Research on Application of Deep Learning in Esophageal Cancer Pathological Detection

Xiang Lin(✉), Zhang Juxiao, Yin Lu, and Ji Wenpei

Nanjing Normal University of Special Education, Nanjing, Jiangsu, China
lin_xiang@njts.edu.cn

Abstract. As the "gold standard" of tumor diagnosis, pathological diagnosis is more reliable than the analytical diagnosis, ultrasound, CT, nuclear magnetic resonance, etc. Detection of esophageal cancer based on pathological slice images is focused in this paper combining with deep learning to intelligently obtain reliable detection results. A data set is built by collecting and labeling pathological slices of esophageal cancer at varying stages for model training and verification. By comparing the performance of multiple models, ResNet50 is chosen as the network model. The model is pre-trained on ImageNet with a public breast cancer data set and transferred to the task of esophageal cancer detection. The original data set is enlarged by data augmentation to improve the accuracy, effectively avoiding over-fitting. Experimental results show the test accuracy achieves 0.950 which demonstrates the feasibility of deep learning on the esophageal cancer detection with pathological slice images.

Keywords: Esophageal Cancer Detection · Deep Learning · Transfer Learning · Data Augment

1 Introduction

Statistical data shows that the incidence rate of cancer has increased year by year in many cities of China in recent years. Taking Huai'an City in Jiangsu province as example, the proportion of cancer deaths among people who die of illness every year accounts for 30% ~ 40%. Among cancer deaths, esophageal cancer ranks the first.

By analysis, the increase of the incidence rate of esophageal cancer in Huai'an citizens is not only caused by environmental factors and increasingly serious chemical pollution, but also related to the unhealthy eating habits. Citizens of Huai'an are inclined to eat salty or pickled food, spicy food and excessive drinking. Nitrite rich in pickled food and unbalanced nutrition are important factors to induce esophageal cancer. In addition, one of the main reasons for the higher and higher mortality rate of esophageal cancer is that most of the cancers are in the middle or late stages when discovering illness, and it is difficult for patients at this stage to obtain fundamentally effective treatment. However, in the early stage esophageal cancer is not easy to be detected accurately for less extent of the lesion, which results in that some patients cannot get accurate diagnosis in time

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2024
Published by Springer Nature Switzerland AG 2024. All Rights Reserved
B. Wang et al. (Eds.): ICMTEL 2023, LNICST 535, pp. 95–105, 2024.
https://doi.org/10.1007/978-3-031-50580-5_9

and miss the best treatment opportunities [1]. If the cancer is detected in the early stage, the survival rates will be greatly improved.

Pathological diagnosis has always been seen as the "gold standard" of tumor diagnosis which is a disease diagnostic means to observe the pathological characteristics of organ tissue structure and cells under the microscope. Pathologists are called "doctors among doctors" [2]. As pathological diagnosis directly acts on cytopathic conditions, it is more reliable than the analytical diagnosis based on medical records and symptoms. And it is more convincing than the clinical diagnosis made by means of ultrasound, CT, nuclear magnetic resonance, etc. However, the number of pathologists is seriously insufficient, less than 10,000. Furthermore most of them are concentrated in economically developed areas. In some remote and poor areas, there is hardly a pathologist. Tumor is a high incidence disease. In addition, the incidence rate of cancer has little correlation with regions. The number of cases that pathologists need to deal with is close to six times the number of pathologists, and the proportion is still rising. Therefore, it is more effective and helpful for pathologists to hand over the pre- screening of tumor pathological images to computers. It can not only reduce the work load of the pathologist, but also improve the accuracy and efficiency of detection [3].

In this paper, pathological section images which are more reliable for tumor detection are used for esophageal cancer detection combining with deep learning. For the outstanding performance of deep learning in medical image recognition, esophageal cancer pathological section images are timely detected by selecting proper network model and improving its performance according to the specific esophageal cancer data. The main contributions of this paper are as follows:

(1) Select an appropriate deep learning model for esophageal cancer detection;
(2) Build and label our own esophageal cancer dataset;
(3) Introduce transfer learning to reduce training duration and difficulty to improve training efficiency;
(4) Enlarge the dataset by data enhancement technology to upgrade the model performance.

