Received: 17 January 2023 | Revised: 27 February 2023 | Accepted: 28 February 2023 | Published online: 3 March 2023

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2024, Vol. 2(4) 212–220
DOI: 10.47852/bonviewJDSIS3202744

BON VIEW PUBLISHING

ResNet for Histopathologic Cancer Detection, the Deeper, the Better?

Ziying Wang[1,†], Jinghong Gao[1,†], Hangyi Kan[1], Yang Huang[1], Furong Tang[2], Wen Li[3] and Fenglong Yang[4,*]

[1]*School of Medical Imaging, Fujian Medical University, China*

[2]*School of Medicine, Tsinghua University, China*

[3]*Department of Pathology, Fujian Medical University Union Hospital, China*

[4]*School of Medical Technology and Engineering, Fujian Medical University, China*

Abstract: Histopathological image classification has become one of the most challenging tasks for researchers, due to the varied categories and detailed differences within diseases. In this study, we investigate the critical role of network depth in histopathological image classification, utilizing deep residual convolutional neural networks (ResNet). We evaluate the efficacy of two transfer learning strategies using ResNet with varying layers (18, 34, 50, 152) pretrained on ImageNet. Specifically, we analyze whether a deeper network or the fine-tuning of all layers in pre-trained ResNets enhances performance compared to freezing most layers and training only the last layer. Conducted on Kaggle's dataset of 220,025 labeled histopathology patches, our findings reveal that increasing the depth of ResNet does not guarantee better accuracy (ResNet-34 AUC: 0.992 vs. ResNet-152 AUC: 0.989). Instead, dataset-specific semantic features and the cost of training should guide model selection. Furthermore, deep ResNet outperforms traditional logistic regression (ResNet AUC: up to 0.992 vs. logistic regression AUC: 0.775), showcasing superior generalization and robustness. Notably, the strategy of freezing most layers doesn't improve the accuracy and efficiency of transfer learning and the performance of both transfer strategies depends largely on the types of data. Overall, both methods produce satisfactory results in comparison to models trained from scratch or conventional machine learning models.

Keywords: histopathological cancer, image classification, residual neural network, transfer learning

1. Introduction

In comparison with other medical images, histopathology in which a positive biopsy is taken from the tumor is the golden standard for diagnosis. Pathological diagnosis is made by human pathologists who observed stained specimens on glass slides with a microscope. They carry phenotypic information and rich structures that are significant for the diagnostics and medicament of cancer disorders. However, it takes decades to cultivate a senior pathologist and pathological diagnosis by different pathologists is often inefficient and inconsistent. In recent years, capturing the tissue characteristics in digital formats which are also called whole-scan imaging (WSI) opens new horizons for diagnosis in the medical field and various convolutional neural networks (CNNs) assist WSI to diagnose diseases. This will be a relief for the huge workload of pathologists and many hospitals with limited space. Finally, there is no doubt that computer-aided diagnosis of pathology is the ultimate frontier in vision-based disease diagnosis [1, 2]. However, any technology is a double-edged sword, digital pathology is no exception and comes with its challenges [3, 4]; WSI typically generates billion-pixel files, which also require (digital) storage and are not easily analyzed by computer algorithms. In recent years, artificial intelligence is dramatically affecting the medical field.

2. Literature Review

With the accumulation of a large amount of WSI data, to make more efficient use of these data, methods based on traditional machine learning algorithms to analyze and diagnose histopathology images have been proposed [5, 6], and a series of CNNs [7, 8] such as GoogLeNet, VGG, AlexNet, which have been used to analyze WSI digital pathological images [9] and to assist doctors in their work, improve the working efficiency of doctors and reduce the occurrence of misdiagnoses and missed diagnoses. However, traditional machine learning only relies on mathematical formulas and does not study the features of data [10]. When training feature-rich datasets, network performance is significantly

***Corresponding author:** Fenglong Yang, School of Medical Technology and Engineering, Fujian Medical University, China. Email: yangfenglong@fjmu.edu.cn

†These authors contributed to the work equally and should be regarded as co-first authors.

degraded. Unlike traditional machine learning, deep learning consists of multiple layers of neurons to form a deep neural network structure, which is based on the data and learns the characteristics of the data.

A CNN is a neural network structure that has a significant impact on medical image analysis. The deeper the convolutional layers, the more features are extracted [11], but the more the layers, the more unstable the network weight update, followed by gradient disappearance and gradient explosion, which will result in the degradation of network performance. Nowadays, the most popular deep residual network (ResNet) [12], which also belongs to CNNs, can solve these problems by adding residual skip connections to plain CNNs. This operation can even stack the network to more than 1000 layers, while still having good performance.

Nevertheless, training a deep ResNet from scratch is hard. First, it requires vast labeled training data that are difficult to meet in the medical field, because it requires professional pathologists to spend a lot of time and effort to label, even though there are few disease types in the current public datasets [8]. Second, it is often complicated by overfitting and convergence issues. The resolution frequently requires constant adjustments in the architecture or learning parameters of the network to ensure that all layers are learning with comparable speed, which requires a lot of expertise from algorithm engineers to ensure proper convergence. Therefore, using pre-trained ResNet networks and conducting transfer learning to accomplish various tasks of pathological are necessary and helpful [13–15].

