

Article

Deep Learning Techniques for the Classification of Colorectal Cancer Tissue

Min-Jen Tsai *  and Yu-Han Tao

Institute of Information Management, National Yang Ming Chiao Tung University, Hsin-Chu 300, Taiwan; mjtsai@iim.nctu.edu.tw

* Correspondence: mjtsai@cc.nctu.edu.tw

Abstract: It is very important to make an objective evaluation of colorectal cancer histological images. Current approaches are generally based on the use of different combinations of textual features and classifiers to assess the classification performance, or transfer learning to classify different organizational types. However, since histological images contain multiple tissue types and characteristics, classification is still challenging. In this study, we proposed the best classification methodology based on the selected optimizer and modified the parameters of CNN methods. Then, we used deep learning technology to distinguish between healthy and diseased large intestine tissues. Firstly, we trained a neural network and compared the network architecture optimizers. Secondly, we modified the parameters of the network layer to optimize the superior architecture. Finally, we compared our well-trained deep learning methods on two different histological image open datasets, which comprised 5000 H&E images of colorectal cancer. The other dataset was composed of nine organizational categories of 100,000 images with an external validation of 7180 images. The results showed that the accuracy of the recognition of histopathological images was significantly better than that of existing methods. Therefore, this method is expected to have great potential to assist physicians to make clinical diagnoses and reduce the number of disparate assessments based on the use of artificial intelligence to classify colorectal cancer tissue.

Keywords: convolutional neural network; machine learning; deep learning; colorectal cancer



Citation: Tsai, M.-J.; Tao, Y.-H. Deep Learning Techniques for the Classification of Colorectal Cancer Tissue. *Electronics* **2021**, *10*, 1662. <https://doi.org/10.3390/electronics10141662>

Academic Editor: Tadeusz A. Wysocki

Received: 31 May 2021
Accepted: 8 July 2021
Published: 12 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Colorectal cancer (CRC) is the third most common form of cancer, accounting for about 10% of all cases in the world [1]. The results of many studies have shown that a more accurate classification of medical images can effectively determine the development of colorectal cancer [2,3]. Many common tissue types, such as normal colon mucosa (NORM), adipose tissue (ADI), polyps, cancer-associated stroma (STR), and lymphocytes (LYM) can extract prognosticators directly from these hematoxylin and eosin stains (HE stains), which are the principal tissue stains used in histology [2]. Optical colonoscopy is the medical procedure that is usually used to examine a series of abnormalities on the surface of the colon, including their location, morphology and pathological changes to make a clinical diagnosis. This improves the accuracy of the diagnosis and the ability to predict the severity of the disease in order to apply the most appropriate clinical treatment. Nevertheless, although the correct classification of pathological images is an important factor in assisting doctors to precisely identify the best possible treatment, a great deal of time and effort is required to analyze histopathological images, and the evaluation of tissue classification is easily affected by many subjective factors. Subjective evaluation is generally performed by pathologists who manually review the histological slides images of CRC tissue, which remains the standard for cancer diagnosis and staging. However, the training, experience, evaluation condition or time pressure for each pathologist could result different diagnosis judgement. Hence, the universal automatic classification of CRC pathological tissue slide images for fair evaluation has an important clinical significance.

Pathology slides provide an enormous amount of information, which has been quantified through digital pathology and classic machine learning techniques over the years [4]. Previous research has been based on machine learning approaches for judging the cell classification in the histological slides of tumor tissue. The classification of histopathological images using artificial intelligence not only improves the accuracy and efficiency of the classification, but also enables doctors to make timely decisions in terms of clinical treatment [5,6]. However, most of the proposed experimental methods rely on manual feature labels, which is the main limitation of traditional textual analysis approaches. Therefore, deep learning has been introduced in the last few years to solve this and other limitations. Deep learning is a new technology that is considered to be an evolution of machine learning, since it uses multiple layers of neural networks to learn and progressively extract higher-level features in order to reduce human intervention in the recognition of different classes in the images. It is also effective in classifying non-image data, such as speech recognition, social network filtering, and medical image analysis, and its advanced approach not only reduces the need for human intervention, but it can also automatically achieve results that are comparable to or surpassing those of humans.

Convolutional neural networks (CNN) [7,8] recently showed effective results in classifying images in the field of deep learning where a neural network might have dozens or hundreds of layers to learn containing images with different features. A convolutional layer composed of a small-sized kernel to generate advanced features applies weights to the inputs and directs them through an activation function as the output. The main advantage of using CNN compared to a traditional neural network is that it reduces the model parameters for more accurate results.

With this in mind, we aimed to use deep learning technology to identify medical images to increase the accuracy of the identification due to the automatic classification of tumor types. This involved the achievement of the following objectives:

- a. To compare the classification accuracy rate with different CNN models.
- b. To find the best performance of deep learning techniques.
- c. To compare the results of this method with those of existing techniques.

This paper consists of a systematic study of deep learning and its application for the classification of pathological images. Past studies of deep learning will be reviewed in Section 2, while the approach of deep learning models will be described in Section 3. Details of the experiment will be provided in Section 4, and the paper will be concluded in Section 5 with proposals for possible future investigations in this field.

2. Related Works and Deep Learning Methodology

Some of the prior studies in relation to the automatic classification of histopathological images will be described and discussed in this section with a further explanation of how deep learning works. This will be followed by a presentation of the proposed method to conduct the current research.

2.1. Related Works

Digital technology is currently used extensively to classify medical images, as evidenced by the results of several methods of histopathological image classification shown in Table 1. Kather [2] used a range of textual descriptors to analyze a multi-class problem of tumor epithelium and simple stroma in 5000 histological images. He proposed four classification methods: (1) the k-nearest neighbors algorithm (k-NN), (2) employ an SVM decision function in an attempt to classify all categories, (3) assemble decision tree models using the RUSBoost method, and (4) use a 10-fold cross validation to train the classifiers, without an explicit stratification approach. The results indicated that SVM was the best classification method, which achieved 87.4% accuracy over eight classes. Lately, the classification of tumor types has been found to be more accurate using the CNN classification method. Tsai [9] applied the CNN architecture of a deep learning technique to detect pneumonia

from Chest X-rays and achieved an accuracy rate within 82.1% by using feature selection and the CNN.

Table 1. Related Research.

Literature	Research Objective	Approach	Classification Technique	Accuracy Rate (%)
[2]	Multi-class texture analysis in colorectal cancer histology	Texture based methods	One-nearest neighbor, linear SVM, radial-basis function SVM and decision trees.	87.4
[9]	Machine learning-based common radiologist-level pneumonia detection on chest X-rays	Feature selection, CNN	Feature selection and convolutional neural network (CNN)	80.9
[10]	A deep convolutional neural network for segmenting and classifying the epithelial and stromal regions in histopathological images	CNN	CNN network comprised of two convolutional layers, two max-pooling layers, and two fully connected layers followed by a soft-max layer	84
[11]	Classification of tumor epithelium and stroma by exploiting image features learned by deep convolutional neural networks	CNN	CNN with GoogLeNet transfer learning strategies	90.2
[12]	Large-scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features	CNN	The framework transfer features extracted from CNN were trained by a large natural image database, ImageNet, to assess the histopathology images and also explore the characteristics in the last hidden layer	97.5
[13]	Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies	CNN	Propose some new geometric features of benign biopsies	96.2
[3]	Predicting survival from colorectal cancer histology slides using deep learning, a retrospective multicenter study	CNN	Evaluated the performance of five different CNN models: VGG19, AlexNet, SqueezeNet, GoogLeNet, Resnet50	98.7

Xu [10] used the CNN model and feature extraction approaches to compare two datasets of breast cancer and colorectal cancer. The two types of tissues in the histological

images were epithelial (EP) and stromal (ST). He used automated segmentation or the classification of color features, which included intensive pixels in different color spaces, and analyzed the tumor microenvironment. In his study, Du [11] proposed that learning the basic features of CNN methods outperformed handcrafted features, and automatically distinguished the epithelial and stromal regions in the breast. In addition, he found that colorectal tumors could be distinguished from tumor tissue using a network architecture layer approach with results that were 84% accurate. Transfer learning is a methodology that consists of deep learning techniques to distinguish the features of leverages images. Du [10] discussed the use of transfer learning methods to accurately distinguish breast or ovarian cancer from histological images and of CNN for fine tuning the feature extractor of images. Additionally, he discussed how to distinguish high-level and low-level features inside the neural network. A deep neural network may have multiple layers, the first of which will learn the low-level features and then the more they progress toward the output layer, the more the layers will learn the high-level features. Du [11] also used a transfer learning approach with GoogLeNet and achieved 90.2% accuracy, suggesting the feasibility of using it to classify the tumor stroma ratio (TSR). Xu et al. [12] improved the activation features of the AlexNet model and proposed the characteristic of visualizing the neurons in the last hidden layer to classify and segment them. Trained by ImageNet, the framework successfully transferred the features extracted from the network into little histopathology images features for training and visualization and a test accuracy rate of 97.5% was reported. Bejnordi et al. [13] proposed deep convolutional neural networks with some new geometric features, and trained the algorithm networks to classify stroma images, including stroma, fat tissues, other situ lesions and to predict the stroma regions. Bejnordi analyzed the stroma between surrounding invasive cancer and situ lesions and achieved a 96.2% accuracy. Additionally, Kather [3] replaced the classification layer and the best accuracy rate was 98.7% with VGG19.

2.2. Advantages and Limitations of Using Machine Learning Approaches

Machine learning teaches computers to simulate and implement human learning behavior based on computational methods to learn knowledge from sample data. It is widely used in various applications such as image, content recommendation, computer vision, etc., in which it is difficult to develop conventional algorithms and solve the above problems to achieve the required tasks [14]. There are two main techniques: supervised learning (used to learn mapping between input and output) and unsupervised learning (involves using a model to extract relationships from data). The goals of machine learning are feature extraction, selection, prediction and recognition. The detailed processes are shown in Figure 1. This technology can automatically learn knowledge based on the data process in order to make accurate reaction, which generally can save a great deal of time.

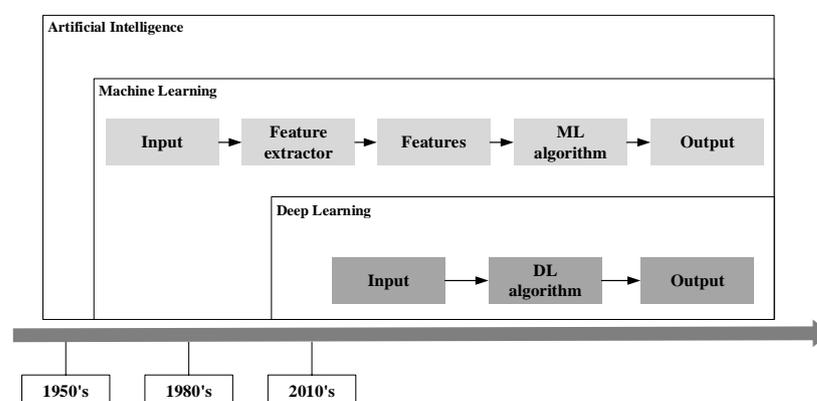


Figure 1. The trend of machine learning.

The deep learning approach generally requires massive amount of data for training which means that the more data there is to train a model, the better it will perform. However, experts are needed for manual identification and labelling of histological images. It subjects to potential time-consumption and high expense issues. Even the underlying method automatically pays attention to discriminative information for better classification, prospective validation studies are still required to firmly establish routine biomarker for clinical use. In short, highly trained pathologists remain the decision-makers during the subjective evaluation for cancer diagnosis. The techniques developed by deep learning can assist the doctors for more accurate projection but not to replace the duty of physicians.

2.3. How Deep Learning Works

Deep Learning is another major subfield of machine learning. Hubel [15] used it to find corresponding relationships between neuron systems based on the cortex cells. Deep learning is inspired by biological nervous systems, and combine multiple nonlinear processing layers and hidden layers to learn features directly from data. Hinton [16,17] proposed that using multi-hidden layers to learn features is conducive to classification, as shown in Figure 2.

