Automated Lung and Colon Cancer Classification Using Histopathological Images

Biomedical Engineering and Computational Biology
Volume 15: 1–8

DOI: 10.1177/11795972241271569

Jie Ji[1*], Jirui Li[2*], Weifeng Zhang[2*], Yiqun Geng[2], Yuejiao Dong[3], Jiexiong Huang[3] and Liangli Hong[3]

[1]Network & Information Center, Shantou University, Shantou, Guangdong, China. [2]Guangdong Provincial International Collaborative Center of Molecular Medicine, Laboratory of Molecular Pathology, Shantou University Medical College, Shantou, China. [3]Department of Pathology, the First Affiliated Hospital of Shantou University Medical College, Shantou, China.

ABSTRACT: Cancer is the leading cause of mortality in the world. And among all cancers lung and colon cancers are 2 of the most common causes of death and morbidity. The aim of this study was to develop an automated lung and colon cancer classification system using histopathological images. An automated lung and colon classification system was developed using histopathological images from the LC25000 dataset. The algorithm development included data splitting, deep neural network model selection, on the fly image augmentation, training and validation. The core of the algorithm was a Swin Transform V2 model, and 5-fold cross validation was used to evaluate model performance. The model performance was evaluated using Accuracy, Kappa, confusion matrix, precision, recall, and F1. Extensive experiments were conducted to compare the performances of different neural networks including both mainstream convolutional neural networks and vision transformers. The Swin Transform V2 model achieved a 1 (100%) on all metrics, which is the first single model to obtain perfect results on this dataset. The Swin Transformer V2 model has the potential to be used to assist pathologists in classifying lung and colon cancers using histopathology images.

KEYWORDS: Lung cancer classification, colon cancer classification, vision transformer, Swin Transformer V2, histopathological images

RECEIVED: November 16, 2023. **ACCEPTED:** July 1, 2024.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by Li Kashing Foundation Cross-Disciplinary Research Program under grant No. 2020LKSFG12B.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Liangli Hong, Department of Pathology, the First Affiliated Hospital of Shantou University Medical College, No.57 Changping Road, Longhu District, Shantou City, Shantou, Guangdong 515041, China. Email: hong_liangli@163.com

Introduction

According to the world health organization (WHO), cancer is the leading cause of mortality in the world. Lung cancer is responsible for 18.4% of cancer-related deaths and 11.6% of all cancer cases. Colon cancer accounts for 9.2% of all cancer-related fatalities globally.[1-5] Adenocarcinoma and squamous cell carcinoma are the most commonly occurring subtypes of lung cancer, and adenocarcinoma is the most common colorectal cancer subtype. Cancer detection at an early stage can significantly reduce the fatality rate. Furthermore, cancer subtype is an important factor for diagnosis and especially treatment plan determination. Assessment of histopathological images by a pathologist is the gold standard for lung and colon cancer diagnosis.[6] The inter-observer variability of lung and colon classification is moderate.[7] Moreover, the number of qualified pathologists is too small to meet the substantial clinical demands, particularly in countries such as China, with a significant population of lung and colon cancer patients.

Artificial intelligence, particularly deep learning algorithm may help pathologists to reduce errors and improve efficiency. Recently, there has been a rise in research interest in automated deep-learning-based lung and colon cancer diagnosis. Most successful studies used histopathology slide images to aid in automated diagnosis.[5] Many of these research used the publicly available LC25000 dataset.[8] Some research only conducted 1 cancer type classification, that is, lung or colon cancer classification.[4] Masud et al designed a small convolutional neural network for lung and colon cancer classification and achieved an accuracy of 96.33%.[9] Mumtaz Ali adopted a multi-input dual-stream capsule network and obtained an accuracy of 99.58%. Tummala et al[10] used an EffcientNetV2 model and obtained an accuracy of 99.97%. Al-Jabbar et al[11] provided a hybrid system with the fusion features of VGG-19 and handcrafted features. The classifier reached a sensitivity of 99.85%, a precision of 100%, an accuracy of 99.64%, a specificity of 100%, and an AUC of 99.86%. A recent study achieved perfect results on the LC25000 dataset.[12] However, in this study the LC25000 was split into 70% training set and 30% testing set only 1 time. Because the test dataset was used for hyper-parameter turning and model selection, there existed a possibility of data leakage. Furthermore, this study adopted a complex framework including 4 CNNs, specifically VGG16, ResNet50, InceptionV3, and DenseNet121, and SVM as the meta-learner conducting model ensemble. The framework cannot be trained end-to-end and the whole classifier consume more computing resource than the simple single model.