Different from other natural images, medical pathological images have privacy, difficult to obtain. Different pathologists make specimens by using different staining methods. And even though the same patient is in the same part of the section, due to the influence of external factors, the WSI images quality will be different. So, all experiments in this paper are done on Kaggle's publicly available datasets, Histopathological Cancer Detection (HCD). Since human eyes are poor at observing millimeter-level features in tissue cells, junior, intermediate, and senior pathologists may draw different conclusions, and this misjudgment often makes patients miss the best treatment time. However, residual neural networks can extract subtle features in tissue cells by stacking very deep network structures. Of course, it can also make mistakes, but it makes mistakes different from human pathologists, which will benefit the pathologists' judgment.

2.1. Theoretical framework

Based on these phenomena as mentioned above, in this paper, we propose transfer learning based on ResNet with 18, 34, 50, and 152 layers pre-trained on the ImageNet database which consists of more than 1.2 million categorized images of 1000+ classes [16, 17]. As shown in Figure 1(c) [12], there are four ResNet network structures and residual block structures. We take two training strategies (Figure 1(a)): (1) fine-tuning all layers and (2) freezing most layers, which are used as feature extractors, training only the last fully connected layer. Then, we explore the influence of different fine-tuning methods ([18], Fine-Tuning Convolutional Neural Networks for Biomedical Image Analysis, n.d.) on the accuracy and answer the question of whether the deeper the ResNet network layers, the better the model performance.

3. Research Methodology

This section discusses topics related to this research, including dataset and image preprocessing which use the color threshold algorithm and augmentation operations.

3.1. Dataset description

All the experiments in this paper were conducted on the Kaggle HCD public dataset, and Kaggle provides non-duplicate pathological images, whose size specifications are 96 pixels, 96 pixels, and 3-channel color images. A positive label indicates that the center 32 × 32 pixels region of a patch contains at least one pixel of tumor tissue. Tumor tissue in the outer region of the patch does not influence the label. Because Kaggle does not provide test labels, only 220,025 images with labels in the training set are used in this paper. As shown in Figure 2(b), the proportion of cancer samples and no cancer samples in the training set's images are close to 1:1, which can be regarded as balanced to assess the classification performance of each category more fairly. Figure 1(c) [12] presents the histopathologic scans of lymph node sections in the datasets.

3.2. Image preprocessing

As shown in Figure 2(a), this is our image preprocessing process, because there are over-bright and over-dark images in the dataset, which is useless information and will affect the judgment of the models. We used the color threshold algorithm to treat values below 10 pixels as over-dark and values above 245 pixels as over-bright. A total of 7 images that met the condition were removed. I divided 220,018 pictures with labels in train in a ratio of about 8:1:1, including 176,232 training sets, 21,783 validation sets, and 22,003 testing sets.

Due to the small number of pathological images, I processed the data with augmentation [19, 20]. Augmentation operations include randomly cutting to the specified 90 × 90 size, random horizontal vertical flip, the three-channel regularization, and standardization of the data. Through data augmentation, the training set accuracy decreases because of the diversity of data features, which is more difficult to train, but the generalization ability of the model will improve.

3.3. Research design

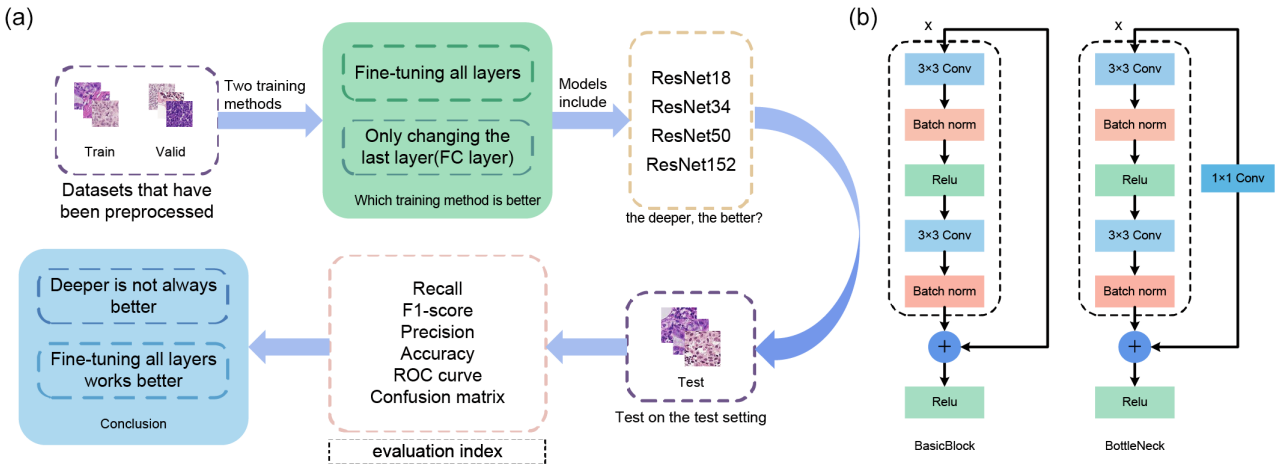
Each experiment was run using GPU RTX 3060, CUDA v11.2 (12 GB of video memory), and the architecture on the PyTorch's official website. Utilizing ResNet pre-trained networks, we then analyze the performance of traditional machine learning classification, the effectiveness of the network when using it just as a feature extractor [21, 22], and when fine-tuning training all layers to the medical imaging domain.

