Rectal Cancer Tumor Image Segmentation Based on CNN Image Feature Recognition

Jifu Feng [1, #], Guangyin Zhou [2, #], Shengwei Zhang [3, #]

[1] School of intelligent medicine, China Medical University, Shenyang, China, 110013

[2] School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China, 211166

[3] School of medicine, University of Electronic Science and Technology of China, Chengdu, China, 610054

* Corresponding Author Email: zhangshengwei1953@163.com

[#]These authors contributed equally.

Abstract. Aiming at the problems of complex gray scale distribution, different local contrast, artifacts, and noise in CT images, this study proposes a new tumor segmentation method for rectal cancer, which employs a CNN model with a specific sequential input strategy to preprocess the images, and by comparing it with a random sequential input strategy, we demonstrate that the proposed method achieves a correctness rate of 70% in complex tumor slices. In addition to having a direct impact on the field of medical imaging, this research provides valuable insights and lessons learned for the wide application of deep learning in medical image processing.

Keywords: Colorectal Cancer Tumor; Convolutional Neural Network; Image Feature Recognition.

1. Introduction

The International Agency for Research on Cancer (IARC) 2020 Cancer Incidence and Mortality Database maintained by the World Health Organization indicates that colorectal cancer ranks third globally in terms of incidence and is the second most common cause of cancer-related deaths.[1]Confirmation of the diagnosis of colorectal cancer is mainly based on colorectoscopy, an invasive test, but early colorectal cancer may only manifest as tiny mucosal lesions and colorectal cancer still carries a risk of missed diagnosis.[2]CT and MRI also play an important role in the diagnosis of colorectal cancer, especially in the early detection of colorectal cancer and the determination of distant metastasis, but due to the complexity of the abdominal organs, the accuracy of CT and MRI is not high and very much relies on the experience of the imaging physician.[3, 4] With the increase in computing power and the improvement of image analysis algorithms, the convolutional neural network (CNN) models based on deep learning techniques have brought the possibility of improving the accuracy and simplicity of CT and MRI diagnosis.

The primary use of convolutional neural networks is the processing and analysis of grid-structured data, including photos and movies. CNNs use convolutional operations and pooling operations to extract features from images and learn a representation of these features through a multilayer neural network. The pooling layer uses a downsampling technique to shrink the size of the feature map, which lowers computation and increases the model's robustness, while the convolutional layer uses a convolution operation to extract features from the input image. Finally, tasks such as classification or regression are performed by fully connected and softmax layers. Deep convolutional networks are one of the most popular deep learning approaches in medical image analysis. They have shown considerable promise in application areas like disease categorization, computer-aided diagnosis, and retrieval.[5, 6]

The accuracy of the convolutional neural networks VGG-16 and Inception-v3, which were employed and evaluated on 279 thyroid tumor images in the Guan et al. study, was approximately 95% and 87.5%, respectively.[7] The accuracy of the VGG-19-based test set (97.34%) was greater than that of InceptionResNet-v2 (94.42%) in Wang et al.'s study using the deep convolutional neural

networks Inception-ResNet-v2 and VGG-19 on histopathology pictures of thyroid cancers.[8]The study of Chen et al. first used a convolutional neural network for the optical diagnosis of colorectal cancer through the 2157 data after training the convolutional neural network, with a sensitivity of 96.3% and specificity of 78.1% when they tested it with 284 cases of colorectal tumors and hyperplastic polyps.[9] By using NBI-Near Focus (NBI-NF) to train a convolutional neural network, Byrne et al. took the field even farther. After testing on 125 small polyps, the model diagnosed 106 of them, reaching the sensitivity of identifying NBI international colorectal endoscopes. The model was trained on 220 colorectal polyp-positive videos.[10]

To address the challenges posed by the poor quality of CT images and the problem of pseudo-representation in the training results, this study proposes a new method for high-precision rectal cancer tumor identification based on a convolutional neural network algorithm. Through real-world case studies, our method effectively addresses these challenges and shows significant efficacy in reducing the region extraction time and decreasing the network convergence time. In addition, we conducted a comprehensive comparative analysis of various convolutional networks, dataset partitioning methods, and learning rates, among other key factors. The results highlight the superior performance of our approach in colorectal tumor image recognition, providing a viable solution for improving diagnostic accuracy and efficiency. In addition to its direct impact on the field of medical imaging, this study provides valuable insights and lessons learned for the wide application of deep learning in medical image processing.

