



Different Machine Learning and Deep Learning Methods for the Classification of Colorectal Cancer Lymph Node Metastasis Images

Jin Li¹, Peng Wang¹, Yang Zhou^{1,2}, Hong Liang¹ and Kuan Luan^{1*}

¹ College of Intelligent System Science and Engineering, Harbin Engineering University, Harbin, China, ² Department of Radiology, Harbin Medical University Cancer Hospital, Harbin, China

OPEN ACCESS

Edited by:

Zhiwei Luo,
Kobe University, Japan

Reviewed by:

Chengzhen Xu,
Huaibei Normal University, China
Yanjun Wei,
University of Texas MD Anderson
Cancer Center, United States

*Correspondence:

Kuan Luan
luankuan@aliyun.com;
luankuan@hrbeu.edu.cn

Specialty section:

This article was submitted to
Bionics and Biomimetics,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 22 October 2020

Accepted: 14 December 2020

Published: 14 January 2021

Citation:

Li J, Wang P, Zhou Y, Liang H and
Luan K (2021) Different Machine
Learning and Deep Learning Methods
for the Classification of Colorectal
Cancer Lymph Node
Metastasis Images.
Front. Bioeng. Biotechnol. 8:620257.
doi: 10.3389/fbioe.2020.620257

The classification of colorectal cancer (CRC) lymph node metastasis (LNM) is a vital clinical issue related to recurrence and design of treatment plans. However, it remains unclear which method is effective in automatically classifying CRC LNM. Hence, this study compared the performance of existing classification methods, i.e., machine learning, deep learning, and deep transfer learning, to identify the most effective method. A total of 3,364 samples (1,646 positive and 1,718 negative) from Harbin Medical University Cancer Hospital were collected. All patches were manually segmented by experienced radiologists, and the image size was based on the lesion to be intercepted. Two classes of global features and one class of local features were extracted from the patches. These features were used in eight machine learning algorithms, while the other models used raw data. Experiment results showed that deep transfer learning was the most effective method with an accuracy of 0.7583 and an area under the curve of 0.7941. Furthermore, to improve the interpretability of the results from the deep learning and deep transfer learning models, the classification heat-map features were used, which displayed the region of feature extraction by superposing with raw data. The research findings are expected to promote the use of effective methods in CRC LNM detection and hence facilitate the design of proper treatment plans.

Keywords: colorectal cancer, lymph node, classification, transfer learning, deep learning

INTRODUCTION

Colorectal cancer (CRC) has a higher recurrence rate than all other cancers (Bray et al., 2018). CRC lymph node metastasis (LNM) is the root cause of CRC recurrence (Ding et al., 2019; Yang and Liu, 2020). CRC patients with LNM have a 5-year survival rate ranging from 50 to 68%, but those without LNM have a higher rate up to 95% (Ishihara et al., 2017; Zhou et al., 2017). Treatment of CRC is also influenced by the presence of LNM. The conventional treatment plan involves endoscopic resection, and surgical resection accompanied by LN dissection is necessary for patients with LNM (Nasu et al., 2013). Hence, it is important to determine the presence of CRC LNM, and to this end, an automatic classification method for CRC LNM should be explored to give a second objective opinion and then assist the radiologist in providing a correct report.

As computer technologies advance in these years, medical imaging classification methods have seen wider adoption. Many new methods based on machine learning (Al-Absi et al., 2012), deep learning (Sun et al., 2016), or deep transfer learning (Pratap and Kokil, 2019), have gradually been applied to medical imaging classification. These methods can provide additional preoperative information to aid radiologists in making proper treatment plans (Carneiro et al., 2017; Lu et al., 2017).

There are many types of machine learning algorithms that can be applied to medical imaging classification, such as support vector machine (SVM) (Burges, 1998), decision trees (DT) (Quinlan, 1986), and naïve Bayes (NB) (Friedman et al., 1997). Vibha et al. (2006) used DT to classify mammograms and obtained a classification accuracy of almost 90%. Inthajak et al. (2011) presented a k-nearest neighbor (KNN) algorithm to categorize medical images. Ahmad et al. (2016) used NB classification to categorize each chest X-ray image to either normal with infection or with fluid in capture features. García-Floriano et al. (2019) proposed a machine learning model based on SVM. This model could classify age-related macular degeneration in fundus images and achieved a higher classification accuracy than many well-regarded state-of-the-art methods. Luo et al. (2018) used SVM to classify human stomach cancer in optical coherence tomography images and obtained a higher classification accuracy than human detection. Although these methods outperformed human radiologists in terms of classification accuracy, they are subject to problems in feature extraction. The features that are often manually extracted fall short of objectivity and will affect the algorithms' performance and hence the classification accuracy. Thus, a method that could learn underlying data features is needed.

Deep learning (Lecun et al., 2015) has achieved stunning success in image classification in the ImageNet Large Scale Visual Recognition Challenge (Krizhevsky et al., 2017) in 2012. There are several landmark studies (Ma et al., 2017; Golatkar et al., 2018) that have promoted the development of deep learning algorithms. Compared to machine learning, deep learning has a vital advantage, i.e., the ability to automatically learn the potential features of data by utilizing the convolution neural network (CNN). A CNN could automatically learn potential features from raw data layer by layer with little or no hands-on intervention. Now, CNNs have already become a study focus, especially in medical imaging (Litjens et al., 2017; Shen et al., 2017). For instance, Song et al. (2017) used a CNN for lung nodule classification on computed tomography (CT). Liu and An (2017) built a classification model for detection of prostate cancer based on deep learning. Despite the good performance of deep learning methods in image classification, however, there are two problems that undermine their wider adoption – the need for massive number of data and high-performance computing devices, like graphic processing units (GPUs). As deep learning needs a large number of data to train and fit the CNN parameters, GPUs are preferred than other equipment to process the images faster, but it incurs a high training cost. Thus, reducing the training cost is the key to solving the problem.

