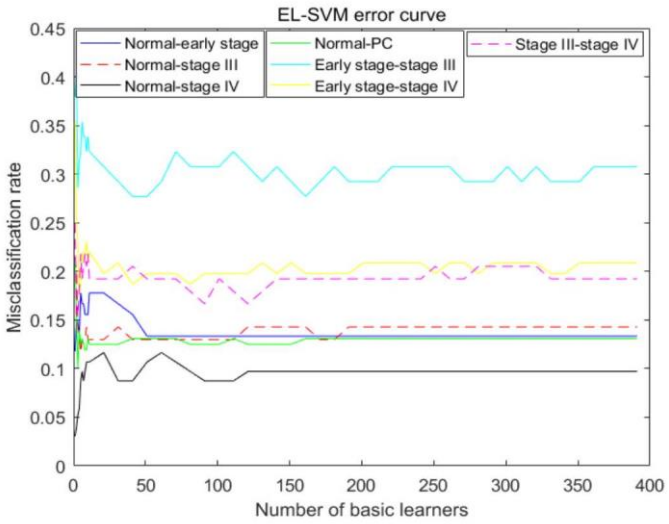


Fig 11. EL-SVM error curve

5. Conflicts of interest

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

6. 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. 789795, 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," International Journal of Clinical and Experimental Pathology, 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 Journal of 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 in Ultrasound Ct and Mri, vol. 40, no. 6, pp. 436-468, Dec, 2019.

[6] S. Schneeweiss, M. Horger, A. Grozinger, K. Nikolaou, D. Ketelsen, R. Syha, and G. Grozinger, "CT-perfusion measurements in pancreatic carcinoma with different kinetic models: Is there a chance for tumour grading based on functional parameters?," Cancer Imaging, vol. 16, Dec, 2016.

[7] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," Caa Cancer Journal for 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?," Clinical & Translational Oncology, vol. 22, no. 1, pp. 81-90, Jan, 2020.

[9] M. Ilic, and I. Ilic, "Epidemiology of pancreatic cancer," World Journal of 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 Journal of 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 Imaging, vol. 17, May 2017.

[12] A. Singh, and A. L. Faulx, "Endoscopic Evaluation in the Workup of Pancreatic Cancer," Surgical Clinics of North America, vol. 96, no. 6, pp. 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?," American Journal of Surgery, 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," American Journal of Gastroenterology, vol. 107, no. 2, pp. 303-310, Feb, 2012.

[15] M. I. Canto, F. Harinck, and R. H. Hruban, "International Cancer of the Pancreas Screening (CAPS) Consortium summit on the management of patients with increased risk for familial pancreatic cancer (vol 62, pg 339, 2013)," Gut, vol. 63, no. 12, pp. 1978-1978, Dec, 2014.

[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 and 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 Archives of Surgery, vol. 393, no. 6, pp. 923927, Nov 2008.

[18] J. Iglesias-Garcia, and J. E. Dominguez-Munoz, "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, Jan 2014.

[19] T. Yoshida, Y. Yamashita, and M. Kitano, "Endoscopic Ultrasound for Early Diagnosis of Pancreatic Cancer," Diagnostics, vol. 9, no. 3, Sep, 2019.