2. Modeling of convolutional neural networks

A convolutional Neural Network (CNN) is a feedforward neural network with convolutional computation and deep structure. [11] The combination of artificial neural networks and deep learning technology gives it a very powerful representation learning ability. [12] The frequent iterative training improves the accuracy of the network by minimizing the loss function. Make its local recognition ability of the image more accurate. Therefore, convolutional neural networks are widely used in the field of visual image processing and recognition. [13]

The classical structure of the Convolutional Neural Network includes convolutional layers, pooling layers, and fully connected layers. The convolutional layer is the core module of the convolutional neural network. The specific content is shown in Figure 1.

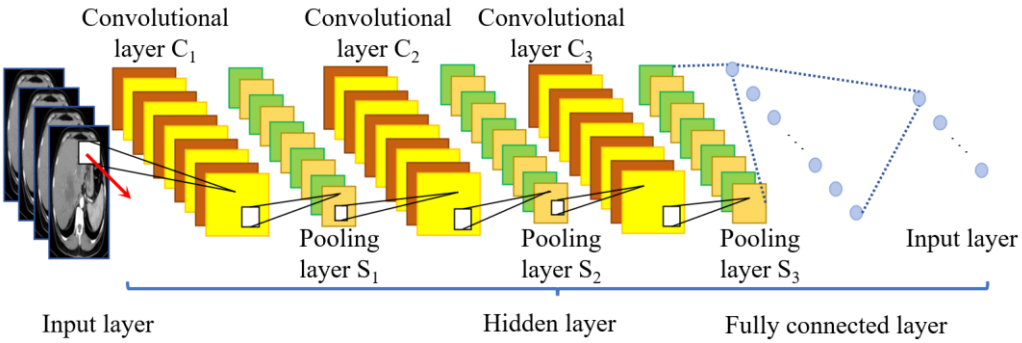


Figure 1. A Schematic diagram of the Convolutional Neural Network

2.1. Convolutional Neural Network Modeling

2.1.1 hidden layer

The hidden layers include convolution, pooling, and fully connected layers.[14] Take both dimensions, making the depth "1". Let the height of the input feature map, and the width be (H_0, W_0), the matrix is $I_{H_0 \times W_0}$; the height of the convolution core or pooling core, the width is (H_1, W_1), the matrix is $K_{H_1 \times W_1}$; the height of the output feature map, the width is(H_2, W_2), the matrix is $O_{H_2 \times W_2}$; the step size is (S_H, S_W). The specific equation is shown in (1).

$$(H_2, W_2) = \left(\frac{H_0 - H_1}{S_H} + 1, \frac{W_0 - W_1}{S_W} + 1 \right) \tag{1}$$

The convolutional layer consists of a set of learnable convolution kernels that slide the steps on the input image for the convolution operations. Each convolution kernel extracts the different features of the input image. The convolutional layer generates feature maps through convolution operation and gradually extracts the high-level features of the image. The specific equation is shown in (2).

$$o_{mn} = K_{H_1 \times W_1} * A_{H_1 \times W_1} + a_{mn} \tag{2}$$

Where o_{mn} denotes the pixel values of m rows and n columns of the feature map output after passing through the convolutional layer, $m \in [1, H_2], n \in [1, W_2]$, both are integers; $A_{H_1 \times W_1}$ denote the pixel values of m rows and n columns of the feature map output from the previous convolutional layer; a_{mn} denotes the offset;

2.1.2 pooling layer

The pooling layer can ensure most of the important information, reduce the dimension of each feature graph, reduce the number of parameters and computation of the network, prevent overfitting, and improve the speed and accuracy of the model.[15]