Kumar et al[13] compared the performances of 2 different 2 different kinds of feature extractors that include 6 handcrafted features extraction techniques and 7 deep neural networks with 4 different machine learning classifiers and found out that deep neural networks features worked much better. And the best result was an accuracy and recall of 98.60%,

* These authors contributed equally to this work.

precision of 98.63%, F1 score of 0.985 and ROC-AUC of 1; Mehmood et al and his group[14] adopted a pretrained neural network (AlexNet) to do LC25000 classification. By adopted Class Selective Image Processing (CSIP), the overall accuracy raised from 89% to 98.8%. The attractive point of this paper is CSIP. Except for CSIP, they used a relatively old network and the performance metrics was lower than those of other studies; Chhillar and Singh[15] proposed a traditional method using handcrafted features plus machine learning, specifically LightGBM with combined features, and obtained an accuracy of 100%. Even though they obtained perfect results on the LC25000 dataset, their handcrafted features are specific to the task, so that their method is difficult to expand to other tasks; Provath et al and his research team[16] design a novel neural network using global context attention module. They showed that the addition of the global context attention module decreases the model's parameter, reduces the computational costs and boost performances. They conducted experiments on 3 datasets LC25000, GLaS, and CRAG dataset. Even though they obtained great results of precision 99.4, sensitivity 99.6, and accuracy: 99.76 on LC25000 dataset, the results are not perfect; Dabass et al.[17] proposed a novel method to do colon histopathological images classification using multiple datasets. Their method includes a stain-invariant pre-processing procedure and a well-designed neural network, which contains enhanced convolutional learning modules, multi-level attention learning module, and transitional modules. Even though they achieved perfect results on the LC25000 dataset, 2-class classification (colon benign and colon malignant classes) instead of 5 class classification. They also[18] invented a novel multi-tasking U-net with hybrid convolutional learning and attention modules to do cancer classification on multiple datasets and achieved an accuracy of 0.9997, recall of 0.9994, precision of 1 on the LC25000 dataset. The novelty of this study is using segmentation models do both segmentation and classification tasks. Just like the previous study, on the LC25000 dataset they do 2-class classification (colon benign and colon malignant classes) instead of 5 class classification.

The field of computer vision has for years been dominated by CNNs (convolutional neural networks).[19-21] However, ViTs (Vision Transformers)[22-26] have recently outperformed CNNs for image classification and some other tasks. Nevertheless, ViTs are hardly successfully used in medical image analysis,[27,28] let alone cancer classification. It probably because the original ViT do not inherently encode inductive biases (prior knowledge) to deal with visual data,[29] they typically require a large amount of training data to figure out the underlying modality-specific rules.[30] By combining self-attention with a hierarchical structure that operates locally at different scales, Swin Transformer[23] have built locality, translational equivariance, and hierarchical scale into ViTs. This will reduce the sample size needed. Swin Transformer models embrace both inductive bias just like CNNs and self-attention to model long range interactions just like ViTs. We hypothesize that in some cases Swin Transformer models outperform both ordinary CNNs and ViTs. Moreover, compared with the original Swin Transformer, Swin Transformer not only has the advantage of scaling up capacity and resolution, which is important for object detection and segmentation tasks, but also can achieve comparable or even better results on classification tasks with ordinary image size.

The aim of this study was to develop an automatic lung and colon cancer classification model using histopathological images. In this study, the publicly available LC25000 dataset[8] was used to develop and validate the algorithm. The algorithm development included data splitting, neural network structure selection, on the fly image augmentation, training and validation.

Materials and Methods

System pipeline

The system pipeline is shown in Figure 1. Given a histopathological image patch cropped from a whole slide image, it was predicted as 1 of 5 classes, that is, lung adenocarcinoma, lung squamous cell carcinoma, benign lung, colon adenocarcinoma and benign colon.

Dataset

The LC25000 dataset,[8] which contains histopathological images of the lung and colon, was used to develop and evaluate the lung and colon cancer classification algorithm. The dataset was organized into 5 classes: lung adenocarcinoma, lung squamous cell carcinoma, benign lung, colon adenocarcinoma and benign colon. HIPAA compliant and validated 750 images of lung tissue (250 benign lung tissues, 250 lung adenocarcinomas, and 250 lung squamous cell carcinomas) and 500 total images of colon tissue (250 benign colon tissues and 250 colon adenocarcinomas) were captured from pathology glass slides.[31] After processing, there were 5000 images for each class in the dataset, which encompasses 25 000 lung and colon images with 768 pixels × 768 pixels.

Data splitting

To avoid the randomness of performance indicators caused by data splitting, a 5-fold cross validation was used to split the dataset. The data splitting process is illustrated in Figure 2. The dataset was split into a training, validation, and testing dataset with a ratio of 60%, 20%, and 20% for 5 times. Every test dataset did not overlap, so did the validation dataset. The training dataset was used to train model parameters, and the validation dataset was used to tune hyper-parameters and select models. The final performance results were obtained by combining results of 5 testing sub-datasets.