Transfer learning, introduced by Pan and Yang (2010), uses the existing knowledge learned from one environment to solve

similar problems in different environments. Pan and Yang (2010) summarized the classification process as well as the pros and cons of transfer learning methods. Transfer learning methods have a lower training cost than their deep learning counterparts as the former does not require as many data as the latter needs for training. In transfer learning, a model pretrained on another large dataset, such as ImageNet, is employed to complete a task through fine-tuning¹ or other methods (Long et al., 2013, 2017; Tzeng et al., 2014), so it has a lower cost than deep learning. Because of these advantages, transfer learning is widely used in medical imaging. Vesal et al. (2018) used two pre-trained models to classify breast cancer histology images and obtained an accuracy of 97.50 and 91.25%. da Nobrega et al. (2018) used a deep transfer learning model (Tan et al., 2018) to classify lung nodules in CT lung images. Dornaika et al. (2019) used transfer learning to estimate age through facial images, and the experiments were carried out on three public databases.

In previous studies, comparisons were made to find the most effective method for specific diseases. Wang L. et al. (2017) used four methods with eight classification schemes to classify ophthalmic images and found that local binary pattern with SVM and wavelet transformation with SVM had the best classification performance, with an accuracy of 87%. Wang H. et al. (2017) used five methods, including random forest, SVM, adaptive boosting, back-propagation artificial neural network (ANN), and CNN to classify mediastinal LNM of non-small cell lung cancer. Lee et al. (2019) used eight deep learning algorithms to differentiate benign and malignant tumors from cervical metastatic LN of thyroid cancer based on preoperative CT images, trying to identify the most suitable model. However, to the authors' knowledge, there are few methods used for CRC LNM classification based on machine learning, deep learning, or deep transfer learning. Furthermore, no previous studies have compared these three methods. Therefore, which method is the most effective one for CRC LNM classification is unclear.

According to the literature, the use of machine learning, deep learning, or transfer learning in medical imaging is rapidly developing and evolving. Using computational approaches to provide additional preoperative information can help doctors design proper treatment plans (Carneiro et al., 2017; Lu et al., 2017). Several papers have been published on CRC (Simjanoska et al., 2013; Ciompi et al., 2017; Nakaya et al., 2017; Bychkov et al., 2018), but few studies have explored the performance of machine learning, deep learning, and transfer learning in CRC LNM classification, and which method has the best performance is yet to be determined, which motivated us to conduct this research. To identify the most effective method for CRC LNM classification, the following approaches were taken in this study.

First, eight machine learning, two deep learning, and one transfer learning classification methods were used to classify CRC LNM images.

Next, the classification results were compared to evaluate the performance of different methods and identify the one with the best classification performance.

¹Transfer learning and Fine-Tuning. Available online at: <https://ru.coursera.org/lecture/machine-learning-duke/transfer-learning-and-fine-tuning-OURo>

Finally, a classification heat-map was used to improve the interpretability of the classification method for CRC LNM.

MATERIALS AND METHODS

Data

Data used in the present study were collected from Harbin Medical University Cancer Hospital. There were 3,364 samples in the dataset, among which 1,646 were positive and 1,718 were negative. The standard of all samples was an LN of a diameter >3 mm. All patients underwent 3.0T magnetic resonance imaging (MRI) before surgery using Philips Achieva, with a 16-channel torso array coil. MRI sagittal T2WI scan sequence was performed with the following parameters: TR/TE 3000/100 ms, the number of signal frequency (NSA) 2, and the layer thickness 4.0 mm, and layer spacing 0.4 mm. The position of the rectal lesions was determined by the sagittal position, which was perpendicular to the intestinal canal lesions, with a transverse T2WI scan: TR/TE 3,824/110 ms, NSA 0–3, layer thickness 3.5 mm, and interval 0.2 mm. According to the sagittal lesion position, patients with parallel pathological changes received coronal T2WI scans: TR/TE 3,824/110 ms, NSA 3, and layer spacing 0.2 mm. Then, the objective LN in the sagittal, transverse, and coronal images were located. All images used in this study were marked as CRC LNs and classified as negative or positive by experienced radiologists. All patches were manually segmented by experienced radiologists, and the image size was based on the lesion to be intercepted. **Figure 1** presents the CRC LN.

Methods

In this study, intensity features were extracted by a gray-level histogram (GLH) (Otsu, 1979), and the textural features were extracted by the gray-level co-occurrence matrix (GLCM) (Haralick et al., 1973). These two features are global features. Furthermore, the scale-invariant feature transform (SIFT) (Lowe, 1999, 2004) operator was used for local features. These methods were implemented in Python.

Eight classical supervised machine learning methods, including AdaBoost (AB) (Freund and Schapire, 1997), DT (Quinlan, 1986), KNN (Cover and Hart, 1967), logistic regression (LR) (Cucchiara, 2012), NB (Friedman et al.,

1997), SVM (Burges, 1998), stochastic gradient descent (SGD) (Ratnayake et al., 2014), and multilayer perceptron (MLP), two deep learning models [LeNet (Lecun et al., 1998) and AlexNet (Krizhevsky et al., 2017)], and one transfer learning model (AlexNet pre-trained model) were studied for CRC LNM classification. The results of these methods were compared.