The rest of the paper is organized in the following manner. Section 3 lists the classical network models and the technology that we will use later. Section 4 outlines esophageal cancer dataset we collected. The verification experiments are given in Sect. 5. Finally a conclusion is drawn in Sect. 6.

2 Related Research

In recent years, deep learning has made breakthrough achievements in various medical fields, especially in the application of medical images [4–6], including radiation oncology diagnosis [7], classification of skin cancer [8], diabetes retinopathy [9], histological classification of biopsy specimens [10], and description of colorectal diseases with cell endoscope [11]. Research on image diagnosis of esophageal cancer based on deep learning has just emerged at home and abroad [12–15]. The use of esophageal endoscopic images combined with deep learning technology to diagnose esophageal cancer can more accurately locate the location of the disease, providing doctors with

an effective means of auxiliary recognition. The deep learning technology was applied to the detection of esophageal cancer CT images [16] and achieved good recognition results. Due to the lack of public pathological picture library, the research on the application of depth learning technology to the classification and detection of pathological images of esophageal cancer is relatively lacking. At present, more mature research on cancer detection for pathological images includes: detection technology for gastric cancer pathological images [17] and detection technology for breast cancer pathological images [18]. On the big data of pathological images, after large-scale and long-term training with deep learning technology, the detection and recognition of tumor pathological images have achieved high accuracy. In 2017, Google used deep learning technology to achieve an accuracy rate of early diagnosis of breast cancer of more than 99% [19], making breakthroughs in the field of the application of deep learning in medical images.

3 Classical Network Models and Transfer Learning

3.1 Classical Network Models

Convolutional neural network (CNN) was first proposed by Lecun of New York University in 1998. At the same time, the first CNN model LeNet-5 [20] is designed to classify handwritten digital images, which is the first time that CNN can be widely used in industrial practice. Its network structure is as follows (Fig. 1):

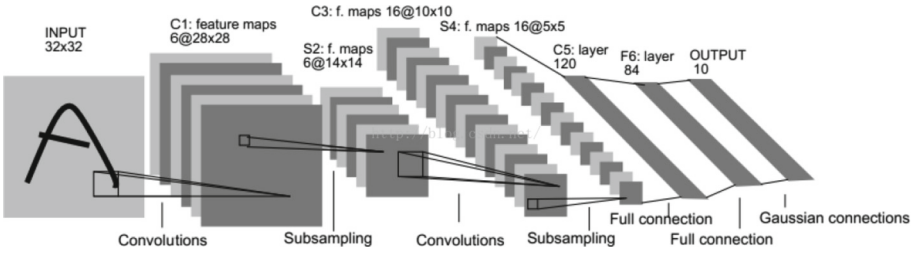


Fig. 1. The structure of LeNet-5 [20]

LeNet-5 is consisted of convolutional layers and a full-connected layer. Local connection and weight sharing greatly reduce network parameters, and the operation of subsampling reduces the data dimension. Convolutional kernels in convolutional layers extract local features, by which images are recognized in full connected layer.

In 2012, AlexNet [21] won the championship of ImageNet with an unprecedented error rate of 15.4%, far lower than the top 5 optimal error rate 26.2% before this, creating a great sensation. AlexNet is a network with 8 layers, whose structure is deeper and more complex than LeNet. Moreover, dropout layer is added behind the full-connected layers to avoid overfitting. It brings about an outstanding performance. The success of AlexNet promotes the development of deep learning.