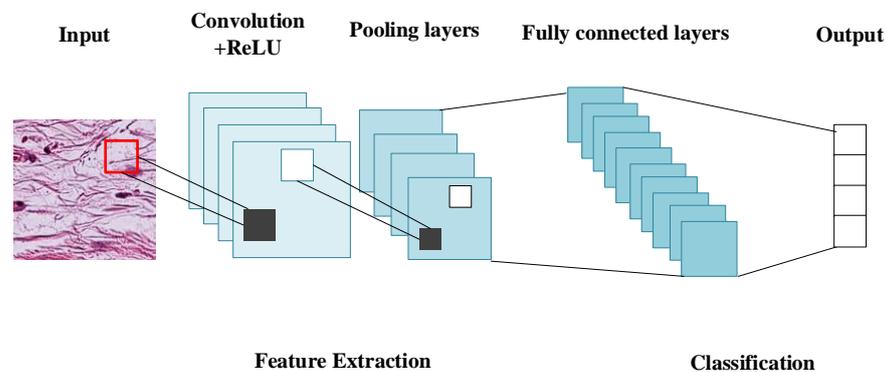


Figure 2. Network architecture of CNN.

The use of deep learning to learn features from multi-hidden layers of a large volume of data enhances the accuracy of predictions and a set of labels can be produced by using a GPU to train this model. Back propagation facilitates statistical regularity. Deep learning is based on the concept of learning from the first layer and automatically learning the features of many images from the combined layers. Each layer uses the output of the previous layer as input and learns to classify new images through to the next layer, and make a prediction.

Many different deep learning models have been developed for image recognition [18] over the past few years, such as histopathological image, facial recognition, and many advanced driver assistance technologies. CNN, which was proposed by Lecun [18,19], has a shared weights multi-neural layer network. The image is directly used as the input, which can reduce the complexity and parameters of the network. Besides, the structure of the network is invariant for image recognition. It is usually composed of two independent neurons, the first of which is an S layer of the characteristic extraction as the number of input connections to the mode, and the neurons are of equal weight. The feature map has displacement invariance, being activated by the small sigmoid function. Another neuron, feature C layer, is the feature extraction layer for anti-deformation. When part of the feature is extracted, its positional relationship with each input neuron connects with the previous layer.

A basic CNN consists of an input layer, output layer, and hidden layers, including ReLU, pooling layers, and fully connected layers:

- (1) **Input layer:** the input layer is the beginning of the artificial neural network, and brings the initial data, which comprises of a number of images, height, width, input channels, etc., into the system for further processing by subsequent layers of artificial neurons.

- (2) Convolutional layers: This layer is used to extract the various features from the input images, such as corners and edges. In this layer, the operation of convolution is performed between the input image and a filter of an image size, which are used to operate the convolution of the input data and set the stride of pixels to scan the full image. Later, this feature map is fed into other layers to learn several other features of the input image.
- (3) Rectified Linear Unit (ReLU): ReLU is the most common activation function in artificial neural networks. It is also known as transfer for better gradient propagation and efficient computation. $f(x) = \max(0, x)$. This function is defined as the positive part of its argument, where x is the input to a neuron, the function used to determine the output of the neural network, and then maps the resulting values between 0 to 1.
- (4) Pooling layers: The pooling layers provide an approach to address this sensitivity to down sampling features by summarizing their presence in patches of the feature map. This capability of local translation invariance has the effect of making the resulting features focus on changes in the position of the featured images.
- (5) Fully connected layers: This layer is the end of the network. It is often accompanied by the classifier to make the classification decision, and can be stacked.

2.4. CNN Architecture and Models

2.4.1. CNN Architecture

The structure of the CNN is designed for different purposes [20]. As seen in common neural networks, the neurons in a fully connected layer can be stacked and related to activations in the previous layer. This puts the filtered image on a higher level and puts it into a vote. With each additional layer, the network can learn more complex combinations of functions, which helps to make better decisions. These votes are expressed as the weight of the connection between each value or category. Therefore, their activation can be calculated by matrix multiplication and bias offset, and the main operations are performed by the backpropagation algorithm and the random gradient descent of the momentum algorithm, and fed back to the optimization method of weight update [20].

CNN is a cascaded filter, the first block of which is dedicated to detecting lower-level features (such as sharp points and surfaces folds), and the subsequent one is aggregated by the previous activation. From the perspective of deep learning, the main advantage of this architecture compared to traditional networks is that it can reduce the parameters of the image. CNN is composed of multiple connected kernels. Continuous layer learning features are gradually improved at the abstract level, and the input information can be represented hierarchically by combining low-level and high-level features. The objective of the fully connected layer is to take the results of the convolution to classify the images.

Currently, CNN [21] is widely used for image recognition by deep learning methods. Convolutional neural networks replace feature extraction, feature selection and classification. Different combinations of convolutional layers contain a series of fixed-size filters, which are used to manipulate the convolution of input data to generate so-called feature value maps (feature maps). During training in the context of convolutional neural networks, these filters can provide useful modules for image recognition, such as line detectors, regular edges and changes in image color. The ReLU layer usually follows the operation of the convolutional layer and provides a non-saturated activation function $f(x) = \max(0, x)$ for the output. According to Krizhevsky's research [22], these equations can be used to train fast-converging neural network convolutions, and can also solve the gradient problem, thereby accelerating the training.

2.4.2. Five Different CNN Models Networks

In this paper, we used five common deep neural networks based on CNN models and proposed an improved classification model for systematic colorectal cancer (CRC) tissue.

AlexNet

AlexNet [22] is a widely-applied deep convolutional neural network, which can still achieve a competitive performance in classification compared to other kinds of networks. In the training stage of the AlexNet model, the input image is resized to 224×224 pixels and fed into the network. The architecture of AlexNet firstly adopts a convolutional layer to perform convolution and max pooling with local response normalization (LRN) using 96 different size 11×11 receptive filters. The max pooling operations are performed with 3×3 filters with a stride size of 2. The same operations are performed in the second layer with 5×5 filters. The 3×3 filters are used in the third, fourth and fifth convolutional layers with 384, 384, and 296 feature maps respectively. The output of the two fully connected (FC) layers is used as an extracted feature vector with dropout followed by a softmax layer at the end.

SqueezeNet

SqueezeNet [23] is a small CNN architecture, which achieves AlexNet-level accuracy on ImageNet with 50x fewer parameters. Additionally, model compression techniques enabled us to compress SqueezeNet to less than 0.5 MB (510× smaller than AlexNet). The SqueezeNet begins with a standalone convolution layer (conv1), followed by 8 Fire modules (fire2–9), ending with a final convolution layer (conv10). We gradually increased the number of filters per fire module from the beginning to the end of the network. SqueezeNet performed max-pooling with a stride of 2 after layers conv1, fire4, fire8, and conv10.

VGGNet

VGGNET [24] was the runner up of the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The main contribution of VGGNET is that it shows that the depth of a network is a critical component to achieve better recognition or classification accuracy in CNNs. The VGG architecture consists of two convolutional layers both of which use the ReLU activation function, which is followed by a single max pooling layer and several fully connected layers, which also use the ReLU activation function. The final layer of the VGGNet model is a Softmax layer for classification. In addition, the size of the convolution filter is changed to a 3×3 filter with a span of 2 in VGG-E.

GoogLeNet

The GoogLeNet [25] main architecture improves the computing resources inside the network model to incorporate inception layers with the objective and reducing complexity. It not only increases the depth of the architectural approach (adding 1×1 convolutional layers to the network) with a different kernel, but also the width of the network. This reduces the number of computation layers to capture sparse correlation patterns.

ResNet

The ResNet [26] has a residual learning framework with ultra-deep networks, and the residual functions can ease networks that did not lose out from a vanishing gradient problem. Unexpectedly, as the depth of the ResNet framework network increases, the accuracy is saturated, but adding more layers causes training errors.

3. Research Method

Deep learning has gained enormous popularity in scientific computing due to CNN, and its algorithms are widely used by industries to solve complex problems. In this study, we observed different network architectures [22–26] for comparison purposes.

3.1. Experimental Steps

The diagram in Figure 3 illustrates the recognition process, which can be divided into three stages. The first stage is model training, the second is finding the superior architecture and parameters, and the third is model testing:

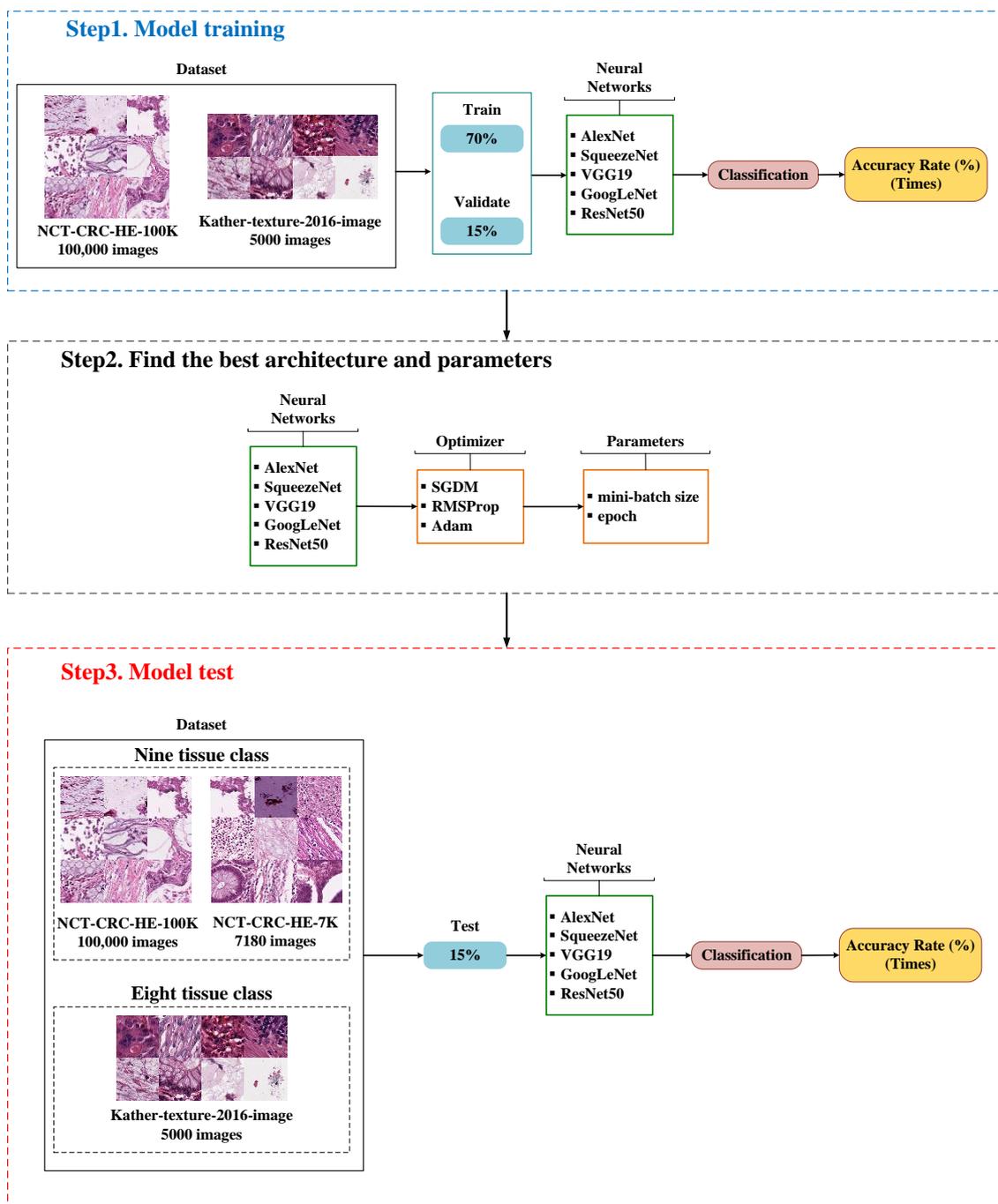


Figure 3. Deep learning structures to identify colorectal cancer tissue in this study.