A detailed introduction to classical machine learning methods was presented in Marsland (2009). Classical machine learning methods were implemented using Python and the scikit-learn package. The optimal parameters of each method were searched via grid search in the parameter space and determined based on the four folds of training samples (Wang H. et al., 2017). AB used 100 decision stumps as weak learners, the learning rate was 0.1, and the maximum split number was equal to 1. DT used scikit-learn package default parameters. The KNN used 10 points of nearest neighbors. LR was a linear classification model, for which the parameters of the norm were l_2 , the optimization algorithm was selected as linear, and iteration was 100. The NB classifier selected Gaussian NB from the scikit-learn package. SVM used a radial basis function as the kernel function, the kernel coefficient was $1e-3$, and the penalty parameter was 1.0. SGD employed hinge as the loss function and l_2 as a penalty, and the iteration was 100. MLP, a form of ANN, was trained with a back-propagation algorithm (Rumelhart et al., 1985). In this method, two hidden layers with 32 and 16 neurons for the first and second layers were used. Epochs were 200, and the learning rate was $1e-3$.

Deep learning methods were implemented by Python and Keras library for Python. LeNet and AlexNet's structures were based on Lecun et al. (1998) and Krizhevsky et al. (2017). A relu activation function was used in the convolution layer, and a sigmoid function in the full connection layer. The optimization function was SGD (Bottou, 2010). The loss function was binary_crossentropy. The learning rate was $1e-3$, and epochs were 200. To avoid overfitting, L2 normalization (Van Laarhoven, 2017) and dropout regularization (Srivastava et al., 2014) were utilized. The methods were running on GPUs (NVIDIA Company GTX 1080ti).

Transfer learning and deep learning share something in common in parameter settings, but are different in training: deep learning starts from scratch, whereas transfer learning uses a pre-trained model. Because the pre-trained model was not trained by ImageNet, it did not need to initialize parameters. As stated in Yosinski et al. (2014), the first three layers were frozen because the layers that extracted the features were general. The other parameters were fine-tuned. In the implementation process, the structure must follow AlexNet then load AlexNet model weight. The fully connected layer activation function was changed to sigmoid. The optimization function was SGD. The loss function was binary_crossentropy, and the learning rate was $1e-4$, and epochs were 200. In this study, methods used to avoid overfitting were L2 normalization and dropout regularization. The dropout was set to 0.5. Finally, transfer learning retrained on GPU was like deep learning.

The explanation of the internal relationship between the input data and label prediction has always been a vital issue (Lipton, 2018) in the CNN-based classification tasks. In this study, a classification heat-map was used to improve the interpretability

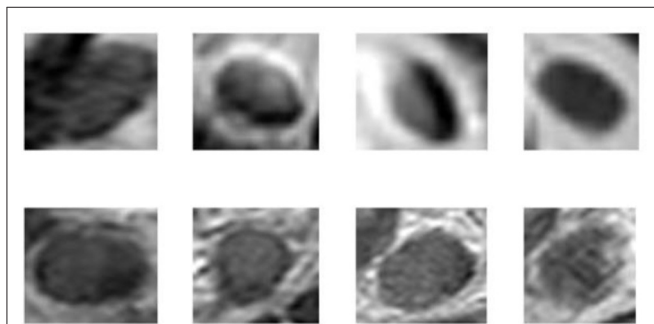


FIGURE 1 | Sample CRC LN images (top row, negative; bottom row, positive).

of the model (Lee et al., 2019). This experiment contained three steps. First, the model was used to display the last convolutional layer feature-map. Then, the feature-map was converted into a heat-map. Last, raw data and heat-map were superimposed into a new image.

RESULTS

Classification Performance

In this study, three kinds of features were employed in eight machine learning methods, respectively. **Table 1** displays the performance indicators of all methods for CRC LNM classification, including accuracy (ACC), area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Based on the ACC and AUC values, the optimal features set of each machine learning method for CRC LNM classification was defined. As shown in **Figures 2, 3**, LR achieved the best performance using the GLH and GLCM features sets. NB obtained the highest ACC and AUC values for the SIFT features set. Deep learning and transfer learning had significantly better performance than machine learning methods in CRC LNM classification. Although deep learning and transfer learning did not use global and local

features sets, the ACC and AUC values of these two methods were still better than machine learning methods (**Table 1**). Hence, transfer learning was identified as the best method for CRC LNM classification in this study.