In 2014, Karen and Andrew of Oxford University built a 16-layer CNN model, which reduced the error rate of Top-5 to 7.3%. This model is VGG [22]. VGG improves the

network performance by adding more layers to make the network structure deeper. It emphasizes the importance of "depth" to the deep learning model with practical effects. VGG basically follows the design idea of AlexNet and implements the word "depth" to the end, which is twice as deep as AlexNet. Unlike the former, VGG uses convolutional kernels of the same size of 3×3. Two convolution kernels of 3×3 is equivalent to a convolutional kernel of 5×5, and three kernels of 3×3 is equivalent to a kernel of 7×7, but the number of parameters is greatly reduced. Comparing three 3×3 kernels with a kernel of 7×7. In a same receptive field, the convolutional kernel has 27 parameters in the former and 49 in the latter. The former not only has 81.5% fewer parameters, but also has two more nonlinear operations, which is more conducive to the learning of the network and can also accelerate the convergence of the network.

In 2015, the Google proposed GoogleNet [23] using the Inception module, which reduced the error rate of Top5 to 6.7%, broke the traditional method of increasing the network depth and width, and transformed full connection and partial convolution modules into sparse connections.

Kaiming He of Microsoft Research Institute and other four Chinese proposed ResNet [24] (Residual Neural Network). This model has won the championship in the large-scale image classification competition ILSVRC2015, with the error rate on Top5 being the best 3.57% at present. At the same time, the number of parameters is less than VGG, but shows an extremely outstanding performance. The main idea of ResNet is to add a direct connection channel to the network, that is, the idea of Highway Network [25].

3.2 Transfer Learning

As all know, the size of training data is the key to training effect of neural networks. The size of the esophageal cancer dataset we collected is much smaller than the requirement of the training of big data. Transfer leaning is introduced to enhance the network performance in this paper.

Transfer Learning is a machine learning method that takes the model developed for Task A as the initial point and reuses it in the process of developing the model for Task B. The focus of transfer implementation is to solve the gap of data characteristics and feature distribution between Source domain and Target domain. Being both cancer pathological images, esophageal cancer and breast cancer has similar data distribution. Before directly training the network with the data we collected, we first train the network on the public breast cancer data set, and then transfer the training results to the esophageal cancer data to solve the problem of insufficient training data.

4 Dataset

4.1 Data Collection

A total of 1524 pathological images were collected from 720 patients in 2011–2017, involving squamous cell carcinoma and adenocarcinoma, as well as pathological section images of chronic inflammation of esophageal and cardiac mucosa. The collected pathological sections were observed and photographed with Olympus BX 50 optical microscope, collected with HMIAS-2000 high-definition full-automatic color image analysis

system and analyzed the positive staining area and staining intensity. The results were interpreted by double blind method and evaluated by two pathologists independently. The whole data set is divided into 5 kinds of different cell carcinoma, a suspected squamous cell carcinoma and 2 kinds of chronic inflammation. In the paper, we focus on the classification of benign and malignant tumors. So all of cell carcinomas including suspected are labeled as malignant results (M), while 2 kinds of chronic inflammation are labeled by benign (B) (as listed in Table 1).

Table 1. Data classification labels and the number of images

Classification Labels	The Number of Images
Poorly differentiated squamous cell carcinoma (B)	123
Moderately and poorly differentiated squamous cell carcinoma (B)	67
Moderately differentiated squamous cell carcinoma (B)	263
Moderately well differentiated squamous cell carcinoma (B)	141
Well differentiated squamous cell carcinoma (B)	113
Chronic inflammation of esophageal mucosa (M)	86
Squamous epithelial hyperplasia, keratosis and chronic inflammation (M)	541
Suspected squamous cell carcinoma (B)	190

4.2 Data Augmentation

To meet the requirements to large sample size of deep learning, data augmentation was used on the digital pathology image library, including rescale, horizontal/vertical rotation, horizontal/vertical translation, horizontal/vertical flip, fill. Table 2 lists the generator parameters and set values of the data augmentation method involved in this paper. Figure 2 shows the comparation of the original and augmented images by transformation.