Model Training

In this step, we used two different datasets. The first dataset we used was NCT-CRC-HE-100K [3] image documents of histological images with a 150×150 -pixel format, which included nine different tissue classes. We divided it into 70% training dataset, 15% validation dataset, and 15% test dataset. Since the number of each type in the original dataset was not the same, we used the number ratio of each type to take the corresponding number of training, validation and test for tissue classification in order to ensure the right proportion. Another dataset, the Kather-texture-2016-image [2] represents a collection of textures of eight tissue categories of human colorectal cancer in 5000 histological images of 150×150 pixels. Each image belongs to exactly one of the eight tissue categories, and the group sizes balanced (625 images per set). We divided it into 70% training dataset, 15%

validation dataset, and 15% test dataset, as shown in Table 2. Secondly, we used 5 different CNN models for training: AlexNet [22], SqueezeNet [23], VGG19 [24], GoogLeNet [25], and Resnet50 [26].

Table 2. CRC dataset of histological images.

Dataset	Diagnosis	Entire		Training		Validate		Testing	
		# WSI	(%)	# WSI	(%)	# WSI	(%)	# WSI	(%)
NCT-CRC-HE-100K [3]	ADI	10,407	10.41	7285	10.41	1561	10.41	1561	10.41
	BACK	10,566	10.57	7396	10.57	1585	10.57	1585	10.57
	DEB	11,512	11.51	8058	11.51	1727	11.51	1727	11.51
	LYM	11,557	11.56	8090	11.56	1734	11.56	1734	11.56
	MUC	8896	8.90	6227	8.90	1334	8.90	1334	8.90
	MUS	13,536	13.54	9475	13.54	2030	13.54	2030	13.54
	NORM	8763	8.76	6134	8.76	1314	8.76	1314	8.76
	STR	10,446	10.45	7312	10.45	1567	10.45	1567	10.45
	TUM	14,317	14.32	10,022	14.32	2148	14.32	2148	14.32
CRC-VAL-HE-7K [3]	ADI	1338	18.64	0	0	0	0	1338	18.64
	BACK	847	11.80	0	0	0	0	847	11.80
	DEB	339	4.72	0	0	0	0	339	4.72
	LYM	634	8.83	0	0	0	0	634	8.83
	MUC	1035	14.42	0	0	0	0	1035	14.42
	MUS	592	8.25	0	0	0	0	592	8.25
	NORM	741	10.32	0	0	0	0	741	10.32
	STR	421	5.86	0	0	0	0	421	5.86
	TUM	1233	17.17	0	0	0	0	1233	17.17
Kather-texture-2016-image [2]	TUMOR	625	78.125	468	12.48	93	12.48	93	12.48
	STROMA	625	78.125	468	12.48	93	12.48	93	12.48
	COMPLEX	625	78.125	468	12.48	93	12.48	93	12.48
	LYMPHO	625	78.125	468	12.48	93	12.48	93	12.48
	DEBRIS	625	78.125	468	12.48	93	12.48	93	12.48
	MUCOSA	625	78.125	468	12.48	93	12.48	93	12.48
	ADIPOSE	625	78.125	468	12.48	93	12.48	93	12.48
	EMPTY	625	78.125	468	12.48	93	12.48	93	12.48

Finding the Superior Architecture and Parameters

To gauge the performance of the network architectures of these CNN models, in the first research experiment we compared three training network method optimizers: stochastic gradient descent with momentum (SGDM), root mean square propagation (RMSProp) and adaptive moment estimation (Adam) with an NCT-CRC-HE-100K [3] dataset. As a result, the root-mean-square prop (RMSprop) method used the training options as input to the argument to select the training network to train, validate and test the CNN model to the highest accuracy, as shown in Table 3. In the second experiment, we trained the 5 different CNN models neural network with the root-mean-square prop (RMSprop) method. We further compared the replaced mini-batch size and epoch to test the models.

Table 3. Comparison of network optimizers (CRC-VAL-HE-100K).

Accuracy Rate (%) (Times)		Mini-Batch Size				
Algorithms	Epoch	8	16	32	64	128
sgdm	30	95.08 (668 min 16 s)	95.21 (709 min 44 s)	95.36 (761 min 6 s)	95.11 (821 min 44 s)	95.17 (889 min 22 s)
	25	95.09 (659 min 31 s)	95.19 (701 min 25 s)	95.35 (748 min 55 s)	95.17 (812 min 53 s)	95.16 (874 min 37 s)
	20	95.02 (647 min 22 s)	95.24 (691 min 15 s)	95.37 (741 min 32 s)	95.19 (809 min 47 s)	95.17 (870 min 49 s)
	15	94.54 (611 min 15 s)	95.24 (689 min 46 s)	95.37 (733 min 27 s)	95.21 (807 min 32 s)	95.17 (869 min 18 s)
	10	94.14 (592 min 46 s)	94.95 (680 min 21 s)	95.31 (726 min 41 s)	95.14 (806 min 58 s)	95.12 (858 min 45 s)
rmsprop	30	96.35 (853 min 24 s)	96.85 (852 min 17 s)	97.09 (805 min 53 s)	97.13 (828 min 44 s)	96.93 (912 min 17 s)
	25	96.34 (850 min 18 s)	96.88 (841 min 34 s)	97.22 (892 min 41 s)	97.14 (897 min 13 s)	96.98 (901 min 28 s)
	20	96.32 (847 min 6 s)	96.88 (838 min 21 s)	97.23 (867 min 44 s)	97.14 (893 min 36 s)	97.01 (900 min 39 s)
	15	96.32 (842 min 11 s)	96.85 (831 min 19 s)	97.22 (871 min 12 s)	97.14 (890 min 11 s)	97.03 (899 min 47 s)
	10	96.32 (838 min 37 s)	96.69 (820 min 47 s)	97.08 (869 min 44 s)	97.01 (884 min 19 s)	97.03 (898 min 7 s)
adam	30	95.28 (814 min 8 s)	95.59 (824 min 33 s)	96.36 (807 min 22 s)	95.38 (830 min 45 s)	94.20 (824 min 37 s)
	25	95.29 (810 min 39 s)	95.41 (822 min 14 s)	96.36 (806 min 35 s)	95.37 (818 min 50 s)	94.17 (820 min 49 s)
	20	95.28 (807 min 52 s)	95.38 (812 min 26 s)	96.37 (802 min 42 s)	95.37 (814 min 59 s)	94.19 (820 min 17 s)
	15	95.27 (796 min 6 s)	95.38 (800 min 32 s)	96.37 (803 min 15 s)	95.39 (809 min 38 s)	94.21 (815 min 29 s)
	10	95.27 (790 min 28 s)	95.38 (798 min 38 s)	95.36 (808 min 14 s)	95.38 (809 min 9 s)	93.21 (811 min 8 s)

Bold symbols represent the maximum values of each column in the tables.

Model Testing

The last stage of our architectural parameters involved classifying the histological images through each CNN model's neural network architecture by training the model to identify the types from different tissue classes. After neural network training all the 100,000 image patches (which were derived from 86 whole-slide images) in the first dataset (NCT-CRC-HE-100K), we used the dataset for testing purposes. Besides, we assessed the accuracy of the tissue classification and the convolutional neural network using an independent external dataset (CRC-VAL-HE-7K), which contained 7180 image patches derived from 25 hematoxylin and eosin (H&E) slides of human CRC tissue. Additionally, we used 70% of the dataset (Kather-texture-2016-image) consisting of 5000 images in eight classes of colorectal cancer tissue for training, 15% for validation and the other 15% for testing. We created a confusion matrix chart of the experimental results and showed the precision of each class using column and row summaries. The percentage of correctly and incorrectly classified observations for the true class were shown in the normalized row,

while the percentage of correctly and incorrectly classified observations for the predicted class were shown in the normalized column.

3.2. Data Availability

3.2.1. Images of Nine Tissue Classes

In this experiment, we used the open histological dataset of nine tissue classes from NCT-CRC-HE-100K for model training. These images were generated by Kather et al. [3], and have 86 hematoxylin and eosin stain (H&E) slides of tissue. The labels of the histological images of the available data were taken from the NCT-UMM website. The example images of the nine tissue classes are listed in Figure 4. The dimensions of all the images were 224×224 pixels ($112 \times 112 \mu\text{m}$), and they were presented to the model network sequentially for training, validation and testing. After training and testing our network framework with the “NCT-CRC-HE-100K”, we also assessed the accuracy of the tissue classification with an external validation set, data description “CRC-VAL-HE-7K”, which contained 7180 image patches for testing purposes only. The nine classes are categorized as following (Figure 4):

- (a) ADI: adipose tissue is mainly composed of adipocytes.
- (b) BACK: histological image background.
- (c) DEB: debris is widely used in histopathology and diagnoses.
- (d) LYM: lymphocytes are the main type of cells found in the lymphatic system.
- (e) MUC: mucus is produced by many tissues in the body, and acts as a protective force.
- (f) MUS: smooth muscle.
- (g) NORM TISSUE: tissues of colon mucosa.
- (h) STR: stroma tissues associated with cancer.
- (i) TUM: epithelium tissues of adenocarcinoma.

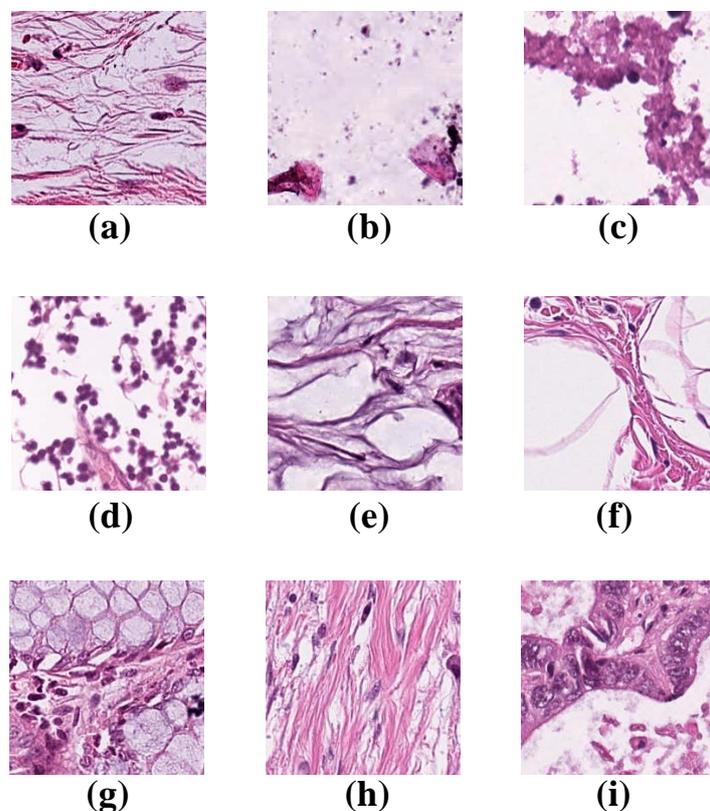


Figure 4. Example images of the nine tissue classes represented in the NCT-CRC-HE-100K dataset. (a–i) defined in Section 3.2.1.

3.2.2. Images of Eight Tissue Classes

We used the open dataset Kather-texture-2016-image to verify the accuracy of our optimized deep neural network architecture in distinguishing other tissue classes. This dataset was collected from the Institute of Histological Images of Pathology of Human Colorectal Cancer taken from the pathology archive by Kather, et al. [2]. The dataset consisted of 5000 non-duplicated histological images of human colorectal cancer (CRC) using hematoxylin and eosin (H&E) and healthy normal tissue images. This dataset created images with 150×150 pixels ($74 \times 74 \mu\text{m}$) each for every RGB color, and contained eight different tissue texture features and original tissue images with a size of 5000 pixels (e.g., Figure 5).