Lesion Classification Heat-Map

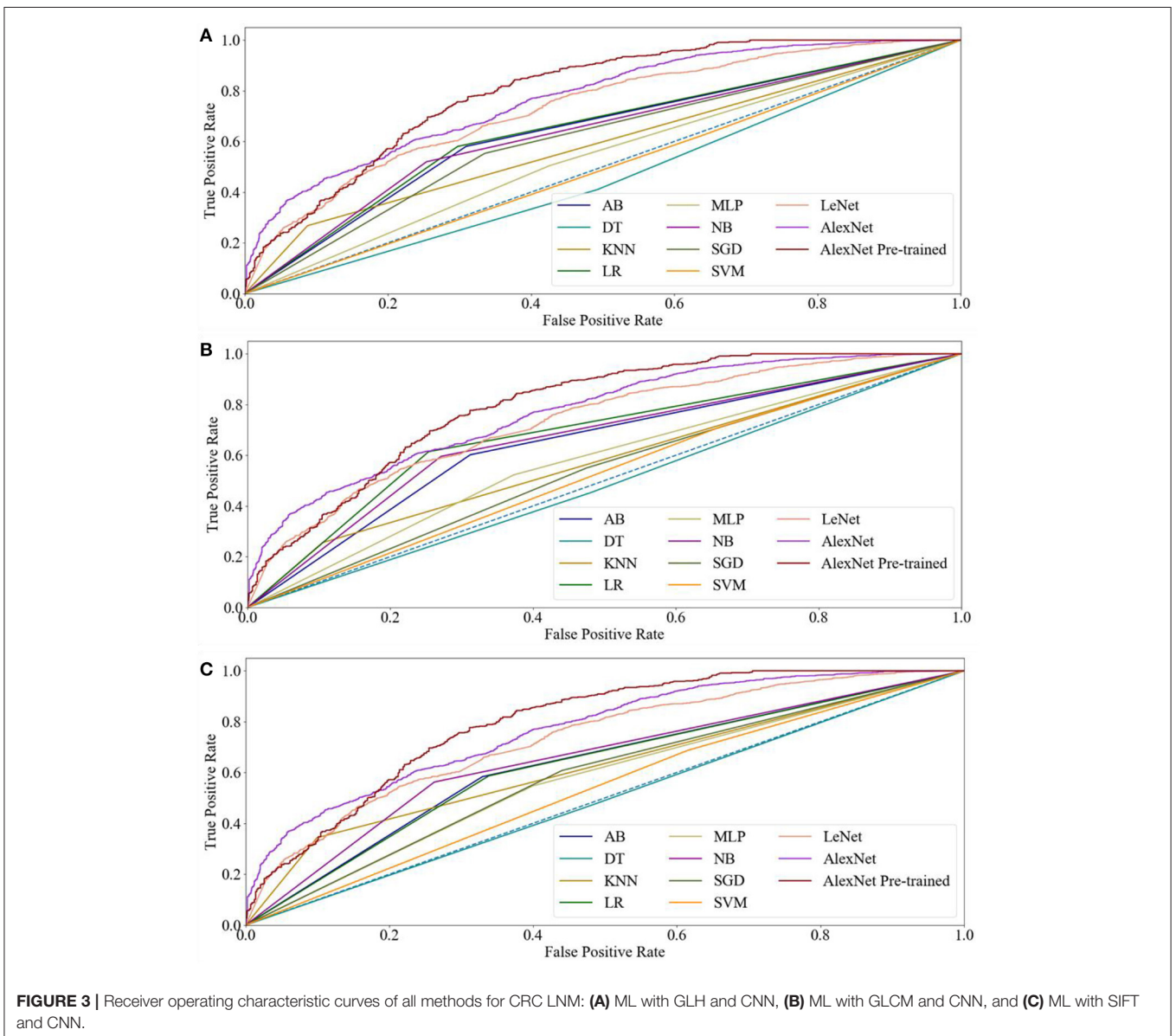
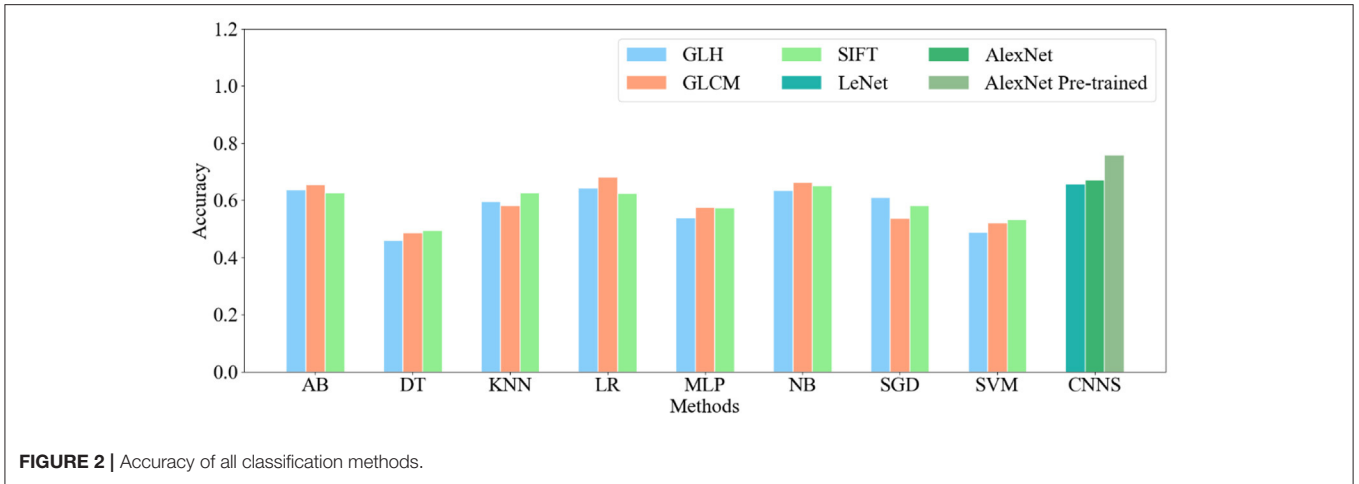
Although CNN has enabled unprecedented breakthroughs in computer vision tasks, interpretability remains unclear (Lipton, 2018). Therefore, to present a classification heat-map (Lee et al., 2019), the interpretability of the CNN model was improved. The classification heat-map identified discriminative regions (Selvaraju et al., 2017). As shown in **Figure 4**, the last convolution layer features heat-map was superimposed on the original MRI so that the location of the actual LN could be compared to the region highlighted by the model (Lee et al., 2019). Red regions represent class information, whereas the others correspond to class evidence.

DISCUSSION

First, as shown by the comparison of machine learning results, one method could yield different results if the selected features differed, and different methods would have different results

TABLE 1 | Performance results of all methods for CRC LNM.