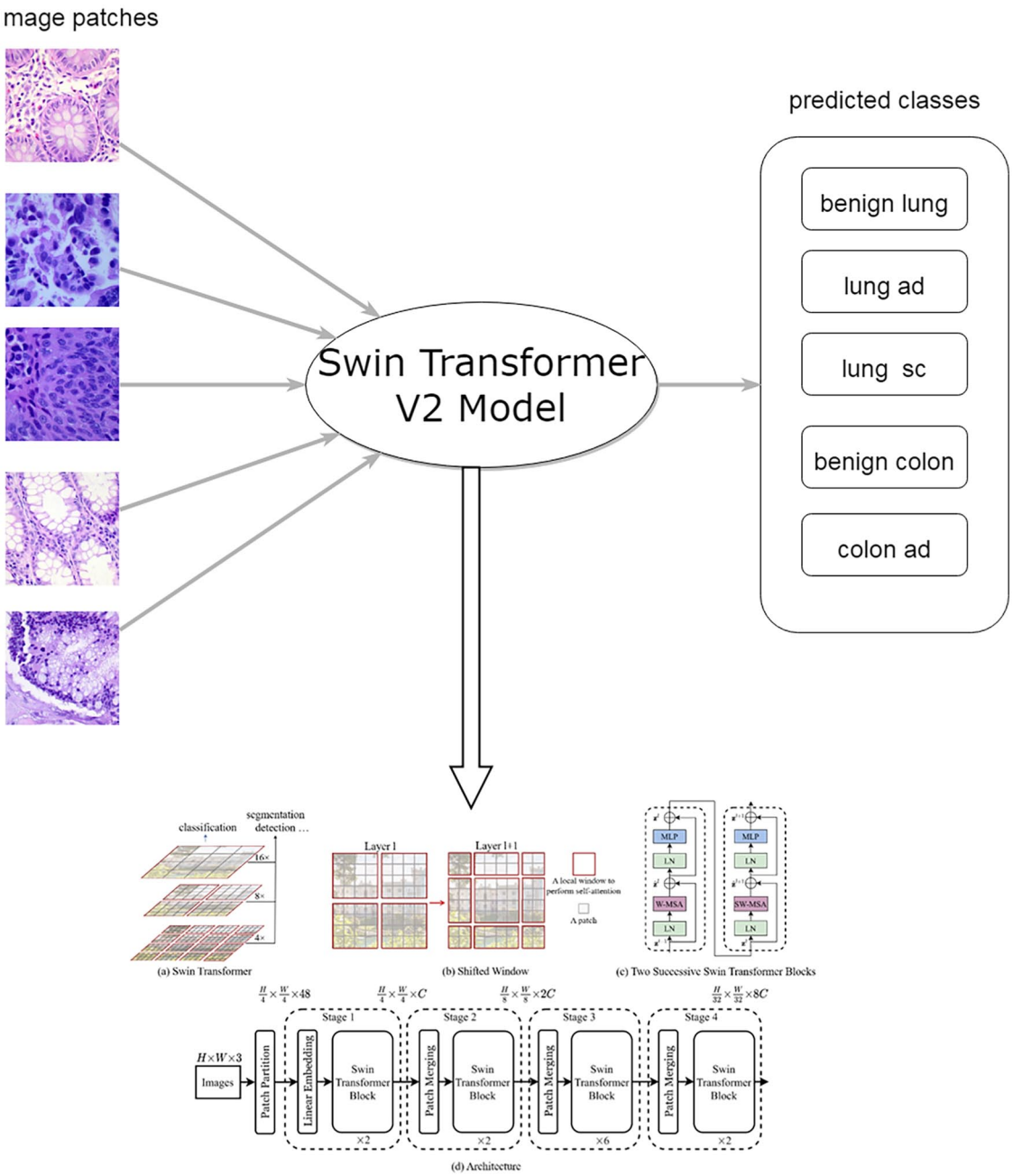

Figure 1. The flowchart of automated lung and colon cancer classification system. The image of Swin Transformer V2 was adopted from https://github.com/microsoft/Swin-Transformer.

Neural networks

The field of computer vision including medical image analysis has for years been dominated by CNNs. However, ViTs[22-26] have recently outperformed CNNs for image classification and some other tasks. Besides the original ViT, other ViTs[22-26] including DeiT (Data efficient image transformer),[25] ConViT (Convolutional-like vision transformer),[26] and Swin Transformer[23,24] have been proposed and obtained excellent performance on some public datasets such as the ImageNet dataset. However, ViTs were rarely successfully used in medical image analysis,[27,28] let alone cancer classification. It was probably because compared with CNNs, most ViTs need a larger dataset to train and the training speed is slower. Hard inductive biases of CNNs enable sample-efficient learning. ViTs rely on more flexible self-attention layers, however, they require costly pre-training on large external datasets or distillation from pre-trained convolutional networks.[25,26]

By combining self-attention with a hierarchical structure that operates locally at different scales, Swin Transformer series,[23] including both Swin Transformer and Swin Transformer V2, have built locality, translational equivariance, and hierarchical scale into ViTs. This may reduce the sample size needed. On the basis of Swin Transformer, Swin Transformer V2[24] made