There are many kinds of pooling methods. The common pooling operations are maximum pooling and average pooling. Select the maximum or average value of each pooling core to compress the feature map and retain the main features. The specific process is as follows (3):

$$o_{mn} = \max \begin{pmatrix} i_{(m-1)\times S_H+1,(n-1)\times S_W+1} & i_{(m-1)\times S_H+1,(n-1)\times S_W+1+1} & \cdots & i_{(m-1)\times S_H+1,(n-1)\times S_W+1+W_1} \\ i_{(m-1)\times S_H+1+1,(n-1)\times S_W+1} & i_{(m-1)\times S_H+1+1,(n-1)\times S_W+1+1} & \cdots & i_{(m-1)\times S_H+1+1,(n-1)\times S_W+1+W_1} \\ \vdots & \vdots & \ddots & \vdots \\ i_{(m-1)\times S_H+1+H_1,(n-1)\times S_W+1} & i_{(m-1)\times S_H+1+H_1,(n-1)\times S_W+1+1} & \cdots & i_{(m-1)\times S_H+1+H_1,(n-1)\times S_W+1+W_1} \end{pmatrix} \tag{3}$$

Where o_{mn} denotes the pixel values of the mth row and nth column of the feature map output through the convolutional layer, $m \in [1, H_2], n \in [1, W_2]$ that are output through the convolution layer and are all integers.

2.1.3 fully connected layer

The fully connected layer is located between the convolution layer or the pooling layer and the output layer.[16] It expands the features extracted from the convolutional layer or pooling layer into a one-dimensional vector x_j and connects to one or more fully connected layers so that all neurons are connected to the next layer, and finally the map transforms into category probability or numerical predictions.

The fully connected layer performs the classification or regression task via the parameter weights w_{ij} and bias b_i . The specific process is as follows:

Let the input n neuron node $X_{n\times 1}$, output m neuron node $Y_{m\times 1}$, and the weight matrix is $W_{m\times n}$. Specifically, as shown in Equation (4).

$$y_i = \sum_{j=1}^{n} w_{ij} * x_j + b_i \tag{4}$$

Where y_i denotes the output of the ith neuron after passing through the fully connected layer; the output of the fully connected layer, $i \in [1, m]$, and is an integer; w_{ij} denotes the weight occupied by the attribute of the input of the j-th neuron before the fully connected layer corresponding to the attribute of the output of the i-th neuron passing through the fully connected layer; x_j denotes the output of the jth neuron of the first layer of the fully connected layer; the input of the fully connected layer, j∈[1,m], and an integer.

2.2. Feature Recognition

Feature recognition refers to the process of using a computer to extract and identify critical information from the raw image data. Through feature identification, the raw data can be transformed into a more representative feature vector, thus simplifying the complexity of the problem and improving the efficiency and accuracy of the model. Feature recognition is mainly achieved through image preprocessing, image gray, and image edge extraction.

2.2.1 Image preprocessing

In image analysis, it is common to encounter low quality image datasets due to technical issues. Since image quality directly affects model accuracy, preprocessing is crucial for feature extraction. This involves operations such as grayscale conversion, normalization, denoising, and enhancement. Gray scale conversion simplifies processing and reduces complexity and storage requirements. Methods include max/min, averaging, and weighted averaging. Normalization converts images to a standard form, which aids in comparison. Techniques include subtracting the minimum pixel value and moment-based normalization.[17]

Interference in image data, termed image noise, can arise during digitization or transmission. Image denoising methods, traditional and deep learning-based, aim to reduce this noise. Traditional methods like mean filtering and wavelet transform process in the image or frequency domain. Deep learning methods such as DnCNN and FFDNet are also employed. [18] Image enhancement enhances features like edge, contour, and contrast, improving visual clarity. Techniques include grayscale transformation, smoothing, and sharpening to refine image quality and highlight important features while minimizing blurring.