	ACC	AUC	Sensitivity	Specificity	PPV	NPV
AB+GLH	0.6369	0.6357	0.6753	0.6431	0.6910	0.5805
AB+GLCM	0.6458	0.6449	0.6431	0.6491	0.6880	0.6018
AB+SIFT	0.6280	0.6270	0.6267	0.6295	0.6706	0.5836
DT+GLH	0.4598	0.4588	0.4728	0.4441	0.5073	0.4103
DT+GLCM	0.4866	0.4859	0.4972	0.4745	0.5190	0.4529
DT+SIFT	0.4955	0.4950	0.5057	0.4844	0.5190	0.4711
KNN+GLH	0.5967	0.5900	0.5650	0.7458	0.7125	0.2675
KNN+GLCM	0.5818	0.5752	0.5562	0.7	0.7051	0.2553
KNN+SIFT	0.6280	0.6222	0.5886	0.7687	0.7009	0.3435
LR+GLH	0.6429	0.6416	0.6359	0.6519	0.7026	0.5805
LR+GLCM	0.6815	0.6802	0.6684	0.6990	0.7464	0.6140
LR+SIFT	0.6250	0.6242	0.6253	0.6246	0.6618	0.5866
MLP+GLH	0.5402	0.5395	0.5472	0.5321	0.5743	0.5046
MLP+GLCM	0.5759	0.5748	0.5780	0.5733	0.6268	0.5228
MLP+SIFT	0.5744	0.5738	0.5803	0.5678	0.6006	0.5471
NB+GLH	0.6354	0.6331	0.6184	0.6628	0.7464	0.5198
NB+GLCM	0.6637	0.6623	0.6527	0.6782	0.7289	0.5957
NB+SIFT	0.6518	0.6500	0.6373	0.6727	0.7376	0.5623
SGD+GLH	0.6101	0.6089	0.6181	0.6128	0.6742	0.5532
SGD+GLCM	0.5372	0.5375	0.5488	0.5262	0.5248	0.5502
SGD+SIFT	0.5833	0.5838	0.5981	0.5698	0.5598	0.6079
SVM+GLH	0.4896	0.4924	0.5000	0.4836	0.3586	0.6261
SVM+GLCM	0.5208	0.5247	0.5500	0.5076	0.3382	0.7112
SVM+SIFT	0.5327	0.5359	0.5617	0.5172	0.3848	0.6869
LeNet	0.6577	0.7305	0.6535	0.6535	0.6535	0.6535
AlexNet	0.6716	0.7696	0.6708	0.6711	0.6714	0.6706
AlexNet pre-trained model	0.7583	0.7941	0.8004	0.7997	0.7992	0.8009



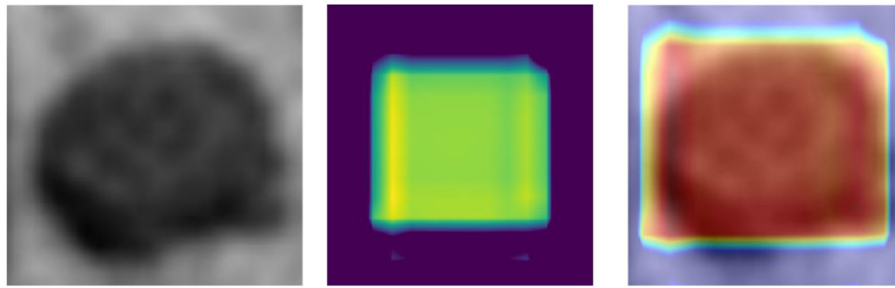


FIGURE 4 | CRC LN classification heat-map (left – original image; middle – feature heat-map; right – superimposed image).

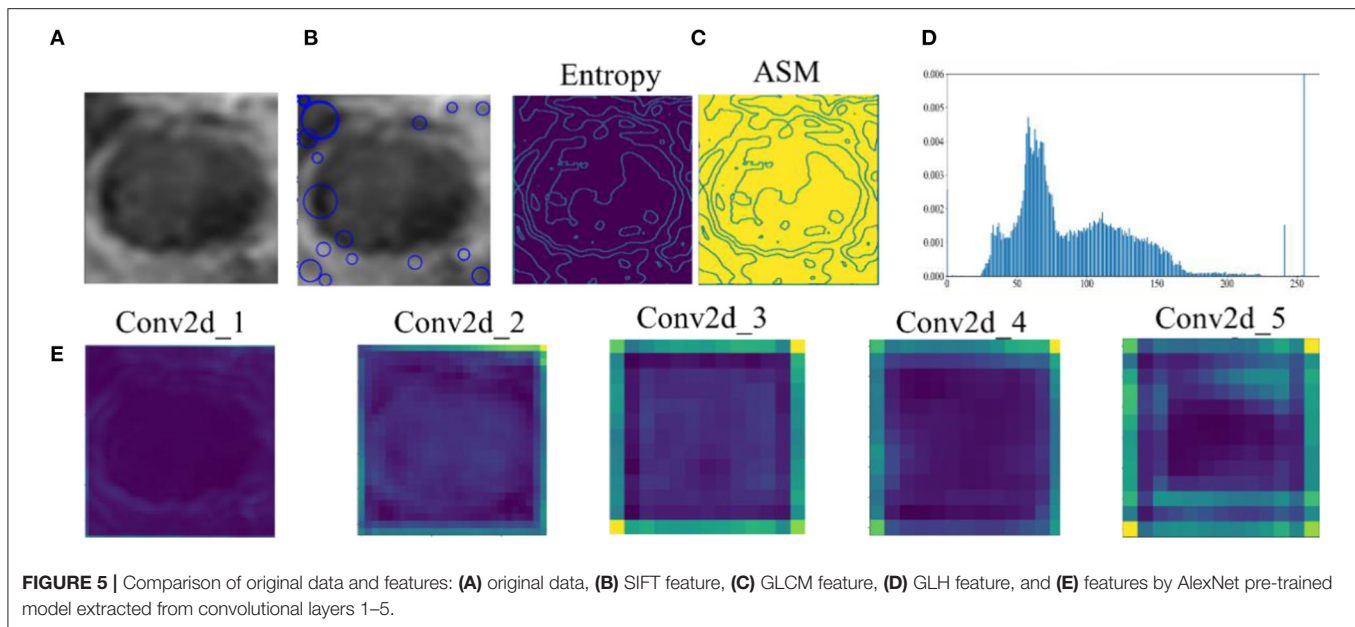
even if the same features were selected. In **Table 1**, when each feature set was used as an input, AB, LR, and NB were better than the other machine learning methods in terms of both ACC and AUC. AB is an ensemble of DT from the view of methodology. AB utilizes multiple weak classifiers which could learn more information from the input and develop into a strong classifier when combined. Hence, AB could become a better classification method based on these weak classifiers. As a linear classification method, LR gives each input feature a weight factor which will have impact on the classification result. Based on the LR methodology, gradient descent iterates were used to find the right factors. Then, predicted values were obtained, and a sigmoid function was used for final classification. NB, with the Bayesian theorem as its foundation, is considered one of the simplest yet most powerful classification methods. A NB classifier calculates the posterior probability of input features as per the prior probability, and the input features must be independent of each other. Other machine learning methods, such as SVM and DT, performed worse. For instance, SVM is based on the kernel function which implicitly maps features into a higher-dimensional feature space and measures the distance between the feature points. Hence, the choice of the kernel function is vital. DT is based on a tree structure, which contains nodes and a directed edge. In general, there is a root node, some internal nodes and leaf nodes in a DT. The leaf nodes are decision results, and other nodes represent the features. Therefore, the higher the feature purity is, the better the classification result is.