Table 2. List of data augmentation parameters

Parameters	Values
rescale	1.0/255
rotation_range	90
width_shift_range	0.3
height_shift_range	0.3
vertical_flip	True
horizontal_flip	True
fill_mode	wrap

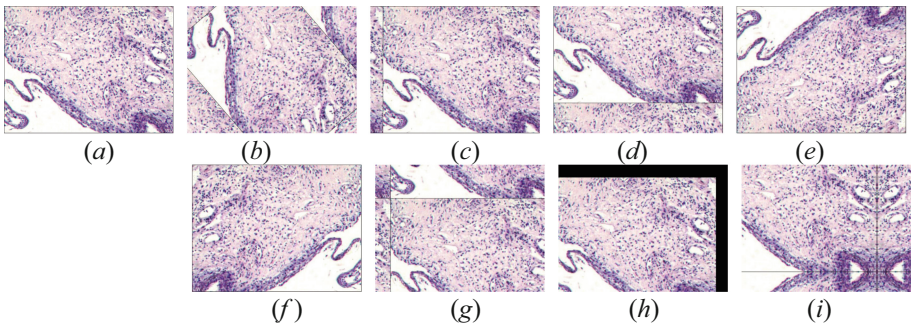


Fig. 2. Original image and images by data augmentation

4.3 BreakHis Breast Cancer Pathological Image Dataset

BreakHis breast cancer pathological image dataset is a common open data set, which is released by Spanhol et al. in 2016, containing 7909 breast tissue pathological images from 82 patients. The data set is divided into 5429 malignant tumor images, 2480 benign tumor images (as listed in Table 3).

Table 3. Distribution of images in BreakHis Dataset

Magnification Factors	Malignant (M)	Benign (B)	Total
40X	1370	625	1995
100X	1437	644	2081
200X	1390	623	2013
400X	1231	588	1820

5 Experiments and Results

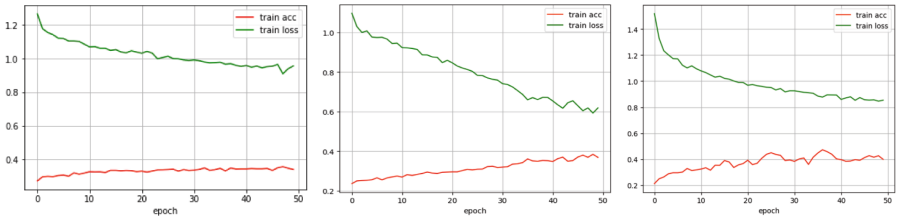
This experiment is run on the platform of a personal computer with Intel Core i5–8300 CPU, NVIDIA GTX 1060 with Max-Q Design, and 16 GB memory, under the operating system of Windows 10.

5.1 Experiment I: Model Selection

For the best classification results, we compared the classic models of deep neural networks, traditional AlexNet, Inception V3 [14] based on GoogleNet, and ResNet,

which have emerged in recent years and achieved excellent results in pathological image diagnosis. The performance of these three networks is preliminarily tested on the public data set BreaKHis breast cancer data set, which is also consisted of pathological slice images. We selected 1995 breast cancer images with magnification of 40 times

and 8 classifications, including 1195 training examples, 400 validation examples and 400 test examples. SGD is used as the optimizer, the learning rate is set to 0.001, the momentum is set to 0.5, the nesterov momentum method is enabled, the batch size is 32, and the number of iterations epochs is set to 50.



(*a*) AlexNet (*b*) InceptionV3 (*c*) ResNet50

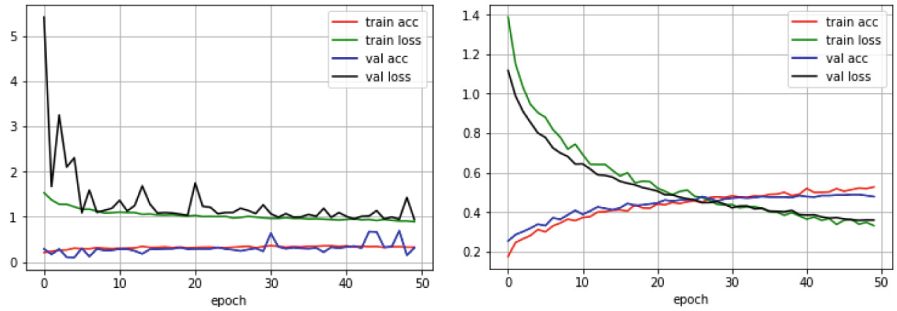
Fig. 3. Training results of AlexNet,InceptionV3 and ResNet50 on BreaKHis