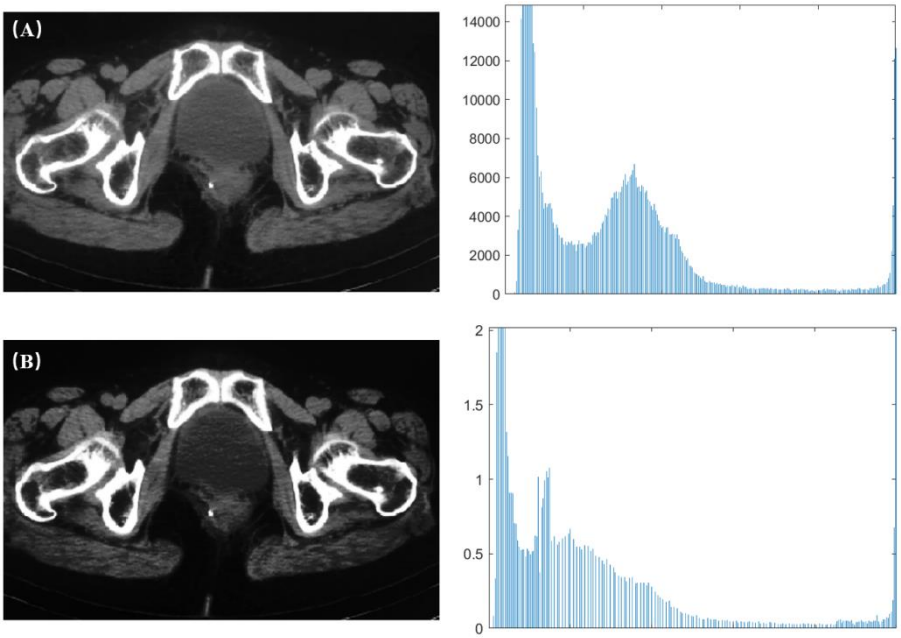


Fig. 2 CT map and histogram of rectal cancer patients after grayed-out enhancement, with (A) and (B) indicating before and after enhancement, respectively.

Specific regions can be enhanced by applying different linear grayscales to different gray ranges of the image using segmented linear transformations. As shown in Fig. 2, the grayed-out enhanced CT map is more interesting for feature differentiation.

2.2.2 Image texture feature extraction

The texture is in the image space, if there must be some grayscale relationship between two pixels separated by a certain distance, that is, the spatial correlation characteristics of the grayscale in the image. [19] Common texture feature extraction methods include gray symbiosis matrix, gray stroke

statistics, Markov model, and so on. In this paper, the texture features are extracted by the grayscale symbiosis matrix. The specific process is as follows:

The gray symbiosis matrix $L(a,b)$, indicates the probability that the distance from another pixel with a gray point of a is (dx,dy) is b; With image $f(x,y)$ with size $M \times N$, gray scale K ; The grayscale covariance matrix satisfying certain spatial relations is shown in Equation (5).

$$L(a,b|d,\theta)=\{(x,y)|f(x,y)=a,\ f(x+dx,y+dy)=b\} \tag{5}$$

where d denotes the relative distance expressed in pixel numbers; θ generally considers four directions, 0°, 45°, 90°, and 135°; *(x, y)* denotes the pixel coordinates in the image.

2.2.3 Image edge feature extraction

The edge of the image generally refers to the part of the image with the largest change rate of local area luminance, and the edge detection is designed to capture the area with a significant change in brightness. Common edge detection methods include the Sobel algorithm, Prewitt algorithm, Robert's algorithm, and Canny algorithm. The Canny algorithm is used in this paper. The Canny algorithm is a method of smoothing and then derivatives. The specific steps are described as follows:

Gaussian blur of the image, the Gaussian function is defined as follows Equation (6):

$$G=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2}{2\sigma^2}} \tag{6}$$

where σ denotes the standard deviation of the Gaussian filter function, which is used to adjust the smoothness of the image.

3. Results and discussion

3.1. Data Set

The data used in this experiment were taken from the Cancer Image Archive (TCIA) database with cancer imaging datasets for various cancer types and various imaging modalities.TCIA provides the publicly available dataset StageII-Colorectal-CT, which consists of 230 stage II colorectal cancer patients with abdominal or pelvic enhancement CT images within 10 days prior to surgery, including DICOM format enhanced CT images. [20] The inclusion criteria were as follows: (i) patients with radical CRC surgery (complete resection of the primary tumor and regional lymph node dissection); (ii) patients with histologically and pathologically confirmed stage ii CRC; and (iii) patients who underwent an enhanced CT examination of the abdomen or pelvis within 10 days prior to surgery and provided a complete enhanced CT scan image. The exclusion criteria were as follows: (i) preoperative treatment (neoadjuvant chemotherapy or radiotherapy); (ii) patients with another concurrent oncologic disease; and (iii) patients who died from acute surgical complications within 1 month after surgery. The purpose of this experiment is to automatically classify CT images of patients with stage II colorectal cancer into two categories: "areas with tumor features" and "areas without tumor features" by machine learning methods. The data used in the experiment are from the StageII-Colorectal-CT dataset in the Cancer Image Archive (TCIA) database. After screening, data from a total of 230 patients were included in the study. In order to train and evaluate the model, the data of these 230 cases were randomly divided into two parts: 70% (161 cases) were used as a training set for model training, while the remaining 30% (69 cases) were used as a test set for evaluating the generalization ability of the model. In this experiment, our goal is to utilize image segmentation techniques as a preprocessing step to aid in the task of classifying images. By separating the tumor region from the non-tumor region in the image, the model is able to focus more accurately on the key features in the image, thus improving the classification accuracy. The samples from the test set were classified into tumor and non-tumor samples to create a sample mapping. The test set had a total of 69 samples, of which 43 samples were classified as tumor and 26 samples were classified as non-