In this study, the comparison results revealed that all other methods employed outperformed the classical machine learning methods for CRC LNM classification. LeNet and AlexNet are deep learning methods, and the AlexNet pre-trained model is a transfer learning model. In deep learning, a CNN is used to extract features from raw data layer by layer. Comparison of LeNet and AlexNet showed that the latter had better performance than the former. The reason is the structure: a deep structure performs better than a shallow one (Bottou et al., 2007; Montufar et al., 2014). The number of parameters also played a part (Tang, 2015): there are more CNN layers in AlexNet which also has more extracted features than LeNet (Lecun et al., 1998; Simonyan and Zisserman, 2014). Hence, even though the same classification function was employed, AlexNet achieves a better result. Although the problem of parameters was considered, more

parameters meant more data to fit. Therefore, AlexNet needs more data, and is more likely to result in overfitting than LeNet. Overfitting hinders medical imaging analysis, and it is often difficult to collect enough medical images, like CRC LN images. In this study, the AlexNet pre-trained model solved the problem of shortage of medical data. As the AlexNet pre-trained model was trained by ImageNet, it did not need extra data for training from scratch. Hence, the pre-trained model parameters could be directly used to train new data with the help of some processing techniques, such as freezing and fine-tuning parameters. As low-level features are general, the parameters of these features could be frozen, and other parameters could be fine-tuned. In transfer learning, the learning rate is often set smaller than that for training from scratch. If the learning rate is set high, the model's parameters would be updated quickly and affect the original weight information (Wang, 2018).

As shown in **Figure 4**, the visualization experiment could show the model focus region of the input image. The classification heat-map represents evidence of the CNN model-based classification and could assist in clinical decision-making by directly identifying the region of interest (Lee et al., 2019).

Feature extraction has direct impacts on the performance of the classification method. **Figure 5** shows the original data and the features extracted by all methods. SIFT are features extracted based on the interest points of the local appearance of the original data. The more the points are, the more the features are. However, there are few points on the CRC LN lesion, whereas some points are on the edge. There is little classification information. The GLCM includes multiple-type features. Entropy and angular second moment features are listed in **Figure 5**. It is not easy to distinguish the region of the lesion. GLH represents the relationship between the occurrence frequency of each gray-level pixel and the gray level. Gray-level pixels could be observed. The central area of the CRC LN lesion is more than the edge area. Nevertheless, GLH does not reflect the features of the central area. The features of the AlexNet pre-trained model are from low to high levels. The features of the first two convolutional layers display the overall contour of the lesion, and others represent the semantic features of a high level. This is helpful for classification. Hence, the features extracted by the pre-trained CNN were better than those extracted by the other methods for CRC LNM classification.



Based on the experiment's results, the observations are as follows: (1) The traditional feature extraction methods are not effective in CRC LNM classification. (2) The pre-trained model of deep learning has strong transferability. Deep transfer learning applied to a small medical image dataset is better than traditional methods, and it does not need to train the model from scratch. (3) The weights of the pre-trained model realize better initialization of the parameters.

CONCLUSION

In conclusion, this study showed that deep transfer learning is better than deep learning and machine learning mainly because the pre-trained CNN extracts features are more discriminative than those extracted by a CNN and artificially-extracted features. Deep transfer learning has been a popular method for image classification in recent years, and it was proved to be the best classification method among all the methods selected in this study. It could extract underlying features from raw data, does not need to select features or use raw data as input, and is less prone to user bias. A pipeline of transfer learning will

REFERENCES

- Ahmad, W. S. H. M. W., Zaki, W. M. D. W., Fauzi, M. F. A., and Tan, W. H. (2016). "Classification of infection and fluid regions in chest x-ray images," in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)* (Gold Coast, QLD: IEEE), 1–5.
- Al-Absi, H. R. H., Samir, B. B., Shaban, K. B., et al. (2012). "Computer aided diagnosis system based on machine learning techniques for lung cancer," in *2012 international conference on computer & information science (ICIS)*, Vol. 1 (Kuala Lumpur: IEEE), 295–300.

be established in future studies, and the optimal deep transfer learning model for CRC LNM classification will be found.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

The project was partially supported by the National Natural Science Foundation of China (Grant Nos. 61773134 and 61803117), the Fundamental Research Funds for the Central Universities (Grant No. HEUCFG201824), and the National Key Scientific Instrument and Equipment Development Projects of China (Grant No. 2012YQ04014010).