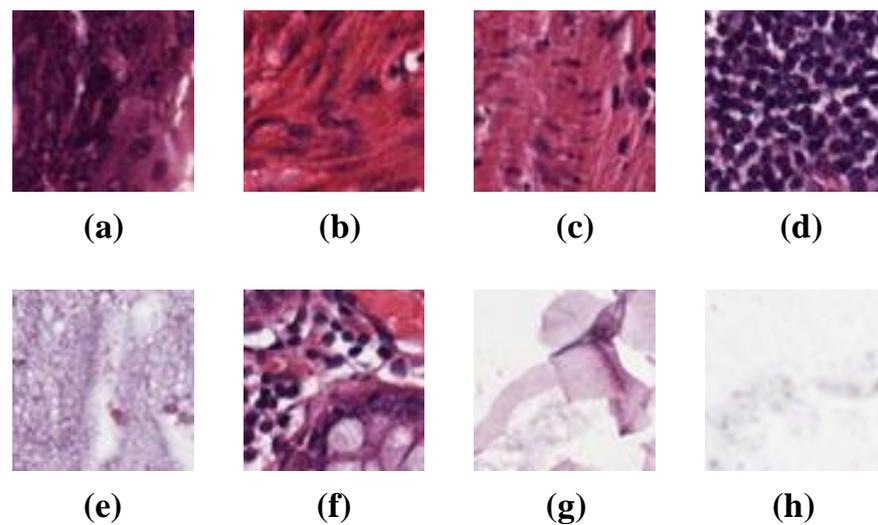


Figure 5. Example images of the eight tissue classes represented in the Kather-texture-2016-image-5000 dataset. (a–h) defined in Section 3.2.2.

The eight classes are categorized as following (Figure 5):

- (a) TUMOR: a tumor is an abnormal new growth of cells.
- (b) STROMA: stroma is the part of a tissue or organ with a structural or connective role.
- (c) COMPLEX: complex stroma contain a single or a few tissue cells.
- (d) LYMPHO: lymphoma is a group of blood malignancies that develop from lymphocytes.
- (e) DEBRIS: debris or H&E stain is one of the principal tissue stains used in histology.
- (f) MUCOSA: mucus is produced by many tissues in the body, and acts as a protective force.
- (g) ADIPOSE: adipose tissue is mainly composed of adipocytes.
- (h) EMPTY: histological image background.

3.3. Software and Tools Platform

In this study, we used the MATLAB of R2020a based on deep neural network architecture to train and test on two Intel workstation computers with high-level NVIDIA QUADRO GeForce GTX 1070 GPUs, and OS Windows 10 64-bit Core i7 i7-7700 with 3.60 GHz Processor (4 Core).

4. Experiments and Discussion

A series of experiments on different convolutional neural networks (CNNs) models were conducted in this study, including AlexNet, SqueezeNet, VGGNet, GoogLeNet and ResNet50. In Experiment I, we compared the accuracy rate of three training network optimizer methods: the stochastic gradient descent with momentum (SGDM), the root mean square propagation (RMSProp), which utilizes the magnitude of recent gradients to normalize the gradients, and the adaptive moment estimation (Adam), which is an

optimization algorithm that can be used for a classical stochastic gradient descent. In addition, parts of the parameters in the network layers were modified, such as the mini-batch size and different epoch. Next, we used our approach to identify the accuracy rate of the colorectal cancer tissue types from the histological images in different open datasets, and the results are presented in the next section.

4.1. Experiment I: Comparing the Accuracy Rate of Network Optimizers

4.1.1. Approach

Load and Explore Image Data

Firstly, we loaded three different open datasets of histological images. The first dataset consisted of 100,000 histological images (NCT-CRC-HE-100K) of colorectal cancer, including nine classes of tissues. The second dataset was (CRC-VAL-HE-7K), which contained 7180 image patches. The last dataset consisted of 5000 histological images of colorectal cancer, including eight different types of tissue. The image store automatically labeled the images based on folder names and efficiently read batches of images while training a convolutional neural network.

Randomly Split the Dataset

Next, we split the image dataset into three data stores: 70% into training data and 15% each into testing and validation, so that none of them overlapped with the others.

Define the Convolutional Neural Network Architecture

At this stage, we used 5 different CNN models for the training dataset: AlexNet [22], SqueezeNet [23], VGG19 [24], GoogLeNet [25], and Resnet50 [26]. The architecture included different convolutional layers, rectified linear units layer (ReLU layer), max-pooling layer, and fully connected layers.

Specify a Set of Options for Training

The network was trained using stochastic gradient descent with three training method optimizers after defining the network structure: stochastic gradient descent with momentum (SGDM), root mean square propagation (RMSProp), adaptive moment estimation (Adam) with an initial learning rate of 0.01 and four training periods on the entire dataset.

Train the Network

Train the network of histological images, and monitor the accuracy rate.

Predict Classification Accuracy

Predict the test data with three open datasets to calculate the final accuracy rate and execution time.

4.1.2. Experimental Results

Since deep learning technique will be adopted in this study, the performance of a CNN model depends on many factors in general. For example, the weight initialization, batch sizes, epochs, learning rates, activation function, optimizer, loss function, network topology, etc. The optimizer selection study of [27] for brain tumor segmentation in magnetic resonance images (MRI) suggests that a good optimizer could be a critical issue for the proposed approach. The authors of [27] listed 10 different state-of-the-art optimizer including: adaptive gradient (Adagrad), adaptive delta (AdaDelta), stochastic gradient descent (SGD), adaptive momentum (Adam), cyclic learning rate (CLR), adaptive max pooling (Adamax), root mean square propagation (RMS Prop), Nesterov adaptive momentum (Nadam), and Nesterov accelerated gradient (NAG) for CNN. The Adam optimizer achieved the best accuracy in study of [27] for MRI. Comprehensive analyses have been performed during this study for those optimizers. Based on the performance of final results, only SGDM,

RMSProp, Adam are listed since their performance are overall better than other optimizers for different network models.

Firstly, the open dataset description (Kather-texture-2016-image), which included 5000 images in eight tissue classes are trained and the results of the experiment are shown in Table 4, and the plot of the confusion matrix in Figure 6.

Table 4. Experiment I: the best results of eight tissue classes (Kather-texture-2016-image).

Accuracy Rate (%) (Times)		Algorithms		
Model	Sgdm	Rmsprop	Adam	
ResNet18	91.89 (17 min 22 s)	92.41 (21 min 15 s)	92.33 (25 min 51 s)	
ResNet50	93.27 (19 min 12 s)	94.18 (28 min 22 s)	94.02 (33 min 56 s)	
ResNet101	92.98 (21 min 19 s)	94.11 (32 min 33 s)	93.99 (43 min 43 s)	
GoogLeNet	91.89 (8 min 25 s)	92.17 (11 min 11 s)	92.13 (15 min 29 s)	
VGG19	88.91 (11 min 27 s)	90.85 (15 min 3 s)	89.94 (17 min 54 s)	
SqueezeNet	85.22 (14 min 57 s)	86.31 (16 min 01 s)	86.11 (19 min 29 s)	
AlexNet	87.93 (17 min 27 s)	88.81 (19 min 36 s)	88.24 (21 min 17 s)	

Bold symbols represent the maximum values of each row in the tables.

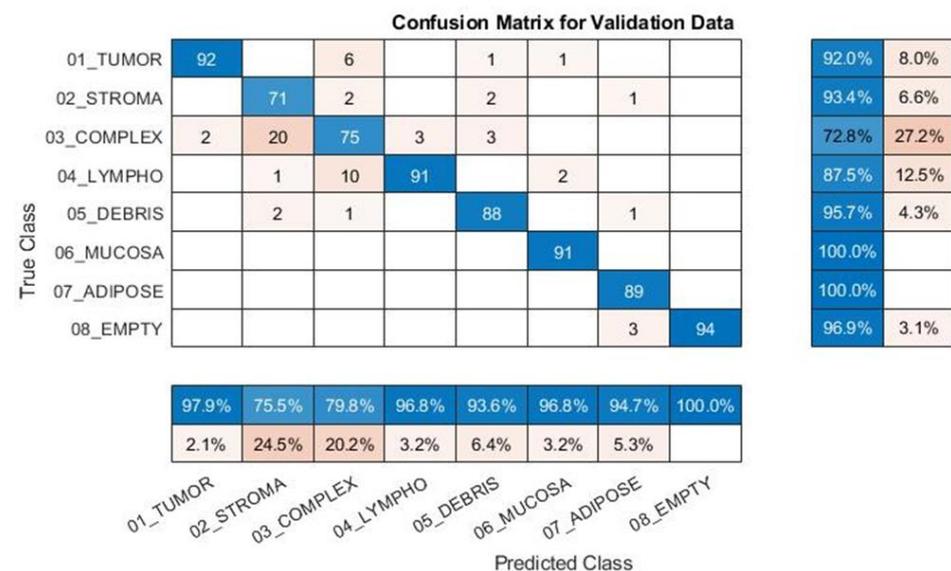


Figure 6. Experiment I: the accuracy of eight classes (Kather-texture-2016-image-5000).

Secondly, we used the data description (NCT-CRC-HE-100K) image documents with a 224 × 224-pixel format to classify the histological images, which included 100,000 images in nine different tissue classes, and display the precision for each class by using column and row summaries to plot the confusion matrix, as shown in Figure 7. In addition, we tested the classification performance with another independent set of 7000 images from different patients (CRC-VAL-HE-7K), and plotted the confusion matrix, as shown in Figure 8. The detailed results are shown in Table 5.

Among the selected optimizer, unlike Adam achieved the highest accuracy for brain tumor segmentation in magnetic resonance images, root-mean-square prop (RMSprop) network optimizer consistently have the highest accuracy rates for colorectal cancer tissues shown in Tables 4 and 5. Therefore, RMSprop optimizer will be adopted in the following experiments.

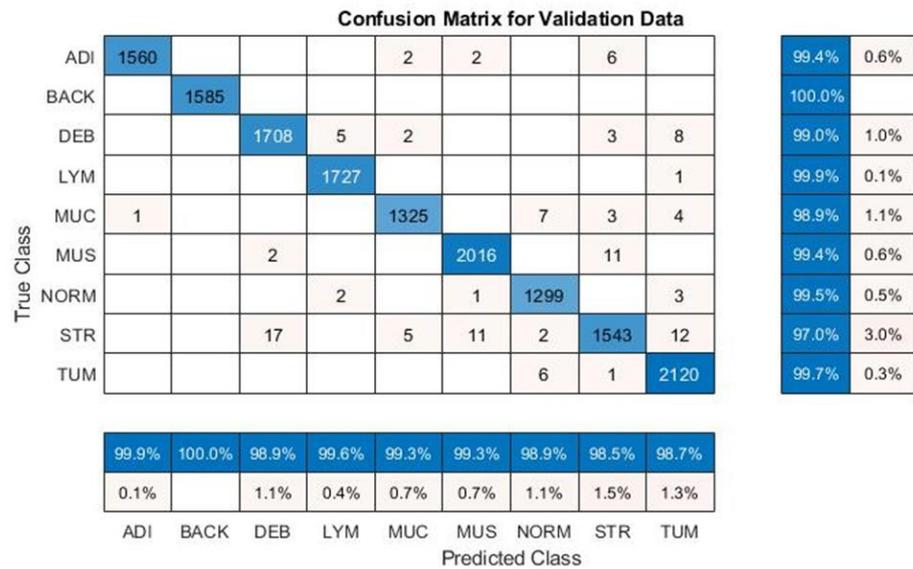


Figure 7. Experiment I: the accuracy of nine classes (CRC-VAL-HE-100K).