tumor. The number of correctly classified tumor samples was 40 and the number of correctly classified non-tumor samples was 29. The correctness rate was 91.3%.

The experiments were performed on the enhanced CT images by grayscaling to simplify the image information and highlight the key structures, using Matlab processing software, version R2022b. After grayscaling, we screened 100 images from the enhanced CT images for uniform resolution processing, with the size of 100*100, to ensure that the images have a consistent spatial pixel density, to improve the quality of the images and the reliability of the comparisons. reliability. The data processing process is shown in Figure 3.

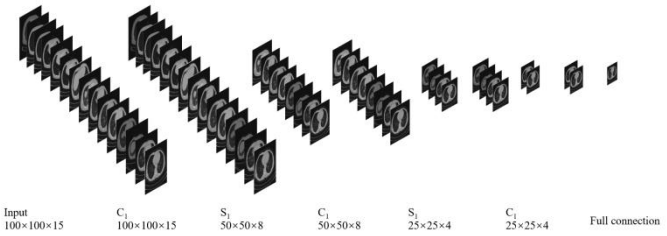


Figure 3. CNN Processing

3.2. Comparative analysis of experiments

Firstly, the CT image of rectal cancer is used for feature recognition cutting, after preprocessing, thresholding, and binarization of the image, the outer boundary of the target is tracked and the continuous region is pseudo-colored and divided into 14 area features, and the ThinnessRatio, AspectRatio, Area, and Eccentricity of each feature is contained in Table 1. Among them, ThinnessRatio, AspectRatio, Area, and Eccentricity are four metrics for the shape and characteristics of the target area cut out of the CT image of rectal cancer. ThinnessRatio measures the slenderness of the object, which is derived by calculating the ratio of the length of the object's boundaries to the area of the square root shape. The equation is calculated as in (1).

$$ThinnessRatio = \frac{4\pi S}{2L} \tag{7}$$

Where S represents the area of the shape and L represents the perimeter of the shape.

AspectRatio reflects the shape of the object and is the ratio of the object's width to its height. The equation is calculated as in (2).

$$AspectRatio = \frac{Length}{Width} \tag{8}$$

Area represents the total number of pixels occupied by the object and is used to measure the size of the object. And Eccentricity describes the extent to which the shape of the object's outline deviates from a circle, derived by calculating the ratio between the long and short axes of the ellipse. The equation is calculated as in (3).

$$Eccentricity = \sqrt{(1-(b^2/a^2))} \tag{9}$$

The computation and analysis of these features help to provide information about the morphological characteristics of rectal cancer images, which in turn provides an important reference for subsequent classification tasks.

Fig. 4 exhibits the The results of feature recognition, it can be seen that the highest classification accuracy of the CNN model with specific order input strategy is 71%, which is higher than the highest accuracy of the CNN model with random order input strategy 63%, which indicates that the specific order input strategy proposed in this paper improves the training accuracy of the model, and thus effectively improves the classification accuracy of rectal cancer images.