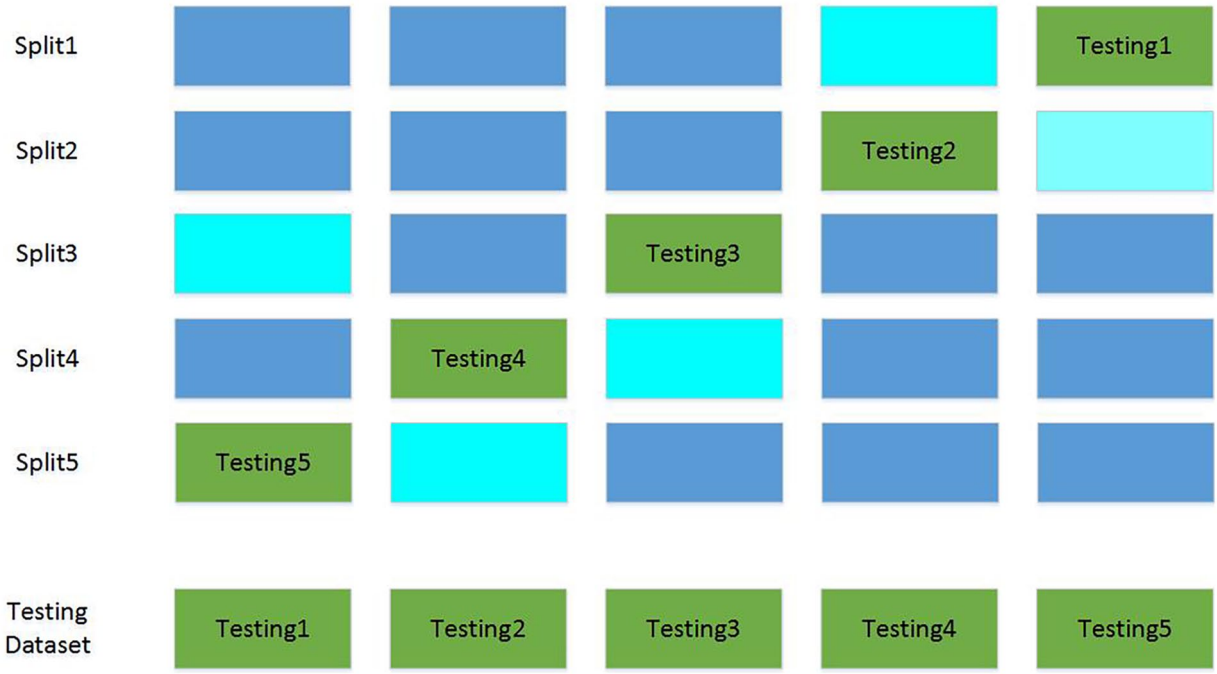

Figure 2. Five-fold cross validation data splitting. Blue, cyan and green stand for training, validation and testing dataset, respectively. The final testing dataset was combined by 5 testing subsets.

the following improvements: pre-layer-normalization changed to post-layer-normalization and dot product replaced by scaled cosine attention. The architecture of Swin Transformer V2 is shown in Figure 2. Swin Transformer V2 can not only scale up capacity and resolution, but also improve training stability and accuracy.

In this study, many mainstream CNNs and ViTs were used to develop the algorithm, and the best model was chosen as the final classifier.

Training strategies

To enlarge the sample size and avoid overfitting, image augmentation was used during training.[32] Compared with image augmentation before training, the on the fly implementation was not only more time-efficient but also more flexible. The image augmentation operation included random horizontal and vertical flipping, random brightness and contrast modifications. The Albumentations[33] library and PyTorch dataset class were used to implement real-time image augmentation. After image augmentation, all pixel values were normalized to (0-1). Technical details about image augmentation can be found in the source code.

Softmax was used as the last layer's activation function, and multi-class cross entropy loss was used as the loss function. For every model, parameters were initialized from the corresponding ImageNet pre-trained model, and then all layers were fine-tuned. During pre-experiments, there was no perceivable performance difference between parameters initialized from the ImageNet-1K models and initialized from ImageNet-21K pre-trained models. Adam[34] was used as the optimizer. Automatic mixed precision training[35] was used to speed up the training and inference process and save GPU memory. Label smoothing ($\varepsilon = 0.1$) was used to calibrate probabilities and improve generalization ability.[36] The batch size was set to 64. The number of epochs was set to 60, and the learning rate was set to 1e-4. During experiments, performances were insensitive to these hyper-parameters.

Because of 5-fold cross validation, every neural network was independently trained on 5 different training datasets. And for every training dataset, the same neural network was trained 3 times, and the model with the minimum validation loss was chosen as the best model. Finally, 5 models were obtained for every neural network, and each for every data splitting.

Inference and performance evaluation

Inference process for 1 image: Given an image, the prediction mathematical formula is as follows:

$$\text{pred\_class} = \text{probs.argmax}\left(\text{axis} = -1\right)$$

Here probs denotes the output of softmax activation of the last layer. And pred_class is the predicted class, which is one of the 5 classes.

Overall performance evaluation: Because of 5-fold cross validation, we have 5 non-overlapping testing datasets. For every neural network, results were calculated separately on each testing dataset using the best model trained on the same data splitting. The final testing results were obtained by combining results of 5 testing datasets.