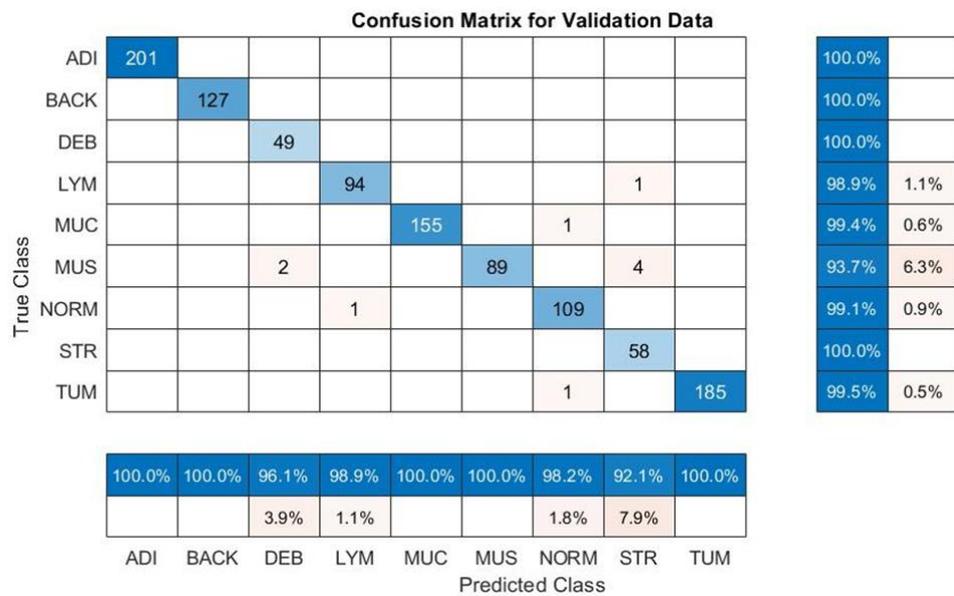


Figure 8. Experiment I: the accuracy of nine classes (CRC-VAL-HE-7K).

Table 5. Experiment I: the best results of nine tissue classes (CRC-VAL-HE-7K).

Accuracy Rate (%) (Times)		Algorithms	
Model	Sgdm	Rmsprop	Adam
ResNet18	94.81 (6 s)	96.33 (11 s)	95.93 (11 s)
ResNet50	96.94 (11 s)	97.22 (10 s)	96.37 (14 s)
ResNet101	95.35 (14 s)	97.14 (13 s)	96.82 (17 s)
GoogLeNet	95.89 (5 s)	97.05 (6 s)	96.18 (5 s)
VGG19	96.12 (3 s)	97.08 (3 s)	97.01 (7 s)
SqueezeNet	95.77 (4 s)	96.04 (6 s)	96.05 (13 s)
AlexNet	96.88 (14 s)	97.02 (17 s)	97.00 (10 s)

Bold symbols represent the maximum values of each row in the table.

4.2. Experiment II: Our Trained Deep Learning Approaches

4.2.1. Approach

Using the same split dataset method as shown in Experiment 1, we trained the neural network of five different CNN models with the most accurate optimizer: root-mean-square prop (RMSprop) and compared it with the replaced mini-batch size and epoch. Parts of the network layer were extracted to a new mode in the model revision process, which was used to extract the image features to modify the parameter. We improved this stage based on five cycles of the minimum batch size per training. We considered different models of convolutional neural networks (CNNs), such as AlexNet, SqueezeNet, VGGNet, GoogLeNet and ResNet for the classification of the pathological images.

The architectural design of the convolutional neural network (CNN) ResNet50 can be seen in Table 6. (ONV + POOL)_max represents the convolutional layer, followed by the use of the maximum generalized pooling layer, and (CONV + POOL)_avg is a pooling layer that follows the generalization of the average.

Table 6. Summary of parameters for ResNet50.

	Name	Type	Activations	Learnable Parameters
1	data $224 \times 224 \times 3$	Image Input	$224 \times 224 \times 3$	-
2	Conv1-7 × 7_s2	Convolution	$112 \times 112 \times 64$	Weights $7 \times 7 \times 3$ $\times 64$ Bias $1 \times 1 \times 64$
3	Conv1-relu_7 × 7	ReLU	$112 \times 112 \times 64$	-
4	Pool-3 × 3_s2	Max Pooling	$56 \times 56 \times 64$	-
5	Pool1-norm1	Cross Channel Normalization	$56 \times 56 \times 64$	-
6	Conv2-3 × 3_reduce	Convolution	$56 \times 56 \times 64$	Weights $1 \times 1 \times 3$ $\times 64$ Bias $1 \times 1 \times 64$

Table 6. Cont.

	Name	Type	Activations	Learnable Parameters
7	Conv2-relu_3 × 3_reduce	ReLU	56 × 56 × 64	-
8	Conv2-3 × 3	Convolution	56 × 56 × 192	Weights 3 × 3 × 64 × 192 Bias 1 × 1 × 192
9	Conv2-relu_3 × 3	ReLU	56 × 56 × 192	-
10	Conv2-norm2	Cross Channel Normalization	56 × 56 × 192	-
11	Pool2-3 × 3_s2	Max Pooling	28 × 28 × 192	-
12	Inception_3a-1 × 1	Convolution	28 × 28 × 64	Weights 1 × 1 × 192 × 64 Bias 1 × 1 × 64
⋮	⋮	⋮	⋮	⋮
144	Output	Classification Output	-	-

4.2.2. Experimental Results

In the second part, we compared five different CNN networks with the replaced mini-batch size and epoch from the network to form a new model. We began by using the data “NCT-CRC-HE-100K” documents of histological images for the model training with a 224 × 224-pixel format for classification, including 100,000 images of nine different tissue classes, and displaying the precision of each class by using column and row summaries to plot the confusion matrix, as shown in Figure 9. The detailed results are shown in Table 7. Secondly, we tested the classification performance in an independent dataset of 7180 images from “CRC-VAL-HE-7K”, and plotted the confusion matrix as shown in Figure 10. The detailed results are shown in Table 8. In addition, we used the open dataset description “Kather-texture-2016-image”, which included 5000 images in eight tissue classes. The experimental results are shown in Table 9, and the plot of the confusion matrix is shown in Figure 11.

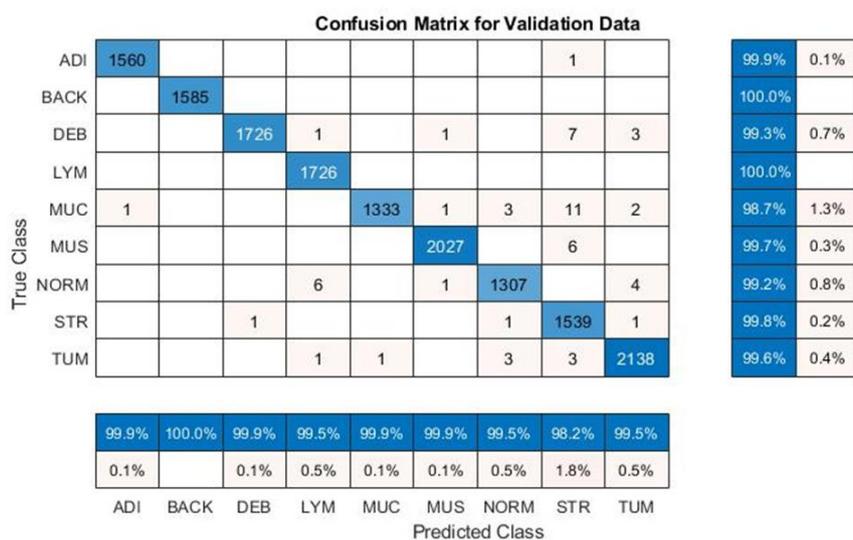


Figure 9. Experiment II: accuracy of nine classes (CRC-VAL-HE-100K).

Table 7. Experiment II: the best results of nine tissue classes (CRC-VAL-HE-100K).

Accuracy Rate (%) (Times)		Mini-Batch Size				
Model	Epoch	8	16	32	64	128
ResNet18	30	98.12 (689 min 43 s)	98.27 (707 min 39 s)	98.60 (709 min 45 s)	98.65 (711 min 45 s)	98.48 (720 min 08 s)
	25	98.09 (671 min 22 s)	98.27 (692 min 44 s)	98.61 (694 min 17 s)	98.64 (698 min 55 s)	98.47 (799 min 49 s)
	20	98.09 (665 min 19 s)	98.24 (680 min 18 s)	98.61 (681 min 22 s)	98.66 (690 min 29 s)	98.49 (796 min 05 s)
	15	98.02 (639 min 28 s)	98.07 (669 min 27 s)	98.61 (672 min 18 s)	98.66 (684 min 05 s)	98.48 (789 min 34 s)
	10	98.01 (622 min 17 s)	98.01 (643 min 19 s)	98.02 (666 min 16 s)	98.44 (771 min 33 s)	98.48 (770 min 31 s)
ResNet50	30	98.35 (757 min 08 s)	99.02 (758 min 31 s)	99.67 (769 min 04 s)	99.46 (790 min 22 s)	99.41 (792 min 12 s)
	25	98.29 (739 min 33 s)	99.02 (756 min 05 s)	99.68 (762 min 18 s)	99.46 (788 min 09 s)	99.45 (789 min 55 s)
	20	98.31 (733 min 29 s)	98.89 (749 min 22 s)	99.69 (756 min 03 s)	99.46 (781 min 26 s)	99.44 (782 min 17 s)
	15	98.21 (719 min 16 s)	98.72 (736 min 05 s)	99.68 (731 min 07 s)	99.46 (768 min 17 s)	99.45 (769 min 38 s)
	10	98.21 (709 min 15 s)	98.77 (728 min 17 s)	99.68 (729 min 19 s)	99.45 (742 min 11 s)	99.45 (761 min 21 s)
ResNet101	30	98.89 (797 min 21 s)	99.03 (792 min 27 s)	99.30 (798 min 41 s)	98.79 (809 min 39 s)	98.50 (829 min 10 s)
	25	98.81 (789 min 17 s)	99.03 (790 min 07 s)	99.31 (793 min 28 s)	98.77 (795 min 48 s)	98.54 (827 min 01 s)
	20	98.79 (780 min 54 s)	99.02 (787 min 52 s)	99.32 (788 min 18 s)	98.81 (794 min 27 s)	98.54 (818 min 35 s)
	15	98.77 (775 min 15 s)	99.02 (778 min 17 s)	99.31 (780 min 26 s)	98.81 (761 min 39 s)	98.59 (815 min 29 s)
	10	98.77 (771 min 17 s)	99.01 (773 min 32 s)	99.30 (776 min 09 s)	98.78 (759 min 05 s)	98.59 (813 min 16 s)
GoogLeNet	30	98.01 (523 min 29 s)	98.53 (529 min 49 s)	98.54 (527 min 49 s)	98.49 (532 min 47 s)	98.43 (654 min 51 s)
	25	97.75 (509 min 17 s)	98.53 (510 min 24 s)	98.56 (525 min 37 s)	98.48 (528 min 44 s)	98.45 (639 min 51 s)
	20	97.58 (501 min 22 s)	98.51 (509 min 17 s)	98.56 (513 min 18 s)	98.49 (516 min 15 s)	98.44 (617 min 42 s)
	15	97.44 (498 min 27 s)	98.47 (506 min 24 s)	98.57 (507 min 09 s)	98.49 (508 min 27 s)	98.45 (612 min 33 s)
	10	97.02 (485 min 54 s)	98.47 (501 min 35 s)	98.56 (503 min 38 s)	98.48 (507 min 44 s)	98.45 (592 min 37 s)
VGG19	30	98.49 (576 min 43 s)	98.49 (578 min 26 s)	98.49 (581 min 52 s)	98.48 (503 min 37 s)	98.41 (509 min 33 s)
	25	98.45 (558 min 17 s)	98.49 (566 min 33 s)	98.51 (577 min 49 s)	98.46 (496 min 47 s)	98.46 (504 min 18 s)
	20	98.45 (537 min 55 s)	98.49 (541 min 34 s)	98.52 (562 min 22 s)	98.43 (470 min 29 s)	98.49 (501 min 24 s)
	15	98.46 (528 min 18 s)	98.49 (527 min 15 s)	98.52 (512 min 06 s)	98.43 (451 min 14 s)	98.46 (489 min 46 s)
	10	98.44 (501 min 09 s)	98.49 (518 min 48 s)	98.51 (512 min 48 s)	98.42 (438 min 42 s)	98.46 (484 min 17 s)
SqueezeNet	30	98.07 (522 min 24 s)	98.21 (524 min 45 s)	98.31 (549 min 42 s)	98.22 (456 min 16 s)	98.29 (486 min 19 s)
	25	98.06 (519 min 18 s)	98.25 (520 min 36 s)	98.27 (540 min 15 s)	98.19 (441 min 18 s)	98.31 (483 min 28 s)
	20	98.07 (511 min 04 s)	98.21 (516 min 49 s)	98.29 (538 min 57 s)	98.18 (438 min 45 s)	98.33 (469 min 34 s)
	15	98.06 (509 min 19 s)	98.19 (510 min 39 s)	98.30 (529 min 42 s)	98.14 (435 min 57 s)	98.42 (463 min 05 s)
	10	98.06 (501 min 31 s)	98.25 (509 min 24 s)	98.26 (519 min 16 s)	98.18 (424 min 19 s)	98.31 (453 min 01 s)
AlexNet	30	97.71 (535 min 45 s)	97.73 (539 min 55 s)	97.77 (541 min 07 s)	97.39 (451 min 38 s)	97.72 (473 min 17 s)
	25	97.69 (532 min 44 s)	97.73 (534 min 49 s)	97.76 (539 min 33 s)	97.28 (447 min 42 s)	97.69 (466 min 35 s)
	20	97.72 (532 min 19 s)	97.70 (533 min 16 s)	97.81 (539 min 17 s)	97.68 (443 min 22 s)	97.91 (457 min 08 s)
	15	97.70 (531 min 28 s)	97.69 (531 min 49 s)	97.77 (534 min 16 s)	97.24 (441 min 17 s)	97.67 (453 min 03 s)
	10	97.71 (530 min 46 s)	97.70 (531 min 04 s)	97.70 (533 min 35 s)	97.23 (438 min 19 s)	97.67 (423 min 39 s)