In the current medical field, Full Convolutional Networks (FCN) in deep learning is widely recognized as a superior algorithm for image segmentation tasks.FCN combines Convolutional Neural Networks (CNNs) and transpositional convolutional operations with the advantages of automatic learning of features, capturing of contextual information, multi-scale processing and end-to-end training. FCN combines convolutional neural network (CNN) and transposed convolutional operations with the advantages of automatic feature learning, capturing contextual information, multi-scale processing and end-to-end training. [21] In this paper, Support Vector Machine (SVM) is used for comparative feature classification, and SVM is performed for rectal cancer CT images that undergo the same preprocessing for RF image feature extraction. The previously mentioned dataset is divided into a training set (80%) and a test set (20%) for training, and the SVM classifier is utilized to learn the relationship between feature images and labels to achieve the classification, and the performance of the SVM classifier is evaluated on the test set, and it is found that it achieves an accuracy of 60%, which is slightly lower than the CNN network proposed in this paper.[22] In this paper, (FCN) is used for feature comparison classification and SVM is performed on rectal cancer CT images that have undergone the same preprocessing to extract RF image features. The previously mentioned dataset is divided into a training set (80%) and a test set (20%) for training, and the FCN classifier is utilized to learn the relationship between feature images and labels to achieve the classification, and the performance of the FCN classifier is evaluated on the test set, and it is found that its accuracy reaches 60%, which is slightly lower than that of the CNN network proposed in this paper. After analysis, it is concluded that FCN can be optimized as a whole by end-to-end training, which makes the model have better generalization performance on larger medical image datasets. However, CNN has better performance in accurate image analysis.

Table 1. Characterization data

Zone	ThinnessRatio	AspectRatio	Area	Eccentricity
1	0.666824	1.579646	623380	0.748586
2	0.286551	1.15	9741	0.887628
3	3.975084	1	4	0
4	0.975629	2	24	0.877058
5	0.422868	1.380952	324	0.874425
6	0.425611	0.558559	4046	0.871931
7	0.538398	0.969231	2725	0.586498
8	0.546246	0.885246	2327	0.535794
9	1.79871	2	8	0.866025
10	0.525441	0.560748	4068	0.871709
11	3.975084	1	4	0
12	0.527856	2	60	0.885718
13	0.40841	1.34188	10220	0.849772
14	3.975084	1	4	0

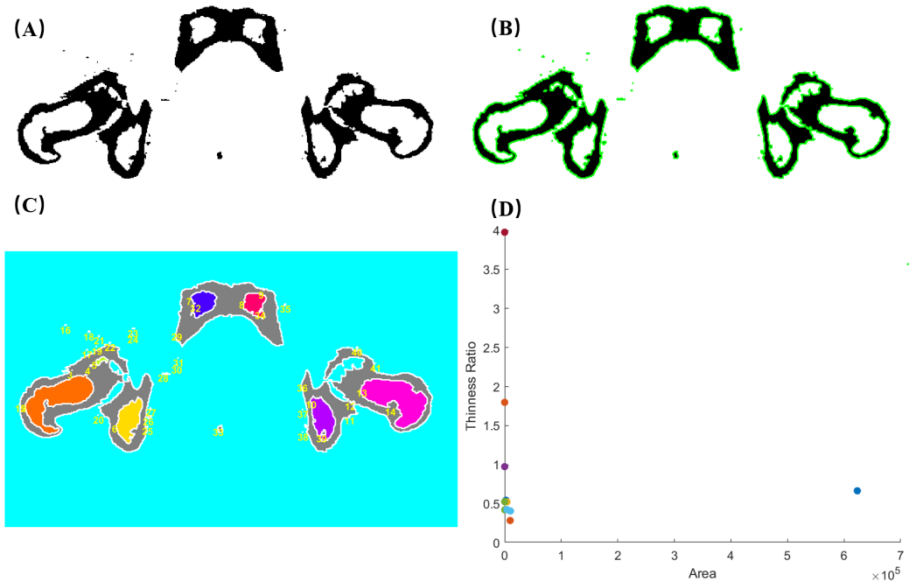


Figure. 4 CT image binarization, feature recognition, threshold classification, and classification statistic graph

4. Conclusions

After pre-processing 100 enhanced CT images of stage II colorectal cancer, the experiment tracked the outer boundary of the target and divided the images into four regional features. The CNN model with the specific order input strategy proposed in this paper has a maximum classification accuracy of 71%, which is significantly better than 63% of the random order input strategy. The results show that the specific order input strategy proposed in this paper can improve the training accuracy of the model, and thus effectively improve the segmentation accuracy of colorectal cancer images. In addition, when compared with the feature recognition of CT images of pancreatic neuroendocrine tumors, it was found that the segmentation accuracy of colorectal cancer images was 72%, while that of pancreatic neuroendocrine tumors was 64%, which is due to the more complex features of adenoid-neuroendocrine tumors and the need for subsequent optimization of the model to improve accuracy.