3.4. Traditional machine learning methods

Machine learning techniques ([23], Medical Imaging Using Machine Learning and Deep Learning Algorithms, n.d.) usually are divided into supervised learning and unsupervised learning. The goal of supervised learning is to conclude a function that can map the input images to their proper labels (e.g. cancer cell images correspond to the cancer labels) well using training data. Labels are related to a WSI or an object in WSIs [24, 25]. The algorithms for supervised learning contain support vector machines [26] and CNNs. However, the goal of unsupervised learning is to

Figure 1

Graphical representation of the workflow and ResNet structures. (a) Workflow of the experiment. (b) Two ResNet structures: BasicBlock is used in ResNet18 and ResNet34, but ResNet50 and ResNet152 use bottleneck. (c) Architectures for ImageNet. Building blocks are shown in brackets, with the numbers of blocks stacked. Down-sampling is performed by conv3_1, conv4_1, and conv5_1 with a stride of 2



(c)

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	[3×3, 64; 3×3, 64] ×2	[3×3, 64; 3×3, 64] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3
conv3_x	28×28	[3×3, 128; 3×3, 128] ×2	[3×3, 128; 3×3, 128] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×8
conv4_x	14×14	[3×3, 256; 3×3, 256] ×2	[3×3, 256; 3×3, 256] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×23	[1×1, 256; 3×3, 256; 1×1, 1024] ×36
conv5_x	7×7	[3×3, 512; 3×3, 512] ×2	[3×3, 512; 3×3, 512] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

conclude a function that can represent hidden structures from unlabeled images. The algorithms for unsupervised learning include k-means [27], principal component analysis, and so on. Since the HCD dataset only has two categories and provides the labels, I chose linear regression in a supervised learning algorithm for the classification [28].

Assuming the probability of the sample x belonging to cancer is $p(x)$, apparently, the probability is $1-p(x)$ when the sample belongs to no cancer. The idea of logistic regression [29] is to use exponential Equation (1) to model cancer:

$$cancer = \frac{p(\overline{x}_i)}{1-p(\overline{x}_i)} = e^{\theta^T \cdot \overline{x}_i} = e^{z_i} \tag{1}$$

The given Equation (1) is characterized by continuity and differentiability across the real number domain, and it always produces positive values. These conditions are necessary for the classification model, as they facilitate the creation of a sigmoid Equation (2), which is obtained by taking the logarithm of the aforementioned function. The sigmoid function is a real-valued function that falls between 0 and 1, and it satisfies the previously mentioned expression.

$$p(\overline{x}_i) = \sigma(\overline{x}_i) = \frac{1}{1+e^{-Z_i}} \tag{2}$$

Finally, as shown in Figure A1, the Receiver operating characteristic curve (ROC) obtained using logistic regression is as follows in the figure: It can be seen that because the linear regression of dichotomy is learned by finding a curve as the dividing line between the positive and negative samples and the data, so the value of area under curve (AUC) is low.

3.5. Fine-tuning the training of all the layers

Considering the limited availability of plenty of weight parameters and of label data in the ResNet, the pre-trained models came from the ImageNet large natural image dataset. Training the ResNet from a set of pre-trained weights is called fine-tuning and has been successfully used in several domains [30–32].

Fine-tuning begins with migrating the weights from the pre-trained network to the network which we are going to train. The final fully connected layer cannot be migrated firsthand, because its number of nodes depends on the number of categories in the dataset. In our

Figure 2

Dataset and image preprocessing. (a) Flow chart of preprocessing. (b) Fan chart of the ratio of cancer to non-cancer. (c) A selection of patches from each histopathologic scans of lymph sections within the Kaggle dataset. The patches are 96 × 96 pixels in size. The first row is the cancer sample, which we call the positive sample, and the second row is the negative sample