Bold symbols represent the maximum values of each column in the tables.

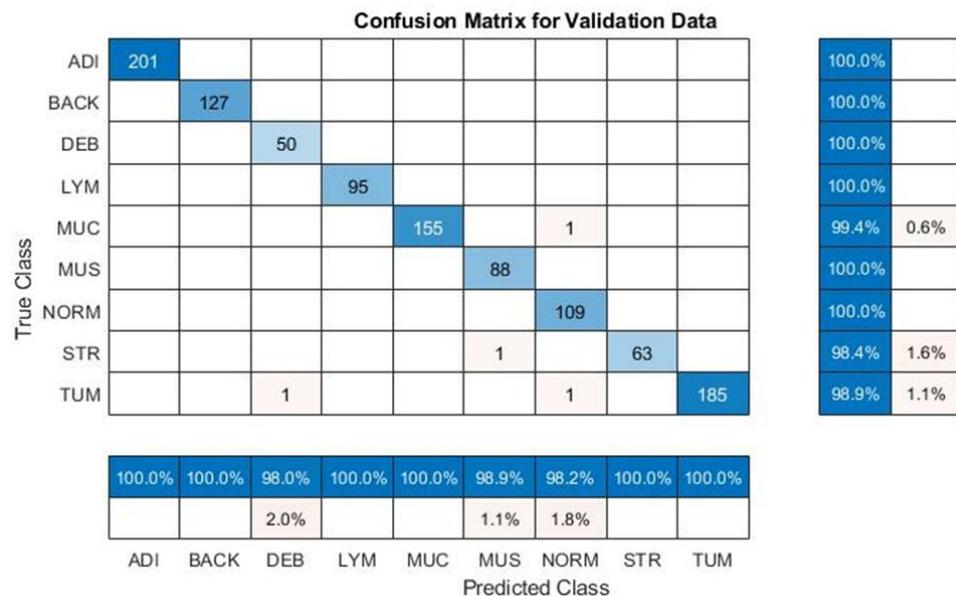


Figure 10. Experiment II: accuracy of nine classes (CRC-VAL-HE-7K).

Table 8. Experiment II: the best result of nine tissue classes (CRC-VAL-HE-7K).

Accuracy Rate (%) (Times)		Mini-Batch Size				
Model	Epoch	8	16	32	64	128
ResNet18	30	98.03 (11 s)	98.07 (14 s)	98.10 (16 s)	98.15 (18 s)	98.08 (18 s)
	25	98.01 (10 s)	98.07 (12 s)	98.11 (13 s)	98.04 (16 s)	98.07 (18 s)
	20	97.91 (7 s)	98.04 (7 s)	98.20 (10 s)	98.16 (12 s)	97.99 (16 s)
	15	97.92 (5 s)	97.97 (6 s)	98.09 (12 s)	98.16 (12 s)	97.11 (14 s)
	10	97.91 (4 s)	97.91 (4 s)	98.08 (8 s)	98.14 (9 s)	97.08 (11 s)
ResNet50	30	98.56 (14 s)	99.01 (16 s)	99.28 (17 s)	99.09 (19 s)	99.10 (20 s)
	25	98.53 (14 s)	99.01 (14 s)	99.31 (15 s)	99.15 (18 s)	99.17 (19 s)
	20	98.51 (7 s)	98.90 (7 s)	99.32 (9 s)	99.19 (12 s)	99.09 (12 s)
	15	98.51 (5 s)	98.88 (7 s)	99.30 (9 s)	99.27 (10 s)	99.09 (12 s)
	10	98.51 (5 s)	98.81 (6 s)	99.30 (7 s)	99.14 (9 s)	99.19 (11 s)
ResNet101	30	98.26 (15 s)	97.81 (16 s)	98.07 (18 s)	98.75 (20 s)	98.56 (20 s)
	25	98.25 (14 s)	97.92 (15 s)	98.29 (16 s)	98.71 (17 s)	98.54 (19 s)
	20	98.29 (14 s)	97.92 (15 s)	98.41 (16 s)	98.76 (17 s)	98.33 (19 s)
	15	98.31 (10 s)	97.83 (11 s)	98.31 (12 s)	98.74 (12 s)	98.19 (13 s)
	10	98.19 (10 s)	97.90 (10 s)	98.30 (11 s)	98.71 (12 s)	98.09 (13 s)
GoogLeNet	30	97.22 (7 s)	98.03 (8 s)	98.04 (8 s)	98.13 (9 s)	98.04 (11 s)
	25	97.34 (7 s)	98.03 (7 s)	98.16 (7 s)	98.17 (8 s)	98.27 (10 s)
	20	97.29 (6 s)	98.01 (7 s)	98.16 (7 s)	98.19 (8 s)	98.23 (9 s)
	15	97.17 (6 s)	97.97 (6 s)	98.07 (7 s)	98.17 (7 s)	98.12 (9 s)
	10	97.06 (6 s)	97.97 (6 s)	98.06 (6 s)	98.18 (7 s)	98.19 (7 s)
VGG19	30	97.22 (5 s)	97.32 (5 s)	97.51 (7 s)	97.55 (7 s)	97.01 (8 s)
	25	97.05 (4 s)	97.11 (4 s)	97.47 (5 s)	97.22 (6 s)	97.16 (6 s)
	20	97.07 (4 s)	97.17 (5 s)	97.51 (5 s)	97.29 (5 s)	97.03 (6 s)
	15	97.19 (4 s)	97.13 (4 s)	97.49 (4 s)	97.21 (5 s)	97.14 (6 s)
	10	97.11 (4 s)	97.14 (4 s)	96.37 (4 s)	97.16 (4 s)	97.07 (5 s)
SqueezeNet	30	96.95 (8 s)	96.13 (8 s)	96.09 (8 s)	97.24 (9 s)	97.32 (11 s)
	25	96.95 (5 s)	96.11 (6 s)	96.07 (6 s)	97.17 (7 s)	97.21 (10 s)
	20	96.91 (5 s)	96.18 (5 s)	96.11 (6 s)	97.12 (7 s)	97.40 (9 s)
	15	96.90 (5 s)	96.17 (5 s)	96.07 (5 s)	97.14 (7 s)	97.43 (9 s)
	10	96.90 (4 s)	96.11 (4 s)	96.01 (5 s)	97.14 (6 s)	97.44 (8 s)
AlexNet	30	95.11 (9 s)	95.27(9 s)	95.29 (11 s)	95.07 (13 s)	95.17 (13 s)
	25	95.05 (9 s)	95.27 (10 s)	95.28 (10 s)	95.08 (12 s)	95.16 (14 s)
	20	95.03 (8 s)	95.25 (10 s)	95.30 (10 s)	95.07 (11 s)	95.18 (11 s)
	15	94.97 (8 s)	95.22 (8 s)	95.28 (9 s)	95.08 (10 s)	95.18 (11 s)
	10	94.94 (7 s)	94.25 (8 s)	95.27 (8 s)	95.07 (10 s)	95.14 (9 s)

Bold symbols represent the maximum values of each column in the tables.

By the same token, based on the experimental results, it can be seen that, when revising the parameters, ResNet50 was found to have achieved the highest accuracy rate at 15 epoch shown in Figure 12a, and 32 mini-batch size for nine classes of CRC images, as shown in Figure 12b. Furthermore, the same parameters of the ResNet50 neural network used for eight types of CRC images achieved a ratio of 94.86% accuracy, as shown in Figure 13a,b. Further extensive experiments have been conducted to verify the efficacy of different variants of the ResNet architecture, such as ResNet18, ResNet50 and ResNet101 [26]. It is also worth noting that an accuracy rate of 99.69% can be achieved using 177 layers of a neural network (ResNet50), which is better than the 98.61% using 71 layers of ResNet18 and the 99.31% using 347 layers of ResNet101. Furthermore, an accuracy rate of 94.86% can be achieved using 177 layers and the same parameters of a ResNet50 neural network for eight classes of CRC images, which is better than the 92.86% using 71 layers of ResNet18 and the 94.16%, using 347 layers of ResNet101. The differences between ResNet18, ResNet50 and ResNet101 are highlighted in Figure 14. It can be seen from the previous experiments

that the best classification accuracy rate can achieved by revising the parameters and using ResNet50.

Table 9. Experiment II: the best result of eight tissue classes (Kather-texture-2016-image).