Performance metrics

The commonly used multi-class classification performance indicators are accuracy, confusion matrix and Kappa coefficient.[37] Besides that, 5 class multi-class classification can be viewed as 5 one-versus-rest binary classifications. For binary classification, the precision, recall(sensitivity), specificity and F1 are popular metrics.[38,39] Considering that previous studies using the same LC25000 dataset adopted these 3 binary classification metrics, to compare our results with theirs, they were also calculated. Because the LC25000 dataset was balanced for every class, the macro average and micro average values for every indicator were identical.

Accuracy is the total number of correct predictions divided by the total number of samples. The confusion matrix is a special case of contingency table, with 2Ds ("actual" and "predicted"). And both accuracy and Kappa can be deduced from the confusion matrix. Precision (also called positive predictive value) is the fraction of true positives among predicted positives. Recall (true positive rate) refers to the probability of a positive test, conditioned on truly being positive. F1 is a special case of Fβ and more specifically is the harmonic mean of precision and recall. All these metrics are bounded between 0 and 1 (perfect). TP, TN, FP, FN stands for true positive, true positive false positive, false negative, respectively. For Kappa annotation, Pr(a) represents the actual observed agreement and Pr(e) represents chance agreement. The mathematical formulas of these metrics are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Kappa} = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F\beta = \frac{1 + \beta^2 TP}{1 + \beta^2 TP + \beta^2 FN + FP}$$

$$F1 = \frac{2TP}{2TP + 2FN + FP}$$

Experimental settings

Hardware: Intel E5-2620 V4 * 2, 256GB Memory, Nvidia GTX 3090 * 2

Software: Ubuntu 20.04, CUDA 11.3, Anaconda 4.10.

The programing language and libraries: Python 3.8, Pytorch 1.12, Torchvision OpenCV, NumPY, Timm, Sklearn, Matplotlib, Pandas, Albumentations, and Tqdm. Detailed information about these software libraries can be found in the file requirements.txt of the source code.

Results

Representative image patches of every class in LC25000 dataset are shown in Figure 3.

The first, second and third image on the first row belongs to class benign lung, lung adenocarcinoma, and lung squamous cell carcinoma, respectively. The first and second image on the second row belongs to class benign colon and colon adenocarcinoma, respectively.

Statistical performance metrics of different models are shown in Table 1.

Performance metrics of some models were very close to 1 and hard to distinguish. However, there existed observable differences in their confusion matrices. The confusion matrices of different models are presented in Supplemental Figure 1. Swin-Transformer V2 model obtained a perfect confusion matrix, which means that the predicted labels and the ground truth labels were identical for all images. Training and validation loss curves are shown in Supplemental Figure 2. For every model, the loss graph was randomly selected from the training process of one data splitting. According to performance metrics presented in Table 1, the Swin Transform V2 was chosen as the final model, which outperformed all other 18 models

Table 2 shows the performance metrics comparisons with other studies using the same dataset. To make a fair comparison, studies conducting only 1 cancer type classification such as lung cancer classification (3 classes) or colon cancer classification (2 classes) were excluded.

Discussion

There existed significant performance differences among CNNs. CNNs including Inception V3, Inception-Resnet V2 and ResnetV2 obtained near perfect results. However, EfficientNet, DenseNet, and MobileNet (both V2 and V3) obtained much worse performance. MobileNets are lightweight models, it is reasonable that they are inferior to other models. However, the performance gap is too big to understand. And we totally do not understand why EfficientNet and DenseNet obtained such poor results. As for ViTs, Swin Transformer series were trained much more stable and faster than other ViTs including the original ViT, Deit, and Convit. There was no obvious training speed difference between CNNs (except for abovementioned bad CNNs) and Swin Transformer series. Most importantly, only Swin Transform V2 achieved perfect results on all metrics, which was much better than those of other models in this study and previous studies using the same dataset.

This study has both strengths and limitations. Strengths: This study adopted the advanced Swin Transform V2 model and