- Bottou, L. (2010). "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010* (Physica-Verlag, HD), 177–186.
- Bottou, L., Chapelle, O., Decoste, D., Weston, J. (2007). "Scaling learning algorithms toward AI," in *Large-Scale Kernel Machines* (MIT Press). p. 321–359.
- Bray, F. I., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* 2, 121–167. doi: 10.1023/A:1009715923555
- Bychkov, D., Linder, N., Turkki, R., Nordling, S., Kovanen, P. E., Verrill, C., et al. (2018). Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci. Rep.* 8:3395. doi: 10.1038/s41598-018-21758-3
- Carneiro, G., Zheng, Y., Xing, F., and Yang, L. (2017). “Review of deep learning methods in mammography, cardiovascular, and microscopy image analysis,” in *Deep Learning and Convolutional Neural Networks for Medical Image Computing* (Cham: Springer), 11–32.
- Ciampi, F., Geessink, O., Bejnordi, B. E., de Souza, G. S., Baidoshvili, A., Litjens, G., et al. (2017). “The importance of stain normalization in colorectal tissue classification with convolutional networks,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (Melbourne, VIC: IEEE), 160–163.
- Cover, T. M., and Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory* 13, 21–27. doi: 10.1109/TIT.1967.1053964
- Cucchiaro, A. (2012). Applied logistic regression. *Technometrics* 34, 358–359. doi: 10.2307/1270048
- da Nóbrega, R. V. M., Peixoto, S. A., da Silva, S. P. P., and Filho, P. P. R. (2018). “Lung nodule classification via deep transfer learning in CT lung images,” in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)* (Karlstad: IEEE), 244–249.
- Ding, L., Liu, G. W., Zhao, B. C., Zhou, Y. P., Li, S., Zhang, Z. D., et al. (2019). Artificial intelligence system of faster region-based convolutional neural network surpassing senior radiologists in evaluation of metastatic lymph nodes of rectal cancer. *Chin. Med. J.* 132, 379–387. doi: 10.1097/CM9.0000000000000095
- Dornaika, F., Argandarreras, I., and Belver, C. (2019). Age estimation in facial images through transfer learning. *Mach. Vis. Appl.* 30, 177–187. doi: 10.1007/s00138-018-0976-1
- Freund, Y., and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning, and an application to boosting. *Conf. Learn. Theory* 55, 119–139. doi: 10.1006/jcss.1997.1504
- Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. *Mach. Learn.* 29, 131–163. doi: 10.1023/A:1007465528199
- García-Florián, A., Ferreira-Santiago, Á., Camacho-Nieto, O., Yáñez-Márquez, C. (2019). A machine learning approach to medical image classification: Detecting age-related macular degeneration in fundus images. *Comp. Electr. Eng.* 75, 218–229. doi: 10.1016/j.compeleceng.2017.11.008
- Golatkar, A., Anand, D., and Sethi, A. (2018). “Classification of breast cancer histology using deep learning,” in *International Conference Image Analysis and Recognition*. (Póvoa de Varzim; Cham: Springer), 837–844.
- Haralick, R. M., Shanmugam, K. S., and Dinstein, I. (1973). Textural features for image classification. *Syst. Man Cybernet.* 3, 610–621. doi: 10.1109/TSMC.1973.4309314
- Inthajak, K., Duanggate, C., Uyyanonvara, B., Makhnov, S. S., and Barman, S. (2011). “Medical image blob detection with feature stability and KNN classification,” in *2011 Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE)* (Nakhon Pathom: IEEE), 128–131.
- Ishihara, S., Kawai, K., Tanaka, T., Kiyomatsu, T., Hata, K., Nozawa, H., et al. (2017). Oncological outcomes of lateral pelvic lymph node metastasis in rectal cancer treated with preoperative chemoradiotherapy. *Dis. Colon Rectum.* 60, 469–476. doi: 10.1097/DCR.0000000000000752
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi: 10.1145/3065386
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521:436. doi: 10.1038/nature14539
- Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791
- Lee, J., Ha, E. J., and Kim, J. H. (2019). Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT. *Eur. Radiol.* 29, 5452–5457. doi: 10.1007/s00330-019-06098-8
- Lipton, Z. C. (2018). The mythos of model interpretability. *ACM Queue* 16:30. doi: 10.1145/3236386.3241340
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciampi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88. doi: 10.1016/j.media.2017.07.005
- Liu, Y., and An, X. (2017). “A classification model for the prostate cancer based on deep learning,” in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, (Shanghai: IEEE), 1–6.
- Long, M., Wang, J., Ding, G., Sun, J., and Yu, P. S. (2013). “Transfer feature learning with joint distribution adaptation,” in *Proceedings of the IEEE International Conference on Computer Vision*, (Sydney, NSW), 2200–2207. doi: 10.1109/ICCV.2013.274
- Long, M., Zhu, H., Wang, J., and Jordan, M. I. (2017). “Deep transfer learning with joint adaptation networks,” in *International Conference on Machine Learning* (Sydney, NSW: PMLR), 2208–2217.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Int. Conf. Comput. Vis.* 2, 1150–1157. doi: 10.1109/ICCV.1999.790410
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 91–110. doi: 10.1023/B:VISI.0000029664.99615.94
- Lu, L., Zheng, Y., Carneiro, G., and Yang, L. (2017). “Deep learning and convolutional neural networks for medical image computing,” in *Advances in Computer Vision and Pattern Recognition* (Cham: Springer International Publishing). doi: 10.1007/978-3-319-42999-1
- Luo, S., Yingwei, F., Liu, H., An, X., Xie, H., Li, P., et al. (2018). “SVM based automatic classification of human stomach cancer with optical coherence tomography images,” in *Conference on Lasers and Electro-Optics* (San Jose, CA). doi: 10.1364/CLEO_AT.2018.JTu2A.99
- Ma, L., Lu, G., Wang, D., et al. (2017). “Deep learning based classification for head and neck cancer detection with hyperspectral imaging in an animal model,” in *Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging* (Orlando, FL: International Society for Optics and Photonics), 101372G.
- Marsland, S. (2009). *Machine Learning: An Algorithmic Perspective*. New York, NY: Chapman and Hall/CRC.
- Montufar, G. F., Pascanu, R., Cho, K., and Bengio, Y. (2014). On the number of linear regions of deep neural networks. *Adv. Neural Inf. Proc. Syst.* 27, 2924–2932. Available online at: <https://arxiv.org/abs/1402.1869v2>
- Nakaya, D., Endo, S., Satori, S., Yoshida, T., Saegusa, M., Ito, T., et al. (2017). Machine learning classification of colorectal cancer using hyperspectral images. *J. Colo. Assoc. Jpn.* 41,99–101. Available online at: https://www.jstage.jst.go.jp/article/jcsaj/41/3+/41_99/_article/-char/en
- Nasu, T., Oku, Y., Takifujii, K., Hotta, T., Yokoyama, S., Matsuda, K., et al. (2013). Predicting lymph node metastasis in early colorectal cancer using the CITED1 expression. *J. Surg. Res.* 185, 136–142. doi: 10.1016/j.jss.2013.05.041
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybernet.* 9, 62–66. doi: 10.1109/TSMC.1979.4310076
- Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191
- Pratap, T., and Kokil, P. (2019). Computer-aided diagnosis of cataract using deep transfer learning. *Biomed. Signal Process. Control.* 53:101533. doi: 10.1016/j.bspc.2019.04.010
- Quinlan, J. R. (1986). Induction of decision trees. *Mach. Learn.* 1, 81–106. doi: 10.1007/BF00116251
- Ratnayake, R. M. V. S., Perera, K. G. U., Wickramanayaka, G. S. K., Gunasekara, C. S., and Samarawickrama, K. C. J. K. (2014). Application of stochastic gradient descent algorithm in evaluating the performance contribution of employees. *IOSR J. Bus. Manage.* 16, 77–80. doi: 10.9790/487X-16637780
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1985). “Learning Internal Representations by Error Propagation,” in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, MIT Press, 1987, p. 318–362.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). “Grad-CAM: visual explanations from deep networks via gradient-based localization,” in *International Conference on Computer Vision* (Venice), 618–626.
- Shen, D., Wu, G., and Suk, H. I. (2017). Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 19, 221–248. doi: 10.1146/annurev-bioeng-071516-044442