Accuracy Rate (%) (Times)		Mini-Batch Size				
Model	Epoch	8	16	32	64	128
ResNet18	30	94.12 (38 min 41 s)	94.14 (39 min 09 s)	94.16 (39 min 28 s)	94.18 (39 min 47 s)	93.24 (40 min 15 s)
	25	94.12 (37 min 22 s)	94.16 (37 min 49 s)	94.11 (38 min 19 s)	94.14 (38 min 49 s)	93.28 (39 min 42 s)
	20	94.11 (37 min 35 s)	94.16 (37 min 21 s)	94.15 (37 min 52 s)	94.16 (38 min 11 s)	93.29 (38 min 39 s)
	15	94.09 (37 min 06 s)	94.17 (35 min 38 s)	94.16 (36 min 43 s)	94.21 (36 min 27 s)	93.27 (37 min 54 s)
	10	94.10 (35 min 49 s)	94.13 (35 min 17 s)	94.13 (36 min 22 s)	94.12 (36 min 19 s)	93.28 (36 min 21 s)
ResNet50	30	94.38 (41 min 11 s)	94.45 (41 min 35 s)	94.85 (43 min 41 s)	94.56 (43 min 52 s)	94.71 (44 min 29 s)
	25	94.38 (40 min 57 s)	94.45 (41 min 09 s)	94.85 (42 min 18 s)	94.54 (43 min 16 s)	94.78 (44 min 48 s)
	20	94.35 (40 min 23 s)	94.43 (40 min 44 s)	94.86 (40 min 22 s)	94.56 (41 min 25 s)	94.72 (43 min 45 s)
	15	94.32 (39 min 43 s)	94.41 (40 min 01 s)	94.86 (40 min 18 s)	94.56 (40 min 07 s)	94.79 (43 min 52 s)
	10	94.33 (39 min 17 s)	94.38 (39 min 44 s)	94.77 (40 min 02 s)	94.55 (40 min 27 s)	94.79 (43 min 31 s)
ResNet101	30	92.59 (41 min 32 s)	92.69 (42 min 54 s)	92.74 (42 min 57 s)	92.66 (43 min 17 s)	91.52 (53 min 41 s)
	25	92.59 (41 min 18 s)	92.69 (41 min 31 s)	92.76 (42 min 26 s)	92.65 (43 min 49 s)	91.50 (53 min 16 s)
	20	92.55 (40 min 55 s)	92.71 (41 min 18 s)	92.76 (41 min 46 s)	92.66 (42 min 59 s)	91.49 (52 min 58 s)
	15	92.48 (40 min 49 s)	92.67 (41 min 01 s)	92.75 (41 min 36 s)	92.66 (41 min 43 s)	91.50 (52 min 19 s)
	10	92.45 (40 min 44 s)	92.69 (40 min 52 s)	92.76 (41 min 18 s)	92.65 (41 min 37 s)	91.50 (51 min 47 s)
GoogLeNet	30	90.61 (30 min 31 s)	91.39 (30 min 45 s)	92.39 (31 min 47 s)	92.36 (31 min 59 s)	92.16 (31 min 30 s)
	25	90.59 (30 min 17 s)	91.39 (30 min 42 s)	92.39 (30 min 53 s)	92.36 (31 min 56 s)	92.17 (30 min 46 s)
	20	90.60 (30 min 01 s)	91.37 (30 min 19 s)	92.42 (30 min 17 s)	92.36 (31 min 44 s)	92.15 (30 min 17 s)
	15	90.60 (29 min 55 s)	91.32 (30 min 11 s)	92.33 (30 min 46 s)	92.36 (31 min 09 s)	92.14 (29 min 52 s)
	10	90.59 (29 min 34 s)	91.33 (29 min 54 s)	92.36 (30 min 34 s)	92.36 (30 min 54 s)	92.15 (29 min 29 s)
VGG19	30	90.91 (26 min 29 s)	90.86 (26 min 17 s)	91.30 (26 min 17 s)	91.29 (27 min 21 s)	90.89 (27 min 29 s)
	25	90.90 (25 min 37 s)	90.86 (25 min 51 s)	91.35 (25 min 59 s)	91.25 (26 min 48 s)	90.87 (27 min 16 s)
	20	90.91 (25 min 11 s)	90.85 (25 min 32 s)	91.37 (25 min 22 s)	91.29 (24 min 27 s)	90.89 (24 min 54 s)
	15	90.91 (24 min 58 s)	90.84 (25 min 08 s)	91.31 (25 min 08 s)	91.29 (24 min 03 s)	90.89 (24 min 22 s)
	10	90.89 (24 min 36 s)	90.84 (24 min 27 s)	91.32 (24 min 43 s)	91.27 (23 min 44 s)	90.89 (24 min 09 s)
SqueezeNet	30	88.09 (25 min 27 s)	88.16 (25 min 44 s)	88.17 (26 min 14 s)	88.24 (26 min 47 s)	87.37 (28 min 49 s)
	25	88.09 (25 min 14 s)	88.17 (25 min 28 s)	88.17 (25 min 52 s)	88.25 (26 min 19 s)	87.29 (27 min 42 s)
	20	88.07 (25 min 02 s)	88.17 (25 min 25 s)	88.17 (25 min 44 s)	88.24 (25 min 58 s)	87.24 (27 min 19 s)
	15	88.09 (24 min 46 s)	88.18 (25 min 09 s)	88.11 (25 min 26 s)	88.27 (25 min 37 s)	87.29 (26 min 27 s)
	10	88.09 (24 min 25 s)	88.17 (24 min 49 s)	88.08 (24 min 38 s)	88.24 (25 min 06 s)	87.19 (25 min 38 s)
AlexNet	30	90.18 (30 min 05 s)	91.94 (31 min 27 s)	91.92 (32 min 44 s)	91.91 (33 min 52 s)	89.40 (45 min 05 s)
	25	90.17 (29 min 45 s)	91.94 (31 min 43 s)	91.94 (31 min 16 s)	91.93 (33 min 27 s)	89.42 (45 min 19 s)
	20	90.18 (29 min 33 s)	91.94 (30 min 22 s)	91.94 (30 min 41 s)	91.91 (32 min 41 s)	89.44 (45 min 37 s)
	15	90.18 (29 min 12 s)	91.94 (29 min 49 s)	91.91 (30 min 22 s)	91.94 (32 min 36 s)	89.41 (46 min 27 s)
	10	90.19 (28 min 58 s)	91.94 (29 min 23 s)	91.87 (29 min 27 s)	91.94 (30 min 11 s)	89.49 (45 min 44 s)

Bold symbols represent the maximum values of each column in the tables.

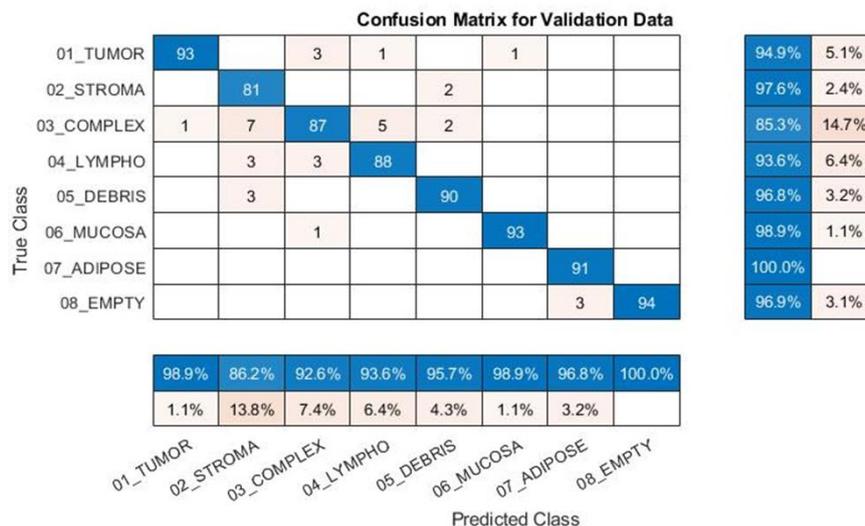


Figure 11. Experiment II: accuracy of eight classes (Kather-texture-2016-image-5000).

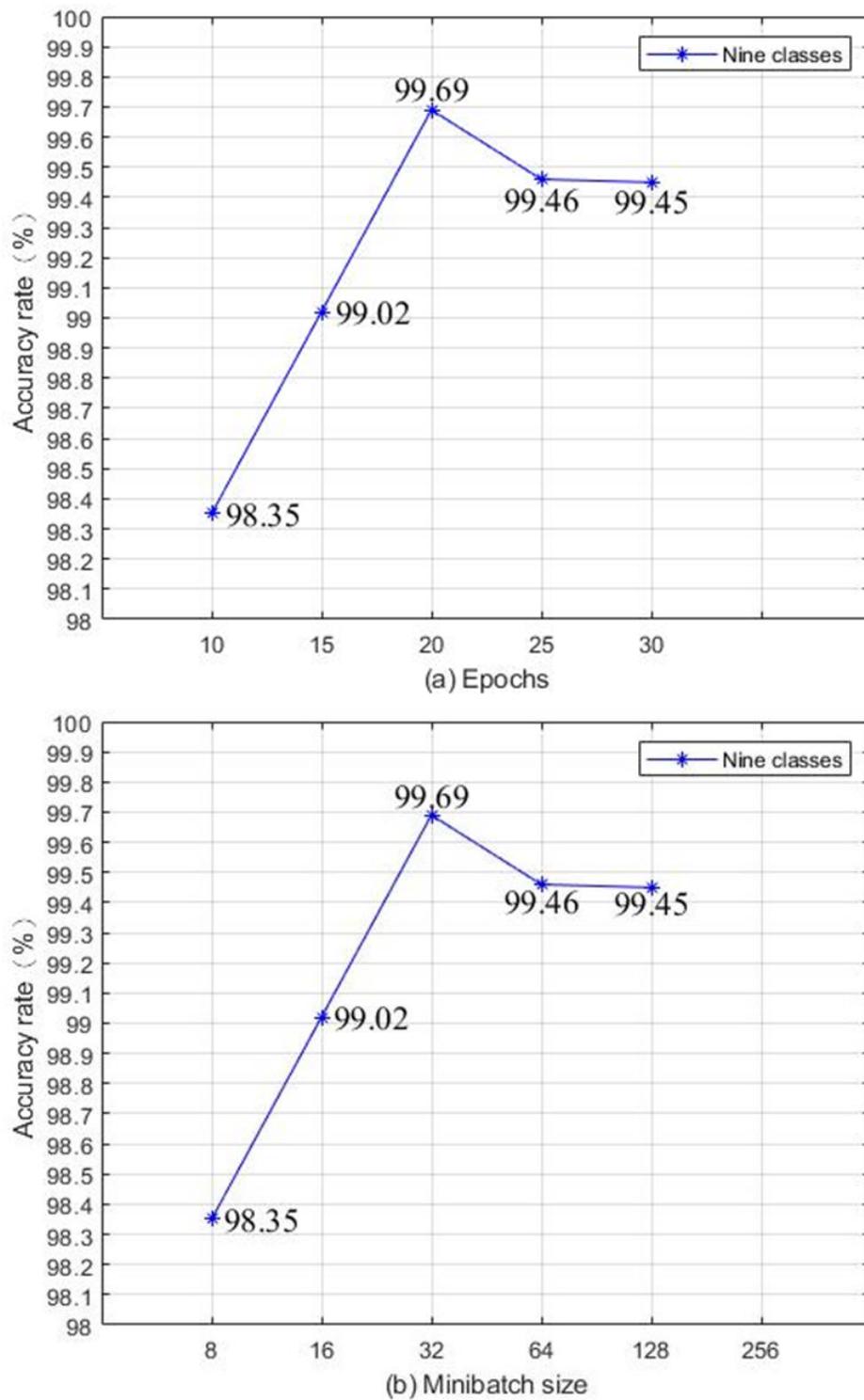


Figure 12. Illustration of the accuracy of nine tissue classes (CRC-VAL-HE-100K) using (a) epoch and (b) mini-batch size.