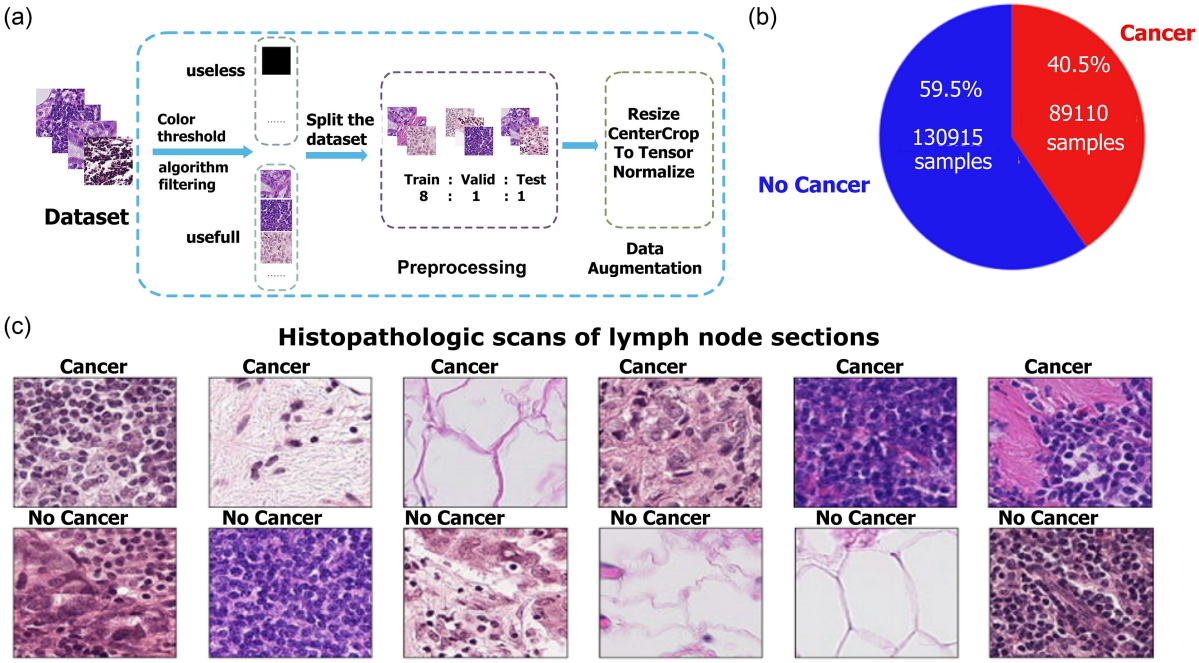


Figure 3

The result of fine-tuning the training of all the ResNet layers. (a) The binary-class confusion matrix on classification of lymphatic pathological tissue by four different ResNet models and the same train methods of fine-tuning training all layers (1 stands for a positive sample, 0 stands for a negative sample). (b) The accuracy of the trained model on the validation set. The abscissa is the number of iterations, and the ordinate is the corresponding accuracy value. For ease of writing, ResNet models for fine-tuning training of all layers are replaced by ResNet18all, ResNet34all, ResNet50all, and ResNet152all in the following paragraphs. (c) AUC curves of the four models on the test set

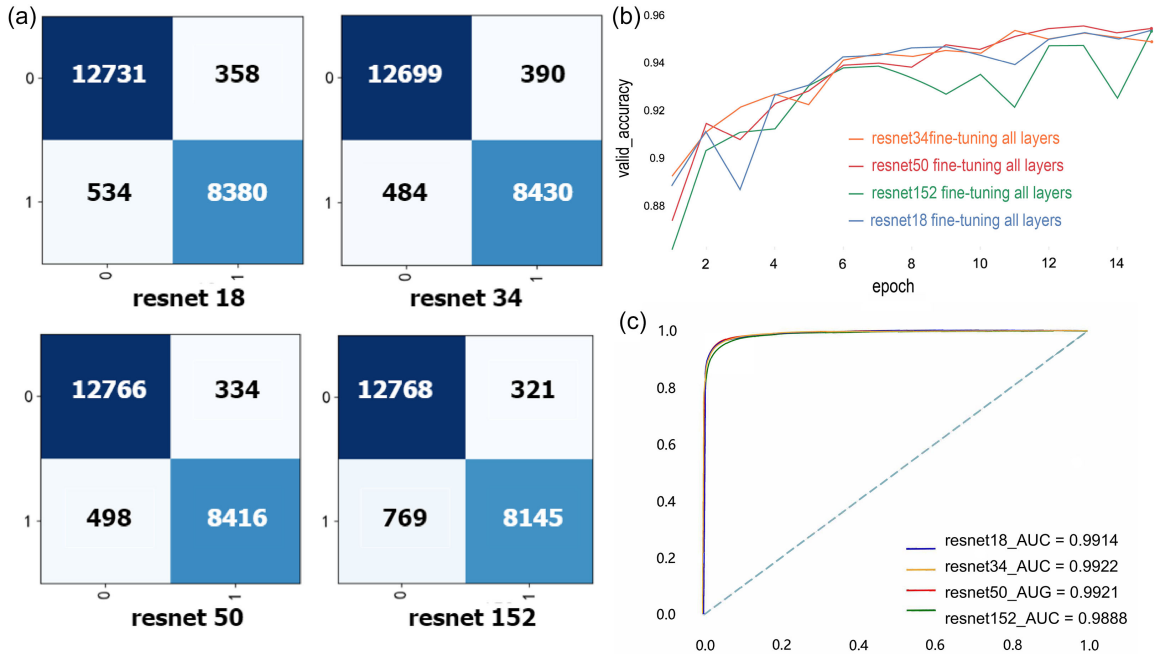
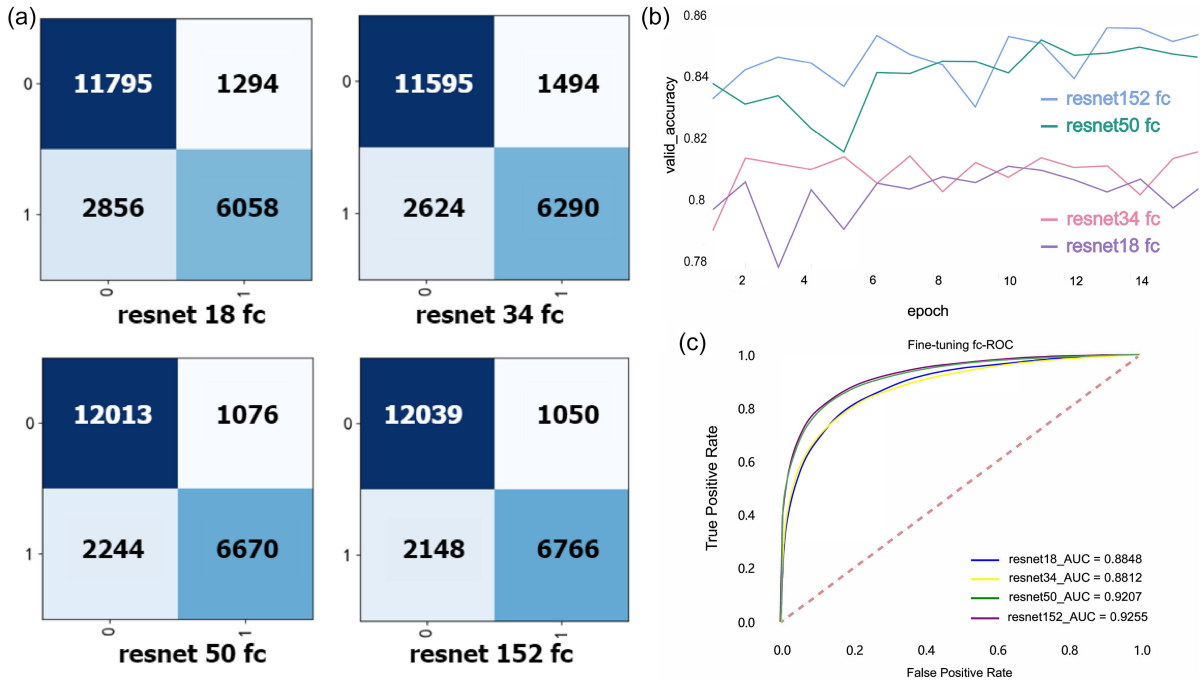