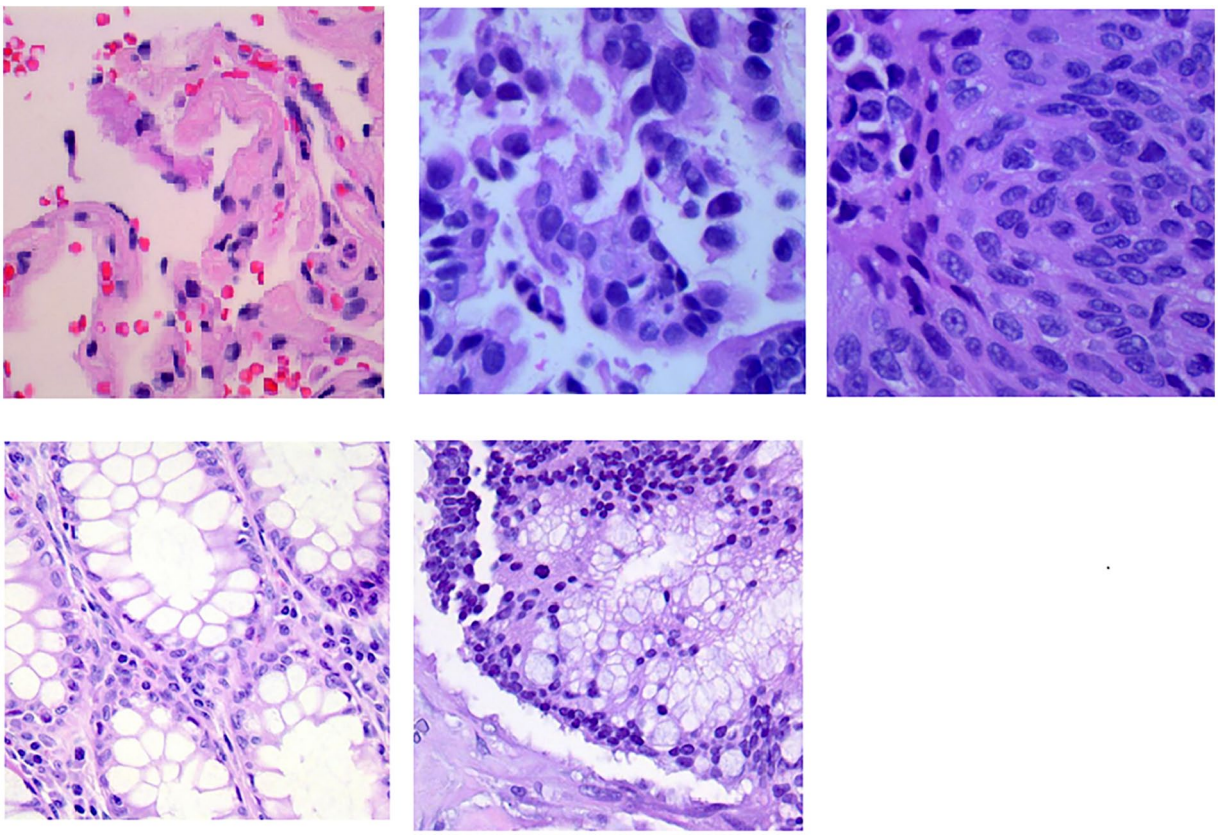

Figure 3. Representative image patches of LC25000 dataset.

achieved perfect results. This study compared the performance and convergence speed of different models including mainstream CNNs and ViTs. These comparison results may be valuable for not only the lung and colon cancer classification but also other medical image analysis tasks. Limitations: First and most importantly, a typical whole slide image classification pipeline includes image patches generation, patches classification and aggregating patch results to obtain the result of the whole slide image. This study only focused on image patch classification because the LC25000 dataset only contains image patches.

The LC25000 is class balanced. However, in real clinical settings class imbalance ratio can reach to 1000.[19] In those cases, class imbalanced should be considered in real clinical setting. Both resampling method and cost sensitive learning method such as weighted cross entropy loss and a combination of these 2 are candidate methods to tackle this problem. Moreover, the LC25000 dataset do not represent all kind of images in the real clinical environment. They not only do not cover most of lung and colon cancer subtypes but also do not include images containing hemorrhage, inflammation, necrotic, and tissue folding areas. The dataset only contains well differentiated samples. The LC25000 dataset did not include any patient or tissue IDs information, so it was impossible to do patient-based data splitting. External validation was not conducted in this study and the dataset used lacks more complex cases Theoretically, the testing dataset may contain data leakage coming from the training dataset. This study only used the LC25000 dataset, the generalization ability of the models was not guaranteed.

In the future, we plan to develop a new lung and colon cancer classification dataset. The whole slide images will be collected from multiple centers and scanned using different protocols. Most importantly, these slides should better represent images in the real clinical environment. These slides should not only cover most of lung and colon cancer subtypes but also contain hemorrhage, inflammation, necrotic, and tissue folding areas. Based on the new dataset, we will conduct a more in-depth research on algorithm development and do external validation to validate model generalization ability. Additionally, exploring the model's performance on more challenging or diverse datasets could offer insights into its robustness and applicability in real-world scenarios. Moreover, we plan to compare performances of tumor detection and classification for pathologist with different level of experience under the conditional of with and without AI assistance. Our study is only a very small step toward AI be used in real clinical practice. So far, there exist tons of research papers on AI in diagnostic pathology; however, very few of them are clinical applicable. In the foreseeable future, the significance of the AI system is not to replace doctors, but to assist doctors. AI systems can deploy to the real clinical environment and act as tireless assistants for pathologists. Before that, it should be proved that pathologists with AI assistant consistently outperform pathologists without AI assistant.[40]

Table 1. Performance metrics of different models.