References

[1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. Ca-a Cancer Journal for Clinicians, 2021, 71(3): 209-49.

[2] DEKKER E, TANIS P J, VLEUGELS J L A, et al. Colorectal cancer [J]. Lancet, 2019, 394(10207): 1467-80.

[3] BEETS-TAN R G H, LAMBREGTS D M J, MAAS M, et al. Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting (vol 28, pg 1465, 2017) [J]. European Radiology, 2018, 28(6): 2711-.

[4] NERAD E, LAHAYE M J, MAAS M, et al. Diagnostic Accuracy of CT for Local Staging of Colon Cancer: A Systematic Review and Meta-Analysis [J]. American Journal of Roentgenology, 2016, 207(5): 984-95.

[5] XIE Y, ZACCAGNA F, RUNDO L, et al. Convolutional Neural Network Techniques for Brain Tumor Classification (from 2015 to 2022): Review, Challenges, and Future Perspectives [J]. Diagnostics, 2022, 12(8).

[6] ANWAR S M, MAJID M, QAYYUM A, et al. Medical Image Analysis using Convolutional Neural Networks: A Review [J]. Journal of Medical Systems, 2018, 42(11).

[7] GUARI Q, WANG Y, PING B, et al. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study [J]. Journal of Cancer, 2019, 10(20): 4876-82.

[8] WANG Y, GUAN Q, LAO I, et al. Using deep convolutional neural networks for multi-classification of thyroid tumor by histopathology: a large-scale pilot study [J]. Annals of Translational Medicine, 2019, 7(18).

[9] CHEN P-J, LIN M-C, LAI M-J, et al. Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis [J]. Gastroenterology, 2018, 154(3): 568-75.

[10] BYRNE M F, CHAPADOS N, SOUDAN F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model [J]. Gut, 2019, 68(1): 94-100.

[11] SHAMSOLMOALI P, JAIN D K, ZAREAPOOR M, et al. High-dimensional multimedia classification using deep CNN and extended residual units [J]. Multimedia Tools and Applications, 2019, 78(17): 23867-82.

[12] HESSLER G, BARINGHAUS K-H. Artificial Intelligence in Drug Design [J]. Molecules, 2018, 23(10).

[13] LAN J, WANG D, SHEN X. Research progress on visual image detection based on convolutional neural network [J]. Chinese Journal of Scientific Instrument, 2020, 41(4): 167-82.

[14] ZHANG T, LI Y, HU H, et al. A Gender Classification Model Based on Cross-connected Convolutional Neural Networks [J]. Acta Automatica Sinica, 2016, 42(6): 858-65.

[15] YIN R, SU S, LI S. Convolutional neural network's image moment regularizing strategy [J]. CAAI Transactions on Intelligent Systems, 2016, 11(1): 43-8.

[16] ZHANG J, SHEN W, WU Z. Recognition of face privacy protection using convolutional neural networks [J]. Journal of Image and Graphics, 2019, 24(5): 744-52.

[17] HEITKOTTER H, VANG K, GAFFNEY M, et al. Evaluation of EZ reflectivity normalization across different OCT devices [J]. Investigative Ophthalmology & Visual Science, 2022, 63(7).

[18] MOU J, SHEN L, WANG H, et al. Finger vein denoising algorithm based on gradient-oriented residual structure and LBP texture loss [J]. Signal Image and Video Processing, 2024, 18(1): 943-52.

[19] LU L, LIU Z, YANG T, et al. Grayscale histogram and texture features of wake vortex image behind circular cylinder [J]. Journal of Hydroelectric Engineering, 2022, 41(9): 1-11.

[20] PRIOR F, SMITH K, SHARMA A, et al. Data Descriptor: The public cancer radiology imaging collections of The Cancer Imaging Archive [J]. Scientific Data, 2017, 4.

[21] ZOU Y, DUANMU J, GAO H. Study on support vector machine regression algorithm based on kernel partial least squares [J]. Computer Engineering and Design, 2010, 31(10): 2290-3.

[22] JINGPING J I am, FEIZHOU Z, YANMEI C. Object Tracking by Using SVM and Trust-Region Method [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2008, 44(5): 810-4.