Figure 4

The result of fine-tuning the training of the last ResNet layer. (a) The binary-class confusion matrix on classification of lymphatic pathological tissue by four different ResNet models and the same train methods of fine-tuning the last layer (1 stands for a positive sample, 0 stands for a negative sample). (b) The accuracy of the trained model on the validation set. The abscissa is the number of iterations, and the ordinate is the corresponding accuracy value. For ease of writing, ResNet models for fine-tuning the last full connected layer are replaced by ResNet18 fc, ResNet34 fc, ResNet50 fc, and ResNet152 fc. (c) AUC curves of the four models on the test set



study, we manage the 2-classification task. Therefore, we replace the last fully connected layer of the pre-trained ResNet with a fully connected layer with only two nodes, initialize the weight parameters of the last layer, and then fine-tune the training of all the layers. We selected four network structures: ResNet18, ResNet34, ResNet50, and ResNet152 to fine-tune all layer training and analyze the results. The fully connected layers perform dimensionality reduction in the ultimately extracted feature maps and the SoftMax layer provides the resultant classification probabilities for every class.

As can be seen in Figure 3(a) of the confusion matrix, the ResNet152 has judged the no cancer samples as the worst of all the models, and the performance is much worse compared with the other models, although ResNet152 predicted the positive sample as the best of all models, only slightly higher than the other models. In Figure 3(c), from the ROC plots of the four models, ResNet152 was the least effective of all the models, with ResNet34 slightly higher than ResNet18 and ResNet34.

3.6. Only change and train the last layer

Generally, the underlying learning of ResNet is suitable for the low-level image features of most visual tasks, such as texture and edges. However, the high-level layer learns high-level semantic features, the receptive field is larger, and the image is more integrated. Although different kinds of datasets usually have similar low-level semantic features, high-level semantic features are different [33–35]. Therefore, for transfer learning, the fine-tuning of the last few layers usually achieves good results. Nevertheless, if there is a large gap between the kind of dataset of the pre-trained model and the dataset we need to train, a fine-tuning of the underlying layer may also be needed. In our experiments, we only change the last layer (here also referred to as the full connection layer) of the four ResNet network structures, change the 1,000 nodes into two nodes, and then initialize their weight parameters, while the rest of the network structure and parameters remains unchanged and acts as feature extractors before training [36, 37].

As shown in Figure 4(a), the confusion matrix derived from training only the last layer and predicting on the testing set shows that although ResNet152 fc has the best prediction effect for both positive samples and negative samples, ResNet50 fc is about to catch up with ResNet152. Considering multiple factors, ResNet152 fc uses several times more computing power and memory than several other models, so people should be reluctant to want ResNet152 fc.