MODEL	ACCURACY	KAPPA	PRECISION	RECALL	F1
Inception-Resnet V2	0.9999	0.9999	0.9999	0.9999	0.9999
Inception V3	1	0.9999	0.9999	0.9999	0.9999
Resnetv2_101 × 1_bitm	1	0.9999	0.9999	0.9999	0.9999
DenseNet121	0.9135	0.8919	0.9139	0.9135	0.9135
EfficientNetB2	0.8299	0.7874	0.8318	0.8299	0.8296
EfficientNetB3	0.7981	0.7476	0.8057	0.7981	0.7983
MobileNetV2_100	0.7510	0.6888	0.7531	0.7510	0.7513
MobileNetV2_100d	0.7513	0.6891	0.7580	0.7513	0.7531
MobileNetV3_large_100	0.7573	0.6966	0.7585	0.7573	0.7574
MobileNetV3_small_075	0.8108	0.7635	0.8146	0.8108	0.8101
ViT_small	0.9945	0.9931	0.9945	0.9945	0.9945
ViT_base	0.9882	0.9852	0.9883	0.9882	0.9882
DeiT_small	0.9904	0.9879	0.9904	0.9904	0.9904
DeiT_base	0.9920	0.9900	0.9920	0.9920	0.9920
ConViT_tiny	0.9856	0.9819	0.9856	0.9856	0.9856
ConViT_small	0.9915	0.9894	0.9915	0.9915	0.9915
ConViT_base	0.9927	0.9909	0.9927	0.9927	0.9927
Swin Transform	0.9999	0.9999	0.9999	0.9999	0.9999
Swin Transform V2(Proposed)	**1**	**1**	**1**	**1**	**1**

Bold values represent the best results.

Table 2. Performance comparisons with other studies that use single model.

REFERENCE	CLASSIFIER	ACCURACY	PRECISION	RECALL	F1
Masud et al[9]	Custom designed CNN	0.9633	0.9639	0.9637	0.9638
Ali and Ali[5]	CapsNts	0.9958	0.9866	0.9906	0.9904
Tummala et al[10]	EffcientNetV2	0.9997	/	/	0.9997
Proposed	Swin Transformer V2	**1**	**1**	**1**	**1**

The symbol "/" indicates that this metric was not provided by this study. Bold values represent the best results.

Conclusions

In this study, we have developed an automated lung and colon cancer classification system using the publicly available LC25000 dataset. Extensive experiments showed that the Swin Transformer V2 model obtained perfect results, which outperformed other models including both mainstream CNNs and ViTs and models of previous studies. The LC25000 dataset is more a benchmark dataset than a dataset that can represent real clinical scenarios. In this study all benchmarks hit saturation (perfect results). From the point of view of benchmark dataset, the significance of comparing the algorithm performance using this dataset has been lost.

In the future, after thorough verification, the Swin Transformer V2 model has the potential to be used to assist pathologists in classifying lung and colon cancer histopathology images in the clinical setting.

Acknowledgements

Not applicable.

Author Contributions

Conceptualization and methodology, Jie Ji, Jirui Li, and Weifeng Zhang; Software, Jie Ji, Weifeng Zhang; Data collection and verification, Weifeng Zhang, Yuejiao Dong, Jiexiong

Huang; Manuscripts drafting, Jie Ji, Liangli Hong; Manuscript review optimization, Yiqun Geng, Weifeng Zhang; Visualization, Jie Ji; Supervision, Liangli Hong; Funding acquisition, Liangli Hong. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The used datasets were obtained from publicly open-source datasets from https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images. The source code and trained models are publicly available at https://github.com/linchundan88/lung_colon_cancer_classification

Informed Consent Statement

Not applicable.

Institutional Review Board Statement

Not applicable.