Figure 3 shows the training results of AlexNet、InceptionV3 and ResNet50 (from left to right) on the breast cancer data set BreaKHis. It can be seen from the training curves of error and accuracy that AlexNet's performance in error and accuracy is far inferior to the other two networks. Although Inception V3 reduces the error to a lower level, its accuracy is not as good as ResNet50. Therefore, ResNet50 outperforms the other two networks in terms of training error and accuracy. For the requirement of accuracy, ResNet50 is selected as the main research network in this paper.

5.2 Experiment II: Transfer Learning

It takes a long time to train the network, especially the deep neural network with complex structure. On the selected ResNet50, we verify the validity of the classification performance of the network on esophageal cancer data with randomly initialized weights on the breast cancer images with magnification of 40 times. Figure 4(a) shows the training processing of ResNet50. The decline rate of training loss is very slow. After 50 batches, it only drops to about 0.85. The training accuracy rate (train acc) tends to remain unchanged in the later period, except for the obvious increase observed in the previous batches. After the training, the test error (loss) was 1.443, and the test accuracy was 0.496. The generalization ability could not meet the requirements of actual practice. Training time is up to about 40 min.

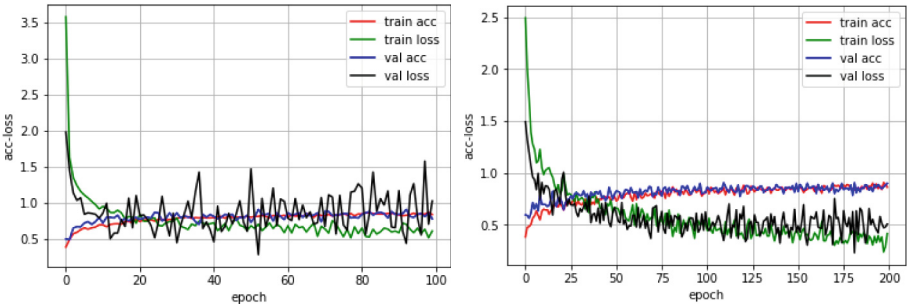
By comparison, we transfer the model of ResNet50 pre-trained on the dataset of ImageNet. On the same data set with the same hyperparameters, the model is re-trained and the results are shown in Fig. 4 (b). It can be seen that after using transfer learning, the decline speed of error and the increase speed of accuracy rate have been greatly improved. After the training, the test error is 0.808, and the test accuracy rate is 0.705, which is much better than the result of randomly initialized weights. When all the convolution layers are frozen, it takes an average of 12 s to train a batch, and the sharing time is about 10.5 min. Compared with the randomly initialization, transfer learning greatly reduces the training time. So more time is spent on fine-tuning the network parameters,



(*a*) Randomly Initialized (*b*) After Transfer Learning

Fig. 4. ResNet50 training with transfer learning

rather than training the network model, which greatly reduces the training difficulty and improves the training efficiency.



(*a*) 100 epochs (*b*) 200 epochs

Fig. 5. ResNet50 training on breast cancer data set

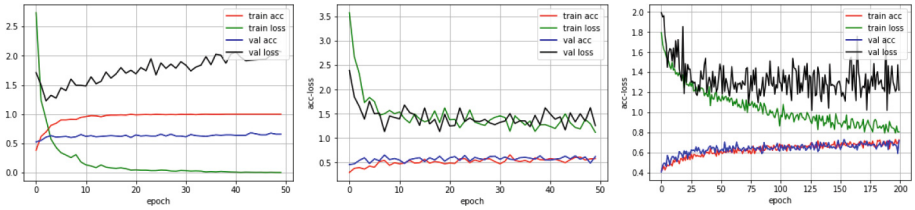
Table 4. Training parameters and accuracy on breast cancer data set

train_b_s	val_b_s	train_step	val_step	epochs	loss	accuracy
8	2	120	40	100	0.791	0.856
8	2	120	50	100	0.384	0.892
8	2	120	50	200	0.424	0.880
8	4	40	40	200	0.492	0.846