4. Result

Our experimental results are summarized in Table A1. The experimental results show that using pre-trained networks and parameters as feature extractors, and fine-tuning all the layers of the pre-trained network, can provide similar results for Kaggle HCD, compared to training from scratch. Training from scratch takes a lot of time and cost. Compared with traditional machine learning algorithms, such as logistic regression, although machine learning is relatively short, transfer learning training with ResNet achieves much higher accuracy than machine learning accuracy.

Different experimental results will happen for different transfer learning and fine-tuning methods [38]. When fine-tuning

training all layers, from Figure 3(b), the four models have almost identical accuracy on the validation set. However, when evaluating the testing set with the ROC curve in Figure 3(c), the deepest network of the ResNet152 takes time, and computing power and memory resources are several times more than the other models, but incredibly, the ResNet152 all is slightly lower than the other model AUC. When training only the last layer (FCN), as shown in Figure 4(c), the AUC of ResNet18fc and ResNet34fc can only be around 0.88, whereas the AUC of ResNet50fc and ResNet152fc can reach 0.92. As shown in the graph of Figure 1(c) [12], probably because BasicBlock was applied to the ResNet18, and ResNet34 models, and BottleNeck to the ResNet50, and ResNet152 models. The same convolution structure is found for ResNet18 and ResNet34, as are ResNet50 and ResNet152, while the other pairwise combinations are different. But overall, trailing only the last layer yielded lower AUC than fine-tuning all layers.

5. Discussion

Surprisingly, comprehensive fine-tuning using pre-trained networks (trained on the ImageNet natural image dataset) can achieve comparable results to training networks from scratch (reference here refers to the histopathological image dataset), which spend a considerable amount of time and resources.

Although the four models had similar accuracy on the validation set, ResNet152 performed poorly on the testing set, indicating that ResNet152 is most likely because the model is more complicated and has more parameters, only tuning the parameters make it work well on the validation set, and it is easy to cause overfitting on the testing set. When different models have the same performance, people tend to prefer to choose simple, smaller models. Because the small model universality is more extensive, it is not easy to occur over fitting, and the cost of the resources is also less.

In another training method, only the last fully connected layer is replaced and trained, the remaining other network layers are frozen as the feature extractor. The ResNet performance is in line with the logic presented in the other papers that "ResNet performance will not become at least worse as the network deepens" [12]. However, the unparalleled precision achieved through comprehensive fine-tuning of all ResNet layers highlights the stark disparities between histopathological images and natural images, ultimately impeding the transfer learning's capacity to effectively learn the distinctive characteristics of histopathological images.

If we have a large enough number of samples, that is, millions of histopathology images, and if we have sufficiently advanced computational equipment for effective training, then ResNet may provide better results than transfer learning. While this claim has a lot of support [9], it remains speculation in sensitive areas such as medical imaging.

6. Conclusion

Classification of histopathological images is a useful and challenging task in the analysis of diagnostic pathology. In transfer learning, training only specific layers and fine-tuning training all layers can provide good accuracy on the HCD dataset. The reason for the relatively low former may be the large feature gap between natural and medical images [39, 40], and the lack of a sufficient number of medical training images. On the one hand, "the deeper the ResNet network, the better the performance" cannot be generalized. There are many factors affecting the network performance, including various training methods and fine-tuning means. On the other hand, model selection cannot blindly pursue the accuracy of the model but also requires comprehensive memory, computational resources, and cost considerations. Our experiments further confirm the potential of ResNets for medical imaging applications because both fine-tuning-specific layer ResNets and fully training all layers ResNets outperformed the corresponding traditional machine learning. Future work will continue to use different network structures and different fine-tuning methods (training bottom layer, training middle layer, training top-level, or layered training) to compare, get more data augmentation, and collect more training samples. In short, transfer learning has great potential for improvement [41, 42].

Funding Support

This research was financially supported by the National Natural Science Foundation of China (62102065, 62271353, 62001311), Joint Funds for the Innovation of Science and Technology, Fujian province (Grant number: 2022J05055), Fujian Medical University Research Foundation of Talented Scholars (XRCZX2022003), and Natural Science Foundation of Sichuan Province (2022NSFSC0926).

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle HCD at https://www.kaggle.com/datasets/drbeane/hcd-cropped.

Author Contribution Statement

Ziying Wang: Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Jinghong Gao:** Methodology, Software, Formal analysis, Writing – original draft, Visualization. **Hangyi Kan:** Validation, Investigation. **Yang Huang:** Formal analysis. **Furong Tang:** Writing – review & editing, Funding acquisition. **Wen Li:** Conceptualization, Resources, Data curation, Supervision, Project administration. **Fenglong Yang:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

References

[1] Aljuaid, H., Alturki, N., Alsubaie, N., Cavallaro, L., & Liotta, A. (2022). Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. *Computer Methods and Programs in Biomedicine*, *223*, 106951. https://doi.org/10.1016/j.cmpb.2022.106951

[2] Sun, H., Zeng, X., Xu, T., Peng, G., & Ma, Y. (2020). Computer-aided diagnosis in histopathological images of the endometrium using a convolutional neural network and attention mechanisms. *IEEE Journal of Biomedical and Health Informatics*, *24*(6), 1664–1676. https://doi.org/10.1109/JBHI.2019.2944977