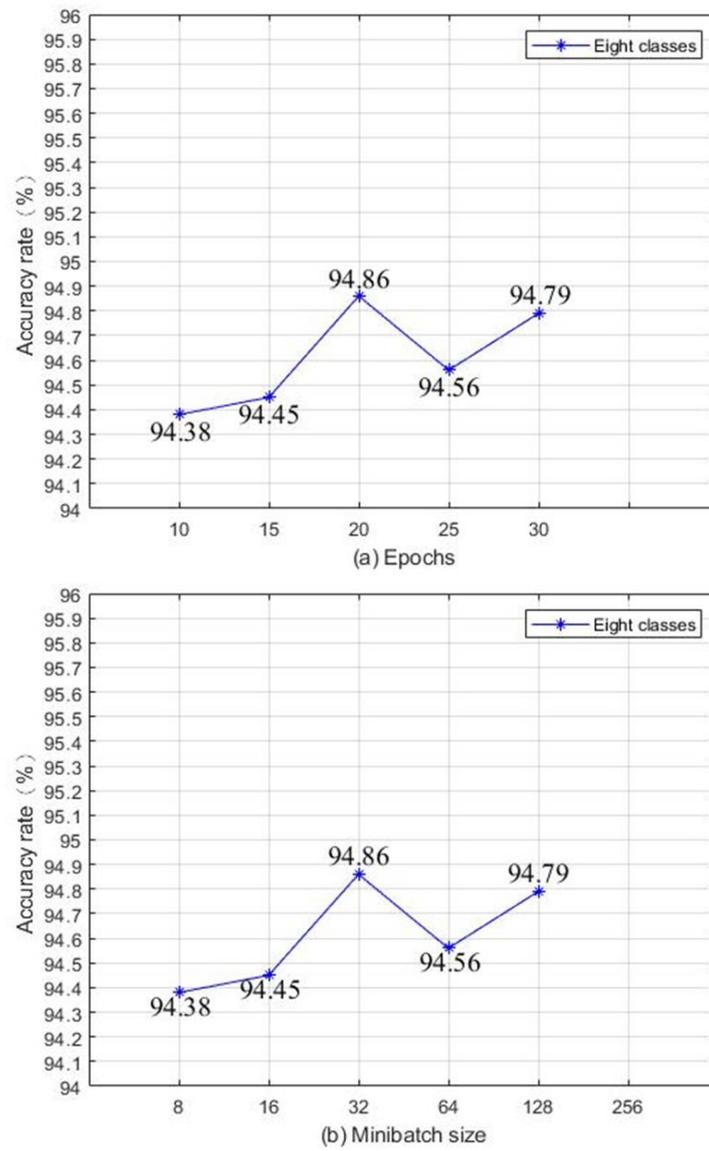


Figure 13. Illustration of the accuracy of eight tissue classes (Kather-texture-2016-image-5000) using (a) epoch and (b) mini-batch size.

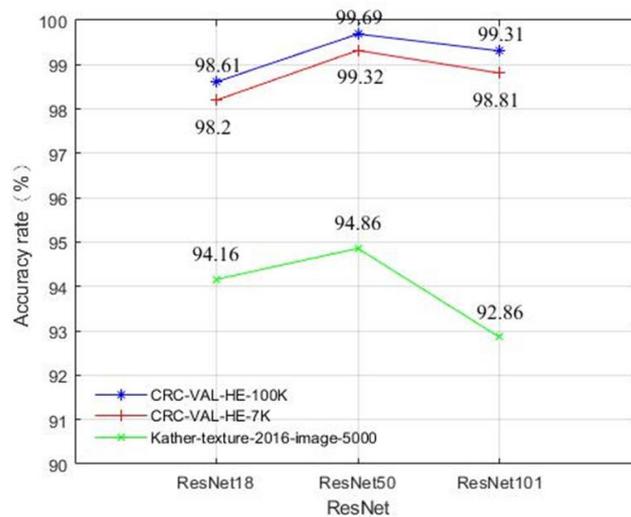


Figure 14. Accuracy rates of ResNet.

4.3. Discussion

After the detailed explanation of the approach and experiments, it is necessary to compare the performance of the proposed techniques with published data. In Reference [13], Kather et al. applied the same NCT-HE-100K data set of 100,000 histological images to train a VGG19 CNN model and tested the classification performance in an independent set of 7180 images from different patients (CRC-VAL-HE-7K). The overall nine-class accuracy was close to 99% in an internal testing set and 94.3% in an external testing set. Unlike the approach in [13], the experimental results of ResNet50 outperform the data of VGG19 in Table 7. We had achieved 99.69% accuracy rate in the same internal testing set and 99.32% in the same external testing set from Figure 14. Through comprehensive and thorough analyses, this study suggests that ResNet50 could be a better deep learning architecture for colorectal cancer tissue than VGG19.

To further validate our claim, the independent data set with eight classes of [2] is also utilized for comparison purpose. Through our study, ResNet50 achieved 94.86% accuracy in Figure 14 and [2] reported the best accuracy rate was 87.4%. Through comprehensive studies and comparison, it is highly suggested that ResNet50 with suggested settings of this study could be the most efficient and accuracy deep learning techniques to classify colorectal cancer tissue.

Since deep neural networks have been adopted in this study as the classifier, the modular design of those models conveniently provides their architecture to specific needs. Many factors could easily be modified like weight initialization, batch sizes, epochs, learning rates, activation function, optimizer, loss function, network topology, etc. to improve the classification accuracy. Among various settings for the superior classification performance, several studies [28–30] have suggested that loss function could be critical to affect the deep learning models and learning efficiency, as well as the classifier robustness to various situation.

In this study, the authors adopted the transfer learning of deep learning architecture for the classification of colorectal cancer tissue, those network models are optimized based on the pre-trained data from ImageNet [22]. Since ImageNet is a large labeled dataset of real-world images, it is one of the most widely used dataset in latest computer vision research and several well-known models are the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners. The loss functions for each network adopted in this study are cross entropy loss function for all network models. The authors would pay extra attention to optimize the selection of loss function for future research in order to further improve the overall accuracy, class imbalance awareness and convergence speed for the classification of colorectal cancer tissue. Therefore, our research could effectively classify the medical images in aiding clinical care and treatment.

5. Conclusions

This study was based on exploring different deep learning models for the recognition of colorectal cancer tissue using CNN. An improved version of deep learning parameters was proposed in this article to improve the accuracy of classification. In order to verify our optimized parameters, we used CRC histological images as the experimental dataset, and compared the ability of the five most commonly-used deep learning network models to accurately distinguish colorectal cancer tissues. Based on the experimental results, our method was superior to the techniques described in the literature and achieved a high recognition rate. In summary, the nine-class accuracy of NCT-HE-100K data set of 100,000 histological images was close to 99% in an internal testing set and 94.3% in an external testing set in [3]. However, the experimental results of ResNet50 in this study achieved 99.69% accuracy rate in the same internal testing set and 99.32% in the same external testing set which outperform the data of VGG19 of [3]. In addition, the independent data set with eight classes of [2] is also utilized for comparison purpose. Consequently, ResNet50 achieved 94.86% accuracy and [2] reported the best accuracy rate was 87.4%. Through comprehensive studies and comparison, it is highly suggested that ResNet50 with

suggested settings of this study could be the most efficient and accuracy deep learning techniques to classify colorectal cancer tissue.

In short, the experimental results demonstrate that artificial intelligence has a broad application in classifying colorectal cancer (CRC) histology images, and it can also enhance doctors' critical thinking skills and enable them to make suitable decisions in the diagnostic process.

Author Contributions: Data curation, Y.-H.T.; writing—review & editing, M.-J.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the National Science Council in Taiwan, Republic of China, grant number MOST 109-2410-H-009-022-MY3.

Acknowledgments: The authors thank to National Center for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs) in Taiwan for providing computational and storage resources.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Egeblad, M.; Nakasone, E.S.; Werb, Z. Tumors as Organs: Complex Tissues that Interface with the Entire Organism. *Dev. Cell* **2010**, *18*, 884–901. [CrossRef] [PubMed]
- Kather, J.N.; Weis, C.-A.; Bianconi, F.; Melchers, S.M.; Schad, L.R.; Gaiser, T.; Marx, A.; Zöllner, F. Multi-class texture analysis in colorectal cancer histology. *Sci. Rep.* **2016**, *6*, 27988. [CrossRef]
- Kather, J.N.; Krisam, J.; Charoentong, P.; Luedde, T.; Herpel, E.; Weis, C.-A.; Gaiser, T.; Marx, A.; Valous, N.A.; Ferber, D.; et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med.* **2019**, *16*, e1002730. [CrossRef]
- Gurcan, M.N.; Boucheron, L.E.; Can, A.; Madabhushi, A.; Rajpoot, N.M.; Yener, B. Histopathological Image Analysis: A Review. *IEEE Rev. Biomed. Eng.* **2009**, *2*, 147–171. [CrossRef] [PubMed]
- Zhang, S.; Metaxas, D. Large-scale medical image analytics: Recent methodologies, applications and future directions. *Med. Image Anal.* **2016**, *33*, 98–101. [CrossRef] [PubMed]
- Zhang, X.; Su, H.; Yang, L.; Zhang, S. Fine-grained histopathological image analysis via robust segmentation and large-scale retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5361–5368.
- Janowczyk, A.; Madabhushi, A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J. Pathol. Inform.* **2016**, *7*, 29. [CrossRef]
- Korbar, B.; Olofson, A.M.; Mirafior, A.P.; Nicka, C.M.; Suriawinata, M.A.; Torresani, L.; Suriawinata, A.A.; Hassanpour, S. Deep learning for classification of colorectal polyps on whole-slide images. *J. Pathol. Inform.* **2017**, *8*. [CrossRef]
- Tsai, M.J.; Tao, Y.H. Machine Learning Based Common Radiologist-Level Pneumonia Detection on Chest X-rays. In Proceedings of the 2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS), Gold Coast, Australia, 16–18 December 2019.
- Xu, J.; Luo, X.; Wang, G.; Gilmore, H.; Madabhushi, A. A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing* **2016**, *191*, 214–223. [CrossRef]
- Du, Y.; Zhang, R.; Zargari, A.; Thai, T.C.; Gunderson, C.C.; Moxley, K.M.; Liu, H.; Zheng, B.; Qiu, Y. Classification of tumor epithelium and stroma by exploiting image features learned by deep convolutional neural networks. *Ann. Biomed. Eng.* **2018**, *46*, 1988–1999. [CrossRef]
- Xu, Y.; Jia, Z.; Wang, L.-B.; Ai, Y.; Zhang, F.; Lai, M.; Chang, E.I.-C. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinform.* **2017**, *18*, 281. [CrossRef]
- Bejnordi, B.E.; Mullooly, M.; Pfeiffer, R.M.; Fan, S.; Vacek, P.M.; Weaver, D.L.; Herschorn, S.; Brinton, L.A.; Van Ginneken, B.; Karssemeijer, N.; et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Mod. Pathol.* **2018**, *31*, 1502–1512. [CrossRef]
- Bowles, M. *Machine Learning in Python: Essential Techniques for Predictive Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
- Hubel, D.H.; Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **1962**, *160*, 106–154. [CrossRef]
- Hinton, G.; Salakhutdinov, R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504–507. [CrossRef]
- Markoff, J. How Many Computers to Identify a Cat. 22 June 2012. Available online: <https://mobile.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html> (accessed on 11 July 2020).
- LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
- Lin, M.; Chen, Q.; Yan, S. Network in Network, Cornell University Library. *arXiv* **2014**, arXiv:1312.4400.

20. Rumelhart, E.; Geoffrey, E.; Ronald, J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
21. Yadav, S.S.; Jadhav, S.M. Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data* **2019**, *6*, 1–18. [[CrossRef](#)]
22. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
23. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
24. Simonyan, K.Z. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
25. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 7–12.
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 1 June 2016; pp. 770–778.
27. Yaqub, M.; Feng, J.; Zia, M.; Arshid, K.; Jia, K.; Rehman, Z.; Mehmood, A. State-of-the-Art CNN Optimizer for Brain Tumor Segmentation in Magnetic Resonance Images. *Brain Sci.* **2020**, *10*, 427. [[CrossRef](#)] [[PubMed](#)]
28. Janocha, K.; Czarnecki, W.M. On Loss Functions for Deep Neural Networks in Classification. *Schedae Inform.* **2016**, *25*, 49–59. [[CrossRef](#)]
29. Manwar, R.; Li, X.; Mahmoodkalayeh, S.; Asano, E.; Zhu, D.; Avanaki, K. Deep learning protocol for improved photoacoustic brain imaging. *J. Biophotonics* **2020**, *13*, e202000212. [[CrossRef](#)] [[PubMed](#)]
30. Yessou, H.; Sumbul, G.; Demir, B. A Comparative Study of Deep Learning Loss Functions for Multi-Label Remote Sensing Image Classification. In Proceedings of the IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020. [[CrossRef](#)]