Likewise the model of ResNet50 is trained on the breast cancer data set first. The trained model is transferred to the training of the classification of esophageal cancer. Figure 5 shows the training results after 100 and 200 epochs. Although the error curve has a large range of oscillations, it does not affect the convergence of training. Table 4 lists the parameters and accuracy of random four running times.The average accuracy is higher than 0.85.

5.3 Experiment III: Data Augmentation

On the model of ResNet50 pre-trained on the breast cancer data set, esophageal cancer data is directly trained on the original images without data augmentation. Figure 6(a) shows the training results. From the training records, it can be observed that the network has a very serious over fitting phenomenon when training about 10 batches. The training accuracy rate (train acc) is infinitely close to 1, but the verification accuracy rate (val acc) is no longer improving. The training error (train loss) is approaching 0 with the growth of the training batches, and the verification error (val loss) starts to rise rapidly after a small decline. After the training, the test error is 2.486, and the test accuracy is 0.634. However, due to the large test error, the test accuracy is not convincing, and the generalization performance is poor. Figure 6 (b) and (c)show the training processing of the transferred network on the augment esophageal cancer dataset after 50 and 200 epochs separately. It is obvious that the over fitting of the model has been effectively suppressed after data enhancement, and the test results are basically consistent with the training case. The test error is 1.582, and the test accuracy is 0.560. However, it can be seen from the test results that the generalization ability of the model is far from enough, and the test accuracy of 0.56 cannot reach the standard that can be used. Although compared with 50 batches, the accuracy rate after 200 epochs is still too low. Over-fitting also occurred in the late stage.

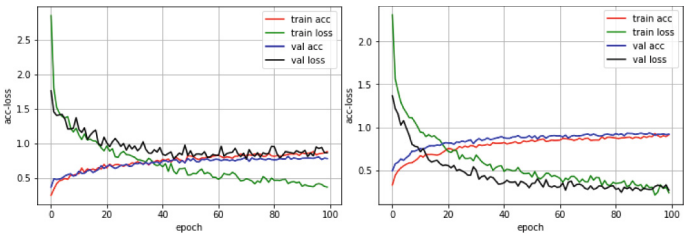


(*a*) original (*b*) 50 epochs (*c*) 200 epochs

Fig. 6. Training on the original and augment esophageal cancer datasets

The main reason for the phenomenon is that the amount of data is too small to meet the requirements of the deep network for large-scale training samples. Data augmentation is used to avoid it. By setting different scaling coefficients, the esophageal cancer dataset is augmented to 2420 (Data augment I) and 4840 (Data augment II).

As shown in Fig. 7 (a), the training results after data augment is greatly improved comparing with the originally un-augmented data set. The training error has dropped to a lower level, and the accuracy has also been greatly improved. From the performance of the verification set, it shows that the network is still over fitting, indicating that the distribution of the dataset is not very ideal. Figure 7 (b) over fitting is solved after large-scale data enhancement, and the result in the validation set basically conform to the training data. The test error is 0.189, and the test accuracy rate (accuracy) is 0.950 (as listed in Table 5). The testing results show that the performance of the network can be effectively optimized by data augmentation, and the classification accuracy can be greatly improved.