[3] Tizhoosh, H. R., & Pantanowitz, L. (2018). Artificial intelligence and digital pathology: Challenges and opportunities. *Journal of Pathology Informatics*, *9*(1), 38. https://doi.org/10.4103/jpi.jpi_53_18

[4] Yao, K., Sadimin, E., Chang, S., Schmolze, D., & Li, Z. (2022). Current applications and challenges of digital pathology in cytopathology. *Human Pathology Reports*, *28*, 300634. https://doi.org/10.1016/j.hpr.2022.300634

[5] Lohani, B. P., & Thirunavukkarasan, M. (2021). A review: Application of machine learning algorithm in medical diagnosis. In *International Conference on Technological Advancements and Innovations*, 378–381. https://doi.org/10.1109/ICTAI53825.2021.9673250

[6] Shaheamlung, G., Kaur, H., & Kaur, M. (2020). A survey on machine learning techniques for the diagnosis of liver disease. In *International Conference on Intelligent Engineering and Management*, 337–341. https://doi.org/10.1109/ICIEM48762.2020.9160097

[7] Arena, P., Basile, A., Bucolo, M., & Fortuna, L. (2003). Image processing for medical diagnosis using CNN. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *497*(1), 174–178. https://doi.org/10.1016/S0168-9002(02)01908-3

[8] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ..., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285–1298. https://doi.org/10.1109/TMI.2016.2528162

[9] Kieffer, B., Babaie, M., Kalra, S., & Tizhoosh, H. R. (2017). Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks. In *Seventh International Conference on Image Processing Theory, Tools and Applications*, 1–6. https://doi.org/10.1109/IPTA.2017.8310149

[10] Deepa, P., & Gunavathi, C. (2022). A systematic review on machine learning and deep learning techniques in cancer survival prediction. *Progress in Biophysics and Molecular Biology*, *174*, 62–71. https://doi.org/10.1016/j.pbiomolbio.2022.07.004

[11] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). Going deeper with convolutions. *arXiv Preprint: 1409.4842*. https://doi.org/10.48550/arXiv.1409.4842

[12] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

[13] Karimi, D., Warfield, S. K., & Gholipour, A. (2021). Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model parameters and learned representations. *Artificial Intelligence in Medicine*, *116*, 102078. https://doi.org/10.1016/j.artmed.2021.102078

[14] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

[15] Yu, X., Wang, J., Hong, Q. Q., Teku, R., Wang, S. H., & Zhang, Y. D. (2022). Transfer learning for medical images analyses: A survey. *Neurocomputing*, *489*, 230–254. https://doi.org/10.1016/j.neucom.2021.08.159

[16] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Li, F. F. (2009). ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[17] Martinez, R., Smith, D., & Trevino, H. (1992). ImageNet: A global distributed database for color image storage, and retrieval in medical imaging systems. In *Proceedings of Fifth Annual IEEE Symposium on Computer-Based Medical Systems*, 710–719. https://doi.org/10.1109/CBMS.1992.244984

[18] Cai, Z., & Peng, C. (2021). A study on training fine-tuning of convolutional neural networks. In *13th International Conference on Knowledge and Smart Technology*, 84–89. https://doi.org/10.1109/KST51265.2021.9415793

[19] Kaur, C., & Garg, U. (2023). Artificial intelligence techniques for cancer detection in medical image processing: A review. *Materials Today: Proceedings*, *81*, 806–809. https://doi.org/10.1016/j.matpr.2021.04.241

[20] Ritter, F., Boskamp, T., Homeyer, A., Laue, H., Schwier, M., Link, F., & Peitgen, H. O. (2011). Medical image analysis. *IEEE Pulse*, *2*(6), 60–70. https://doi.org/10.1109/MPUL.2011.942929

[21] Du, P., Li, E., Xia, J., Samat, A., & Bai, X. (2019). Feature and model level fusion of pretrained CNN for remote sensing scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(8), 2600–2611. https://doi.org/10.1109/JSTARS.2018.2878037

[22] Qadir, H. A., Shin, Y., Solhusvik, J., Bergsland, J., Aabakken, L., & Balasingham, I. (2019). Polyp detection and segmentation using mask R-CNN: Does a deeper feature extractor CNN always perform better? In *13th International Symposium on Medical Information and Communication Technology*, 1–6. https://doi.org/10.1109/ISMICT.2019.8743694

[23] Komura, D., & Ishikawa, S. (2018). Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal*, *16*, 34–42. https://doi.org/10.1016/j.csbj.2018.01.001

[24] Chen, Y., Yang, H., Cheng, Z., Chen, L., Peng, S., Wang, J., ..., & Ke, Z. (2022). A whole-slide image (WSI)-based immunohistochemical feature prediction system improves the subtyping of lung cancer. *Lung Cancer*, *165*, 18–27. https://doi.org/10.1016/j.lungcan.2022.01.005

[25] Millican-Slater, R. A., & Rakha, E. A. (2022). Applications and implications of whole-slide imaging in breast pathology. *Diagnostic Histopathology*, *28*(3), 149–155. https://doi.org/10.1016/j.mpdhp.2021.12.003

[26] Yan, X., & Zhu, H. (2022). A novel robust support vector machine classifier with feature mapping. *Knowledge-Based Systems*, *257*, 109928. https://doi.org/10.1016/j.knosys.2022.109928