- Simjanoska, M., Bogdanova, A. M., and Popeska, Z. (2013). "Bayesian multiclass classification of gene expression colorectal cancer stages," in *International Conference on ICT Innovations* (Heidelberg: Springer), 177–186.
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv [preprint]*. arXiv:1409.1556.
- Song, Q., Zhao, L., Luo, X., and Dou, X. (2017). Using deep learning for classification of lung nodules on computed tomography images. *J. Healthc. Eng.* 2017:831470. doi: 10.1155/2017/8314740
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.
- Sun, W., Zheng, B., and Wei, Q. (2016). Computer aided lung cancer diagnosis with deep learning algorithms. *Med. Imaging Comput. Aided Diagn.* 9785:97850z. doi: 10.1117/12.2216307
- Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). "A survey on deep transfer learning," in *International Conference on Artificial Neural Networks*. Eds V. Kurková, Y. Manolopoulos, B. Hammer, L. Iliadis and I. Maglogiannis (Cham: Springer), 270–279. doi: 10.1007/978-3-030-01424-7_27
- Tang, X. Y. X. H. (2015). The influence of the amount of parameters in different layers on the performance of deep learning models. *Comput. Sci. Appl.* 5, 445–453. doi: 10.12677/CSA.2015.512056
- Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., and Darrell, T. (2014). Deep domain confusion: maximizing for domain invariance. *arXiv [preprint]*. arXiv:1412.3474.
- Van Laarhoven, T. (2017). L2 regularization versus batch and weight normalization. *arXiv [preprint]*. arXiv:1706.05350.
- Vesal, S., Ravikumar, N., Davari, A., Ellmann, S., and Maier, A. (2018). Classification of breast cancer histology images using transfer learning. *arXiv* 1802.09424v1. doi: 10.1007/978-3-319-93000-8_92
- Vibha, L., Harshavardhan, G. M., Pranaw, K., Deepa Shenoy, P., Venugopal, K. R., and Patnaik, L. M. (2006). "Classification of mammograms using decision trees," in *2006 10th International Database Engineering and Applications Symposium (IDEAS'06)* (Delhi: IEEE), 263–266. doi: 10.1109/IDEAS.2006.14
- Wang, H., Zhou, Z., Li, Y., Chen, Z., Lu, P., Wang, W., et al. (2017). Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET/CT images. *EJNMMI Res.* 7:11. doi: 10.1186/s13550-017-0260-9
- Wang, J., et al. (2018). *Transfer Learning Tutorial*. Available online at: <https://github.com/jindongwang/transferlearning-tutorial>
- Wang, L., Zhang, K., Liu, X., Long, E., Jiang, J., An, Y., et al. (2017). Comparative analysis of image classification methods for automatic diagnosis of ophthalmic images. *Sci. Rep.* 7:41545. doi: 10.1038/srep41545
- Yang, Z., and Liu, Z. (2020). The efficacy of 18F-FDG PET/CT-based diagnostic model in the diagnosis of colorectal cancer regional lymph node metastasis. *Saudi J. Biol. Sci.* 27, 805–811. doi: 10.1016/j.sjbs.2019.12.017
- Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems*, 3320–3328.
- Zhou, L., Wang, J. Z., Wang, J. T., Wu, Y. J., Chen, H., Wang, W. B., et al. (2017). Correlation analysis of MR/CT on colorectal cancer lymph node metastasis characteristics and prognosis. *Eur. Rev. Med. Pharmacol. Sci.* 21, 1219–1225.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Li, Wang, Zhou, Liang and Luan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.