(*a*) Data augment I (*b*) Data augment II

Fig. 7. Training after data augment

Table 5. Training parameters and test loss with accuracy on esophageal cancer data set

train_b_s	val_b_s	train_step	val_step	epochs	loss	accuracy
8	2	120	40	100	0.337	0.901
8	2	120	50	100	0.218	0.922
8	2	120	50	200	0.203	0.945
8	4	40	40	200	0.189	0.950

6 Conclusion

Classification of esophageal cancer pathological images is mainly researched in this paper combining with deep learning. Comparing the results of 3 classical models: AlexNet, InceptionV3, and ResNet50, ResNet50 is chosen as the training model for the classification for outperforming performance. Transfer learning is introduced to reduce the duration of the training instead of randomly initialization. Rather than training on the original data, data augmentation is used to improve the accuracy and effectively avoid over fitting. The test error on verification set is 0.189, and the test accuracy is 0.950. The classification results verify the effectiveness of the network model in the augmented dataset. However, the dilemma of insufficient data still exists, and only eight classified models are slightly inadequate for practical application. In the future, we would continue collecting and enlarging the esophageal cancer pathological image data set to meet the demand for large amounts of data for deep training. Moreover, we would also make an attempt to design proper deep network structures or integrate multiple networks with good performances to achieve a better pathological detection of esophageal cancer.

References

1. Tsai, T.J., et al.: Intelligent identification of early esophageal cancer by band-selective hyperspectral imaging. Cancers **14**(17), 4292 (2022)
2. Wang, Z., et al.: Three feature streams based on a convolutional neural network for early esophageal cancer identification. Multimedia Tools and Applications, 1–18 (2022)
3. Niazi, M.K.K., et al.: Digital pathology and artificial intelligence. Lancet Oncol. **20**(5), 253–261 (2019)

4. Takeuchi, M., et al.: Performance of a deep learning-based identification system for esophageal cancer from CT images. Esophagus **18**, 612–620 (2021)
5. Zhu, H., et al.: An evolutionary attention-based network for medical image classification. Int. J. Neural Sys. 2350010 (2022)
6. Huang, L.M., et al.: Artificial intelligence technique in detection of early esophageal cancer. World J. Gastroenterol. **26**(39), 5959 (2020)
7. Bibault, J.E., et al.: Big data and machine learning in radiation oncology: state of the art and future prospects. Cancer Lett. **382**(1), 110–117 (2016)
8. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115 (2017)
9. Gulshan, V., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA **316**(22), 2402–2410 (2016)
10. Yoshida, H., et al.: Automated histological classification of whole-slide images of gastric biopsy specimens. Gastric Cancer **21**(2), 249–257 (2018)
11. Misawa, M., et al.: Accuracy of computer-aided diagnosis based on narrow-band imaging endocytoscopy for diagnosing colorectal lesions: comparison with experts. Int. J. Comput. Assist. Radiol. Surg. **12**(5), 757–766 (2017)
12. Yang, C.K., et al.: Deep convolutional neural network-based positron emission tomography analysis predicts esophageal cancer outcome. J. Clini. Medi. **8**(6), 844 (2019)
13. Van Riel, S., Van Der Sommen, F., Zinger, S., et al.: Automatic detection of early esophageal cancer with CNNS using transfer learning. In: 25th IEEE International Conference on Image Processing (ICIP), pp. 1383–1387 (2018)
14. Ren, Z., Zhang, Y., Wang, S.: A hybrid framework for lung cancer classification. Electronics (Basel) **11**(10), 1614 (2022)
15. Ren, Z., Zhang, Y., Wang, S.: LCDAE: data augmented ensemble framework for lung cancer classification. Technol. Cancer Res. Treat. **21**, 15330338221124372 (2022)
16. Horie, Y., Yoshio, T., Aoyama, K., et al.: Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. Gastrointest. Endosc. **89**(1), 25–32 (2019)
17. Kather, J.N., et al.: Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat. Med. **25**(7), 1054–1056 (2019)
18. Al-Haija, Q.A., Adebanjo, A.: Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network. In: 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1–7 (2020)
19. Liu, Y., Gadepalli, K., et al.: Detecting cancer metastases on gigapixel pathology images. arXiv preprint. 1703.02442 (2017)
20. LeCun, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
21. Krizhevsky, A., et al.: Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 1097–1105 (2012)
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint, pp. 1409–1556 (2014)
23. Szegedy, C., Liu, W., Jia, Y., et al.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9 (2015)
24. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)
25. Srivastava, R.K., et al.: Highway networks. arXiv preprint. 1505.00387 (2015)