[27] Wang, F., Wang, Q., Nie, F., Li, Z., Yu, W., & Ren, F. (2020). A linear multivariate binary decision tree classifier based on K-means splitting. *Pattern Recognition*, *107*, 107521. https://doi.org/10.1016/j.patcog.2020.107521

[28] Rashidi, H. H., Tran, N. K., Betts, E. V., Howell, L. P., & Green, R. (2019). Artificial intelligence and machine learning in pathology: The present landscape of supervised methods. *Academic Pathology*, *6*, 2374289519873088. https://doi.org/10.1177/2374289519873088

[29] Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, *49*(11), 1225–1231. https://doi.org/10.1016/S0895-4356(96)00002-9

[30] Azizpour, H., Razavian, A. S., Sullivan, J., Maki, A., & Carlsson, S. (2014). Factors of transferability for a generic ConvNet representation. *arXiv Preprint:1406.5774*. https://doi.org/10.48550/arXiv.1406.5774

[31] Penatti, O. A. B., Nogueira, K., & dos Santos, J. A. (2015). Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 44–51. https://doi.org/10.1109/CVPRW.2015.7301382

[32] Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 806–813.

[33] Eigen, D., Rolfe, J., Fergus, R., & LeCun, Y. (2013). Understanding deep architectures using a recursive convolutional network. *arXiv Preprint:1312.1847*. https://doi.org/10.48550/arXiv.1312.1847

[34] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *The 3rd International Conference on Learning Representations*.

[35] Zeiler, M. D., & Fergus, R. (2013). Visualizing and understanding convolutional networks. *arXiv Preprint:1311.2901*. https://doi.org/10.48550/arXiv.1311.2901

[36] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2016). Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(1), 142–158. https://doi.org/10.1109/TPAMI.2015.2437384

[37] Yu, D., & Seltzer, M. L. (2011). Improved bottleneck features using pretrained deep neural networks. In *Twelfth Annual Conference of the International Speech Communication Association*.

[38] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, *35*(5), 1299–1312. https://doi.org/10.1109/TMI.2016.2535302

[39] Cheplygina, V. (2019). Cats or CAT scans: Transfer learning from natural or medical image source data sets? *Current Opinion in Biomedical Engineering*, *9*, 21–27. https://doi.org/10.1016/j.cobme.2018.12.005

[40] Pandey, B., Kumar Pandey, D., Pratap Mishra, B., & Rhmann, W. (2022). A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions. *Journal of King Saud University - Computer and Information Sciences*, *34*(8), 5083–5099. https://doi.org/10.1016/j.jksuci.2021.01.007

[41] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

[42] Li, W., Huang, R., Li, J., Liao, Y., Chen, Z., He, G., . . . , & Gryllias, K. (2022). A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. *Mechanical Systems and Signal Processing*, *167*, 108487. https://doi.org/10.1016/j.ymssp.2021.108487

How to Cite: Wang, Z., Gao, J., Kan, H., Huang, Y., Tang, F., Li, W., & Yang, F. (2024). ResNet for Histopathologic Cancer Detection, the Deeper, the Better? *Journal of Data Science and Intelligent Systems*, *2*(4), 212–220. https://doi.org/10.47852/bonviewJDSIS3202744

Appendix

Figure A1
Logistic regression algorithm in traditional machine learning was used to test the testing set to get the AUC curve

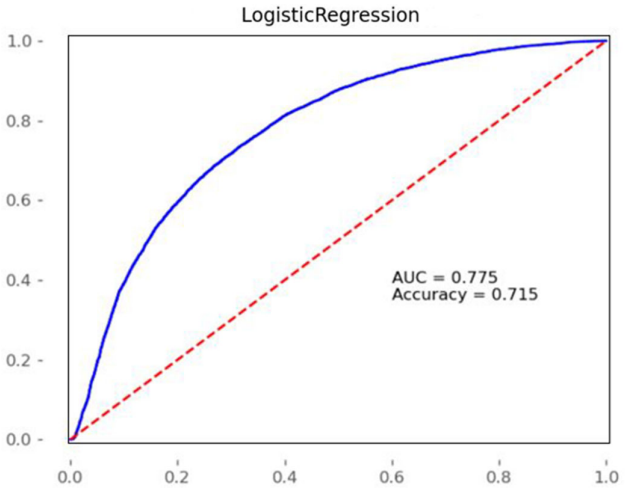


Table A1
The table compares the different evaluation indexes of the four models of the two training methods and each training method on the test set

	Precision	Recall	F1-score	Accuracy
ResNet18all	0.95903	0.94009	0.94946	0.94009
ResNet34all	0.95578	0.94570	0.95071	0.94570
ResNet50all	0.96183	0.94413	0.95290	0.94413
ResNet152all	0.96208	0.91373	0.93728	0.91373
ResNet18fc	0.82399	0.67961	0.74487	0.67961
ResNet34fc	0.80807	0.70563	0.75338	0.70563
ResNet50fc	0.86109	0.74826	0.80072	0.74826
ResNet152fc	0.86566	0.75903	0.80885	0.75903