ORCID iD

Weifeng Zhang https://orcid.org/0000-0001-9273-7197

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Jemal A, Ward EM, Johnson CJ, et al. Annual report to the nation on the status of cancer, 1975–2014, featuring survival. *J Natl Cancer Inst*. 2017;109:3-12.
2. World Health Organization. Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer.
3. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. 2023;73:17-48.
4. Mangal S, Chaurasia A, Khajanchi A. Convolution neural networks for diagnosing colon and lung cancer histopathological images 2020 September 01, 2020 [arXiv:2009.03878p.].https://ui.adsabs.harvard.edu/abs/2020arXiv200903878M.
5. Ali M, Ali R. Multi-input dual-stream capsule network for improved lung and colon cancer classification. *Diagnostics*. 2021;11:1485.
6. Travis WD. Pathology of lung cancer. *Clin Chest Med*. 2011;32:669-692.
7. Steinfort DP, Russell PA, Tsui A, et al. Interobserver agreement in determining non-small cell lung cancer subtype in specimens acquired by EBUS-TBNA. *Eur Respir J*. 2012;40:699-705.
8. Borkowski A, Bui M, Brannon Thomas L, et al. Lung and colon cancer histopathological image dataset (LC25000). 2019. [arXiv:1912.12142 p.]. https://ui.adsabs.harvard.edu/abs/2019arXiv191212142B.
9. Masud M, Sikder N, Nahid A-A, Bairagi AK, AlZain MA. A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. *Sensors*. 2021;21:748.
10. Tummala S, Kadry S, Nadeem A, Rauf HT, Gul N. An explainable classification method based on complex scaling in histopathology images for lung and colon cancer. *Diagnostics*. 2023;13:1594.
11. Al-Jabbar M, Alshahrani M, Senan EM, Ahmed IA. Histopathological analysis for detecting lung and colon cancer malignancies using hybrid systems with fused features. *Bioengineering*. 2023;10:383.
12. Gabralla LA, Hussien AM, AlMohimeed A, et al. Automated diagnosis for colon cancer diseases using stacking transformer models and explainable artificial intelligence. *Diagnostics*. 2023;13:2939.
13. Kumar N, Sharma M, Singh VP, Madan C, Mehandia S. An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images. *Biomed Signal Process Control*. 2022;75:103596.
14. Mehmood S, Ghazal TM, Khan MA, et al. Malignancy detection in lung and colon histopathology images using transfer learning with class selective image processing. *IEEE Access*. 2022;10:25657-25668.
15. Chhillar I, Singh A. A feature engineering-based machine learning technique to detect and classify lung and colon cancer from histopathological images. *Med Biol Eng Comput*. 2024;62:913-924.
16. Provath MA-M, Deb K, Dhar PK, Shimamura T. Classification of lung and colon cancer histopathological images using global context attention based convolutional neural network. *IEEE Access*. 2023;11:110164-110183.
17. Dabass M, Vashisth S, Vig R. A convolution neural network with multi-level convolutional and attention learning for classification of cancer grades and tissue structures in colon histopathological images. *Comput Biol Med*. 2022;147:1-5.
18. Dabass M, Vashisth S, Vig R. MTU: a multi-tasking U-net with hybrid convolutional learning and attention modules for cancer classification and gland segmentation in colon histopathological images. *Comput Biol Med*. 2022;150:106095.
19. Cen L-P, Ji J, Lin J-W, et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nat Commun*. 2021;12:2-11.
20. Wang J, Ji J, Zhang M, et al. Automated Explainable Multidimensional Deep Learning Platform of retinal images for retinopathy of prematurity screening. *JAMA Netw Open*. 2021;4:e218758.
21. Zhang G, Lin J-W, Wang J, et al. Automated multidimensional deep learning platform for referable diabetic retinopathy detection: a multicentre, retrospective study. *BMJ Open*. 2022;12:e060155.
22. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv e-prints. 2020:arXiv:2010.11929.
23. Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows. 2021. [arXiv:2103.14030 p.]. https://ui.adsabs.harvard.edu/abs/2021arXiv210314030L.
24. Liu Z, Hu H, Lin Y, et al. Swin transformer V2: scaling up capacity and resolution. arXiv e-prints. 2021:arXiv:2111.09883.
25. Touvron H, Cord M, Douze M, et al. Training data-efficient image transformers & distillation through attention. arXiv e-prints. 2020:arXiv:2012.12877.
26. d'Ascoli S, Touvron H, Leavitt M, et al. ConViT: improving vision transformers with soft convolutional inductive biases. 2021. [arXiv:2103.10697 p.]. https://ui.adsabs.harvard.edu/abs/2021arXiv210310697D.
27. Shamshad F, Khan S, Waqas Zamir S, et al. Transformers in medical imaging: a survey. 2022. [arXiv:2201.09873 p.]. https://ui.adsabs.harvard.edu/abs/2022arXiv220109873S.
28. Han K, Wang Y, Chen H, et al. A survey on vision transformer. 2020. [arXiv:2012.12556 p.]. https://ui.adsabs.harvard.edu/abs/2020arXiv201212556H.
29. Cordonnier JB, Loukas A, Jaggi M. On the relationship between self-attention and convolutional layers. 2019. [arXiv:1911.03584 p.]. https://ui.adsabs.harvard.edu/abs/2019arXiv191103584C.
30. Khan S, Naseer M, Hayat M, et al. Transformers in vision: a survey. *ACM Comput Surv*. 2021;54:1-28.
31. Borkowski A, Wilson C, Borkowski S, et al. Comparing artificial intelligence platforms for histopathologic cancer diagnosis. *Federal practitioner*. 2019;36:456-463.
32. Shorten C, Khoshgoftaar TM, Furht B. Text data augmentation for deep learning. *J Big Data*. 2021;8:101.
33. Buslaev A, Iglovikov VI, Khvedchenya E, et al. Albumentations: fast and flexible image augmentations. *Information*. 2020;11:125.
34. Kingma D, Ba J. Adam: a method for stochastic optimization. arXiv e-prints. 2014:arXiv:1412.6980.
35. Micikevicius P, Narang S, Alben J, et al. Mixed precision training. arXiv e-prints. 2017:arXiv:1710.03740.
36. Guo C, Pleiss G, Sun Y, Weinberger K. On calibration of modern neural networks. ArXiv; 2017. https://arxiv.org/pdf/1706.04599.pdf.
37. McHugh ML. Interrater reliability: the kappa statistic. *Biochem -med*. 2012;22:276-282.
38. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56:45-50.
39. Goutte C, Gaussier E (eds). A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. European Conference on Information Retrieval Advances in Information Retrieval; 2005; Berlin, Heidelberg: Springer Berlin Heidelberg.
40. Bharati S, Mondal M, Podder P. A review on explainable artificial intelligence for healthcare: why, how, and when? *IEEE Trans Artif Intell*. 2023.